PLANEARY NEBULAE: NEW CHALLENGES IN THE 21ST CENTURY

SUN KWOK
Institute of Astronomy and Astrophysics, Academia Sinica, Taipei, Taiwan
E-mail: kwok@asiaa.sinica.edu.tw
(Received February 1, 2005; Accepted March 15, 2005)

ABSTRACT

Although planetary nebulae (PNe) have been discovered for over 200 years, it was not until 30 years ago that we arrived at a basic understanding of their origin and evolution. Even today, with observations covering the entire electromagnetic spectrum from radio to X-ray, there are still many unanswered questions on their structure and morphology. In this review, we summarize recent theoretical and observational advances in PNe research, and discuss the roles of PNe in the chemical (atomic, molecular, and solid-state) enrichment of the galaxy and as tracers of the large scale structure of the Universe.

Key words: ISM: planetary nebulae: general — binaries:symbiotic— stars: AGB and post-AGB

1. INTRODUCTION

Planetary nebulae (PNe) represent a short (~10^4 yr) phase of stellar evolution between the asymptotic giant branch (AGB) and white dwarfs. The central stars of PNe are remnants of the electron-degenerate C-O cores of their AGB progenitors, having lost most of their H envelopes due to mass loss on the AGB. They maintain their energy output through H-shell burning and evolve with constant luminosity from low to high temperature across the H-R diagram. When their entire H envelope is consumed by a combination of nuclear burning and mass loss, their luminosities begin to decrease and the central star gradually cool to become white dwarfs. This basic scenario was outlined by Paczynski (1971), and was confirmed by detailed evolutionary tracks calculations by Schönbberger (1979), Iben (1984), and Wood & Faulkner (1986).

Although the ionized regions of PNe are easily observable due to strong recombination lines of H and He and collisionally excited lines of metals in the optical region, recent infrared and mm/submm observations have revealed the presence of neutral material in the form of molecules and dust. We now know that PNe possess the ionized, neutral atomic, molecular, and solid state forms of matter, and are contain regions of widely different densities, temperatures and morphological structures. A summary of our modern observational view and theoretical understanding of the PNe phenomenon can be found in Kwok (2000).

(a) What Stars Will Become PNe?

Since PNe represent material ejected during the AGB phase and later ionized by the increasingly hot central star, the existence of PNe requires that the dynamical (expansion) age of the nebula be comparable

to the evolutionary age of the central star. A star with higher core mass leaves the AGB with a smaller H envelope but burns at a higher rate. As a result, the transition time between the AGB and PN stage is highly dependent on the core mass. While a high-mass star will evolve so fast that it only illuminates the nebula for a very short time, a star with low core mass will evolve too slowly to ionize the circumstellar nebula before it dissipates into the ISM. Consequently, not all stars that evolve past the AGB will become PNe.

Due to strong mass loss on the AGB, stars with main-sequence masses as high as 8 $M_\odot$ can lose their entire H envelope before the core mass reaches the Chandrasekhar limit of 1.4 $M_\odot$. From the initial mass function, one can estimate that 95% of all stars that have evolved from the main sequence during the lifetime of the Galaxy will end their lives as white dwarfs. A majority (30-90%, Drilling and Schönberner 1985) of these stars will have passed through the PN stage. From evolutionary models, the minimum core mass required for PNe is 0.58 $M_\odot$. This translates to a main-sequence mass of 1.5 $M_\odot$, based on current initial mass-final mass relationship (Weidemann 2000). Although these values are still uncertain, we believe that all Popoulation I stars with initial masses between 1.5 and 8 $M_\odot$ will become PNe.

(b) PN in Different Galactic Environments

PNe are traditionally discovered by their spectral characteristics in emission-line surveys or by their morphology in imaging surveys. The current number of known PNe in the Galaxy is approximately 2000 (Parker et al. 2003). Due to galactic extinction and incompleteness of surveys, the total PNe population is much higher. Possible missing in catalogues are large, nearby, low-surface-brightness PNe, or PNe suffering from large amount of circumstellar extinction. For example, recent infrared survey of the Galactic plane by the Spitzer Space Telescope (SST) have identified some very red PNe which are faint in the visible (Fig. 2). Based on the observed local density of PNe and an assumed scale height, the total population in the Galaxy is estimated to be about 14,000. However, this number is highly uncertain due to the inaccurate distances of local PNe. Dividing this total population by the lifetime of PNe, one can estimate that the birth rate in the Galaxy is $\sim$ 1 PN per year.

Since PNe are the result of AGB mass loss, the birth rate of PNe could be a function of metallicity and may differ in different galactic environments. Narrow-band imaging in the [OIII] line has identified PNe in galaxies as far away as the Virgo cluster, and spectroscopic observations have shown that PNe in external galaxies (e.g. Cen A) are similar to Galactic PNe (Walsh et al. 1999). High-resolution imaging by the HST of PNe in the Magellanic Clouds found that they have the same kinds of morphology as Galactic PNe, suggesting that the PNe have undergone similar shaping processes. Extensive surveys of globular clusters have revealed only 4 PNe in globular clusters. This is not unexpected because globular clusters are old and consist of mainly low mass stars. The question why PNe (e.g. K648 in M15) are present at all has led to the suggestion that they are the result of merged binaries (Bond & Alves 2001). Since almost all of our theoretical understanding of PN evolution is based on a single star scenario, the role of binaries play in PN evolution is still underdeveloped.

II. DYNAMICAL EVOLUTION OF PN

While an impulsive ejection appears intuitively obvious for the formation of PNe, the discovery of large-scale mass loss on the AGB led to the realization of the importance of AGB mass loss on the origin of PNe and the formulation of the interacting stellar winds model (Kwok et al. 1978, Kwok 1982). In addition to the well-observed shell, the ISW model predicts the following components: a fast wind from the central star, a low-density halo representing the unshocked AGB wind, and a high-temperature bubble representing the shocked fast wind. With the launch of the IUE satellite in 1978, P Cygni profiles implying wind velocities of several thousands km s⁻¹ were found in many central stars of PNe, confirming that fast winds are indeed common in PNe. Evidence for the existence of haloes outside the PN shells were found by CCD imaging, and by the detections of dust and molecular envelopes by infrared and millimeter-wave observations. Extended diffuse X-ray emission from the hot bubble was detected by ROSAT (Leahy et al. 2000) and by CHANDRA observations (Kostner et al. 2001, Chu et al. 2001).

A PN is a dynamical system whose evolution is tightly coupled to the evolution of the central star through a changing rate of stellar Lyman continuum output and photoionization, and by a changing mass loss rate and wind velocity from the central star and wind interaction. The appearance and the structure of PN therefore reflect the coupled dynamical and ioniza-

Fig. 2.— A possible PN 329.6-00.4 observed by the GLIMPSE survey of the SST. The four images are 3.6 $\mu$m (top left), 4.5 $\mu$m (top right), 5.8 $\mu$m (bottom left), and 8 $\mu$m (bottom right).
tional evolution of the nebula. The time-dependent nature of PN evolution was incorporated into many of the 1-D spherically symmetric treatments of the ISW model (e.g., Schmid-Voigt & Köppen 1987a; b; Marten & Schönberner 1991; Frank 1994; Mellema 1994). These models reveal a complex structure of density, velocity, and temperature as the nebula evolves, leading to fresh interpretations of the observed properties of PNe, including mass, kinematics age, and abundances (Perinotto et al. 1998, 2004).

III. MORPHOLOGY OF PN

High dynamical range imaging observations of PNe with CCD detectors have revealed a much more complicated structure of PN than the classical picture of a shell plus central star. In addition to a bright shell, low surface brightness outer structures can be seen. These structures were labelled “Type I” and “Type II” outer shells by Chu et al. (1987), and “inner shell”, intermediate shell”, and “halo” by Guerrero et al. (2000). The most comprehensive description of PNe was given by Frank et al. (1990), who used terms “inner core”, “bright rim”, “shell”, “crown”, “edge”, “halo”, and “limb-brightened halo” to describe the morphological features observed. While such multiple shell structures are difficult to understand in the classical model, they can be reproduced when the evolution of the central star is incorporated in the interacting winds model (Mellema 1994, Steffen et al. 1998, Corradi et al. 2000). In a 1-D model, the “rim” can be identified as the high-density shell compressed by the hot bubble, the “shell” as the extent of the ionization front, and the halo as the remnant of the AGB wind (Schönberner et al. 2005). It is important to note that these structures can change with time as the star evolves. For example, “recombination haloes” can emerge as the stellar luminosity begins to decline (Tylenda 1983).

(a) Morphological Classification of PNe

Beginning with the work of Curtis (1918), there have been many attempts in classifying the morphologies of PNe (e.g., Stanghellini et al. 1993, Manchado et al. 1996). However, all classification schemes suffer from the following problems:

- Sensitivity dependence: a deeper exposure can reveal fainter structures which change the classification of the PNe. For example, the waist of a bipolar nebula could be classified as elliptical if the bipolar lobes are too faint to be detected. NGC 650-1, Sh 1-89, and SaWe 3 are some of the cases where their bipolar nature were discovered as the result of deep CCD imaging (Hua, 1997; Hua et al. 1998).
- Species dependence: the morphology of PNe observed in lines of different ions is not necessarily the same, as the result of ionization structures and stratification effects.
- Projection effects: morphology classifications describe the two-dimensional apparent structures, not the intrinsic structures of the PNe.

It is clear that the examination of the apparent morphology alone is not sufficient to obtain the true intrinsic structure of PN. Kinematic data are necessary to separate various components projected on the same positions in the sky.

(b) The Intrinsic Structure of PNe

There have been several efforts to account for the variety of PN morphologies by different views of a single, unified, basic three-dimensional structure. Khromov & Kohoutek (1968) explained the morphology of PN in terms of a open-ended cylinder projected onto the sky, and Masson (1989, 1990) and Aaquist & Kwok (1996) employed an ellipsoidal shell (ES) model in which the PN morphology is determined by an ellipsoidal shell with both radial and angular density gradients ionized by a central star to different depths in different directions. Simulated images for 110 PN were produced by Zhang & Kwok (1998) using the ES model.

In the first order, the 2-D structure of a PN consists of the following components:

- A low-density spherical halo representing a remnant of the AGB wind.
- An ionization-bounded, dust-obscured torus representing an equatorial outflow during the late stages of AGB evolution.
- Two density-bounded polar lobes representing cavities created by fast outflows and the subsequent photoionization of the circumstellar material.

Under this model, a PN viewed near pole on will appear elliptical, with the torus seen as a shell and the lobes seen as envelopes surrounding the shell. Detailed kinematic studies have shown that several of the well-observed ring-like PN are in fact bipolar (NGC 6720, Bryce et al. 1994; NGC 7027, Latter et al. 2000; NGC 3132, Monteiro et al. 2000). When viewed near edge on, the PN will appear to have a bipolar or butterfly shape. Although the fraction of bipolar PNe based on apparent morphology is relatively small (~15%), the true fraction could be much higher.

(c) Microstructures

Recent observations, in particular high-resolution images obtained with the HST, have discovered a number of microstructures beyond the basic bipolar structure.

- FLIESRs and jets: FLIESRs are pairs of small, bright knots of low excitation gas found along the major axes of PNe (Balick et al. 1998). Linear structures (jets) can be seen in two of the corners of the [NII] image of NGC 6543. The existence of...
these features suggests that the fast outflow could be collimated rather than spherical.

- Point-symmetric structures: point-symmetric pairs of knots in an S-shape structure, or sometimes referred to as bipolar, rotating, episodic jets (BRET), have been seen in a number of PNe (e.g., KpPn8, López et al. 1995; NGC 6884, Miranda et al. 1999). Some PNe have been found to have more than one polar axis, suggesting that the outflow direction has changed with time (e.g., NGC 2440, López et al. 1998; M1-37 and He2-47, Sahai 2000). Two examples of quadrupolar PNe are shown in Figs. 3 and 6.

- Collimated outflows:

  some PNe (e.g., M2-9, Schwarz et al. 1997) and PPNe (e.g., Hen 3-401, Sahai et al. 1999) have extreme bipolar (cylindrical) shapes, suggesting that their morphology is shaped by a collimated outflow (Fig. 4). The direct imaging of bipolar lobes emerging from a circumstellar disk in the PPN IRAS 17106–3046 (Kwok et al. 2000) suggests that disks could play a role in the collimation of the bipolar flows.

- Rings and arcs: concentric circular arcs have been observed in both PNe (e.g., Hb 5, NGC 6543, NGC 7027) and PPNe (e.g., AFGL 2688, IRAS 17150–3224, Fig. 5). These arcs are of almost perfectly circular in shape, and have relatively uniform separations of $\sim 10^2$ yr (Kwok et al. 2001b). Similar arcs have been detected in the carbon star IRC+10216, suggesting that these features originate in the AGB phase. The coexistence of these perfectly circular features with bipolar lobes suggests that the arcs are projections of undisturbed spherical shells on the sky. Possible mechanisms for the creation of such arcs include dynamical instability in the gas-dust coupling in the AGB outflow (Deguchi 1997), perturbation by a binary companion (Mastrodemos & Morris 1999), and magnetic cycle (Soker 2000, Garcia-Segura et al. 2001).

Two-dimensional rings perpendicular to the bipolar axis have been found in several PNe, including MyCn18 (Sahai et al. 1999b) and NGC 6881 (Fig. 6, Kwok & Su 2005). The origin of these rings is not understood.

(d) Shaping of PNe and the Origin of the Asymmetry

While the ISW model has been shown to be able to amplify the asymmetry of the AGB wind to create the variety of morphologies of PN (Kahn & West 1985, Balick 1987, Mellema & Frank 1995), the origin of the asymmetry remains to be identified. Millimeter interferometric observations of the molecular envelopes of AGB stars have always found the envelopes to be spherically symmetric, suggesting the AGB mass loss

Fig. 3.—Two pairs of bipolar lobes can be seen in the [N ii] image of NGC 2440 taken at the CFHT.

Fig. 4.—Top: He2-320 is an example of PNe with highly collimated lobes. Bottom: A disk can clearly be seen in scattered light in the PPN IRAS 17106–3046.
over the dynamical lifetime (~ $10^4$ yr) is approximately spherical. Recent imaging of proto-planetary nebulae (PPNe) have found many PPNe to possess bipolar morphology, implying that the shaping process occurs during the PPN phase, before the photoionization of the nebulae (Kwok et al. 1996, 1998). Observations of PPNe therefore hold the key of understanding of the shaping process.

Some PPNe and young PNe have highly focused lobes, suggesting that the fast wind is being collimated by a disk (Fig. 4). Spectroscopic monitoring has found a number of post-AGB stars and PNe to possess binary central stars (van Winckel et al. 1999, De Marco et al. 2004). To what extent binary central stars play a role in the shaping and collimation process is one of the most difficult questions we face.

Hydrodynamic models have shown that FLLERs and BRETs could be naturally produced by the ISW process if the mass loss rate and velocity of the fast wind are functions of both time and direction (Stefen et al. 2001). The change in outflow direction could be the result of rotation and magnetic fields (Garcia-Segovia et al. 1999).

IV. WHAT IS A PLANETARY NEBULA?

The possibility that binary central stars can play a role in the shaping of PNe has raised the questions whether some of the bipolar nebulae that we observed are actually PNe, or they are symbiotic stars. Before we can answer this question, we have to first give a precise definition to PNe.

Observationally, PNe are traditionally defined as optical nebulae with a central star with some degree of symmetry and a strong emission-line spectrum with no or very weak continuum. This definition does not take into account modern observational properties such as dust emission in the infrared, X-ray continuum from hot gas, emission lines from molecules, etc. PNe have a complex physical environment, ranging from temperature of ~ $10^4$ K in the molecular and dust region, ~ $10^4$ K in the ionized region, to > $10^5$ K in the hot bubble, and radiate throughout the electromagnetic spectrum from radio to X-ray (Fig. 1). However, even an expanded observational definition is not enough because symbiotic stars share many of these same properties.

It is clear that we need to incorporate a theoretical element into our definition. For example, we can define PNe as ionized circumstellar shells showing some degree of symmetry surrounding a hot, compact star evolving from the AGB to the white dwarf phase (Kwok 2000). This definition will distinguish PNe from symbiotic stars or novae, which are binary systems undergoing mass exchange.

While the hydrogen envelope mass of the a PN central star is being depleted by nuclear burning and mass loss, and therefore constantly evolving, a symbiotic star maintains its energy source through accretion, and is
therefore stationary in evolution (Paczyński & Rudak 1980). It should be noted, however, the a post-outburst nova or symbiotic nova also evolves to the blue similar to a PN. Unless the outburst (II ignition) is observed, it may also be difficult to distinguish a symbiotic nova from a PN.

Although PNe and symbiotic stars share many common observational properties (optical emission lines and radio continuum emissions due to photoionized gas, infrared emission due to dust, etc.), they can be distinguished observationally because symbiotic stars have higher excitation lines as the result of accretion, periodic photometric variability due to the pulsation of the cool component, and molecular absorption features produced in the cool star’s atmosphere. This, however, raises some difficulty for cases of PNe that have binary central stars and high-excitation lines. For example, the central star of the bipolar PN NGC 6302 has a G-type companion. It has very high excitation lines ([Mg II], [Mg III], etc.), which, if arise from a photoionized gas, will require a very hot central star. Or is it possible that these high-excitation lines are the result of accretion, which could mean that NGC 6302 is a symbiotic star. This example illustrates the difficulty of classification, as indicated by the many other examples (M2-9, OH231.8+4.2, etc).

We should note that PN is a more fundamental phenomenon because single stars can evolve to PNe. The compact component of a symbiotic star has probably gone through the PN phase and is now a white dwarf. When the cool component becomes an AGB star, wind accretion can cause the white dwarf to become a symbiotic nova, creating a circumstellar nebula through photoionization and the interacting winds process. After the symbiotic phase, the cool component can also evolve to a PN.

Is the fraction with asymmetric nebulosity higher in symbiotic stars than in PNe? Are the most extreme examples of asymmetric PNe really symbiotic stars? The answers to these questions may be relevant to the origin of asymmetry in PNe.

V. CHEMICAL ENRICHMENT OF THE ISM

Since PNe are descendents of thermal-pulsing AGB stars, they are long known to play a major role in the chemical enrichment of the ISM. As emission-line objects, PNe serve a natural laboratory for the observations of C, N, and many s-process elements. The recent availability of ultraviolet and infrared spectroscopic observations through the FUSE and ISO missions has resulted in an increasing degree of sophistication and accuracy in the determination of the abundance of these elements. The development of millimeter-wave spectroscopy has led to the detection of rotational transitions of over 50 molecules (including carbon chains as large as HCHN and cyclic molecules such as C3H2) in the circumstellar envelopes of AGB stars and PNe, suggesting that these objects are also major sources of molecular enrichment of the ISM. The detection of HCO+ in PNe demonstrates the importance of photochemistry and shows that gas-phase chemistry is still actively ongoing in the molecular envelopes of PNe (Hasegawa et al. 2000, Hasegawa & Kwok 2001). Infrared spectroscopy from ISO has made possible the observation of stretching and bending modes of molecules, leading to the detection a variety of new species including diacetylene, triacetylene and benzene in PPNe (Cernicharo et al. 2001a,b). These results clearly show that the synthesis of organic molecules are occuring through the AGB-PN transition.

Solid-state compounds in the form of amorphous silicates and silicon carbides are commonly found in AGB stars, and these grains are expected to survive through the PNe stage (Kwok 1982). However, recent infrared observations have revealed that, in addition to these species, the dust envelopes of PPNe and PNe contain other inorganic and organic compounds (Fig. 7). The infrared spectra of PNe show strong emission features at 3.3, 6.2, 7.7, and 11.3 μm due to aromatic hydrocarbons, and aliphatic features at 3.4 and 6.9 μm are seen in the spectra of PPNe. The broad emission plateaus at 8 and 12 μm have also been identified as due to a variety of aliphatic sidegroups attached to aromatic rings (Kwok et al. 2001a). There are also emission features at 21 and 30 μm which origins are yet to be identified. These discoveries, together with the detection of crystalline silicates in PNe, suggests that the solid-state component is also under change in the PNe environment (Kwok 2004).

![Fig. 7](ISO SWS01 spectrum of the PN CN1-1 showing the presence of crystalline silicate features as well as aromatic infrared bands at 3.3, 6.2, 7.7, 8.6 and 11.3 μm.]

There is now strong evidence that meteorites contain grains of presolar origin. The detection of silicon carbide in meteorites (Bernatowicz et al. 1987) clearly suggests that solid state materials made in AGB stars can pass through the ISM and survive the formation of the solar system. This was followed by the
detection of presolar grains of corundum and spinels (Nittler et al. 1997), which probably originate from oxygen-rich AGB stars. The similarity between the 3.4 μm features seen in meteorites and galactic dust also suggests a common origin due to CH₂ and CH₃ side groups (Pendleton 1999). The fact that many organic compounds including aliphatic and aromatic hydrocarbons, alcohols, amines, amides, and carbonyl compounds are found in meteorites raises the question what fraction of these organic compounds is stellar in origin.

VI. PN AS GALAXY PROBES

The technique of PN luminosity function (PNLF) pioneered by Jacoby (1989) and Ciardullo et al. (1989) has turned out to be one of the most robust method to determine extragalactic distances. Comparison with Cepheid distances shows that the two techniques yield excellent agreement, at least for large, metal-rich galaxies. Interestingly, the success of the PNLF imposes useful constraints on our understanding of the PN evolution, initial mass-final mass relationship, and the initial mass function of PN progenitors.

The utility of PNe as tracer of dark matter distribution in elliptical galaxies is now firmly established since the early work of Hui et al. (1995). Even for spiral galaxies where the existence of dark matter halos can be inferred from HI rotational curves, PNe can separate the mass contributions from the bulge, disk, and haloes.

The detection of intracluster PN (IPN, Arnaboldi et al. 1996) opened the possibility of using PNe to infer the dynamical history of the cluster. The detection of IPN also allows us to estimate the total intracluster stellar population. Since the fraction of matter in baryonic form is a significant parameter in any cosmological model, the observations of IPN has obvious implications on the origin of the Universe.

VII. SUMMARY

In spite of the long history of PNe as objects of astronomical observations, we are just beginning to understand their origin, structures, and evolution. As a short phase of stellar evolution linking the AGB and white dwarf stages, PNe (and PPNe) evolve over a factor of 10-100 in temperature and a factor a 100 in luminosity. PNe are unique among astrophysical objects in the richness of physical and chemical processes present. These processes lead to radiations throughout the electromagnetic spectrum, from radio to X-ray. They also serve as the ideal laboratory for the study of radiative, mechanical, and chemical interactions between stars and their environment.

The high-dynamic range images of PNe have revealed new details in the structures of PNe. The lessons we learn in the interpretation of these features by dynamical models will have significant implications in our understandings of Wolf-Rayet nebulae, supernovae, young stellar objects, and active galactic nuclei.

The chemical abundances in PNe reflect the history of nucleosynthesis on the AGB, as well as the galactic environment from which they are born. The study of PNe therefore provides important clues on both stellar and galactic evolution. As standard candles and dynamical tracers, PNe serve as useful tools for the determination of the large scale structure of the universe, the distribution of dark matter, and the measure of the total baryonic mass in the universe. The possibility that the organic content of meteorites may be linked to PNe ejecta suggests that PNe are relevant from the largest structure of the universe to the smallest bodies in the solar system.

ACKNOWLEDGEMENTS

I thank Kevin Volk for the ISO spectrum of CN1-1. The images of NGC 2440 and NGC 6881 were taken at the CFHT and HST, respectively in collaboration with Kate Y.L. Su.

I would like to thank the local organizing committee for a very enjoyable meeting. Research of S.K. is supported by the Academia Sinica and the National Science Council of Taiwan. S.K. also acknowledges support of the Natural Science and Engineering Research Council of Canada, the Canadian Space Agency, and the Canada Council for the Arts.

REFERENCES


© Korean Astronomical Society • Provided by the NASA Astrophysics Data System
Latter, W.B., Dayal, A., Bieging, J.H., Meakin, C., Hora, J.L., Kelly, D.M.
Paczynski, B. 1971, Acta Astr., 21, 417
Sahai, R. et al. 1999b, AJ, 118, 468