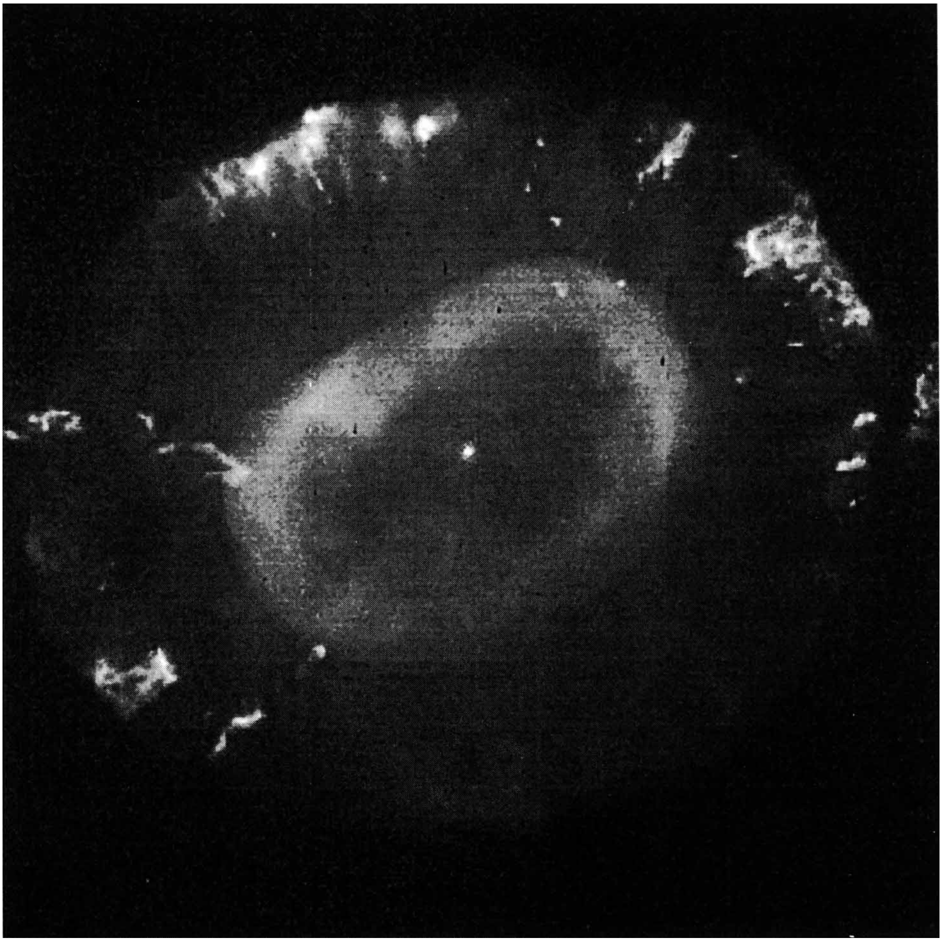
I. INTRODUCTION

https://doi.org/10.1017/S0074180900129651 Published online by Cambridge University Press



NGC 7662 observed with the Planetary Camera of the Hubble Space Telescope in the light of [N II]658.3nm. In order of distance from the *stellar nucleus*, the main features are the sharp-edged elliptical *bright rim*, the smooth surrounding *shell*, and the pair of low-ionization *FLIER clusters* seen between 1-3 and 7-9 o'clock. Additional microstructures found near the edge of the shell between 10-12 o'clock do not have the characteristic high velocities of FLIERs, and hence are denoted "LIERs".
Credit: *Narrowband HST images of Microstructures in Planetary Nebulae*,
Bruce Balick, J. Alexander, A. Hajian, Yervant Terzian, M. Perinotto, and P. Patriarchi.

https://doi.org/10.1017/S0074180900129651 Published online by Cambridge University Press

EARLY DAYS OF PLANETARY NEBULAR THEORY

LAWRENCE H. ALLER
UCLA Physics and Astronomy-Astronomy Division

The spectra of gaseous nebulae differed strikingly from those of stars which were well understood since 1922 thanks to the work of Saha. Gaseous nebulae exhibited exotic bright line spectra characterized by strange emissions of unknown origin as well as familiar lines of hydrogen and helium. The strongest lines in most nebulae fell at 4959A and 5007A. They were originally attributed to an unknown element. First came the interpretation of the lines of H.

The primary mechanism for the excitation of H lines is photoionization from the ground level followed by recombination on all levels. Zanstra [1926] realized that in an optically thick nebula the number of quanta in the Balmer series equaled the number of quanta emitted by the central star beyond the Lyman limit. Astronomers therefore had a unique and very powerful method for finding the temperature of the central star. You compared the flux in the star at the wavelengths of the Balmer lines with the nebular Balmer line fluxes. Thus, Zanstra was able to show that these stars were very hot, indeed. Some of them had temperatures higher than those proposed for any then-known stars.

Zanstra's theory embodied the principle that in an optically thick nebula the number of transitions 1-n must equal the number of transitions n-l. Since the recombination rates were known for each level and the transition probabilities, n-n', likewise, one could set up the equations of statistical equilibrium, solve for the population in each level, n, and derive the relative intensities of the Balmer lines (see e.g. Menzel and Baker (1937). Plaskett (1928), Carroll (1930), and Cillie (1932) tried such calculations, but the number of levels employed was too small. The first realistic calculations were those of Baker and Menzel (1938). Indeed, when accurate observations were obtained and theory further improved, the fit with Zanstra's theory was good. Another point was that it was a nuisance to measure both star and nebula for many lines. Now one had to measure the star and nebula at one line, usually H_β, and employ the theory of the Balmer decrement.

H. J. Habing and H. J. G. L. M. Lamers (eds.), Planetary Nebulae, 3–9.

Furthermore for very hot stars one could use ionized helium and obtain both H and HeII Zanstra temperatures.

The strongest lines in most nebulae were the green lines at 4959A and 5007A, sometimes attributed to an unknown element called nebulium. There were many other unidentifiable lines which did not behave in any coherent fashion from nebula to nebula. The fundamental argument against nebulium was there was no place for it among light and/or abundant elements. Russell concluded that these lines must be due to some familiar element radiating under unusual conditions, an insight which was soon confirmed by Bowen (1928).

In his studies of the spectrum and energy levels of O^{++}, Bowen found significant clues from the spacing of the levels of the ground p^3 configuration, ${}^3P_{2,1,0}$, ${}^1D_2$, and ${}^1S_0$. The 4959, 5007 lines corresponded exactly to ${}^1D_2$ - ${}^3P_1$, and ${}^1D_2$ - ${}^3P_2$ respectively. The ${}^1S_0$ - ${}^1D_2$ transition corresponded to yet another line found in nebular spectra. These lines violated the Laporte parity rule; hence they had low transition probabilities. They attained great strength in nebulae because of the immense size of these objects. Bowen realized they appeared because the electrons liberated by the photo-ionization of H had enough energy to excite the ${}^1S$ and ${}^1D$ levels of the ground configuration (2.5 to 5 ev but not enough to excite the levels responsible for the permitted lines). Thus, the dice were loaded against the lines of the ordinarily observed spectrum. Lines produced by the primary mechanism, i.e. photo-ionization and recombination, would be of the order of ten thousand times weaker. With such odds, it is no wonder that nebular spectra are exotic.

Additional forbidden lines were soon identified, especially by Menzel, Payne, and Boyce and by Swings. There was a strong motivation for theoretical studies; these involved an examination of physical processes on one hand and on atomic parameters on the other. Equations for a steady state and energy equilibrium (Menzel et al. 1938) led to a relation between the radiation field and the electron temperature for a pure hydrogen nebula with no collisional effects (Baker et al. 1938). The electron temperature lags slowly behind that corresponding to the radiation field, reaching 57,000K when the former reaches 80,000K. When the theory is extended to include the dissipation of energy in the forbidden lines, the results are drastically modified (Menzel and Aller 1941a). No matter how hot the exciting star, the electron temperature is limited to less than 20,000K when allowance is made for improved collisional cross-sections. Spitzer later used an energy balance argument for just the continuum to come out with closely similar results.

Forbidden lines contain a wealth of information pertaining to the diagnostics and ionic concentrations. To utilize the intensities of these lines

https://doi.org/10.1017/S0074180900129651 Published online by Cambridge University Press

we must know the A values for the transitions involved as well as the target areas for the collisional excitation of the levels. Forbidden lines include magnetic dipole transitions which depend only on the angular properties of the wave functions and the deviation from LS coupling. Electric quadrupole transitions require a knowledge of the radial wave function, R(r), as well, and were more difficult to obtain. Early calculations were superseded by the calculations of Pasternack, (1940) and of Shortley et al. (1941) who tabulated the A-values for magnetic dipole transitions, or coefficients of radial quantum integrals for electric quadrupole transitions as a function of a parameter which measures the deviation from LS coupling.

In low density nebulae, the ratio of the ^1S - ^1D (auroral type transition) to ^1D - ^3P (nebular type transitions) depends in the first approximation on the electron temperature for most ions. We have to know the collisional cross-sections for jumps between the ^1S, ^1D and ^3P levels. Hebb and Menzel (1940) calculated the collisional cross-sections between the levels of the p^2 ground configuration of OIII, while Menzel et al. (1941) determined the electron temperatures of a number of planetary nebulae wherein the line intensities were measured by photographic photometry. The temperatures ranged from 6,000 to 10,000K with most of them falling near 8,000K. Similar calculations were carried out for N^+ and O^+.

Unfortunately, these early cross sections were erroneous. They violated what is called the Bohr-Peierls-Placek theorem and were systematically too large. The derived electron temperatures were too low and the ionic densities were too small. Happily, Seaton (1953, 1954) rectified the deficiency and supplied reliable cross-sections for a large number of ions. Since the auroral/nebular line ratio depends on both temperature and density, especially for N^+, by comparing the [NII] and [OIII] lines, one could obtain both electron temperature and density provided they were produced in exactly the same strata. Monochromatic images show that for most planetaries [NII] and [OIII] originate primarily in different strata. For most p^2 and p^4 ions the auroral to nebular line ratio is sensitive to the electron temperature but NOT to the density. Seaton's clarification of the collisional cross section problem was the decisive break-through which made possible accurate calculations of electron densities and temperatures and ionic concentrations.

For a p^3 configuration such as O^+ or S^+, it turned out that both electron temperature and density can be found for the same ion. Pasternack's A-values for the transition from the $^2D_{3/2}$ and $^2D_{5/2}$ levels to the ground $^4S_{3/2}$ level did not give the correct intensity ratios for the 3726 and 3729 nebular type lines of O^+. The answer was found in a suggestion by Van Vleck that one should take into account the second order spin orbit effect and the interaction of the spin of one electron with the orbit of another.

https://doi.org/10.1017/S0074180900129651 Published online by Cambridge University Press

This gave a 3726/3729 ratio much more nearly in accord with observations. Aller, Ufford, and Van Vleck (1949), who made the detailed calculations and comparison with the observations, noted that the agreement was best for the high density nebula IC 4997IC 4997 and poorest for low density objects like NGC 40NGC 40. If the target areas for the collisional excitation of the ^2D term from the ground ^4S term could be calculated, we would have a clue to the density in the region where the [OII] lines are formed. Both the temperature and the density could be obtained by comparing the auroral type transitions at 7320,7330 with the nebular "3727" pair when the necessary target areas were found and after observed intensities were corrected for effects of interstellar extinction. Subsequently [SII], [ClIII], and [ArIV] were handled in a similar fashion, but calculation of the necessary wave functions is difficult. Densities found by the forbidden line method often exceed those estimated from the surface brightness method, suggesting that there are density fluctuations on a scale smaller than optical resolution.

The distances of individual galactic planetary nebulae are difficult to establish accurately. We shall not explore the matter here. One popular method is to use a relation between the surface brightness in H_β, the angular size, the mass of the ionized shell and the distance. For example, Minkowski and Aller (1954) used this approach to set limits on the distance of the Owl nebulaOwl nebula NGC 3587NGC 3587 and concluded it was of little value since we did not know the mass of the ionized shell. Shklovsky, (1956) by assuming a fixed mass for the ionized shell, set up a distance scale. Expansion rates of the nebular shell and spectroscopic parallaxes of the central star may be used to get nebular distances.

By 1945, most of the ground work had been laid for a quantitative analysis of a gaseous nebular spectrum. Recombination mechanisms seemed to be understood at least for H and He. The theory of collisional excitation of forbidden lines was quantified; what remained was the quantum mechanical calculation of collision strengths, a problem which was first solved by Seaton in the early fifties. An attempt was made to assess the chemical composition of planetaries by using the Pasternack A-values, Hebb-Menzel cross-sections or values estimated from the same on the advice of Massey, and the then-available observational data (Aller and Menzel 1945). The contribution of unobserved stages of ionization was made by empirical procedures. To within the glaring uncertainties, the nebulae did not seem to differ significally from what was then known of stellar chemical composition. The physical theory was incomplete in that no account was taken of charge exchange and dielectronic recombination. When these refinements were added, and appropriate recombination coefficients and target areas for collisional excitation and improved observational material became available, the compositions of planetary nebulae were well enough established to show that significant differences existed from one object to another and often between them and ordinary oxygen-rich stars.

The original observational data with which theoretical predictions had to be compared were all obtained by photographic photometry or by calibrated eye estimates as in the comprehensive survey of nebular spectra by Bowen and Wyse (1939) and by Wyse (1942). The obvious inadequacies of the photographic plate, in particular its non-linear response to light intensity, made it especially inadequate for nebular spectrophotometry where the measurable lines range over a factor of a thousand in intensity. Thus the line intensities in NGC 7027NGC 7027 by Aller, Bowen, & Minkowski (1955) have a serious scale error arising from this cause. Photo-electric photometry was applied to the strong lines, 5007, 4959 and 4851 by MacRae and Stock (1954) and by Liller and Aller (1954), but for the weaker lines such as 4363 [OIII] one still had to rely on photographic photometry. Only with the development of the spectrum scanner in the latter part of the fifties, so that all important diagnostic lines could be measured photoelectrically, was it possible to obtain accurate electron temperatures. Spectrum scanners had poor spectral resolution. A judicious combination of scanner data and high dispersion photographic spectrophotometry proved useful for some purposes until the development of charge-coupled devices rendered the photographic plate nearly obsolete.

The spectra of high excitation planetaries show a number of strong OIII lines, which result from cascade from the highly excited 2p3d $^3P_2$ level. Bowen (1935) noticed the wavelength of the transition from this level to the $2p^2\ ^3P_2$ level of the ground configuration almost coincided with the 303.780A Lyman alpha of ionized helium. The lines appeared in nebulae with strong HeII 4686. The subsequent cascading of the atoms through levels of the 2p3p and 2p3s configurations produced the strong OIII lines observed in these planetaries.

The evidence in favor of the Bowen fluorescent mechanism was overwhelming. Not only did one observe the predicted lines and none others (with then available equipment) but all the expected lines were present. Furthermore, their relative intensities were in harmony with expectation within the errors of observation and theory, Menzel and Aller (1941b). Transition probabilities were estimated by Slater's rules (1930). By assuming all 2p3s and 2p3d levels were populated by cascade, it was possible to compute the relative intensities of all Bowen lines and show them to be consistent with then available data.

The plasma of a gaseous nebula departs far from thermodynamic equilibrium, but it had generally been assumed that the velocity distribution of the electrons was closely Maxwellian, although it had not been rigorously proven. Hagihara (1941) suggested that the velocity distribution may

https://doi.org/10.1017/S0074180900129651 Published online by Cambridge University Press

depart substantially from a Maxwellian. If so, much of the work on the physical theory of gaseous nebulae would be vitiated.

Processes acting to destroy a Mawellian distribution are recombination which selectively favors the slower electrons, free-free emissions, and collisional excitation of low-lying levels, mostly metastable ones, and to a lesser degree superelastic ones, The main process acting to restore a Maxwellian distribution is collisions between electrons. Electrostatic forces reshuffle kinetic energy in a random fashion. The energies of the electrons are redistributed about once a second by these encounters. On the other hand an inelastic collision with an ion occurs on the order of perhaps once a month at nebular densities. The interval between the photo-ionization of an H atom and the recapture of an electron is of the order of ten years. Thus deviations of the electron velocity distribution from the Maxwellian distribution is utterly negligible! (Bohm & Aller 1947)

The early theoretical efforts to understand the physical processes occuring in gaseous nebulae may have blazed the trail but accurate calculations were required for recombination coefficients, transition probabilities, and collision strengths for ions such as those oF N, O, Ne, Cl, S. Ar. The latter computations turned out to be particularly troublesome because of resonance effects. In addition, attention had to be paid to certain processes overlooked in the earlier work such as charge exchange and dielectronic recombination.

Furthermore, planetary nebulae were regarded as static structures; attention to dynamical effects such as shock waves came later as did the realization of the relationship between these objects and stellar evolution, a driving concept in modern work in this subject. None of these advances would have been possible without the extension of the observable spectral range to the ultraviolet, infra-red, and radio-frequency ranges. Advances in detector technology, image converters, the Lallemand electronic camera, the image tube scanner, and the CCD provided an increase in spectral range, sensitivity, and accuracy. With adaptive optics, direct images of high spatial resolution are obtainable.

The planetary nebulae enthusiasts of the thirties and forties were working in a field that seem detatched from most of astronomy. Who could have anticipated what a central role in stellar evolution these objects were destined to occupy.

References

Aller L.H. and Menzel D.H. 1945, ApJ. 102, 239.
Aller L.H., Ufford W., and VanVleck J.H. 1949, ApJ. 109, 42.
Baker J.G. and Menzel D.H. 1938, ApJ. 88, 52.
Baker J.G., Menzel D.H., and Aller L.H. 1938, ApJ. 88, 422.

https://doi.org/10.1017/S0074180900129651 Published online by Cambridge University Press

Bowen I.S. 1928, ApJ., 67, 1935, ApJ. 88, 115.
Bowen I.S. and Wyse A.B. 1939, Lick Obs. Bull 19, 1.
Bohm D. and Aller L.H. 1947, ApJ. 105, 1.
Carroll J.A. 1930, MNRAS. 90, 588.
Cillie G. 1932, MNRAS. 92, 820.
Hagihara Y. 1941, Tokyo Astr. Bull. 542, 571.
Hebb M.H. and Menzel D.H. 1940, ApJ. 92, 408.
Liller W. and Aller L.H. 1954, ApJ. 120, 48.
MacRae D. and Stock J. 1954, Nature 173, 589.
Menzel D.H. and Aller L.H. 1941a, ApJ. 94, 30 1941b, ApJ. 94, 438.
Menzel D.H. and Baker J.G. 1937, ApJ. 86, 70.
Menzel D.H, Aller L.H., and Baker,J.G. 1938, ApJ. 88, 313.
Minkowski R. and Aller L.H. 1954, ApJ. 120, 261.
Pasternack S. 1940, ApJ. 92, 129.
Plaskett H.H. 1928, Harvard Circular #335.
Seaton M.J. 1953, Ann d'Ap. 17, 74 1954, MNRAS. 118, 154.
Shklovsky J.S. 1956, AJ. (U.S.S.R) 33, 222, 315.
Shortley G., Aller L.H., Baker J.G., and Menzel D.H. 1941, ApJ. 98, 178.
Slater J. 1930, Phys.Rev. 57, 36.
Stevenson A.F. 1932, Proc.Roy.Soc. Ser A, 137, 298.
Wyse A.B. 1942, ApJ. 95, 356.
Zanstra H. 1926, Phys.Rev.(2) 27, 64.

https://doi.org/10.1017/S0074180900129651 Published online by Cambridge University Press