

Spectroscopy of H II regions

Revised, 14 April 2021, DMW

1. Nebulium	1
2. H II regions	3
3. Recombination lines of hydrogen.....	3
4. Extinction and its correction.....	7
5. Forbidden lines.....	8
6. Collisional de-excitation.....	11
7. Detailed balance in thermal equilibrium, and collisional excitation.....	14
8. Critical density for collisional de-excitation	14
9. Multilevel collisional excitation	15
10. Derivation of physical parameters from line-intensity observations.....	17
10.1. Uniform ionized cloud, 1-D	18
10.2. More complex, 1-D and 3-D models	24
11. CHIANTI and Cloudy.....	24

1. Nebulium

The brightest gaseous nebulae in the sky are made of starlight-ionized gas. They are subdivided into **H II regions**: sprawling, irregular clouds associated with massive young stars and their formation regions; and **planetary nebulae**: compact, symmetrical shells of gas shed from dying, non-massive stars. Both varieties were subjects of the very first astronomical spectroscopic studies, soon after absorption lines in the Solar spectrum were identified, in 1860, with emission lines in laboratory spectra of hot vapors of chemically-purified elements, by Robert Bunsen and Gustav Kirchhoff ¹. This was the first proof that the Sun contains all the elements found on Earth. One of the first important followups to the Bunsen-Kirchhoff work was the identification on the Sun of the theretofore-undiscovered element Helium in 1868, which took a few more decades to find on Earth.

¹ Kirchhoff, G.R. 1860, [Annalen der Physik, 185, 275](#). Same Kirchhoff whose rules you use to solve electrical circuit problems; same Bunsen whose burner heats your chemistry experiments. This was a hugely important result with influence throughout the sciences and the humanities. Statues of both Bunsen and Kirchhoff are found in Heidelberg, outside the buildings housing their departments: the highest accolade a scientist could receive, during the times before Nobel prizes, and after absolute monarchies.

In contrast to the Sun, other stars, and the nebulae we now call galaxies, H II regions and planetary nebula show only emission-line spectra. The emission includes many bright lines unrepresented in any laboratory spectra. A set of particularly strong and pervasive unidentified nebular lines at 436 nm, 496 nm, and 501 nm was first seen around 1864 by William Huggins. With Helium as the precedent, and there being no good reason to assume that celestial objects consist *only* of the terrestrial cast of elements, Huggins eventually took these lines to be the spectral signature of an element, Nebulium, that is found neither on Earth nor in the stars. This suggestion even survived the invention of the periodic table by Meyer and Mendeleev (1869) ², which at the time of Huggins' work was rapidly filling up with elements found in Earth or Sun. Failing to find a place in the table for Nebulium, people began discussing the possibility of "protoelements" of which Nebulium might be an example: that the nebulae are the progenitors of stars and planets; that nebulae are made of protoelements; and that the stars and planets are made from the elements, after those form *from* the protoelements. At the time, too little was known about the nature of atomic nuclei to rule out such possibilities.

This issue was still around in the mid-1920s for Ira Bowen to resolve. Bowen measured many very-high-resolution vacuum ultraviolet spectra of second-row elements as a function of ionization state, and used the spectra to characterize the energy levels of these ions. Thereby he discovered that the energy differences between the lowest-energy, **metastable**, states of several ordinary ions – some of whose permitted lines had been detected, albeit faintly, in nebulae – matched up perfectly with all the spectral lines ascribed to Nebulium. Bowen noted that the forbidden nature of transitions between these metastable states, and the large rate of collisions even under the hardest vacuum that could be produced, would make it essentially impossible to detect the transitions directly in the laboratory. He also noted that, by the same token, these states could radiate brightly under the more rarified conditions already known by hydrogen recombination-line observations to exist in nebulae. Finally he noted that the brightest Nebulium lines, with wavelengths 436.3 nm, 495.9 nm, and 500.7 nm in air, were strongly correlated in strength with recombination lines of He⁺. Bowen had found the wavelengths of these lines to match metastable energy-level differences in O⁺⁺; this ion has excitation and ionization potentials similar to He⁺ and should be expected to subsist in the same parts of nebulae. ³

Thus, after a six-decade scientific life, Nebulium suddenly suffered the same fate as the aether. Soon, in the hands of such scientists as Zanstra, Menzel, Aller, and Bowen ⁴, the

² Found most accessibly in the Dover reprint collection: Mendeleev, D.I. 2013, [Selected writings, 1869-1905 \(New York: Dover\)](#), paper #1. Mendeleev was Kirchhoff's grad student.

³ Bowen, I.S. 1927, [PASP, 39, 295](#) and [Nature, 120, 473](#).

⁴ In her revolutionary work on the relative abundances of elements in stellar atmospheres (PhD thesis, Harvard, 1925, and [PNAS, 11, 192](#)), Cecilia Payne uncharacteristically bowed in one respect to the prejudices of senior astronomers, prominently Henry Norris Russell's. Though it was a robust result of her

forbidden lines emitted by nebulae took their current place as the best probes of physical conditions and element abundance in the interstellar medium.

2. H II regions

Although visible light from young massive stars may escape in certain directions, their UV emission rarely does. Neutral hydrogen and especially dust are just too good at absorbing it. O stars have sufficient UV luminosity to dissociate molecules and ionize atoms for several light years around. In turn, these atoms and ions can emit light with a much better chance than UV of escaping to a distant observer. Result: the temperature of the ionized gas is determined by the balance between UV heating, and the cooling emission by atoms, ions, and dust. The detection of this emission permits us to seek a solution to the inverse problem: what the density, temperature, ionization state, and element abundances are, in the dissected gas that once belonged to the O star's natal molecular cloud.

This is a sport in which observational records are still broken by dramatic margins, despite 150 years of competition. Recently, integral-field spectrographs ⁵, such as [MUSE](#) on one of ESO's 8-meter components of the Very Large Telescope in Chile, have produced essentially perfect data sets on the brightest H II regions. The MUSE results on the central 6'×9' of M 42 ⁶, for example, cover $\lambda = 460 - 937$ nm, at high dynamic range and approximately Nyquist spatial sampling of the best seeing, with 0.2"×0.2"×0.085 nm spaxels. In this datacube all the major features and lines in the nebula are spatially and spectrally resolved. Several abundant elements are detected in all their significant ionization states. This yields a full accounting of temperature and density, ionization state, element abundance, bulk motion, and distinction between the H II region and the Herbig-Haro objects associated with younger stars behind the nebula. It took a total of only *five minutes* of exposure time to do all this. Figures 1-2 are samples of the results.

3. Recombination lines of hydrogen

Once ionized, atoms recombine, usually into a highly excited state; once recombined, atoms emit a cascade of spectral lines, called **recombination lines**, as they decay to their

analysis that H and He were vastly more abundant in stars than the other elements, despite not being so on Earth, Payne concluded that "...the stellar abundance deduced for these elements is improbably high, and is almost certainly not real." It must have cheered her up that, soon afterward, her result was strongly supported by those on gaseous nebulae. Bowen's conclusion, for example (1935, [ApJ, 81, 1](#)): "It is therefore evident that all elements except H and He are very rare in the nebulae."

⁵ Invented by Bowen (1938): [ApJ, 88, 113](#).

⁶ Wellbacher, P.M. et al. 2015, [A&A, 582, 114](#); McCleod, A.F. et al. 2016, [MNRAS, 455, 407](#).

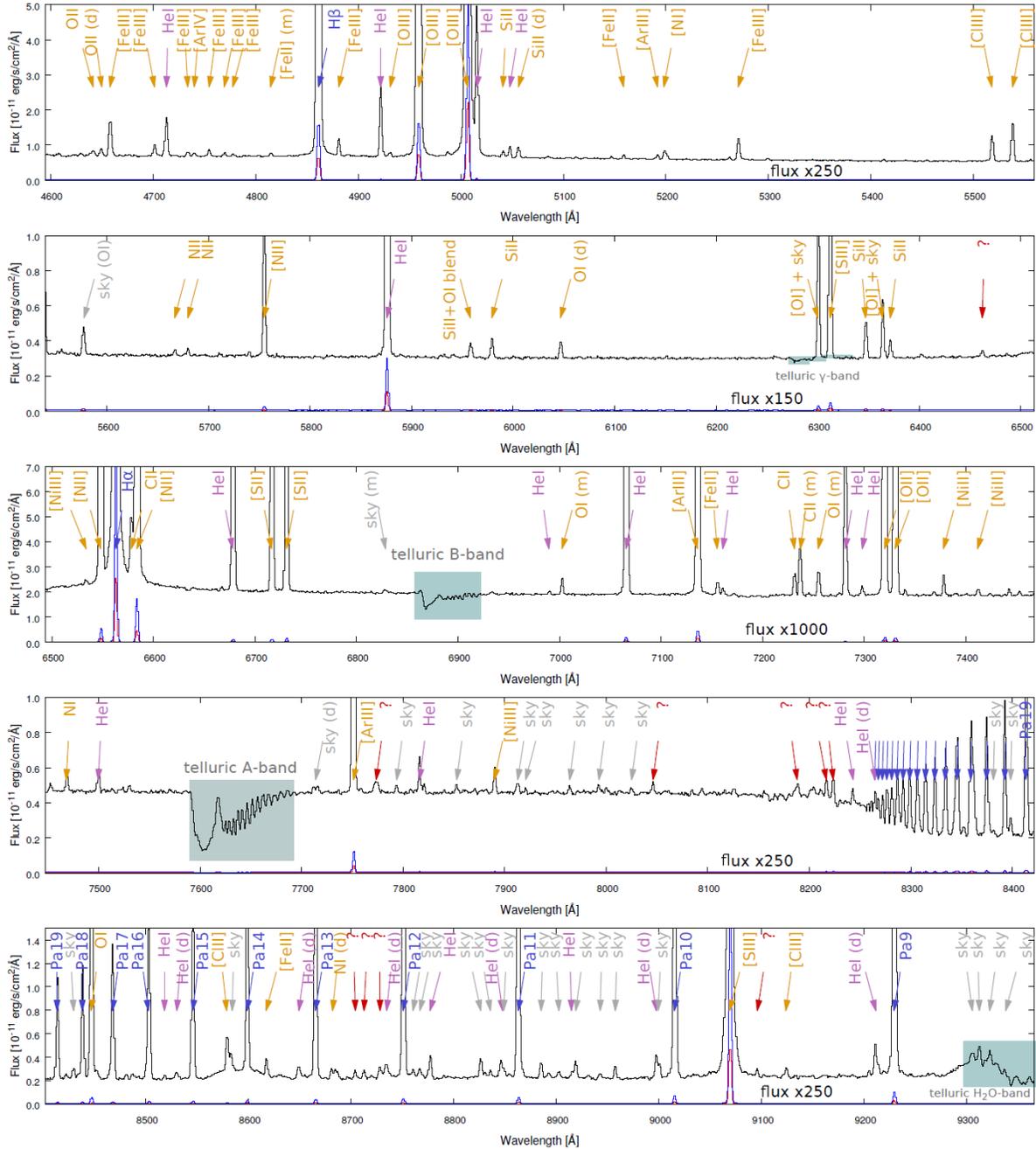


Figure 1: MUSE spectra ⁶ of a couple of east-west stripes of M 42, shown in blue and red; and a magnified version of the latter spectrum, in black with scale factors given in each frame, to show fainter lines. The molecular bands labelled A and B are from oxygen in Earth’s atmosphere. Ten “U lines” (unidentified) are also present, surprisingly numerous at this late date. They are probably not due to Nebulium 2.0; most are unidentified due to multiple plausible identifications.



Figure 2: MUSE spectral-line composite images of M 42 ⁶. Both are corrected for extinction, using images of the Balmer decrement. Left: RGB = [S II] 673.1 nm, [N II] 658.4 nm, H β . Right: RGB = [O III] 500.7 nm, [O II] 732.0 nm, [O I] 630.0 nm.

ground state. The spectral lines emitted are all **permitted**, a.k.a **electric-dipole** ⁷, transitions. The radiation rates – the average number of spontaneous transitions per unit time, a.k.a **Einstein A-coefficients**, of such transitions are of order once per microsecond – much faster than the rate at which atoms or ions collide with each other in interstellar gas, as we will see below. So the vast majority of neutral H atoms are in the ground state, and the emission rate for a given recombination line depends only on the rate of recombination, times a factor that accounts for the branching among different allowed

⁷ Too quickly: In first-order, time-dependent perturbation theory, one learns that the rate of spontaneous transitions between states j and i is proportional to the square modulus of the matrix element between these states of the interaction Hamiltonian V , $|\langle j|V|i\rangle|^2$. In turn the interaction Hamiltonian V for light and an atom contains a factor $e^{-i\kappa\cdot r}$ that represents a plane wave of light; here $|\kappa| = 2\pi/\lambda$ is the wavenumber of the light. The wavelength of light for ultraviolet wavelengths and longward is many orders of magnitude larger than the extent of an atom's electron states in r , so it is appropriate to expand the exponential in a power series ($e^{-i\kappa\cdot r} = 1 - i\kappa\cdot r + \dots$) and keep only the lowest-order term that gives a nonvanishing matrix element $\langle j|V|i\rangle$. Together with the rest of the interaction Hamiltonian, the first term in this series expansion gives rise to a matrix element proportional to $q_e \sum_n \langle j|r_n|i\rangle$, which happens to be the total electric dipole moment of the atom's electrons. Transitions between states j and i for which this term gives the rate are therefore called electric dipole transitions. Conservation of angular momentum and the spin-1 nature of photons – or, if you prefer, the orthogonality of spherical harmonics – dictate the selection rule for electric dipole transitions in single-electron atoms: $\Delta\ell = \pm 1$, where ℓ is the orbital angular momentum quantum number.

Table 1: visible and infrared hydrogen recombination lines, $T = 10^4$ K.

λ , vacuum (μm)						$(hc/\lambda)\alpha_{ji}$ ($\text{erg cm}^{-3} \text{sec}^{-1}$)					
	Balmer	Paschen	Brackett	Pfund	Humphreys		Balmer	Paschen	Brackett	Pfund	Humphreys
j	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	j	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$
10	0.37970	0.90125	1.73575	3.03756	5.12588	10	6.55E-27	2.28E-27	1.13E-27	6.56E-28	4.18E-28
9	0.38344	0.92266	1.81693	3.29521	5.90501	9	9.03E-27	3.14E-27	1.56E-27	9.10E-28	5.79E-28
8	0.38881	0.95435	1.94404	3.73853	7.49843	8	1.30E-26	4.52E-27	2.26E-27	1.32E-27	8.31E-28
7	0.39691	1.00467	2.16495	4.65126	12.36519	7	1.96E-26	6.85E-27	3.44E-27	2.00E-27	1.21E-27
6	0.41007	1.09352	2.62445	7.45582		6	3.20E-26	1.12E-26	5.62E-27	3.15E-27	
5	0.43394	1.28147	4.05008			5	5.78E-26	2.02E-26	9.91E-27		
4	0.48601	1.87461				4	1.24E-25	4.18E-26			
3	0.65611					3	3.54E-25				

paths in the cascade. All but recombinations directly into the ground ($n = 1$) state, that is: this results in the emission of a photon with energy greater than the ionization energy, that are absorbed, practically on the spot, by a neighboring neutral hydrogen atom. Recombinations that immediately lead to ionizations don't really count as recombinations: they are left out of the rate coefficients we use for H II region analysis. ⁸

The rate of recombination also depends on the densities of electrons and ions that are available to recombine. Recombination and branching, leading to a transition between states j and i , are usually characterized by an **effective recombination rate coefficient** α_{ji} , such that the rate at which these transitions happen per unit volume is given by $\alpha_{ji}n_en_{\text{ion}}$. The α s are calculated (quantum-mechanically), and are known with very high accuracy for simple, well-characterized atoms like hydrogen. They depend weakly on temperature and even more weakly on density.

Given an effective recombination rate, the total power per unit volume, \mathcal{J}_{ji} , emitted in spectral line radiation for states j and i , is simply the product of this rate and the photon energy:

$$\mathcal{J}_{ji} = \frac{hc}{\lambda_{ji}} \alpha_{ji} n_e n_{\text{ion}} \quad . \quad (1)$$

Table 1 is a list of wavelengths ⁹ and recombination rate coefficients ¹⁰ times hc/λ for hydrogen lines connecting lower-energy states, at temperature $T = 10^4$ K and low

⁸ The omission from rate coefficients of recombinations directly into the ground state is called **Case B recombination**; this case will apply to all the rate coefficients we list in the following.

⁹ https://physics.nist.gov/PhysRefData/ASD/lines_form.html

¹⁰ Hummer, D.G. & Storey, P.J. 1987 *MNRAS*, **224**, 801; Storey, P.J. & Sochi, T. 2013, *MNRAS*, **440**, 2581. In the latter reference the results of the former are extended to the kappa electron energy distribution. The easiest way to get the former reference's data in electronic form is to use the latter reference at the largest values of kappa.

density. Transitions to the $2s$ or $2p$ state, called the **Balmer series**, lie at visible wavelengths and are labelled $H\alpha$, $H\beta$, ... from long wavelength to shorter.

We presume the atoms emit isotropically: uniformly into 4π steradians. We are safe in assuming that nebulae are optically thin (transparent) to recombination-line radiation, except for those that connect to the ground, $n = 1$ state ¹¹. With these caveats the recombination line intensity I_{ji} (power per unit area and solid angle) at the surface of the nebula is

$$I_{ji} = \frac{1}{4\pi} \int \mathcal{J}_{ji} ds = \frac{hc}{4\pi\lambda_{ji}} \int \alpha_{ji} n_e n_{\text{ion}} ds \quad , \quad (2)$$

where the integration is over distance s along the line of sight through the nebula. In the case of hydrogen, the ion density is that of free protons, n_p . Since α_{ji} depends only weakly on density and temperature along the line-of-sight distance s , it can be taken out of the integral to good approximation. The remaining integral $\int n_e n_p ds$ is often called the nebular **emission measure**.

4. Extinction and its correction

The intensity received by a distant observer suffers an additional factor of $e^{-\tau(\lambda_{ji})}$: extinction, primarily due to absorption by cold foreground dust. This must be corrected before analysis. If two hydrogen recombination lines are observed, from upper states that can be assumed to be populated only by recombination ¹², the extinction can be determined accurately. This is also insensitive to the details of the nebula, due to the weak dependence of the α_s on density and temperature. The two longest-wavelength Balmer series lines, $H\alpha$ and $H\beta$, often serve this purpose. Their intensity ratio, called the **Balmer decrement**, is $I(H\alpha)/I(H\beta) \cong 2.86$ without extinction; any upward departure from this ratio can be ascribed to extinction, and serves to measure the reddening, $e^{-\tau(H\alpha)+\tau(H\beta)}$. From this plus the experimental results on interstellar dust one can obtain the foreground extinction, as in Figure 3.

¹¹ Because of hydrogen's large abundance and the pileup of the population in the ground state, nebulae are always extremely optically thick (opaque) in transitions to the ground, $1s$ state: the **Lyman series**.

¹² Pure recombination and Case B apply practically universally to photoionized nebulae. In the next memo of this series, on spectral line emission from shock-excited gas, we will meet conditions in which collisional excitation is significant for some states of H; Case B intensity ratios involving these states do not apply. Annoyingly the most prominent example is the upper state of the brightest line from shocks, $H\alpha$.

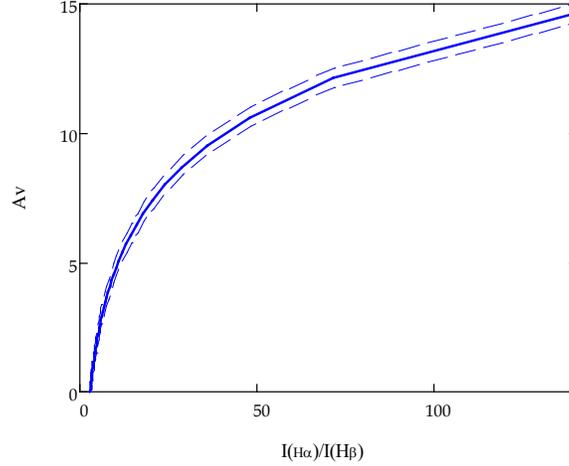


Figure 3: A_V in magnitudes, plotted against Balmer decrement, $I(H\alpha)/I(H\beta)$, with hypothetical $\pm 10\%$ uncertainty in the decrement indicated by the dashed lines. The decrement is evaluated at low density and $T = 10^4 \text{K}$, as in Table 1. Extinction and dust grain properties are from Weingartner & Draine¹³ for a total-to-selective extinction ratio of $R_V \equiv A_V/E(B-V) = 5.5$.

5. Forbidden lines

In a multielectron atom or ion, the lowest-energy electronic configuration gives rise to several **multiplets**: groups of states with the same total orbital and spin angular momentum, separated from the other multiplets by energies of order 0.1 – 1 eV.¹⁴ In the

¹³ Weingartner, J.,C. & Draine, B.T. 2001, [ApJ, 548, 296](#).

¹⁴ Also too quickly: The classification scheme normally used for the states of elements H through Fe is **Russell-Saunders**, or **LS, coupling**. In this scheme, one adds electron spins and directions to get a total electron spin quantum number S , their orbital angular momenta to get a total orbital angular momentum quantum number L , and takes the vector sum of S and L to get the total angular momentum quantum number J . For example, O^{++} has electronic configuration $[\text{He}]2s^2 2p^2$. The spins of the two p electrons, either up (\uparrow) or down (\downarrow), can be arranged into either in a **triplet** ($|\uparrow\uparrow\rangle, |\downarrow\downarrow\rangle, \text{ or } \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)$) or a **singlet** ($\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$), so O^{++} 's total spin quantum number S can be either 1 or zero respectively. Each electron has one unit of orbital angular momentum, so the total orbital angular momentum quantum number L can be zero, 1, or 2. The spin triplet state has even symmetry: interchanging the electron coordinates does not change the sign of the spin part of the wavefunction. By the same token the odd-symmetry, spin-singlet state changes sign on electron-coordinate interchange. However, the total wavefunction must change sign on interchange of the electrons if they are to obey Fermi-Dirac statistics. The $L = 1$ orbital-angular-momentum state changes sign on interchange, as the Legendre polynomial to which it is proportional, P_1 , is an odd function of the spatial coordinates. The others do not. Thus the $S = 1$ states can only have $L = 1$, while the $L = 0$ and 2 states must have $S = 0$. The choices for total angular

lighter-weight atoms, in which the indices of total orbital and spin angular momenta, L and S , are good quantum numbers¹⁵, transitions among these multiplets, and the states within them, are **forbidden**¹⁶. Their Einstein A-coefficients are many orders of magnitude smaller than those of lines of similar wavelength linking different electronic configurations, such as recombination lines. Forbidden transitions are of magnetic-dipole or electric-quadrupole nature, rather than electric dipole like recombination lines¹⁷. States for which all transitions to lower-energy states are forbidden, are called **metastable**. Figure 3 is a diagram of the lowest-energy multiplets of O^{++} and S^+ , two ions common in H II regions, along with the brighter visible and infrared lines connecting their states. All these states are metastable, and linked only by forbidden transitions, as is common among multielectron atoms – as Bowen found, and as described in Section 1.

For elements of the first three rows, or their ions, electronic wavefunctions are well characterized. The matrix elements for magnetic dipole and electric quadrupole transitions, and thus their A-coefficients, can be calculated accurately. If the density of an ion in state j is n_j , then the power in transition $j \rightarrow i$ emitted from a unit volume is

$$\mathcal{J}_{ji} = \frac{hc}{\lambda_{ji}} A_{ji} n_j \quad . \quad (3)$$

momentum quantum number J are zero, one, and two for the $L = 1$ states, while the $L = 0$ and 2 states must have $J = 0$ and 2 respectively. It is customary to label these states $^{2S+1}L_J$, and to use the electron-orbital-reminiscent letters S , P , and D to stand for $L = 0, 1$ and 2. So the five states of $[\text{He}]2s^2 2p^2$ are 1S_0 , 1D_2 , 3P_2 , 3P_1 , and 3P_0 .

¹⁵ A good quantum number corresponds to a rigorously conserved quantity. Total angular momentum is of course such a quantity. L and S are often pretty good quantum numbers, but in principle their corresponding angular momenta are not individually conserved. LS coupling becomes less useful past the first three rows of the periodic table. Past Fe or so, it usually works better to use **j - j coupling**, for which one combines the spin and orbital angular momenta of each electron into a “subtotal” angular momentum j , and then finds the total electronic angular momentum J from the vector sum of all the individual js .

¹⁶ Forbidden lines are indicated with square brackets enclosing the identity and spectral series (ionization state) of the emitting species. Thus the forbidden transition between $O^{++} \ ^1S_0$ and 1D_2 goes as $[\text{O III}] \ 436.3$ nm, and the permitted transition between hydrogen’s $n = 3$ and $n = 2$ states (H α) is H I 656.3 nm.

¹⁷ In strict LS coupling, the selection rules for electric dipole transitions are $\Delta S = 0$; $\Delta L = \pm 1$, or 0; and $\Delta J = \pm 1$, or 0, but no $J = 0 \rightarrow 0$. Magnetic dipole transitions have no separate selection rule for S – that is, any ΔS is OK for such a transition – but require $\Delta L = 0$, $\Delta J = \pm 1$. For electric quadrupole transitions only J has a selection rule: $\Delta J = \pm 2, \pm 1, 0$, but no $0 \rightarrow 0$, $1/2 \rightarrow 1/2$, or $0 \rightarrow 1$.

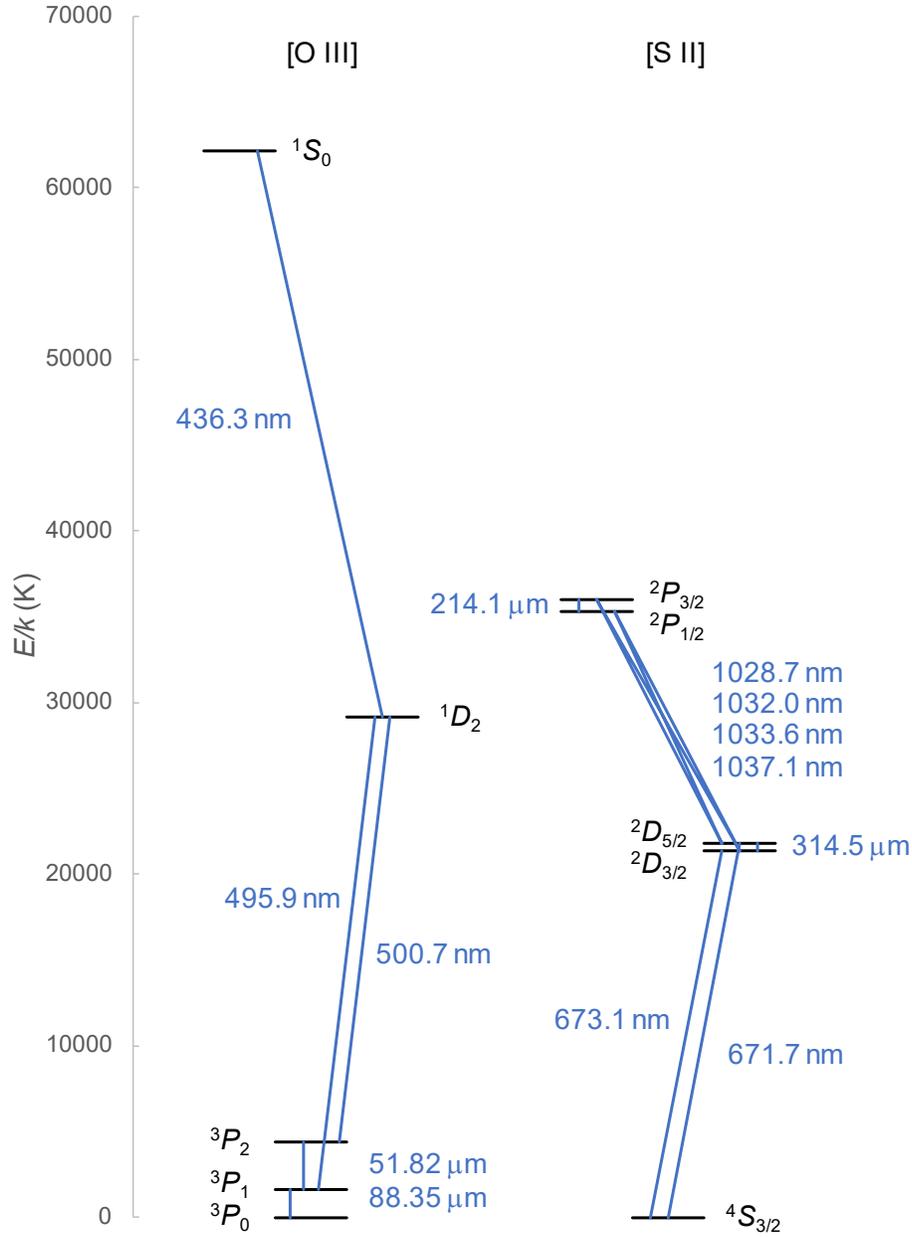


Figure 4: lowest five states of O⁺⁺ and S⁺, all of which are metastable, and the stronger of the forbidden transitions connecting them.

Here A_{ji} is the Einstein A-coefficient for the transition¹⁸. As they are for recombination lines, nebulae are optically thin (transparent) to forbidden-line radiation. In analogy with Equation (2), the intensity emitted from the nebula, at the nebula's surface, is

¹⁸ Forbidden transitions being too weak to measure in the laboratory, their A-coefficients must be calculated from the electronic wavefunctions of the ions⁷, which in turn can be constrained accurately by experimental determination of the energies of these states. The simpler the atom or ion, the more accurately its wavefunctions are known. For elements of the first three rows of the periodic table, like O and S,

$$I_{ji} = \frac{hc}{4\pi\lambda_{ji}} A_{ji} \int n_j ds \quad . \quad (4)$$

The integral of density along the line of sight, $N_j = \int n_j ds$, is often called the **column density** of the ion in state j . We don't know yet how n_j relates to the total density n of the ion; this is the subject of the next few sections.

As before, the intensity received by a distant observer suffers an additional factor of $e^{-\tau(\lambda_{ji})}$ from extinction by cold foreground dust. None of the forbidden lines that are bright in H II regions are particularly useful in determining extinction, but we will meet some in the next memo in this series, on [Fe II] emission from shocked gas.

6. Collisional de-excitation

Unlike recombination lines, forbidden lines are emitted because their upper states are excited by collisions, mostly with electrons within the ionized gas. This makes the forbidden lines crucially important for the thermal state of nebulae: thermal energy is converted thereby, efficiently, to radiation which escapes directly to outer space, cooling the nebula.

Let us consider an oversimplified version of a collision between a positive ion with charge Ze and an electron, incident on the ion at speed v and aimed a distance b from the ion. Suppose that, if the electron got sufficiently close to the ion, it could transfer some of its kinetic energy to the ion and cause it to change states: an inelastic collision. Much lighter than the ion, the electron moves much faster; the ion is essentially at rest during the interaction. Electrostatic attraction bends the electron's path toward the ion¹⁹, as shown in Figure 3. It passes within a distance r_0 of the ion, now travelling at speed v_0 . Energy and angular momentum conservation give us

$$\begin{aligned} \frac{1}{2} m_e v^2 &= \frac{1}{2} m_e v_0^2 - \frac{Ze^2}{r_0} \quad \text{and} \\ m_e v b &= m_e v_0 r_0 \quad . \end{aligned} \quad (5)$$

calculations of A-coefficients have changed very little over the last three decades despite much more complex calculations, and revolutionary changes in computing power that have made them possible, so the A-coefficient accuracy is probably very good.

¹⁹ A proton in the same medium would have the same thermal energy as the electron, but would be steered away from the ion; because of the exponential radial dependence of atomic wavefunctions, the resulting collisional transition rate would be vastly smaller than that with electrons. That's why you'll only ever see electron collisions discussed in the context of H II regions.

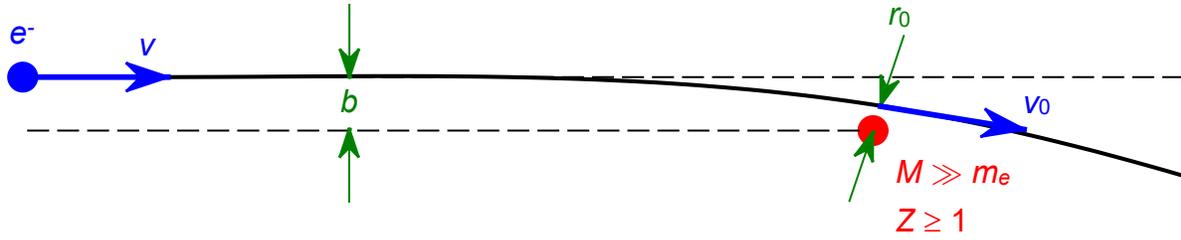


Figure 5: electron-ion collision in ionized gas.

Use the latter expression to eliminate v_0 from the former expression, and rearrange to solve for $\sigma = \pi b^2$:

$$\sigma \equiv \pi b^2 = \pi r_0^2 \left(1 + \frac{2Ze^2}{mv^2 r_0} \right) . \quad (6)$$

For this collision to result in a transition among the ion's states, r_0 must be small enough to give sufficient overlap of wavefunctions. If we take this to be of order the Bohr radius, $r_0 = a_0 = \hbar^2 / m_e e^2$, or smaller, and if we take the initial electron kinetic energy to be the thermal energy, $3kT/2$, at a typical H II region temperature of $T = 10^4$ K, we see that $2Ze^2 / mv^2 r_0 \gtrsim 20Z \gg 1$, so the 1 in the brackets above can be neglected:

$$\sigma \approx \frac{2\pi Z e^2 a_0}{m_e v^2} = \frac{\pi \hbar^2}{m_e^2 v^2} 2Z . \quad (7)$$

The **cross section** σ is the effective area of the atom for intercepting electrons and undergoing that transition. If the electron density is n_e and the electrons move at typical speed v , then the flux of electrons (electrons per unit area) in one direction is $n_e v$. If the density of ions is n , then the rate R at which these inelastic collisions occur per unit volume is

$$R = n_e v \sigma n = n_e n v \frac{\pi \hbar^2}{m_e^2 v^2} 2Z . \quad (8)$$

When one casts this situation as a quantum-mechanical problem – which turns out to be very challenging – the resulting cross section takes a similar form:

$$\sigma_{ji} = \frac{\pi \hbar^2}{m_e^2 v^2} \frac{\Omega_{ji}(v)}{g_j} , \quad (9)$$

where $\Omega_{ji}(v)$ is the dimensionless **collision strength** that embodies the integral of the time-dependent matrix elements; j and i are the states of the ion; and g_j is the statistical weight of state j . $g_j = 2J + 1$ if the state has total angular momentum quantum number J , as this is the number of values of J_z . Usually one calculates just the strengths for collisional de-excitations: j is the higher energy of the two states. The number of collisionally-induced transitions $j \rightarrow i$ per unit volume, $R_{ji}(v)$, is

$$R_{ji}(v) = n_e n_j v \sigma_{ji}(v) = n_e n_j v \frac{\pi \hbar^2}{m_e^2 v^2} \frac{\Omega_{ji}(v)}{g_j} . \quad (10)$$

In thermal equilibrium at temperature T ²⁰, the Maxwell-Boltzmann distribution applies to electron speeds:

$$f(v) = \left(\frac{m_e}{2\pi kT} \right)^{3/2} 4\pi v^2 e^{-m_e v^2 / 2kT} , \quad (11)$$

so the net rate of transitions per unit volume is

$$\langle R_{ji} \rangle = n_e n_j \int_0^\infty v \sigma_{ji}(v) f(v) dv = n_e n_j \sqrt{\frac{2\pi}{kT}} \frac{\hbar^2}{m_e^{3/2}} \frac{\Upsilon_{ji}(T)}{g_j} , \quad (12)$$

where $\Upsilon_{ji}(T)$ is the dimensionless *effective* collision strength. The $\Upsilon_{ji}(T)$ are what gets tabulated in the literature, on a grid of temperatures in a useful range, dense enough to interpolate accurately in codes.

²⁰ Always ask, “the temperature of what?” and “can they even have a temperature?” Here we mean the temperature of the electrons and protons (ionized hydrogens). Their motions are in thermal equilibrium because they efficiently share their kinetic energy and momentum *via* collisions like that in Figure 5. In turn this is because collisions among them are much more frequent than any other process in which they participate. For example: suppose an ionized cloud of diameter $r = 1$ parsec had density $n_e \approx n_p = 1 \text{ cm}^{-3}$ and typical electron speed $v = 1 \text{ km sec}^{-1}$. The proton cross section for electron collisions would be $\sigma = 8.4 \times 10^{-10} \text{ cm}^2$, and the typical time between collisions $1/n_e \sigma v \approx 10^4 \text{ sec}$. This is very short, astrophysically speaking: the cloud-crossing time, for example, is $r/v \approx 10^6 \text{ years}$, more than nine orders of magnitude longer. So electrons and protons certainly collide often enough that their motions represent thermal equilibrium, and we can describe them with a temperature.

7. Detailed balance in thermal equilibrium, and collisional excitation

The next steps in the use of collisional cross sections are to calculate the metastable-state populations that result from the combined effect of collisional de-excitation, excitation, and radiation. This starts with definition of a collisional de-excitation rate coefficient, $\gamma_{ji}(T)$:

$$\gamma_{ji}(T) = \sqrt{\frac{2\pi}{kT}} \frac{\hbar^2}{m_e^{3/2}} \frac{Y_{ji}(T)}{g_j} = \frac{8.629 \times 10^{-6} \text{ cm}^3 \text{ sec}^{-1} \text{ K}^{1/2}}{\sqrt{T}} \frac{Y_{ji}(T)}{g_j} . \quad (13)$$

Upward, collisional excitation, rate coefficients are obtained from the downward ones by application of detailed balance in thermal equilibrium, as follows. If in addition to thermal equilibrium prevailing among translational motions, the states were populated according to thermal equilibrium at the same temperature as the free electrons, there would be **detailed balance** among the collisional transitions between any two states: $\langle R_{ij} \rangle = \langle R_{ji} \rangle$. With $Z(T)$ as the partition function (total number of electronic states; not to be confused with ion charge as above) and n as the total density of this ion, detailed balance would imply

$$\begin{aligned} n_e n_i^{TE} \gamma_{ij}(T) &= n_e n_j^{TE} \gamma_{ji}(T) \\ n_e \frac{n g_i e^{-E_i/kT}}{Z(T)} \gamma_{ij}(T) &= n_e \frac{n g_j e^{-E_j/kT}}{Z(T)} \gamma_{ji}(T) \\ \gamma_{ij}(T) &= \gamma_{ji}(T) \frac{g_j}{g_i} e^{-\Delta E_{ji}/kT} \end{aligned} \quad (14)$$

Since the densities have dropped out of the expression, the **resulting relation between γ_{ij} and γ_{ji} applies at any density**, whether the states are populated thermally or not. So it suffices for the quantum calculations to cover the de-excitations, and for us to obtain the excitations easily from detailed balance.

8. Critical density for collisional de-excitation

A useful marker of the electron-density sensitivity of a collisionally-excited spectral line is the **critical density** n_C of its upper state, defined as the ratio of sums of A-coefficients and collisional excitation rate coefficients involving this state:

$$n_{Cj}(T) = \frac{\sum_{i < j} A_{ji}}{\sum_{i \neq j} \gamma_{ji}(T)} . \quad (15)$$

That is, the critical density is the electron density for which radiative and collisional decay rates are equal. Table 2 is a list of critical densities for the states in Figure 4, for electron collisions at $T = 10000$ K.

Ground states don't decay, of course, so they don't have critical densities.

Table 2: critical densities for S^+ and O^{++}

S^+		O^{++}	
State	n_C (cm^{-3})	State	n_C (cm^{-3})
$^2D_{3/2}$	2.82×10^3	3P_1	5.01×10^2
$^2D_{5/2}$	1.26×10^3	3P_2	3.53×10^3
$^2P_{1/2}$	1.19×10^6	1D_2	6.92×10^5
$^2P_{3/2}$	2.24×10^6	1S_0	2.43×10^7

9. Multilevel collisional excitation

This section can be skipped at first reading. The material is not difficult, but the linear algebra is tedious. Code is available on request which may eliminate the need to stare at the equations wondering how you got the array indices wrong.

The partition of the population of ions over the various states is determined under the assumption that the densities have reached a **steady state**²¹. Assuming only optically-thin media, the rate of transitions *out* of state j equals the rate of transitions *in*:

$$n_j \sum_{i \neq j} n_e \gamma_{ji} + n_j \sum_{i < j} A_{ji} = \sum_{i \neq j} n_i n_e \gamma_{ij} + \sum_{i > j} n_i A_{ij} \quad . \quad (16)$$

This is a system of equations that we must solve for the unknown n_j , in order to calculate line intensities (Equation (4)). Since $\gamma_{jj} = 0$, and since $A_{ji} = 0$ unless state i has lower energy than state j , we will assume that the states are indexed in order of increasing energy, and allow all the sums to extend over all bound states. The densities in all the levels add up to the density of the ion:

²¹ Steady state is not the same as thermal equilibrium. Some or all the states of the ion in question may have populations very far from their thermal equilibrium values, at the temperature of the surrounding electron gas.

$$\sum_j n_j = n \quad . \quad (17)$$

Write $f_j = n_j/n$ and rearrange slightly to get

$$f_j \sum_i (n_e \gamma_{ji} + A_{ji}) - \sum_i f_i (n_e \gamma_{ij} + A_{ij}) = 0 \quad , \quad (18)$$

$$\sum_j f_j = 1 \quad . \quad (19)$$

The second term in Equation (18) is a matrix multiplication: the inner product of a second-rank tensor, $n_e \gamma_{ij} + A_{ij}$, with a vector, f_i . The first term is of the form $f_j C_j$, which could also be thought of as a matrix multiplication if C_j were the diagonal elements of a matrix with no non-zero off-diagonal elements. Thus we can use a Kronecker delta to represent the diagonal matrix, and rewrite Equation (18) as

$$\sum_i f_i \delta_{ij} \sum_k (n_e \gamma_{jk} + A_{jk}) - \sum_i f_i (n_e \gamma_{ij} + A_{ij}) = 0 \quad , \quad (20)$$

or
$$\sum_i f_i D_{ij} = 0 \quad , \quad D_{ij} = \delta_{ij} \sum_k (n_e \gamma_{jk} + A_{jk}) - n_e \gamma_{ij} - A_{ij} \quad . \quad (21)$$

This is a linear system in the unknowns f_i , as many equations as unknowns. However, the equations are not linearly independent: clearly the coefficient matrix D_{ij} has zero determinant. We can substitute, for one of these equations, the normalization condition, Equation (19). Suppose we replace the p th one, creating a new coefficient matrix D'_{ij} :

$$D'_{pj} = 1 \quad , \quad D'_{i \neq p, j} = D_{ij} \quad . \quad (22)$$

With this substitution, Equation (21) becomes

$$\sum_i f_i D'_{ij} = \delta_{jp} \quad . \quad (23)$$

There are many efficient ways to solve this system for f_j . One *inefficient* way to do so is simply to calculate the inverse of the coefficient matrix and multiply through:

$$f_k = \sum_j \delta_{pj} D'_{jk}{}^{-1} = D'_{pk}{}^{-1} \quad (24)$$

For the temperatures and densities encountered in H II regions, one may neglect the population in excited electron configurations, leaving only the metastable states of the lowest-energy configuration: five states or so, as in Figure 4. It's not prohibitively slow to invert that small a coefficient matrix; most of us do it that way, using matrix-inversion routines native to our programming environment.

Figure 6 is a set of solutions of Equation (23) for the metastable states of S^+ and O^{++} . Note in this figure that the curves all level off at high density, when the states have reached their thermal-equilibrium population fractions. Note also that each curve levels off at an electron density a little larger than the critical density listed in Table 2.

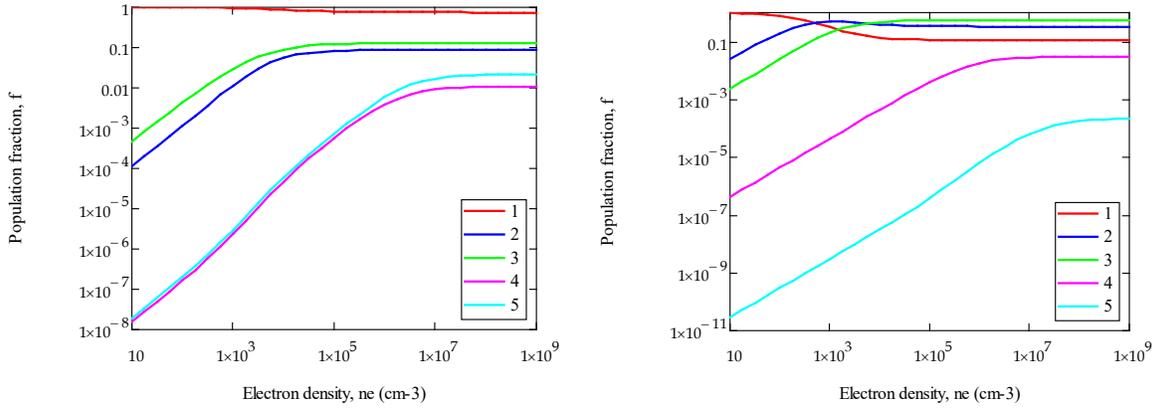


Figure 6: population fractions f_i for the metastable states of S^+ (left) and O^{++} (right) at $T = 10^4$ K, as a function of electron density n_e . The curves are labelled in increasing order of the energy of the states.

10. Derivation of physical parameters from line-intensity observations

We can measure recombination and forbidden-line intensities from H II regions. Using these we can extract the density, temperature, ionization state, and element abundances in the nebula, and the foreground extinction, from Equations (2), (4), and (24):

$$\begin{aligned}
 I_{ji} &= e^{-\tau(\lambda_{ji})} \frac{hc}{4\pi\lambda_{ji}} \int \alpha_{ji} n_e n_p ds \quad \text{for recombination lines, and} \\
 &= e^{-\tau(\lambda_{ji})} \frac{hc}{4\pi\lambda_{ji}} A_{ji} \int f_j n ds \quad \text{for forbidden lines,}
 \end{aligned} \tag{25}$$

where, as in the previous section, n is the density of the ion producing the forbidden line, and f_j is the fraction of the ions in the upper state j . We write the foreground extinction

factor in the sensible exponential form, though of course it is conventional to express extinction in magnitudes.

There are several approaches to the use of Equations (25) with visible and near-infrared data. One can get surprisingly far, though, with the simplest approach that would spring to mind: a uniform plane-parallel slab of ionized gas, in general a different one for every angular resolution element in your images or along your spectrograph slit.

10.1. Uniform ionized cloud, 1-D

With the one D along the line of sight, this assumption simplifies the integrals:

$$\begin{aligned}
 I_{ji} &= e^{-\tau(\lambda_{ji})} \frac{hc}{4\pi\lambda_{ji}} \alpha_{ji} n_e N_p \quad \text{for recombination lines, and} \\
 &= e^{-\tau(\lambda_{ji})} \frac{hc}{4\pi\lambda_{ji}} A_{ji} f_j N \quad \text{for forbidden lines,}
 \end{aligned}
 \tag{26}$$

where the N s are column densities of protons and ions: density integrated along the line of sight, as in section 5 above. We will assume in the following that I_{ji} refers to total line flux, but the analysis does not differ substantially if the line profiles are spectrally resolved. The steps, in order:

- a. **Determine the foreground extinction**, using a pair of hydrogen recombination lines: the Balmer decrement, for example, as in Figure 3. H II regions are associated with molecular clouds, so the extinguishing dust grains are usually on the high end of the range of total/selective extinction ratio R . In Figure 3 we used the industry-standard spectra of interstellar dust by Weingartner & Draine ¹³, with $R_V \equiv A_V/E(B-V) = 5.5$. A routine to interpolate the dust spectrum and calculate the extinction, as in that figure, is available on request. Figure 7 is an image of the extinction toward M 42, derived from MUSE data from the Balmer decrement.
- b. **Determine the electron density and temperature**. The idea here is to base determination of one of the parameters mostly in a way that is insensitive to the other. For electron density you need two lines of the same species whose upper states have similar energy but different critical densities, and not too different from the actual electron density n_e . You get bonus points for the use of pairs of lines in which all states have energies $\ll kT$, such as the far-infrared lines among the 3P states of O^{++} . In such cases the results are nearly independent of T . For temperature you need two lines of the same species for which the upper-state energies differ by at least a hefty fraction of kT , where T is the actual temperature.

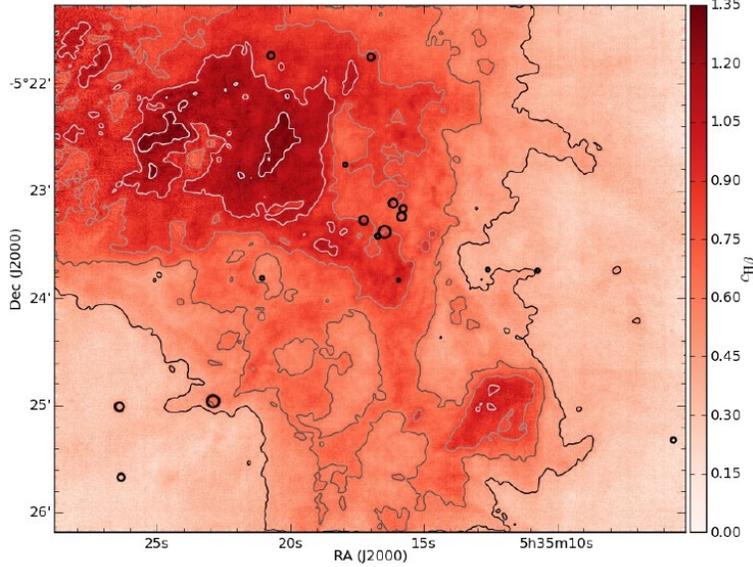


Figure 7: extinction toward M 42, from MUSE observations of the Balmer decrement δ , expressed as $c_{H\beta} = \tau_{H\beta} / \ln(10)$. The full range of $c_{H\beta}$ corresponds to $A_V = 0.9 - 2.7$ mag.

This process is illustrated well by **nomograms**: contours of line-intensity ratios, calculated with Equations (26) and (24), plotted in $n_e - T$ space. Two example nomograms appear in Figure 8 and Figure 9. With n_e on the horizontal axis, intensity ratios with nearly vertical contours are good for measuring n_e nearly independently of T , as with the visible [S II] 672 nm/673 nm ratio and especially with the far-infrared [O III] 51.82 μm /88.35 μm ratio, for $n_e = 10^2 \text{ cm}^{-3} - 10^4 \text{ cm}^{-3}$. Intensity ratios with nearly horizontal contours are good for measuring T nearly independently of n_e , as with [S II] 1032 nm/671.7 nm at low densities and [O III] 436.3 nm/500.7 nm up to $n_e \sim 10^5 \text{ cm}^{-3}$.

If a species has just three states in which all the population resides to good approximation, such as the 3P states of O^{++} at densities below about 10^4 cm^{-3} (Figure 6), the three equations to which Equation (23) reduces have an easy algebraic solution for electron density ²²:

$$n_e = \frac{\gamma_{01}(A_{21} + A_{20}) + \gamma_{02}(A_{21} + rA_{10})}{\gamma_{01}(r\gamma_{12} - \gamma_{21} - \gamma_{20}) + \gamma_{02}(r[\gamma_{12} + \gamma_{10}] - \gamma_{21})}, \quad (27)$$

²² Storey, J.W.V., Watson, D.M. & Townes, C.H. 1979, [ApJ](#), 233, 109.

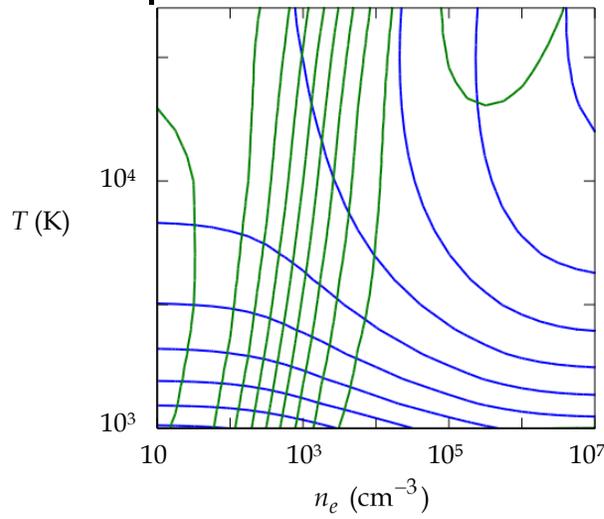


Figure 8: contours of logarithms of [S II] line-intensity ratios, 671.7 nm/673.1 nm in green and 1032 nm/671.7 nm in blue, plotted against n_e and T . From lower left to upper right, contour levels are 0.15 to -0.3 in steps of -0.05, and -6 to 2 in steps of 1, respectively.

where $r = I_{10}\lambda_{10}A_{21}/I_{21}\lambda_{21}A_{10}$. Observations of forbidden lines in lowest-energy 3P terms are useful in this regard, despite being available only with relatively poor angular resolution.

The MUSE observations of M 42 ⁶ have been reduced in this fashion, in essence, by iteration of the following steps. From Equation (26) the authors determined the upper-state population fractions f_j and their uncertainties for pairs of lines with nearly vertical nomogram contours. Then, by solution to Equations (23), in the same manner as to produce Figure 6, they determined density n_e , and the uncertainty therein. Repeating the process for line pairs of the same species with nearly horizontal nomogram yields the temperature T . In the MUSE data set the authors found a comforting close similarity in the density and temperature results for line pairs from species with different abundance and ionization state, which may be taken to indicate that the approximation of a uniform 1-D slab for each pixel is pretty good.

Sample results for density and temperature from the MUSE data analysis appear in Figure 10 and Figure 11.

- c. **Determine ionic and elemental relative abundances.** Once you know extinction, density, and temperature, you know everything but the column densities in Equation (26). The relative abundances are the ratios of column density, usually of an ionic line

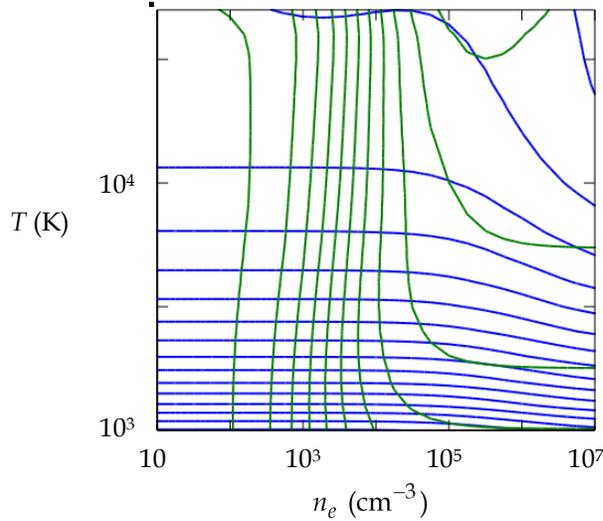


Figure 9: contours of logarithms of [O III] line-intensity ratios, $51.82 \mu\text{m}/88.35 \mu\text{m}$ in green and $436.3 \text{ nm}/500.7 \text{ nm}$ in blue, plotted against n_e and T . From lower left to upper right, contour levels are 1 to 11 in steps of 1, and -14 to -1 in steps of 1, respectively.

to a hydrogen recombination line. If the observations yield relative abundances for all significant ionization stages of a given element, you can obtain the relative abundances of that element. One of the cleanest examples of this sort of result is also one of the most astrophysically important: oxygen. O has an ionization potential very close to hydrogen's, 13.6 eV, so it is photoionized wherever hydrogen is. O^{++} has an ionization potential of 54.9 eV, too large even for hot O stars to produce abundantly in photon form. So images of the column densities N_{O^+} , $N_{\text{O}^{++}}$, and $N_p = N_{\text{H}^+}$ cover these species throughout most H II regions. Their ratios, O^+/H^+ and O^{++}/H^+ for short, are useful in constraining the properties of the O star which gives rise to the ionizing ultraviolet; this can be especially useful when the star suffers greater extinction than the spectral line emitting regions. The sum $\text{O}/\text{H} = \text{O}^+/\text{H}^+ + \text{O}^{++}/\text{H}^+$ is one of the most accurate oxygen relative abundances available.

Figure 12 and Figure 13 are MUSE images ⁶ of the relative ionic and elemental abundances of oxygen and sulfur in M 42. Note that the higher ionization states tend to be more abundant closer to the four Trapezium stars, θ^1 Ori A-D, rather than the O stars on the other side of the bright bar, θ^2 Ori A-C. Thus none of the three of θ^2 Ori A-C play a significant role in ionizing M 42. Note also that the elemental abundance images exhibit much less variation across M 42 than the ionic abundances.

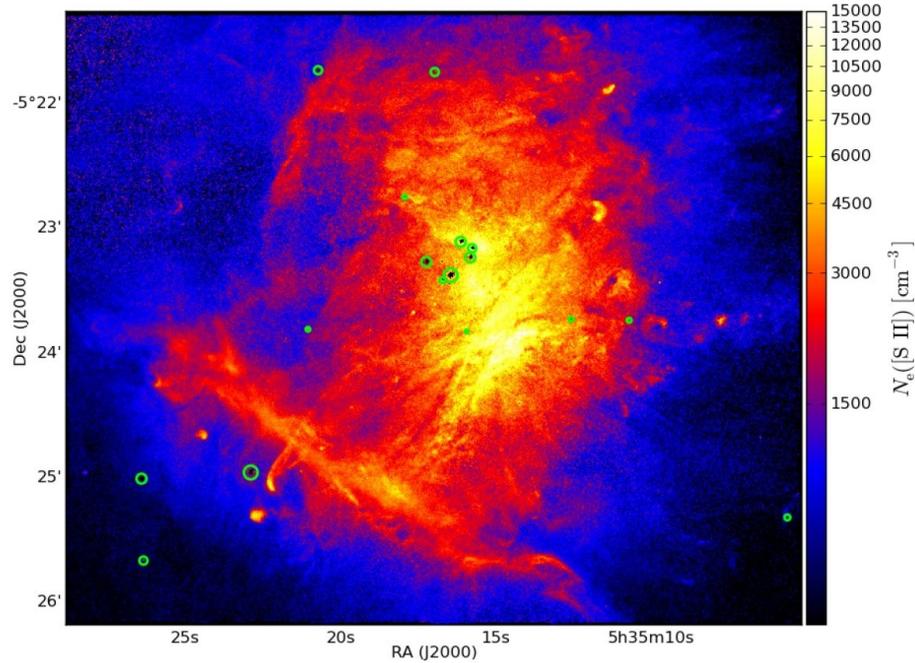


Figure 10: image of electron density in M 42 ⁶, derived from images of [S II] 671.7 nm and 673.1 nm.

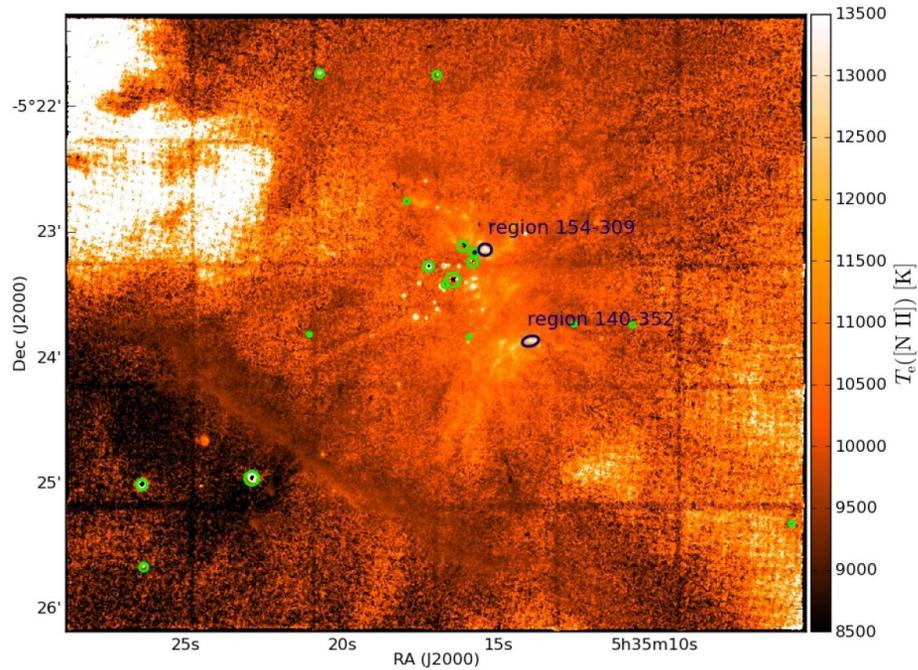


Figure 11: image of temperature in M 42 ⁶, derived from images of [N II] 575.5 nm and 654.8 nm. N⁺ and O⁺⁺ are isoelectronic; these lines are the [N II] counterparts of [O III]'s 436.3 nm and 500.7 nm lines.

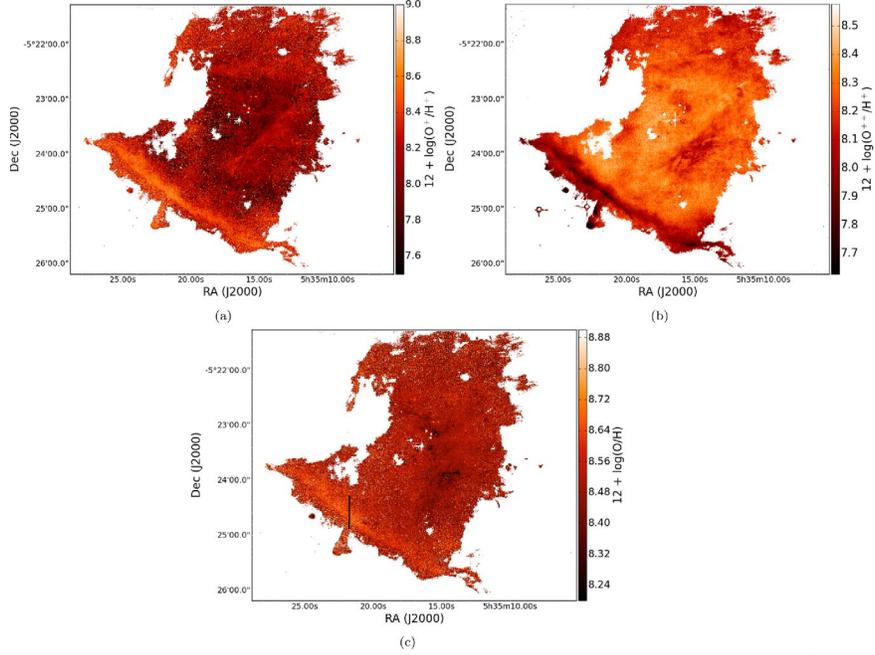


Figure 12: MUSE images ⁶ of logarithms of oxygen abundance, from [O II], [O III], and H α after extinction correction, density and temperature determination. O⁺/H⁺ at upper left, O⁺⁺/H⁺ at upper right, O/H below.

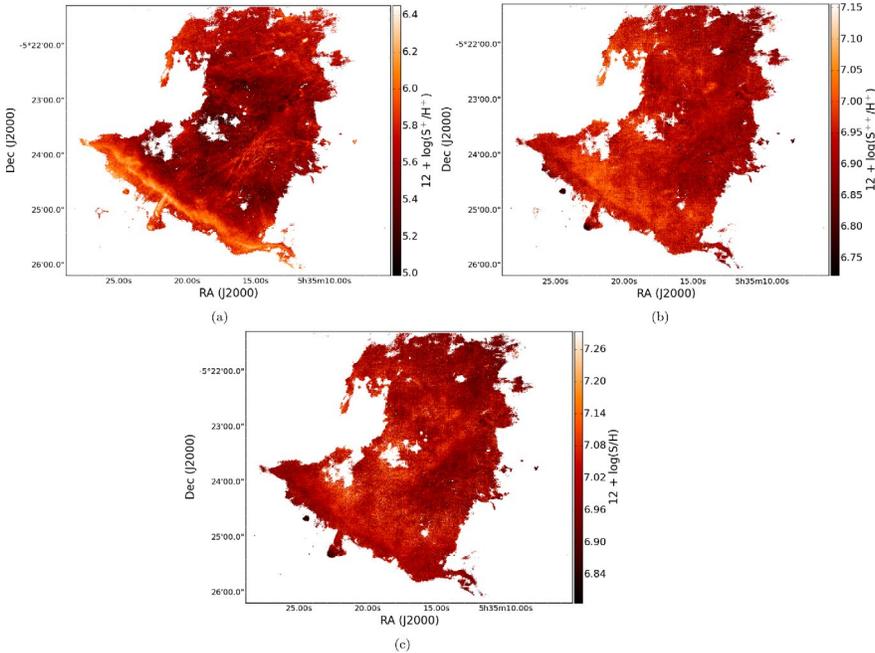


Figure 13: MUSE images ⁶ of logarithms of sulfur abundance, from [S II], [S III], and H α after extinction correction, density and temperature determination. S⁺/H⁺ at upper left, S⁺⁺/H⁺ at upper right, S/H below.

10.2. More complex, 1-D and 3-D models

Nothing prevents you from using other things you know about your target to generate more complex models – except, maybe, for the amorphous shape of most bright H II regions. 3-D modelling goes better in compact H II regions – those found deeply embedded in molecular clouds, created by the youngest OB stars – and in planetary nebulae, both of which often appear at least axisymmetric, if not actually spherical. In either case, the procedure would be to dissect the structure into small blocks, within which the temperature, density, and ion abundances can be taken to be uniform and given by your model. Then one can use Equations (24) to find the population fractions, and Equations (25) to generate the spectral line intensities. Add extinction if you like. Embed all this in a routine in which n_e , T , and abundances are free parameters, to fit to observations.

There is also a virtue in considering a medium that differs from uniform by a small perturbation. Peimbert found ²³ that fits to H II region spectra improved when he included temperature fluctuations in a uniform medium. He characterized these fluctuations in terms of a temperature determination from a given line intensity ratio like [O III] 436.3 nm/500.7 nm by modelling the intensity-to-temperature conversion with a power law, and then expanding in a Taylor series about the mean temperature. The first-order term, in this scheme, is typically 10-20% of the lowest order, meaning that this is the amplitude of the temperature perturbation. This worked well in the days beam sizes tens of arcsec in diameter. Nowadays at least the components of these fluctuations in the plane of the sky are fully resolved in nearby H II regions, so the procedure is probably not as beneficial as it used to be, but is still an arrow to keep in one's quiver for modelling more distant regions in this and other galaxies.

11. CHIANTI and Cloudy

Once one understands the basic principles elaborated above, modellers of H II regions and planetary nebulae – you are one, now – have two widely-used resources to reduce the work of assembling the necessary atomic-physical parameters and generating model H II region spectra.

CHIANTI ²⁴ is a comprehensive database of wavelengths, energies, A-coefficients, and collisional strengths calculated over a useful range of temperatures. The site includes all the atomic data, bundled into uniformly-formatted files, with links to the original publications of the calculations. It also includes code written in your choice of IDL or Python, to aid in accurate computation of upward and downward collisional rate coefficients from the collision strengths, and to make simple calculations of emitted

²³ Peimbert, M. 1967, [ApJ, 150, 825](#).

²⁴ <https://www.chiantidatabase.org/>; Dere, K.P., et al. 2019, [ApJS, 241, 2](#).

spectra. A couple of versions ago, CHIANTI was adopted as the atomic-physics source for two industry-standard nebular codes, MAPPINGS for 1-D shocks and Cloudy for photoionized regions. One might quibble with minutiae such as the calculations for e -Fe⁺ collisions the CHIANTI-masters chose to include in the database, but in most respects the data are reliable. It saves an awful lot of work to have this standardized access to atomic data.

Cloudy ²⁵, invented by and with development still led by Gary Ferland, is a C++-based tool for synthesizing spectra of photoionized regions. It is one of the oldest and best documented openly-accessible simulation codes in astrophysics, and is nicely interfaced to the leading database sources for atomic physics and stellar atmospheres. One of the most widely-used graduate-level astronomy textbooks, Osterbrock's & Ferland's *[Astrophysics of gaseous nebulae and active galactic nuclei](#)* (2nd edition, 2006, a.k.a. AGN3) can be regarded as Cloudy's user manual. It takes a while to become an expert in Cloudy, but one can get good results long before rising to black-belt status. Highly recommended for anybody who deals with any of the subjects covered in AGN3, which is probably most of us.

²⁵ <https://www.nublado.org/>; Ferland, G.J., et al. 2017, *RMxAA*, 53, 385.