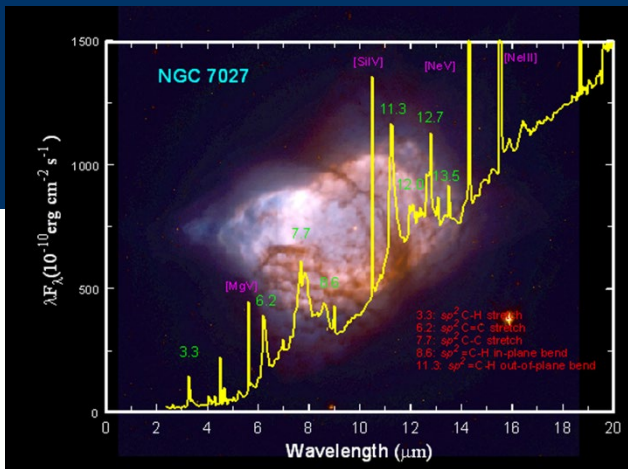


Planetary Nebulae Research: past, present, and future

Sun Kwok

The University of British Columbia

APRIM, Hong Kong
May 7, 2026



Early history

- 1764: **Messier** discovery of M27
- 1864: spectroscopy of NGC 6543 by **Huggins** a gaseous nebula, not a collection of stars
- 1922: **Hubble**'s correlation between magnitude of the central star and size of nebula → energy of nebula derived from central star



Beginning of astrophysics

- 1926: **Menzel**'s theory that the nebula is ionized by Lyman continuum photons from the central star
- 1927: **Zanstra**'s derivation of the temperature of the central star from the $H\beta$ flux $\rightarrow$ central star much hotter than normal stars
- 1928: **Bowen** identified the unidentified nebular lines as due to metastable states of N^+ , O^+ , and O^{++} .

Not nebulum, just ordinary matter under unfamiliar conditions

Identification of PN

- Discovery of new PN from extended nebular objects: imaging (Abell 1966, Kohoutek 1967) and objective prism surveys (Minkowski 1964, Henize 1967)
- Narrow band surveys: 3800 PN in Galaxy (Parker 2022)
- Atomic spectroscopy and nebular physics: photoionization, recombination lines of H and He, collisionally excited lines of metals, chemical abundance (Aller)
- PN in the context of stellar evolution

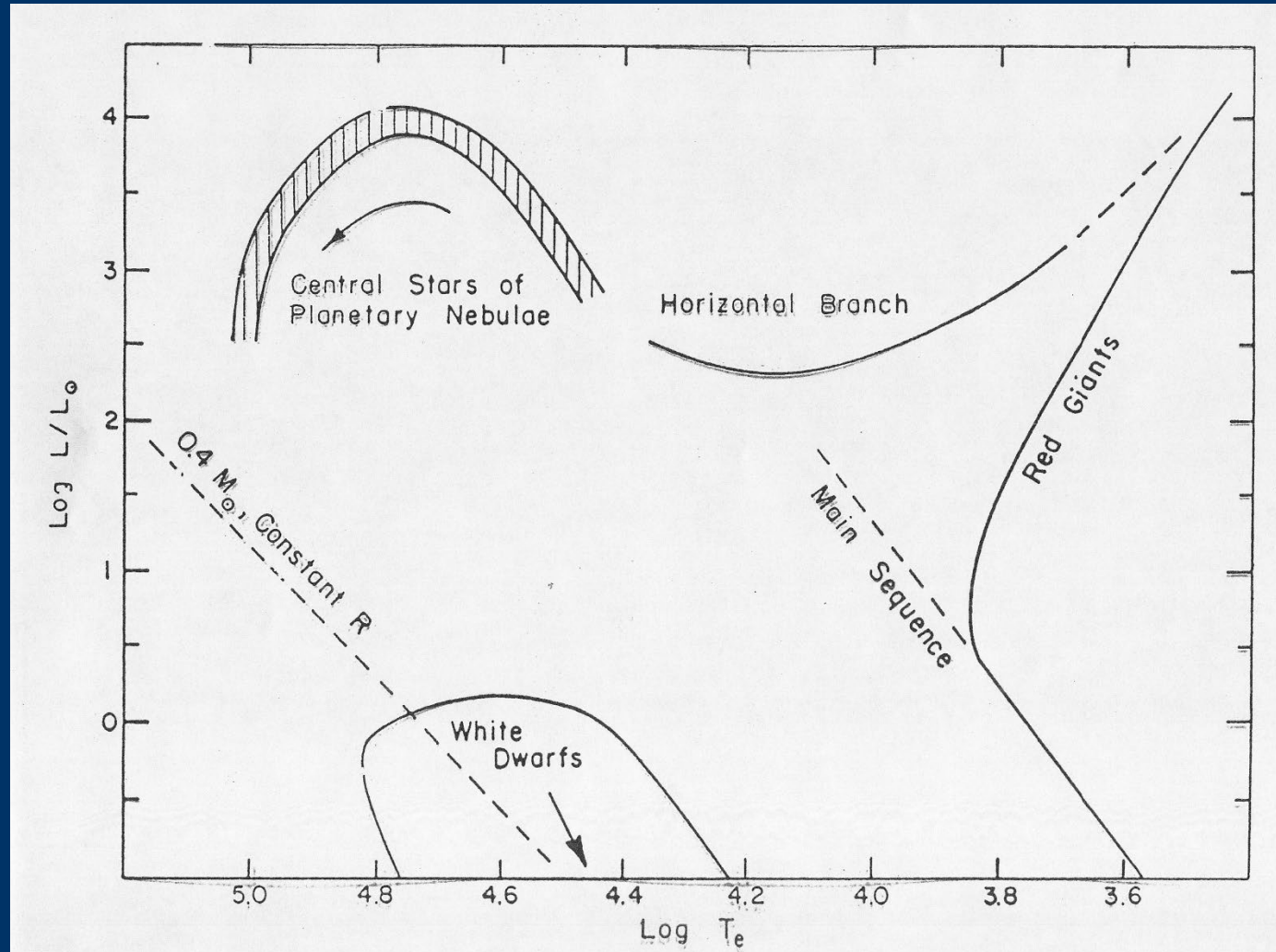
PN as a phase of stellar evolution

- In the early 20th century when stars were believed to evolve from high to low temperatures along the Main Sequence. Being hot, PN were thought to be young stars
- Galactic distribution of PN shows that PN belong to the old star population (Curtis 1918)
- PN are descendants of red giants and progenitors of white dwarfs (Shklovsky 1956)
- expansion velocities of PN similar to escape velocities of red giants and the nebulae are ejected envelopes of red giants (Abell and Goldreich 1966)

Placement of PN on the HR diagram

- Apparent magnitude of central stars
- Temperature from Zanstra method
- Distances from nebular fluxes using the Shklovsky method (assuming nebular mass of $0.2 M_{\odot}$)
- HR diagrams: O'Dell (1963), Harman & Seaton (1964)

The Harman-Seaton Sequence



See summary by Osterbrock 1973

Models of evolution

- Harman-Seaton sequence suggests that PN are formed at the end of the HB, followed by a rapid rise in temperature and luminosity (Roxburgh 1967, *Nature*, 215, 838).

“Theoretical models agree tolerably well with the observed data the overall properties of planetary nebulae are now well understood”

(Nature, September 30, 1967)



Cannot be more wrong

Reconciliation between dynamical and evolutionary timescales

Such a rapid change in temperature and luminosity within the time scale dictated by the dynamical age ($\sim 10^3$ yr) is theoretically difficult to achieve and early models all failed (Salpeter 1968, Deinzer & Hansen 1968, Kutter, Savedoff, Schuerman 1969, Shaviv 1978)

Problems with observed HR diagram

- Missing flux: leaking UV (nebulae not ionization bounded), emission in IR
- High temperature (30-300 K) central star often faint in visible
- Contamination by nebular emission
- Uncertain distance (Shklovsky method not reliable)

Wrong luminosity and temperature, observed tracks cannot be reproduced by any nuclear burning model

Rescued by a simple theory

- Paczynski (1971) started with a static model of putting H in the envelope above an electron-degenerate core
- He allows the star to evolve by reducing the mass in the H envelope.
- He found that the star will NOT evolve until the envelope mass is lower than $10^{-3} M_{\odot}$
- When it does move, the H envelope mass is so low that the core mass is unchanged

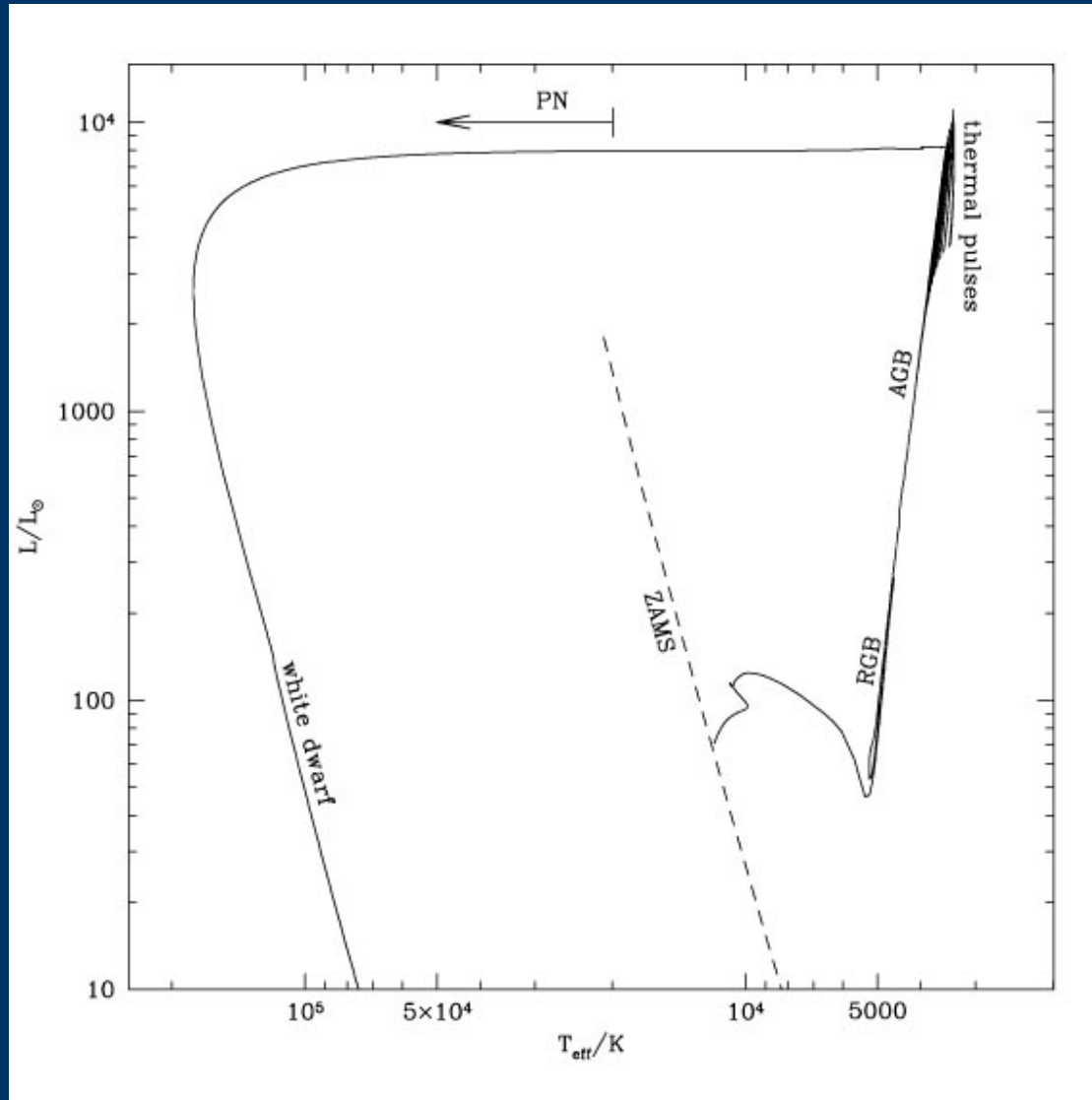
Star evolves at constant luminosity

A new understanding of PN evolution

- H shell burning on an electron degenerate C-O core
(Paczynski 1971)
- Increase in T_{eff} the result of thinning of H envelope due to H burning and mass loss
- Central star evolves at constant luminosity
- Evolutionary timescale highly sensitive to core mass

$$t \propto M_c^{-8}$$

H shell burning track



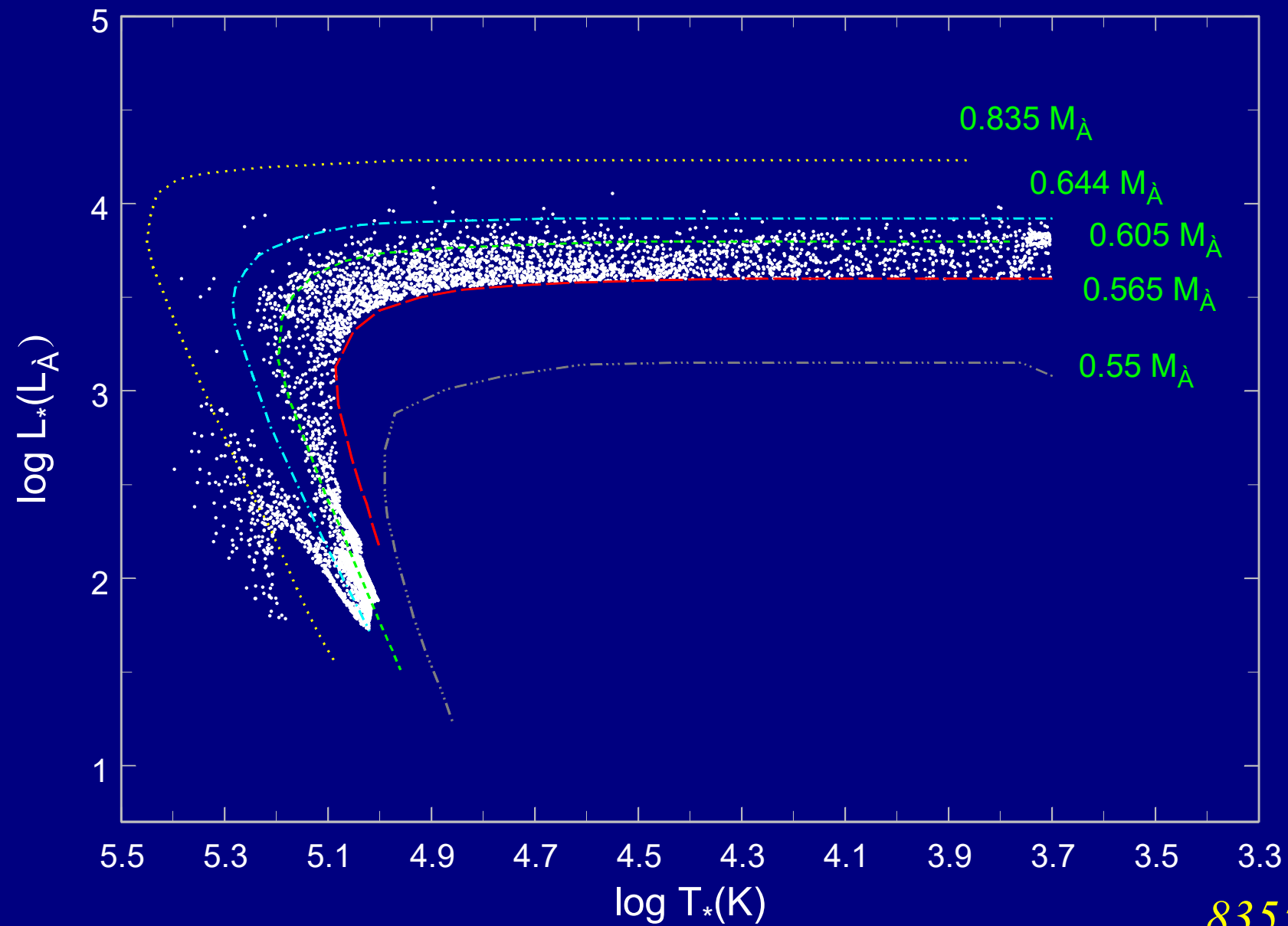
3 $M_{\odot}$ track by Blöcker

The existence of PN

- Existence of PN requires the dynamical age of the nebula be comparable to the evolutionary time of the central star
- Transition time highly dependent on core mass
- high mass PN: lifetime too short
- low mass PN: nebula dissipated before ionization

$$M_c \sim 0.6 M_{\odot}$$

A time dependent dynamical phenomenon



8355 stars

Formation of the nebula

by ejecting the H envelope

- Dynamical instabilities: Roxburgh 1967, Lucy 1967, Paczynski & Ziolkowski 1968
- Pulsational instabilities: Kutter & Sparks 1974, Wood 1974, Tuchman et al. 1979
- Envelope oscillations due to thermal instabilities in the core: Smith & Rose 1972
- Radiation pressure: Faulkner 1970, Finzi & Wolf 1971, Finzi et al. 1974
- Thermal pulses: Harm & Schwarzschild 1975, Trimble & Sackman 1978

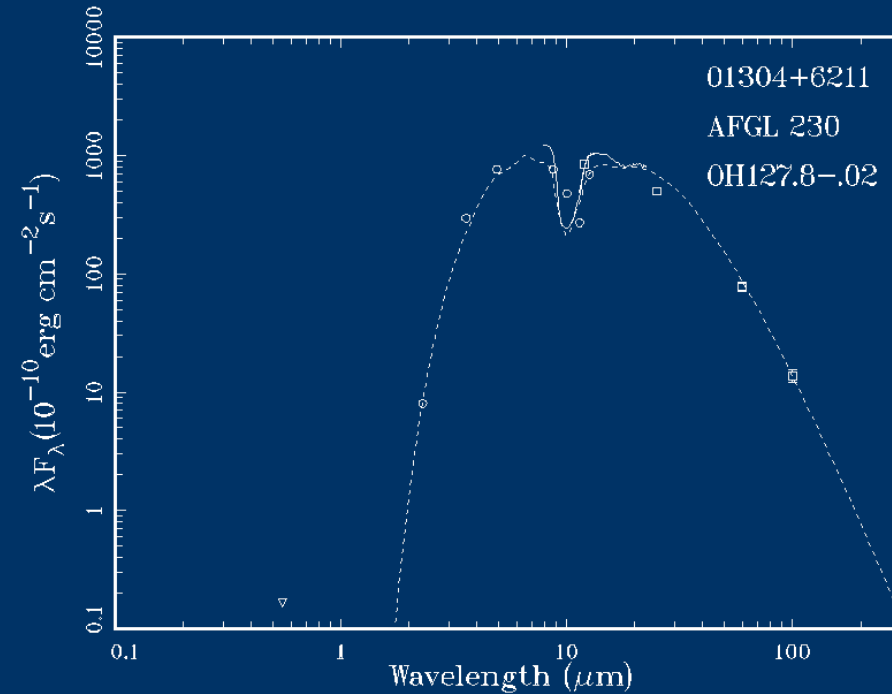
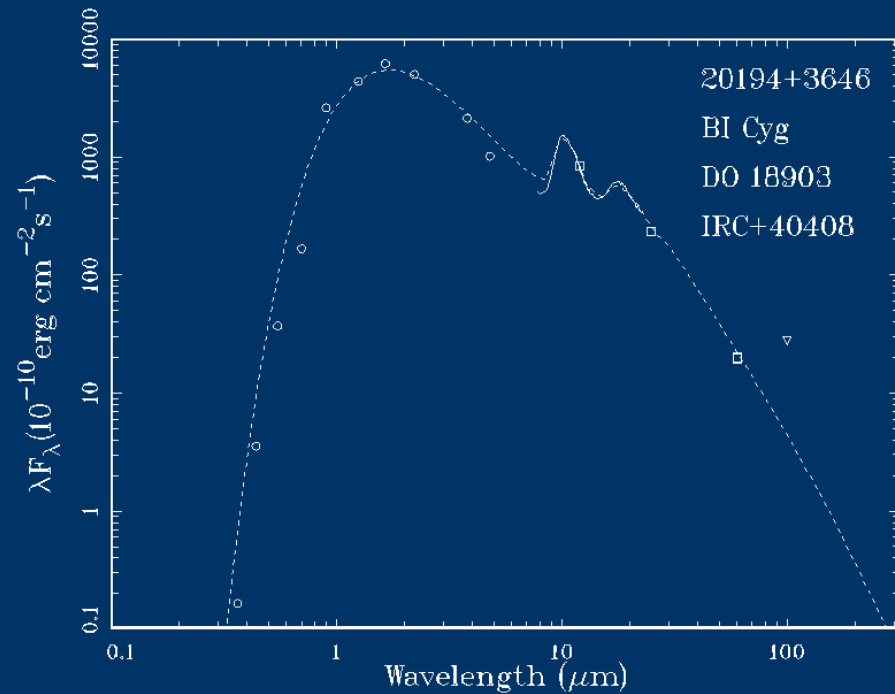
Difficulties with sudden ejection models

- Backfill: additional pressure is needed to maintain the expansion of the nebula and prevent it from falling back onto the central star (Capriotti 1973, Wentzel 1976, Ferch & Salpeter 1975, Mathews 1966)
- Photoionization heating (Sofia & Hunter 1967)

Too much or too little

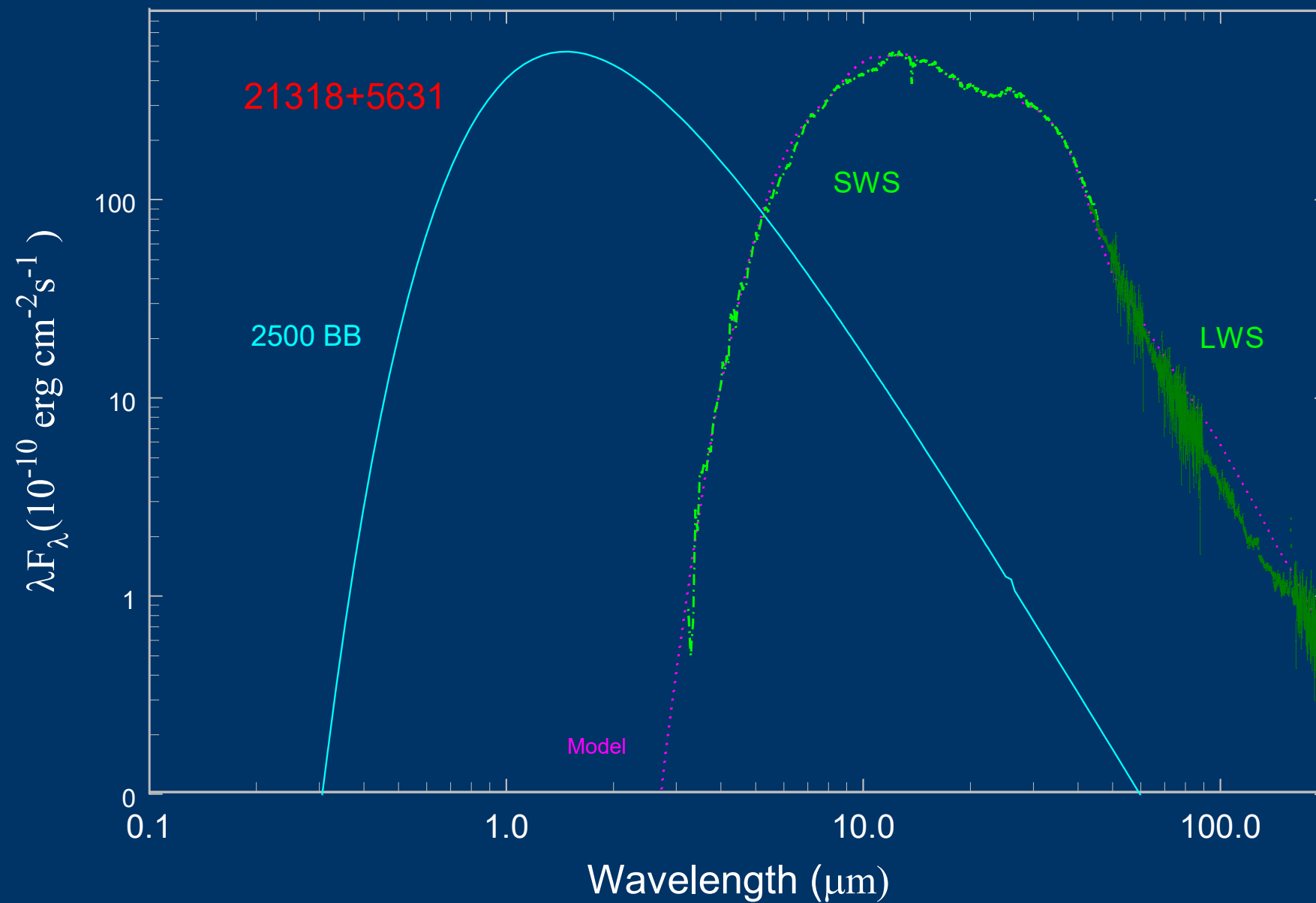
- Paczynski shows that the central star will not move until $M_e < 10^{-3} M_\odot$
- No sudden ejection theory can be this precise in the ejected mass
- If $0.199 M_\odot$ of $0.2 M_\odot$ is ejected, the star will remain red until the never-ionized ejected envelope has dispersed into the ISM.

Development of IR astronomy led to the discovery of mass loss on the AGB



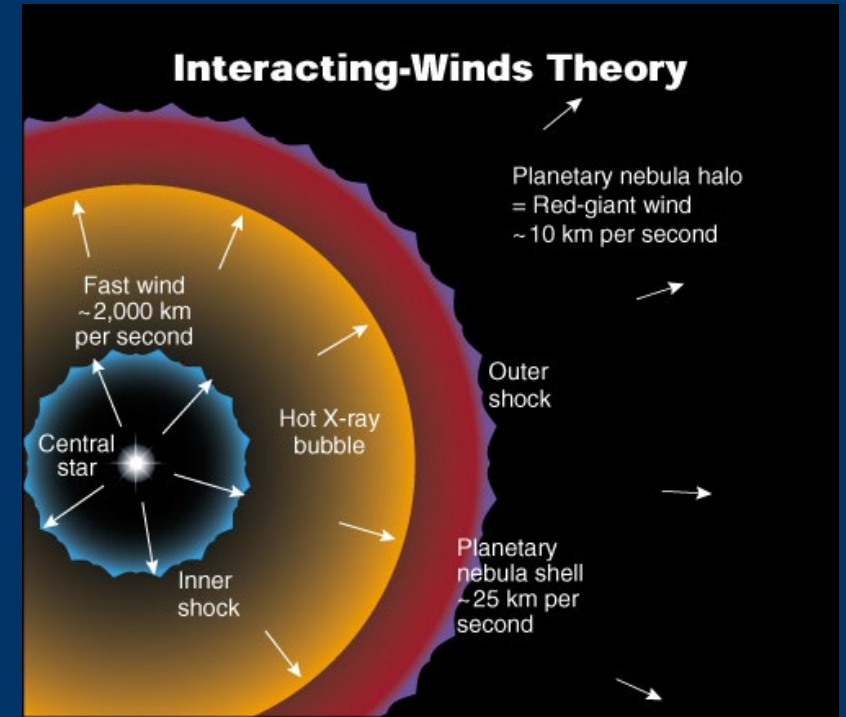
Thousands of AGB stars have been observed by IR and mm observations to lose mass slowly and steadily over several hundred thousand years

Circumstellar envelope completely obscures the central star



Interacting winds theory

- Remnant of circumstellar envelope ejected during the AGB phase must have an effect on the formation of PN
- Nebular mass provided by the AGB wind
- Expansion is driven by thermal pressure from the shocked fast wind



Kwok et al. 1978, Kwok 1982

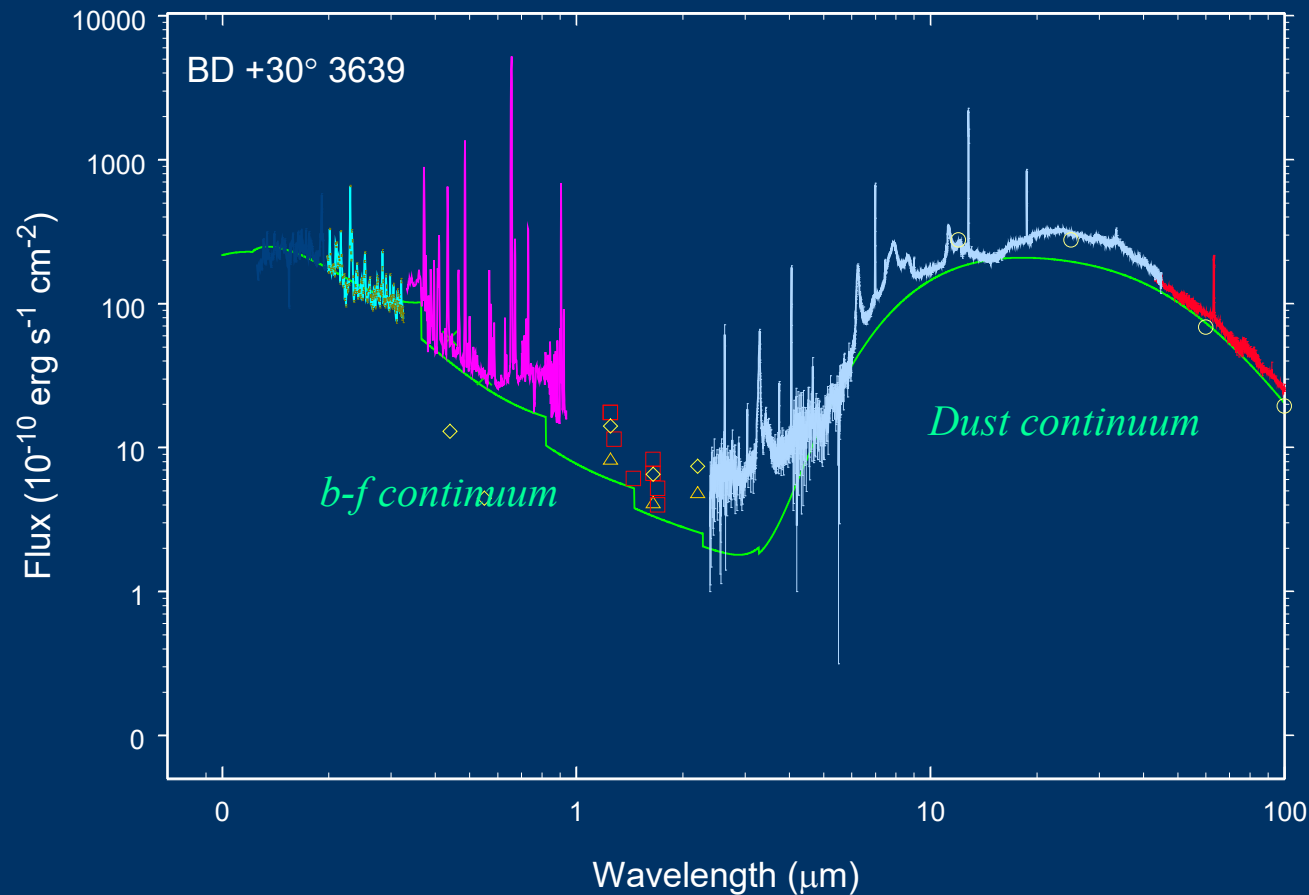
The slow wind depletes the H envelope, the fast wind compresses, shape, and accelerate the slow wind

Predictions of the ISW model

- Fast winds from central star (*IUE, FUSE*) (Heap et al. 1978)
- remnant dust envelope from AGB (*IRAS, ISO*)
- molecular envelopes (CO, HCN)
- Optical haloes (*CCD*) (Jewitt et al. 1986, Chu et al. 1987)
- x-ray bubbles (*ROSAT, Chandra*)

*Confirmed by new observations in the infrared,
submm, UV and X-ray*

Remnant AGB dust envelope in PN

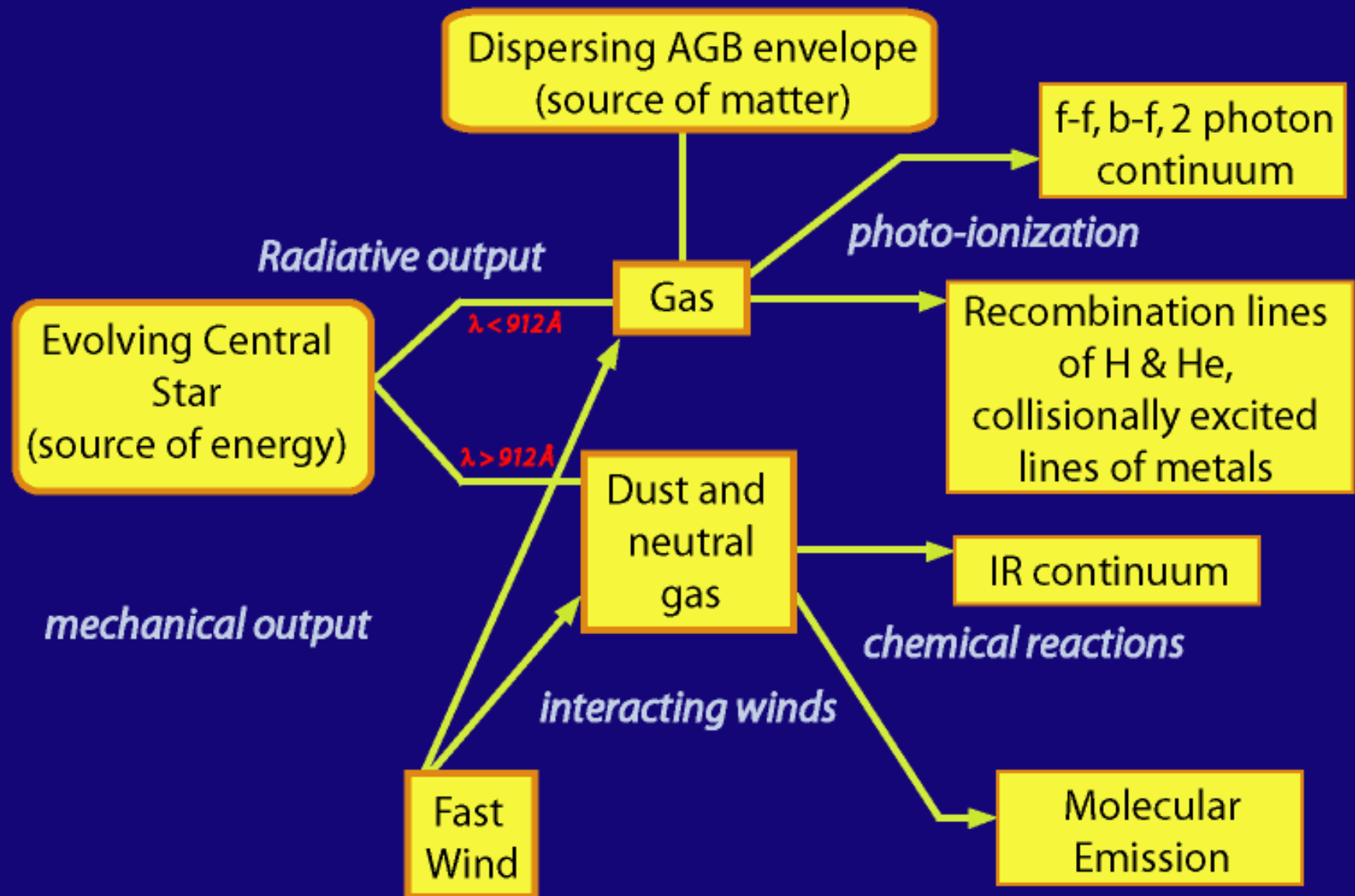


*In young PN, ~1/3 of
flux emitted in IR
(Zhang & Kwok
1991)*



PN is a dynamical system whose evolution is tightly coupled to the evolution of the central star

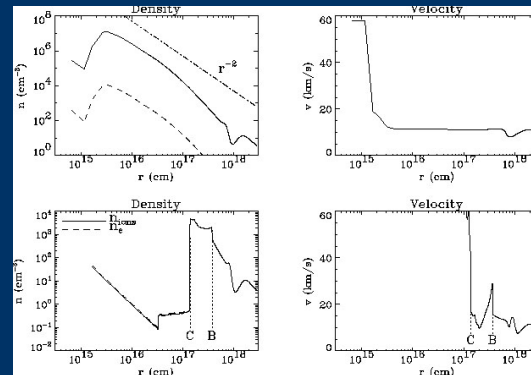
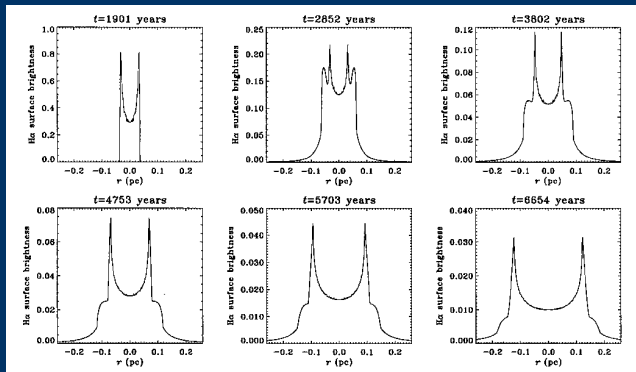
- $L_*(t)$, $T_*(t)$ H or He burning
 - $m(L_*, T_*)$, $v(L_*, T_*)$ mass loss mechanism
 - photoionization ($N_{\text{Ly cont}}(t)$)
 - dust continuum transfer
 - dynamics: interacting winds
 - $\Rightarrow n_e(r, t)$, $V_{\text{shell}}(r, t)$, $\Delta R/R$, etc.
- Schmidt-Voigt & Koppen (1987), Marten & Schonberner (1991), Frank (1994), Mellema (1994),
 - 2-D models: Kahn & West (1985), Balick (1987), Icke (1988), Soker & Livio (1989), Frank et al. (1993), Frank & Mellema (1994), Mellema & Frank (1995), Mellema (1997)



Crowns, haloes, and multiple shells



Can only be understood with a consistent treatment of dynamics and stellar evolution



Mellema (1994)

Steffen et al. (1998), Corradi et al. (2000)

Recombination haloes? (Tytenda)

Kinematic age not reliable (Schönberner)

Proto-planetary nebulae

- Objects in transition between AGB and PN stages
- Difficult to identify because the nebulae of PPN are neutral and do not have emission lines
- ~30 PPN were discovered from their IR properties as the result of follow up of IRAS survey (Kwok 1993, *Ann. Rev. Astr. Ap.*, 31, 63)

PPN as imaged by the HST



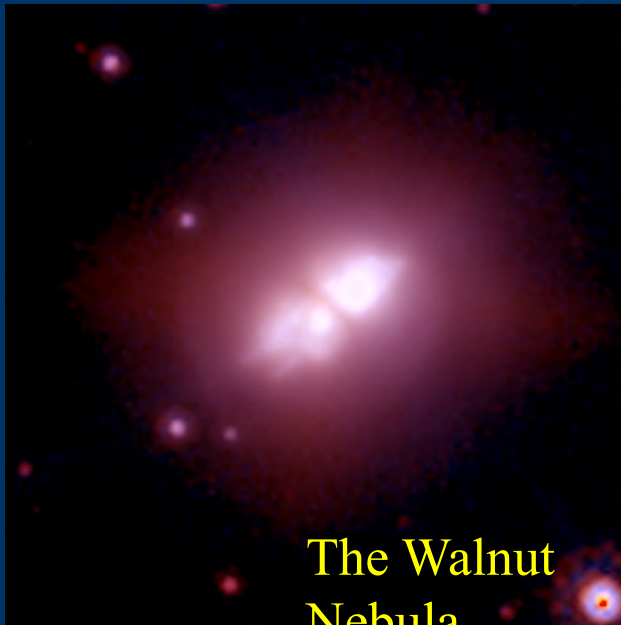
The
Silkworm
Nebula



The Water Lily Nebula



The Cotton
Candy
Nebula



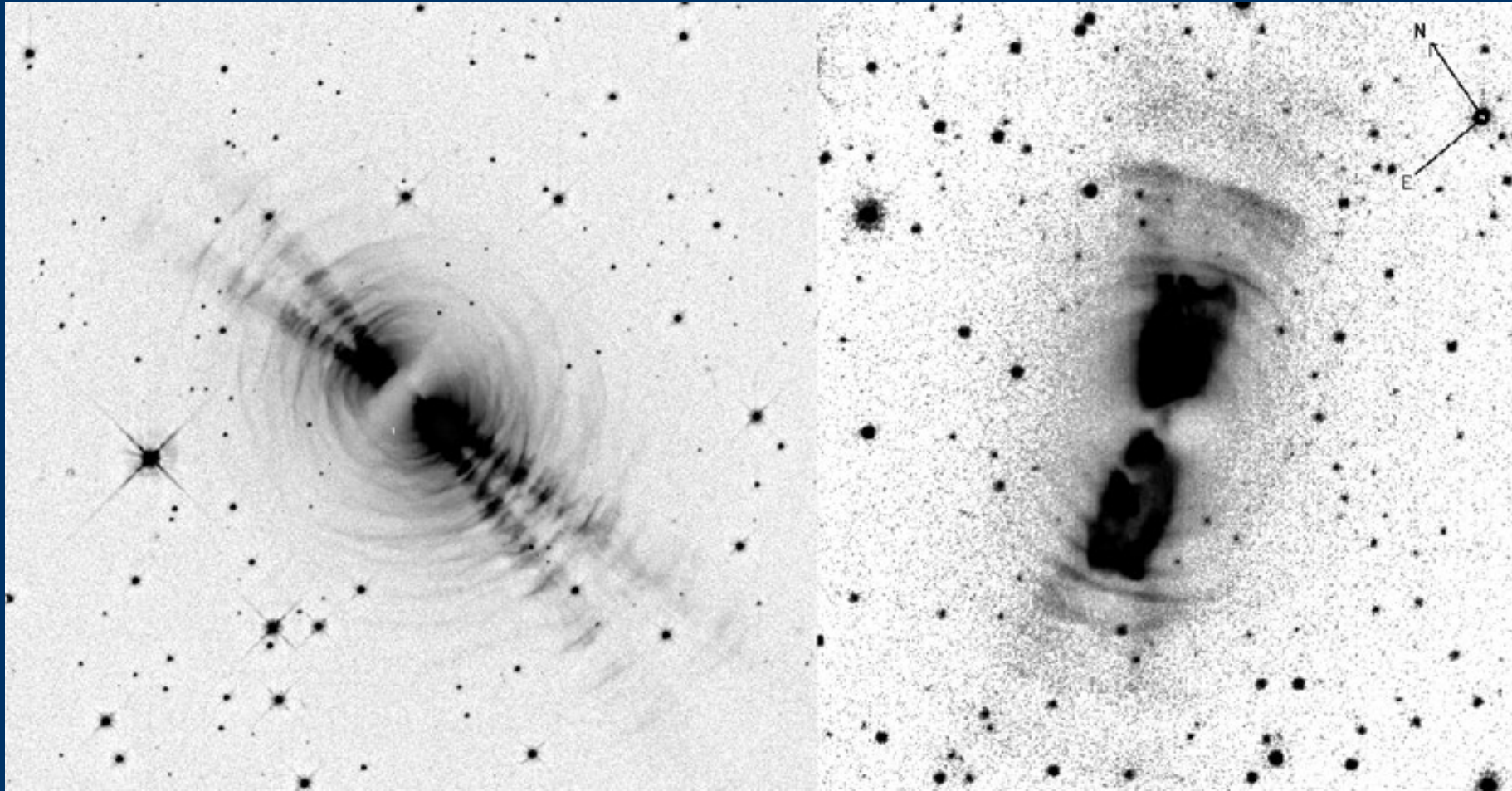
The Walnut
Nebula

Reflected starlight, not emission!



The Spindle Nebula

Concentric arcs



3-D thin shells projected onto the sky,

Kwok et al. 1998

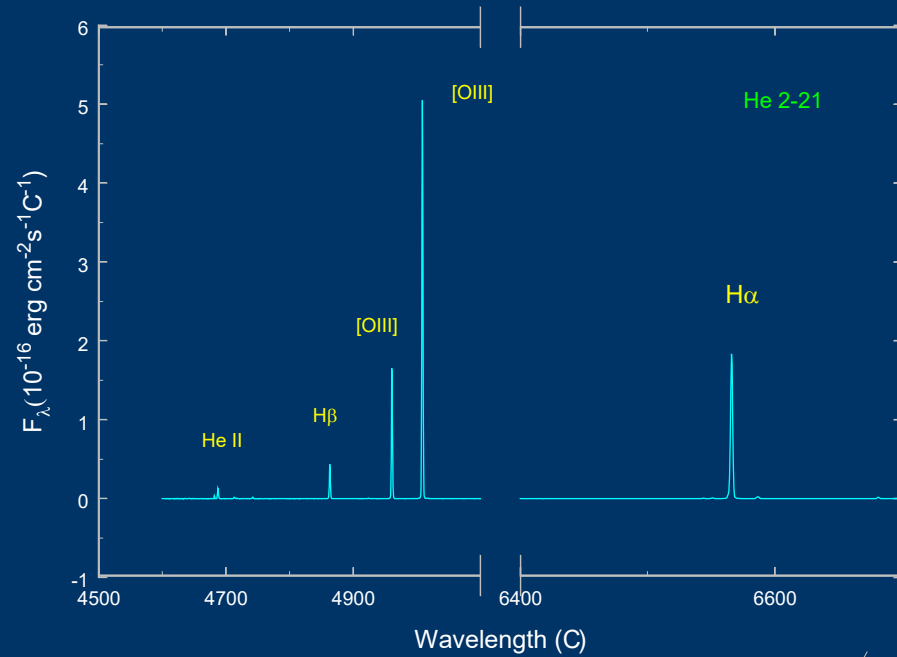
Observational definition of PN

- Nebular morphology: some degree of symmetry
- Spectrum: emission line spectrum with little or no continuum, $H\alpha$ to [OIII] ratio, expansion velocity
- Central star: temperature, luminosity, gravity

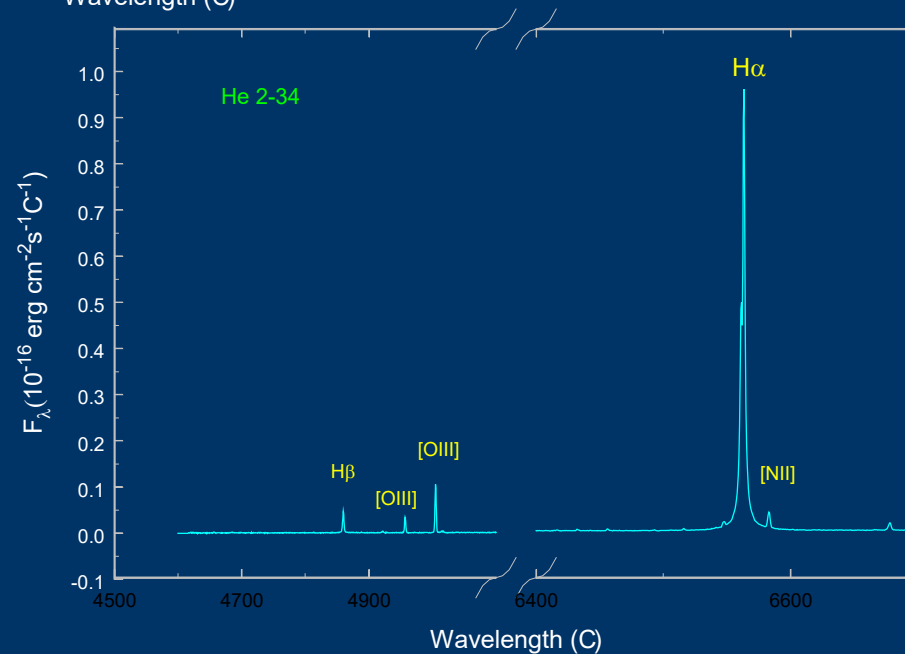
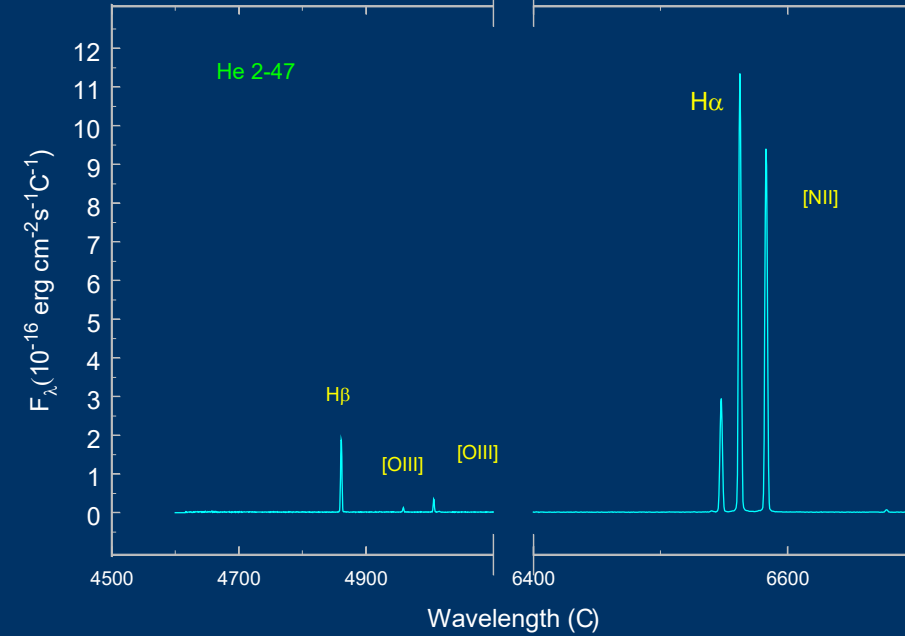
Confusion with other objects

- Emission line galaxies, reflection nebulae, HII regions, Wolf-Rayet nebulae, novae, symbiotic stars, M stars,

High excitation PN



Low excitation PN



Symbiotic star
(Mira implied by
photospheric H₂O and
variability, Feast et al.
1983)

PN and symbiotics

- Symbiotic stars and novae share many of the same observational properties with PN
- However, they are very different in evolutionary status
- Evolution in a binary system is driven by the mass transfer rate and H burning rate (Paczynski & Rudak 1980)
- Evolution can be static (symbiotics) or backwards (novae)

Thanks to David Allen, symbiotic stars is a well understood phenomenon. We should not confuse them with PN

A definition based on evolution?

- Ionized circumstellar shells showing some degree of symmetry surrounding a hot, compact star evolving from the AGB and white dwarf phases.

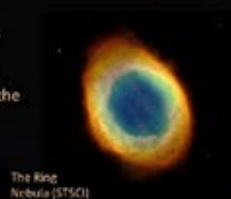
Morphological Classifications

- Curtis (1918): helical, annular, disk, amorphous
- Perek and Kohoutek (1967): stellar, disk, irregular, ring, anomalous
- Stanghellini et al. (1993): elliptical, bipolar, point-symmetric, irregular, stellar

What is the origin of the diverse morphology?

Hubble Space Telescope Images of Planetary Nebulae

At the end of a star's life, it wraps itself in a cocoon by spilling out gas and dust. Sometime later, a butterfly-like nebula emerges from the cocoon and develops into what astronomers call a planetary nebula. Recent observations by the Hubble Space Telescope have revealed the details of this transformation.



The Ring Nebula (M57)



IC 4663



The Egg Nebula



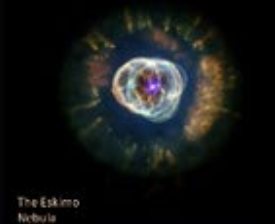
The Cotton Candy Nebula



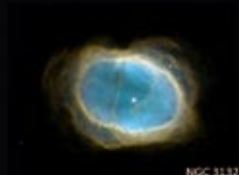
NGC 6826



The Saturn Nebula



The Eskimo Nebula



NGC 3132



Hubble 5

Planetary nebulae come in many different shapes and colors. Some are ring like (e.g. the Ring Nebula and NGC 3132), and some have butterfly shapes like Hb 5, M2-9, NGC 2346 and NGC 6537 below.



M2-9



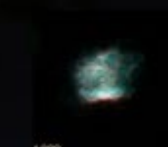
NGC 6543



M3 1



NGC 5307



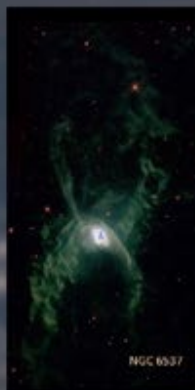
NGC 7440



Hb2-320



NGC 6280



NGC 6537

Our modern understanding of planetary nebulae began with Bohdan Paczynski's suggestion that they descend from asymptotic giant branch (AGB) stars and that the central stars of planetary nebulae are powered by hydrogen fusion. In 1978, Sun Kwok, Chris Purton and Pim FitzGerald explained the formation of the nebula by an interacting stellar winds (ISW) process where a fast wind acts as a snow plow to sweep up matter ejected in the previous AGB phase to create a dense shell. The ISW model was generalized by Bruce Balick in 1987 to explain the variety of morphologies observed in planetary nebulae.



NGC 2346 (STSCII)



MyCn18

These images were processed at the Space Astronomy Laboratory of the University of Calgary using data taken with the Hubble Space Telescope Wide Field Planetary Camera 2. For more information on the Space Astronomy Observatory, see www.iras.ucalgary.ca

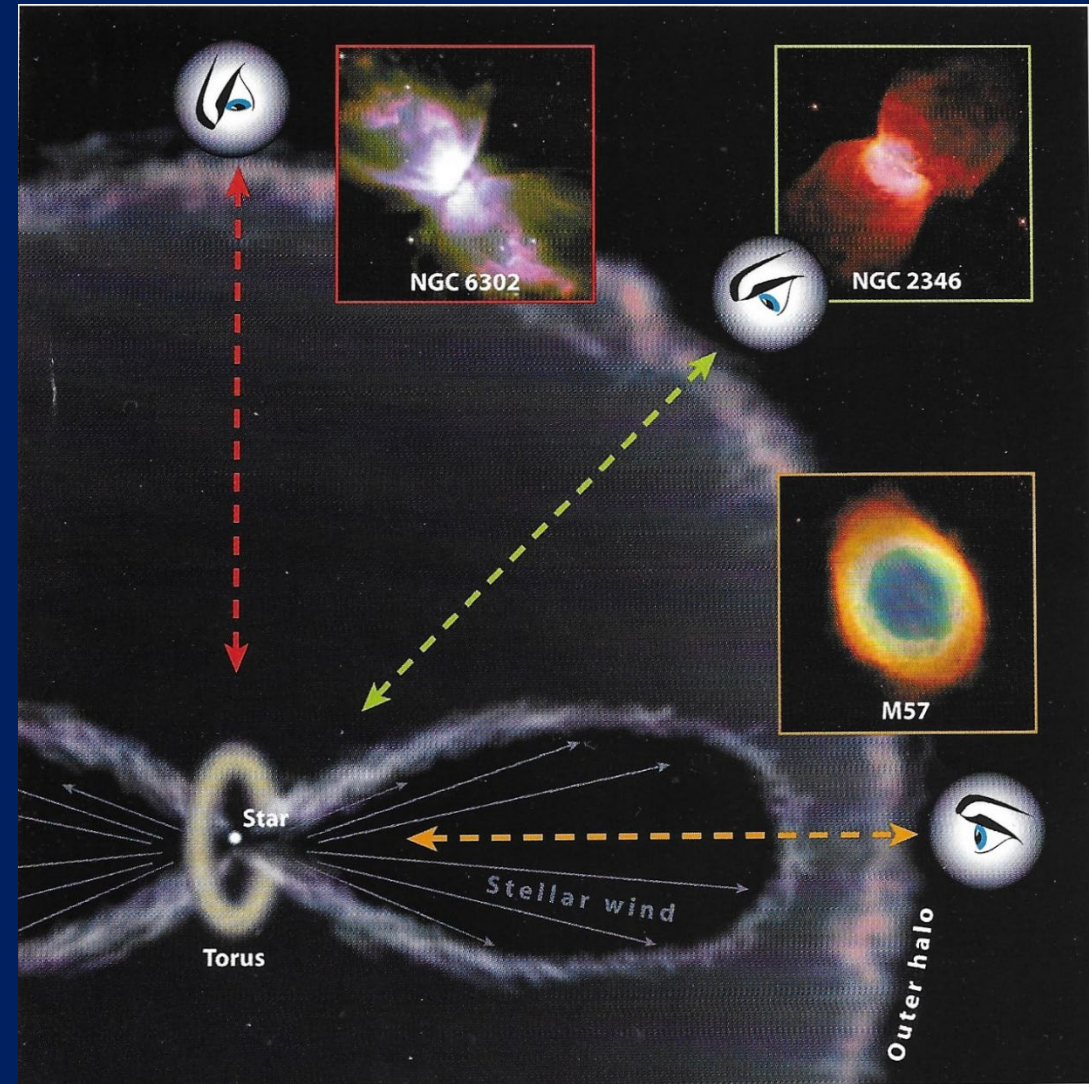
Further readings

- Kwok, S. A Modern View of Planetary Nebulae, *Sky & Telescope*, July 1996
- Balick, B. The Shapes of Planetary Nebulae, *American Scientist*, July 1996
- Kwok, S. Stellar Metamorphosis, *Sky & Telescope*, October 1998
- Kwok, S. The Origin and Evolution of Planetary Nebulae, Cambridge University Press, 2000

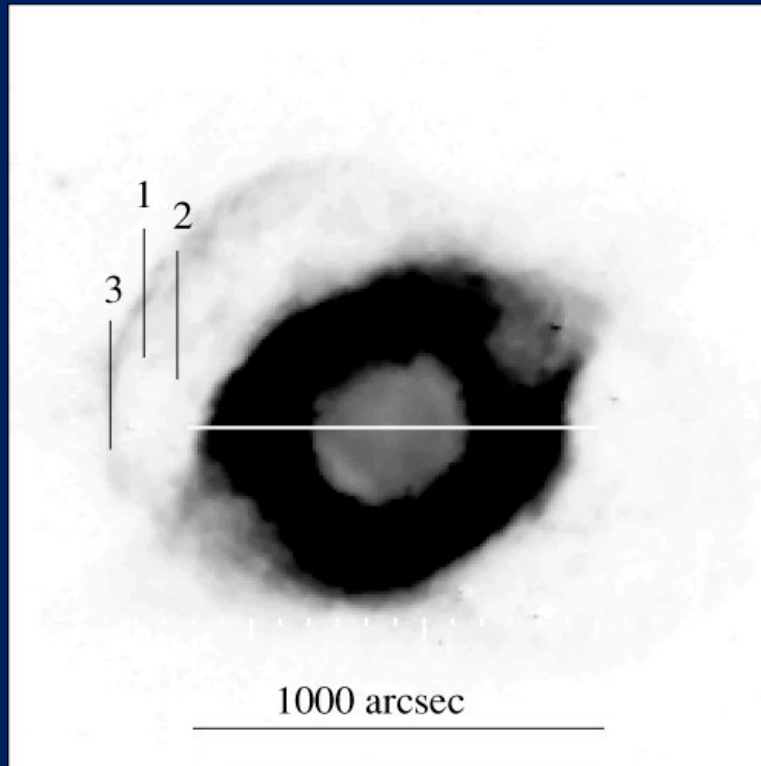


Our morphological classifications are biased by

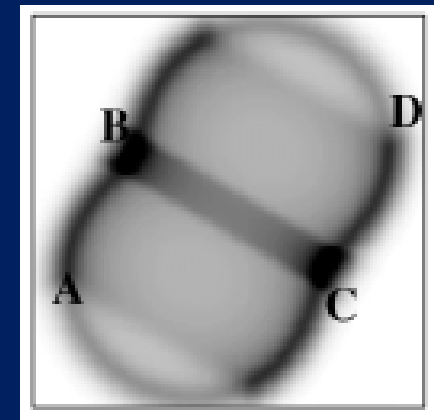
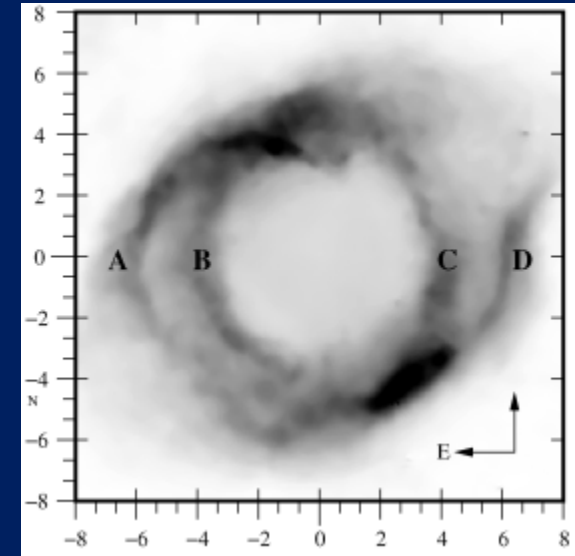
- Orientation effects
- visual brightness: strong recombination lines and collisionally excited lines
- Dynamic range: if we go deep enough we see more/different structures
- Field of view



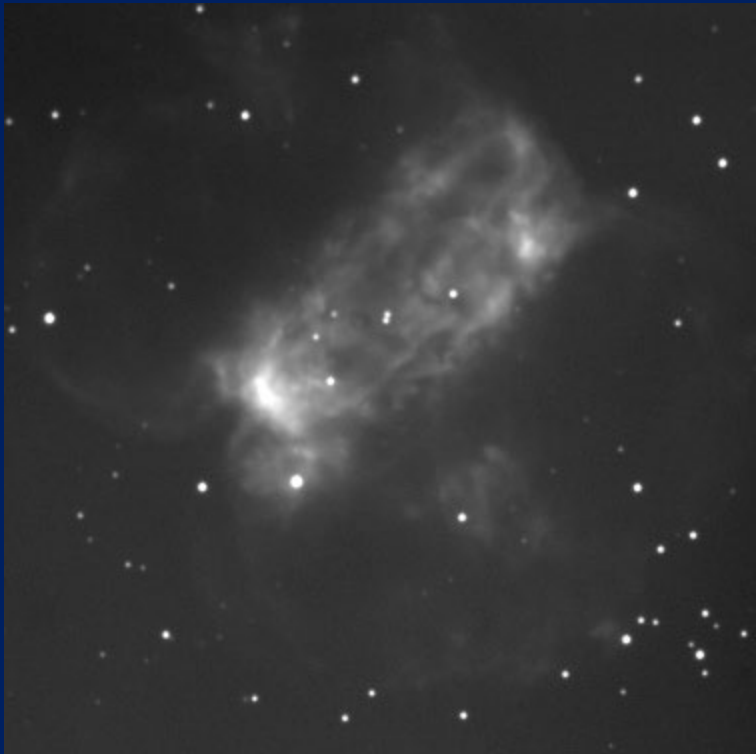
The Helix Nebula as a bipolar nebula inclined at 37°



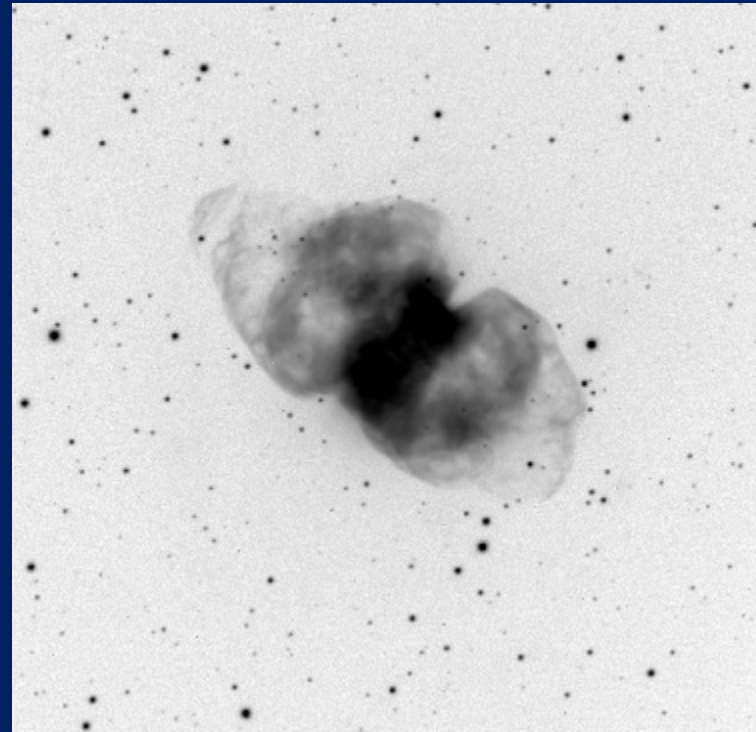
Meaburn et al. (2005)



How many PN are bipolar?



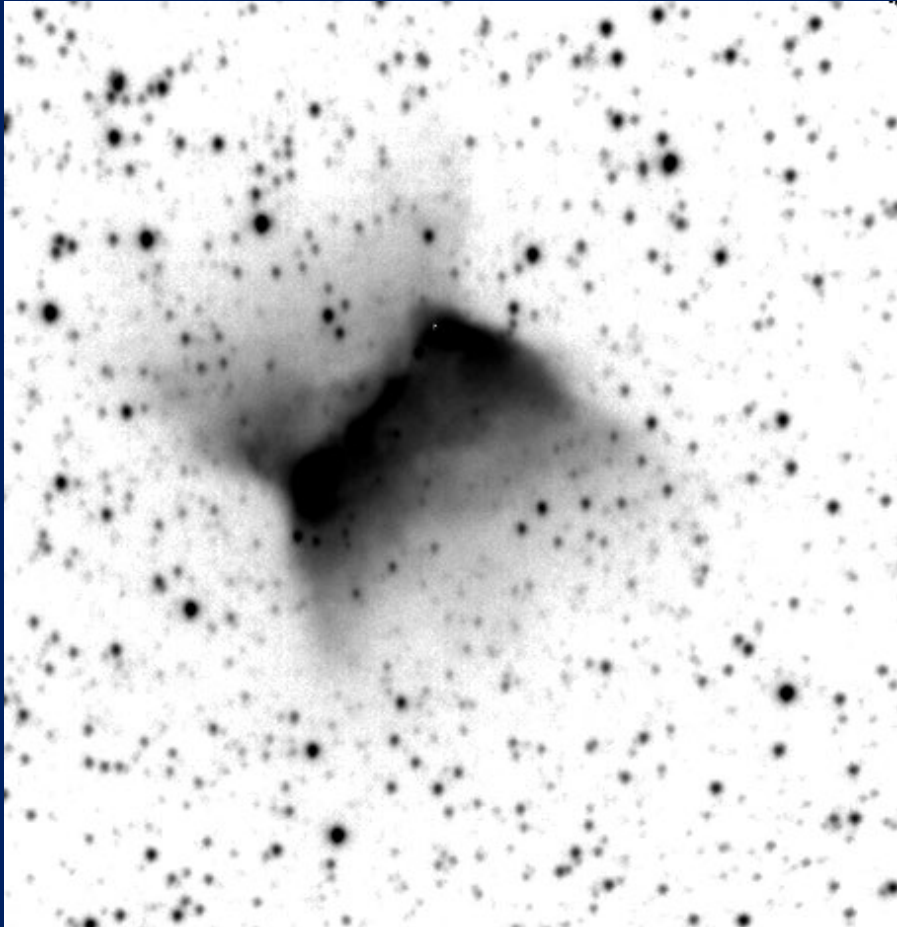
NGC 650-1 (M76)



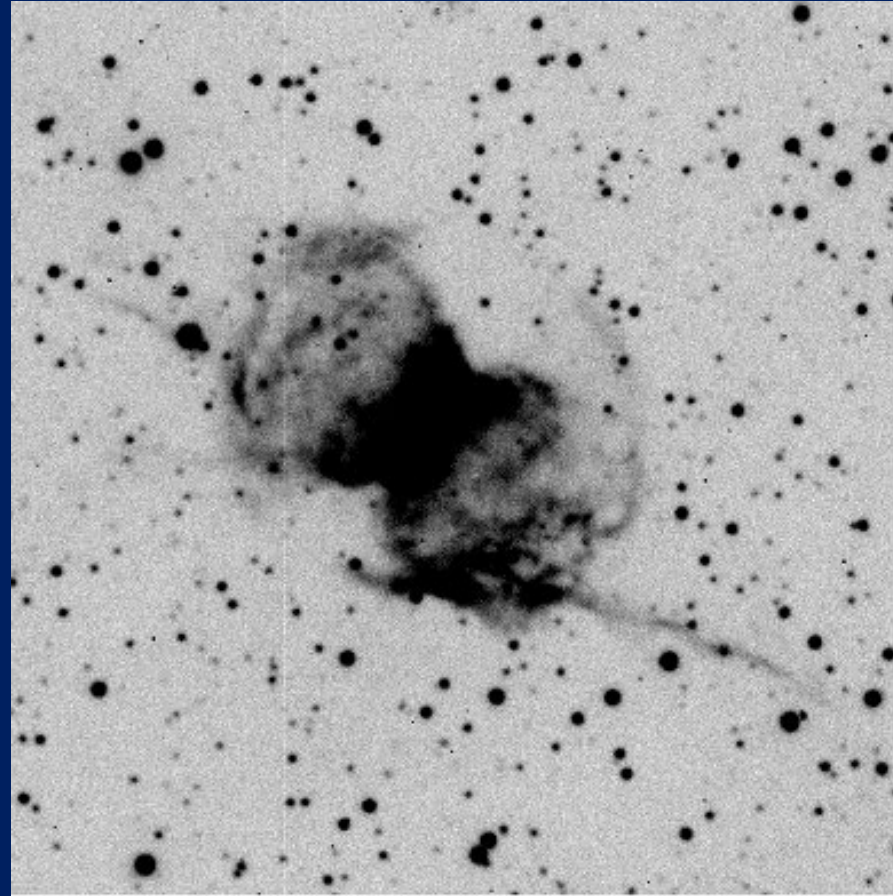
George Jacoby

Deep imaging

Trung Hua

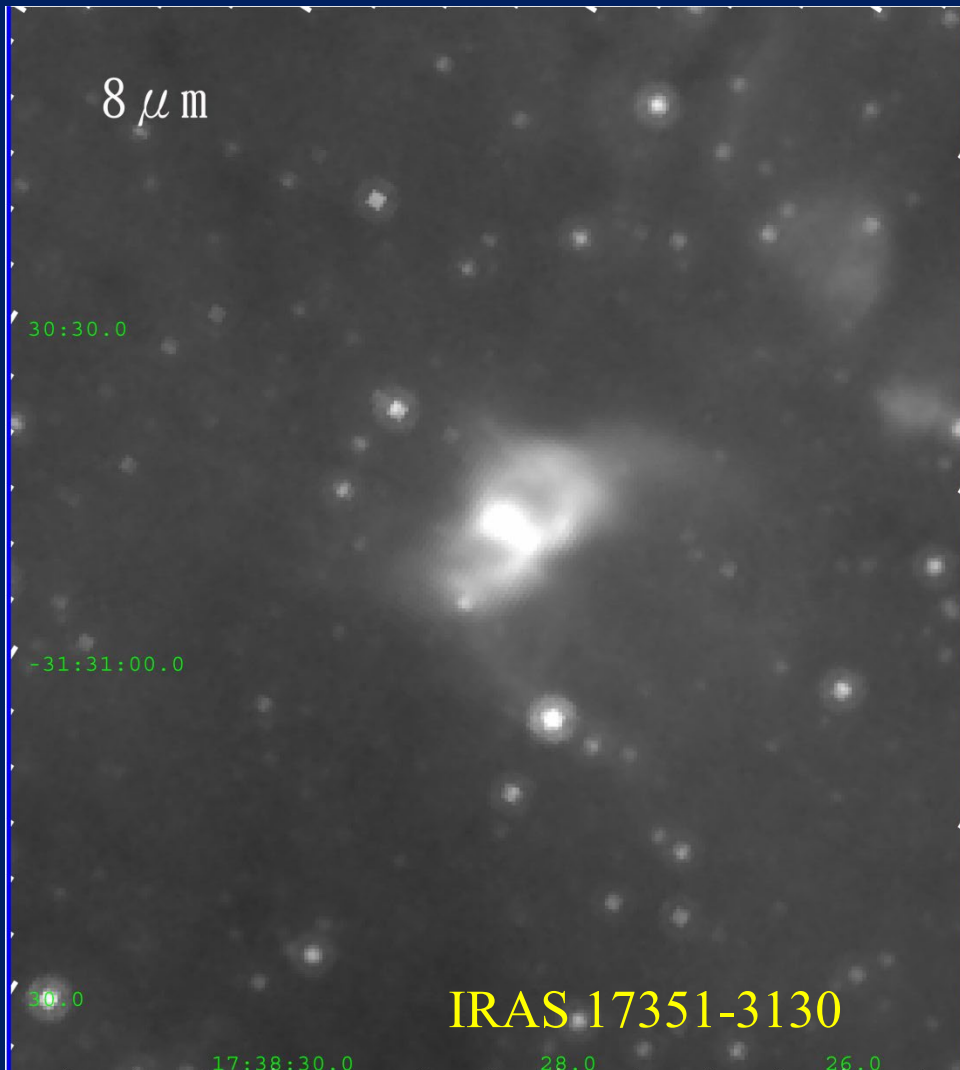


SaWe3



Sh1-89

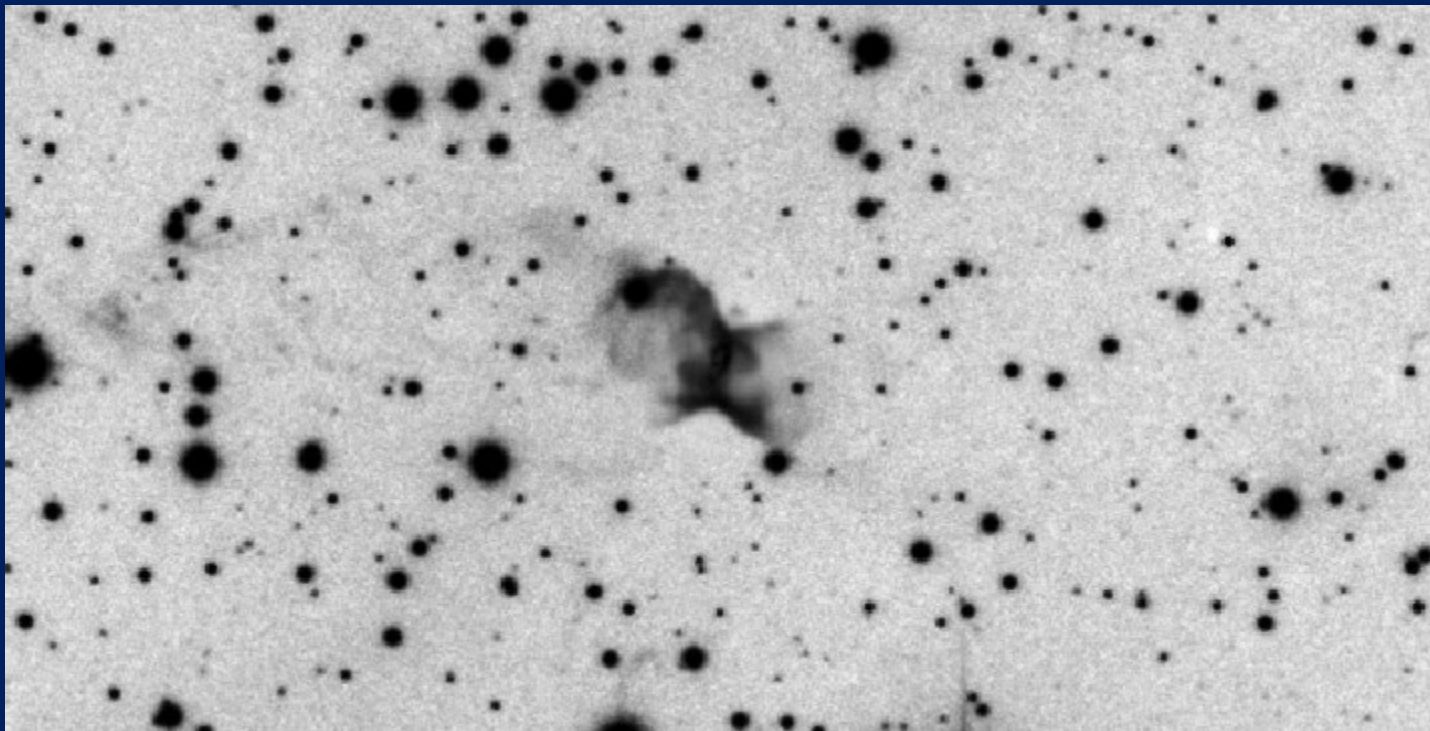
Morphology and large scale structures



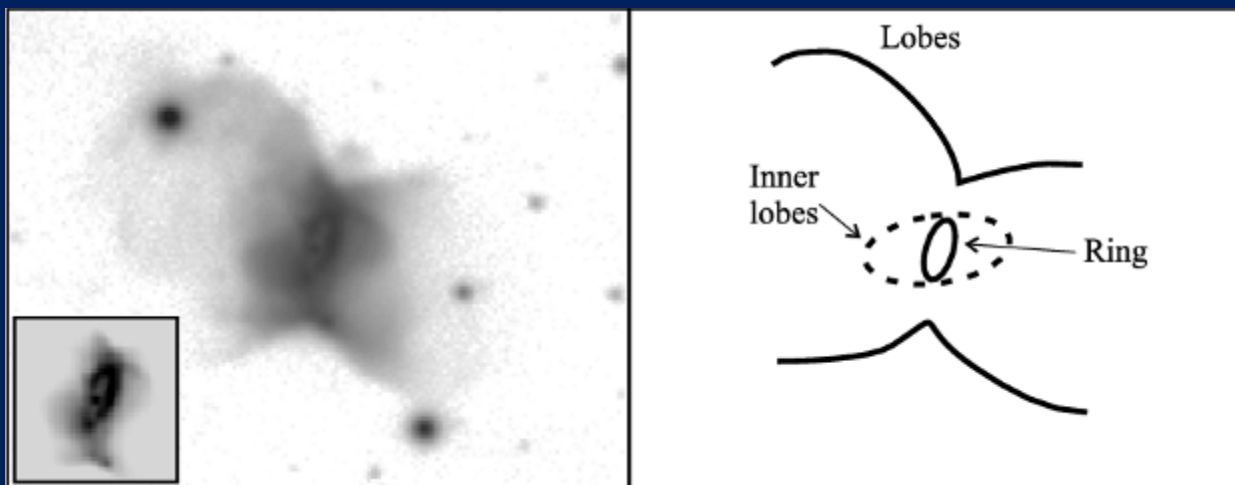
Bipolar morphology revealed by
dust in the lobes

*Traditional
morphological
classification is based
on optical apparent
structures and does not
tell the entire story*

GLIMPSE 8 μ m

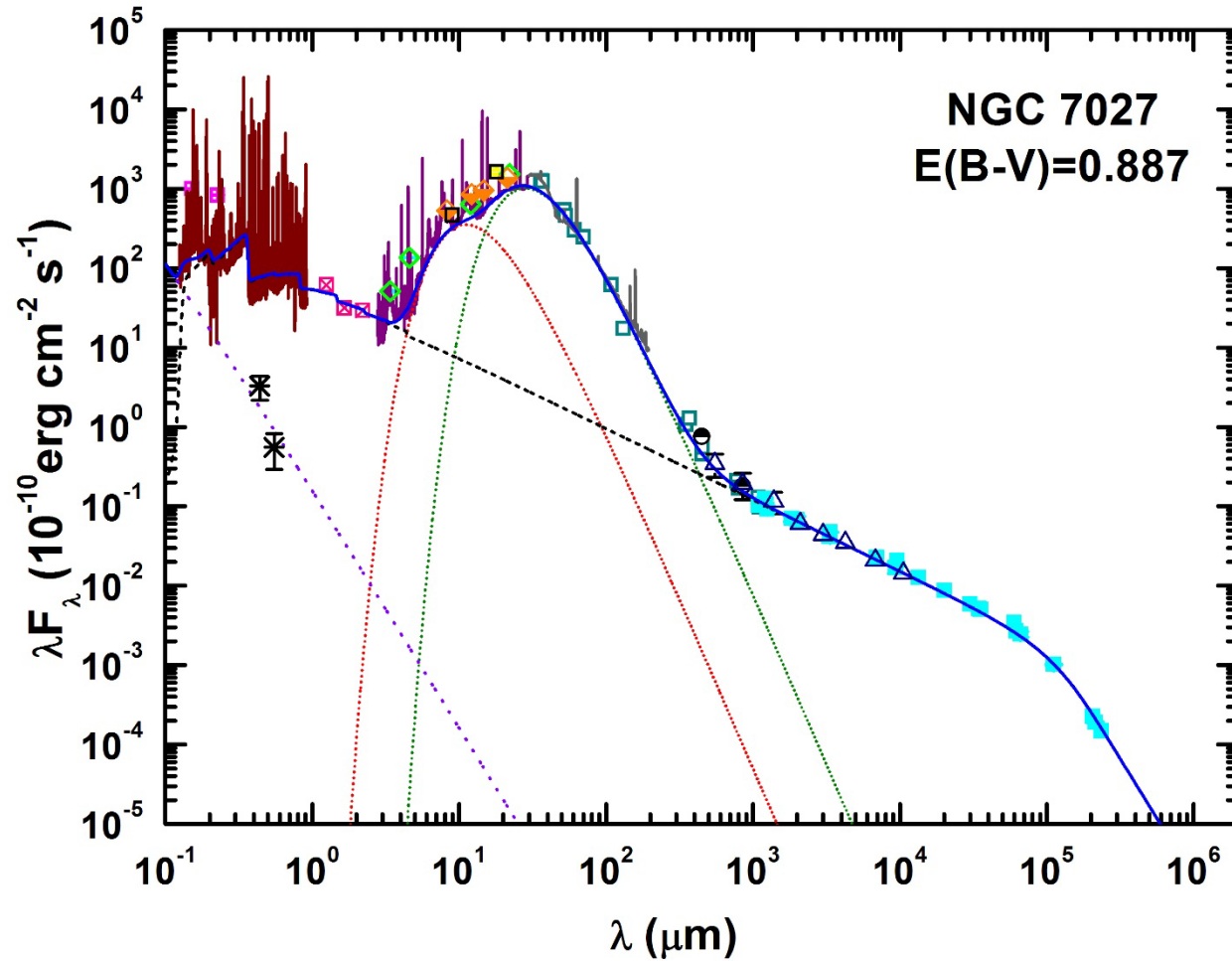


IPHAS PN-1

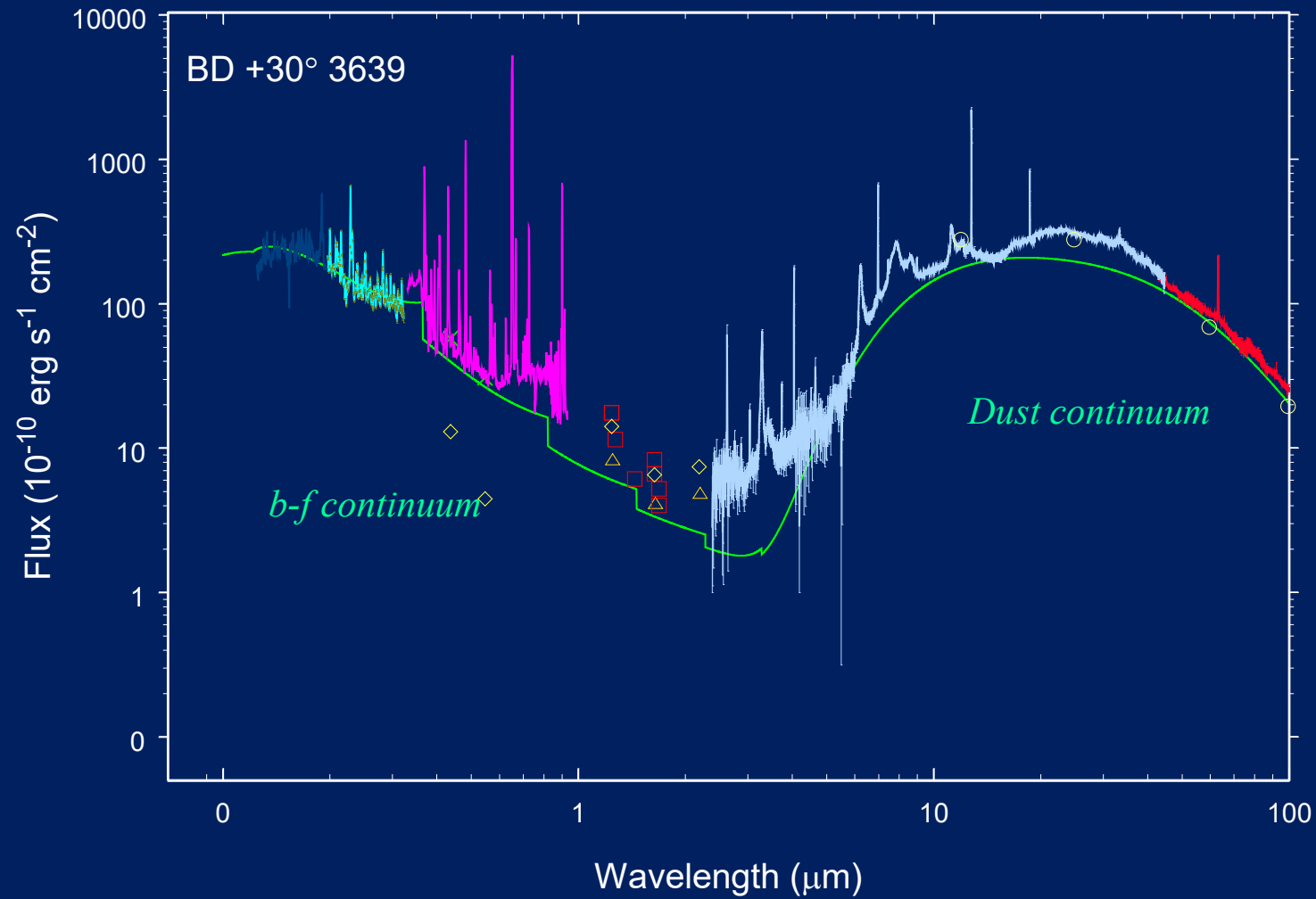


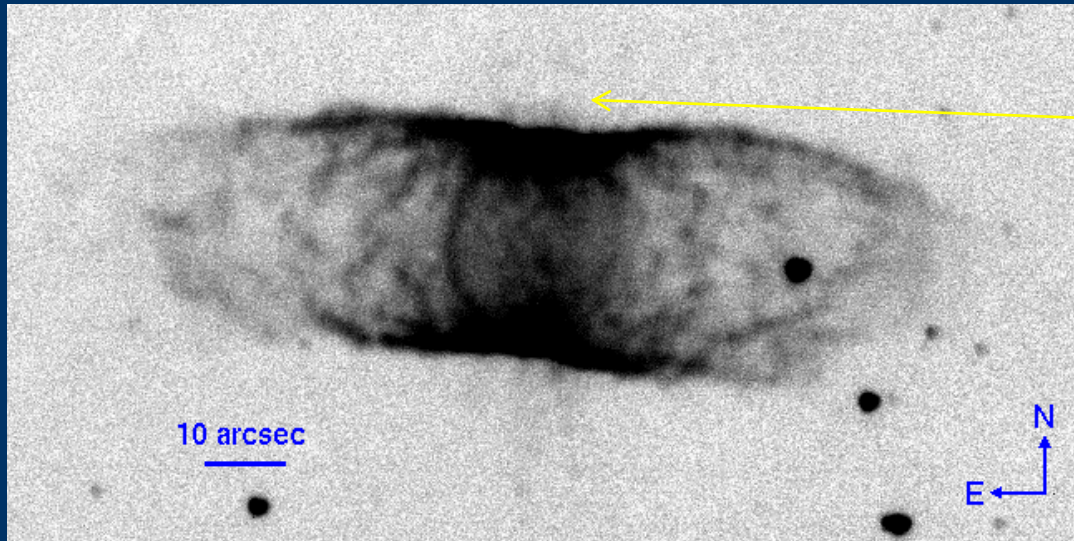
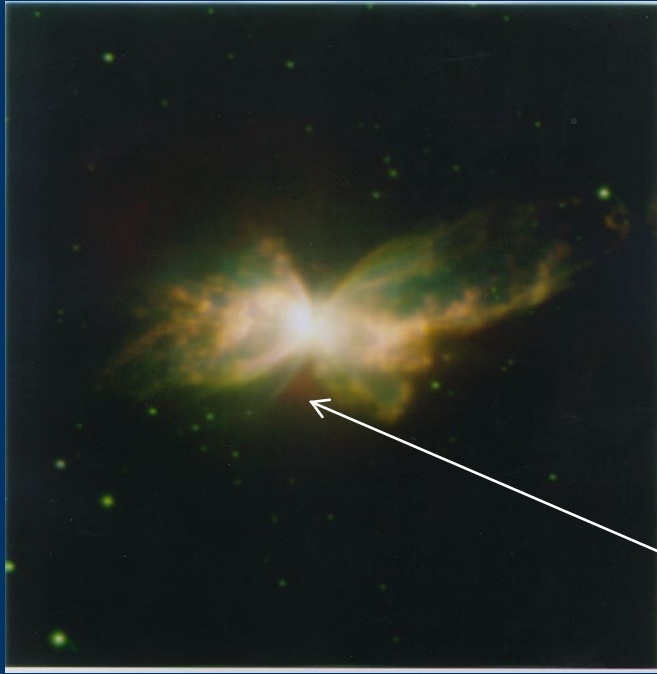
Mampaso et al. 2006

The role of invisible matter in planetary nebulae



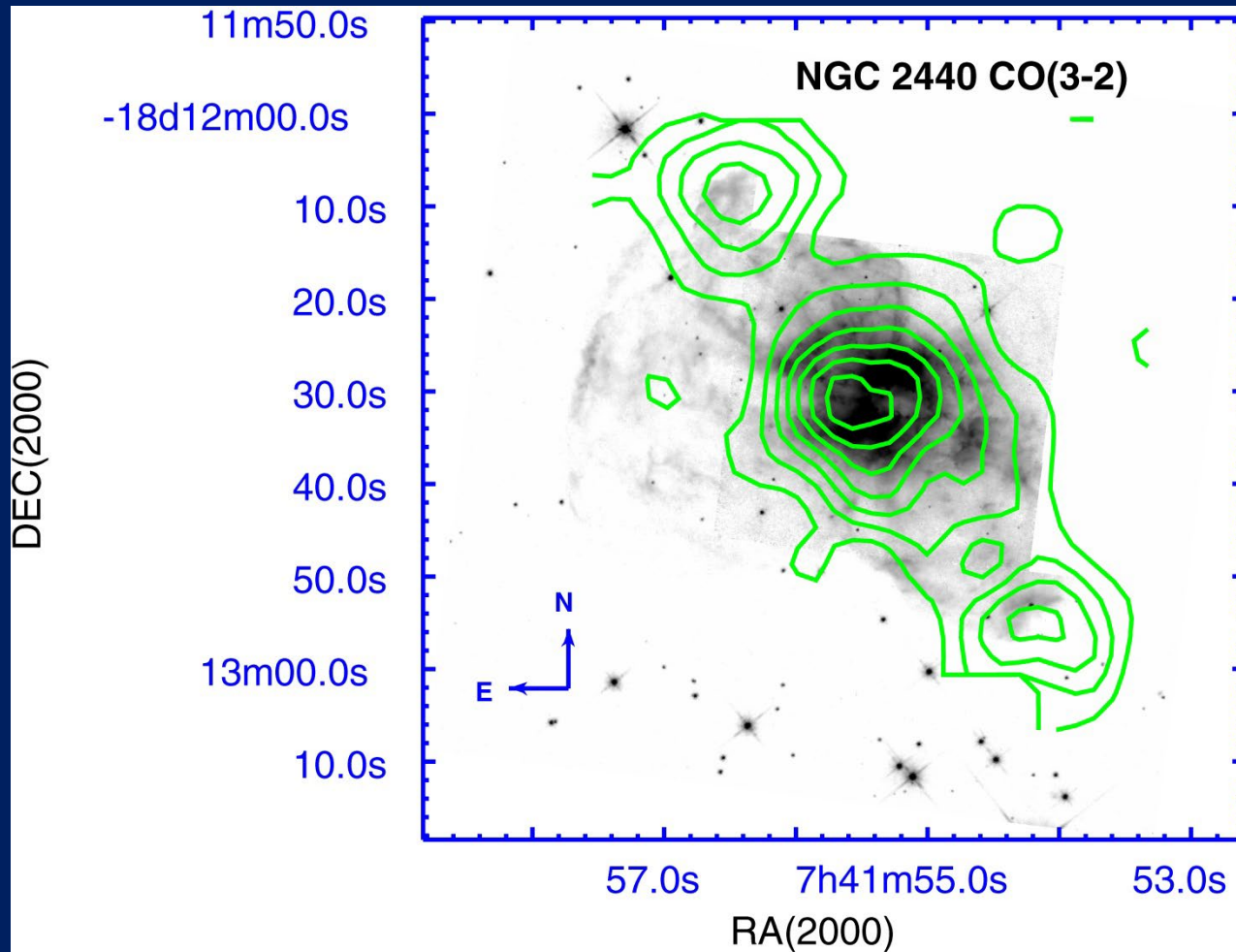
Significant fraction of energy emitted by the dust component





Confined by unseen material. What we see is not what there is.

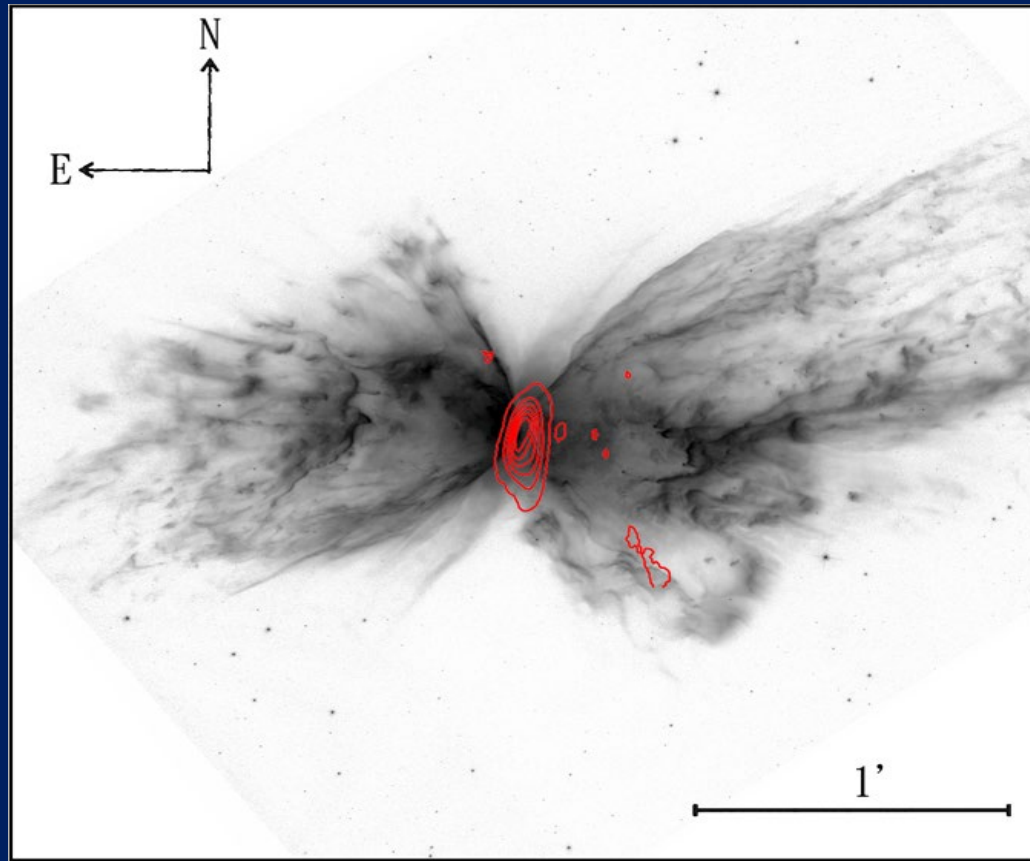
Molecular bipolar flow



Molecules trace
the lobes

Wang, Hasegawa &
Kwok 2008, *ApJ*,
273, 264

Molecular torus



SMA: Trung et al. 2008

Including the Ring, the Helix, the Dumbbell....



But many PN are bipolar, where do they come from?



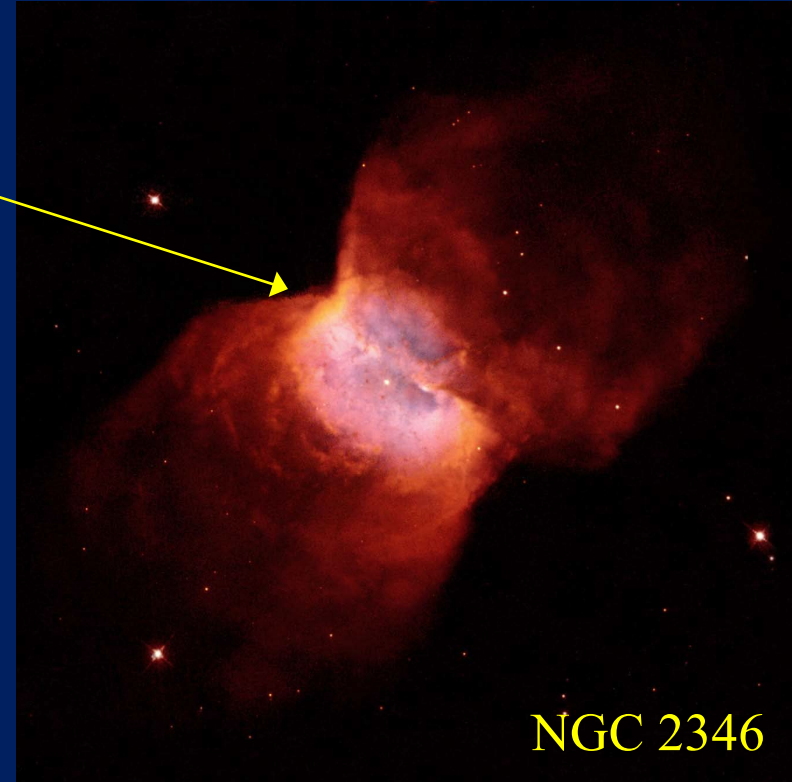
Our Traditional Concept of
Planetary Nebulae is
Dominated by their Optical
Images

*Although PN are bright because
of recombination lines of H and
He and collisionally excited lines
of metals, the ionized region in
fact represents a very small
fraction of the total mass of PN*

Indirect evidence for “dark matter”



*Confined by
dust torus?*



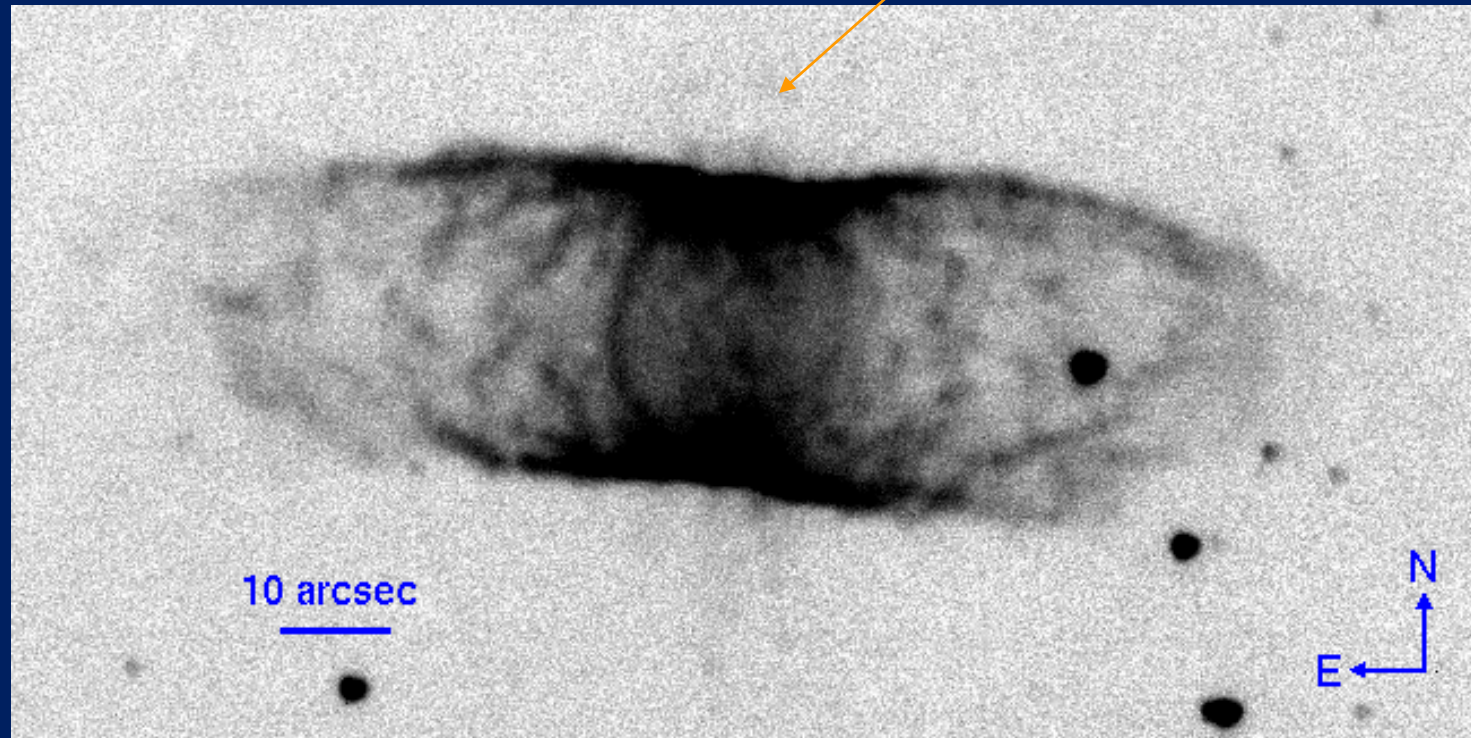
Eastern lobe



*The sharp boundaries suggest
that the optical lobes are
confined by external medium*

Sharp, well defined lobe boundaries

Unseen material



IC 4406 H₂

Bipolar cavity confined by external medium



Hen3-401



IRAS 17106-3046



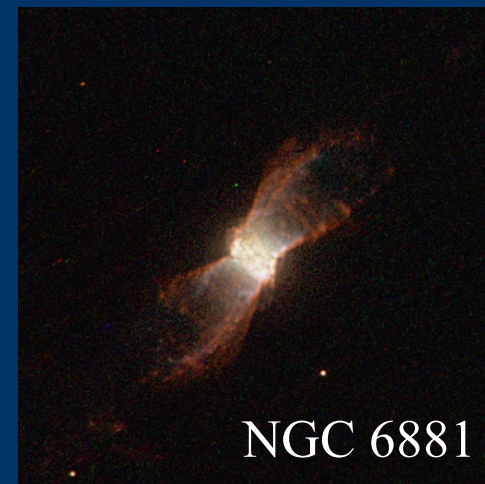
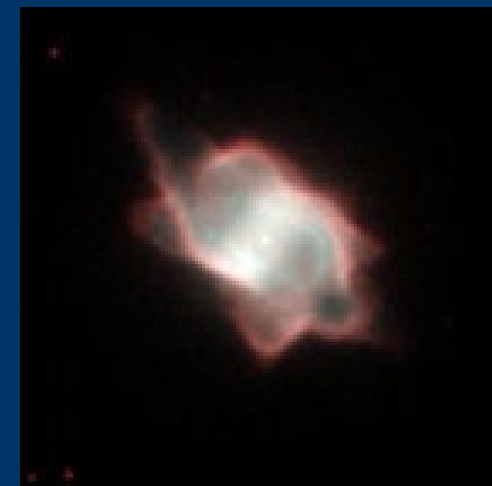
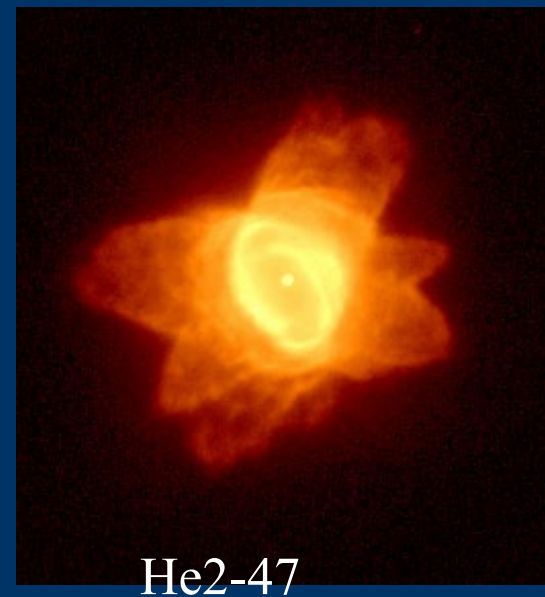
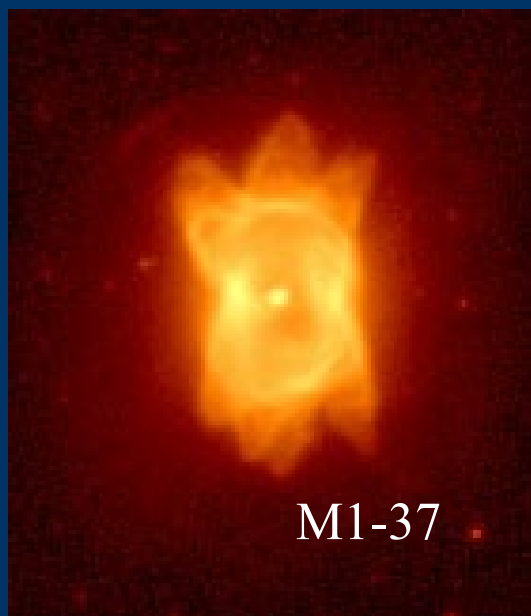
Hen2-320

Bipolar lobes are not regions of massive ejection, but low-density cavities carved out of the neutral circumstellar envelope

Multipolar nebulae

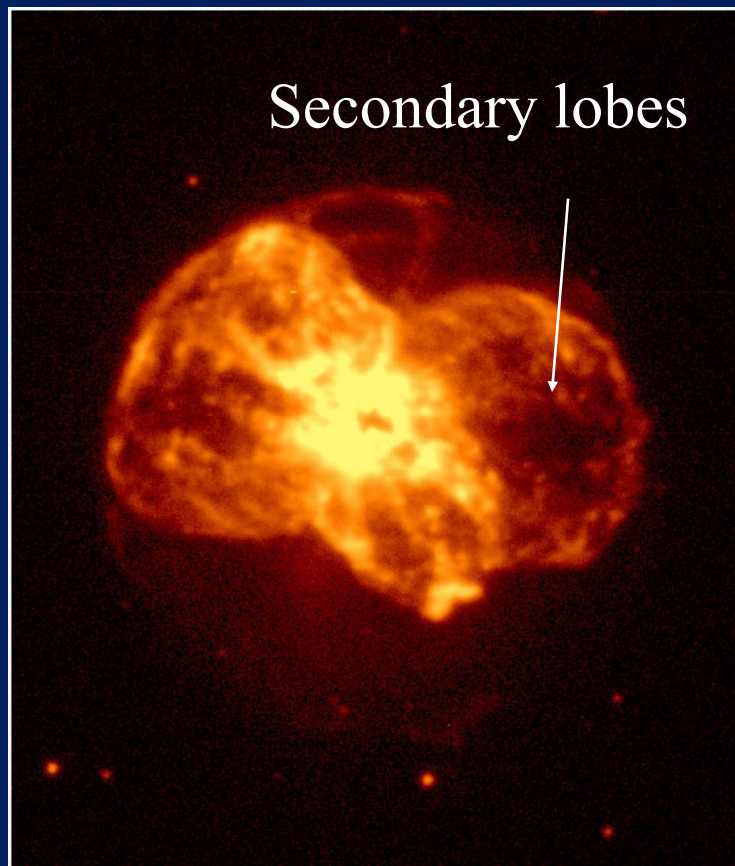
- Revealed by high dynamic-range imaging
- Quadrupolar nebulae (Manchado et al. 1996)
- Polypolar (López et al. 1998)
- Multipolar lobes of equal sizes: He 2-47, M1-37, He 2-113 (Sahai 2000 a,b)

Multi-polar Nebulae

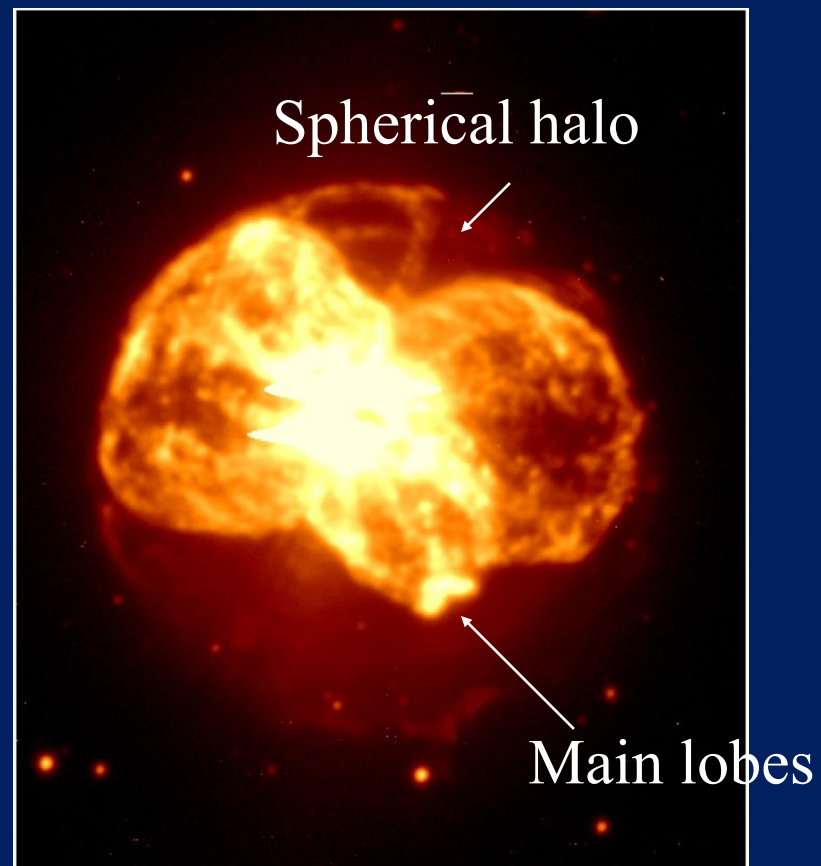


Multiple Lobes in NGC 2440

NII images taken at CFHT, Dec 16, 2002

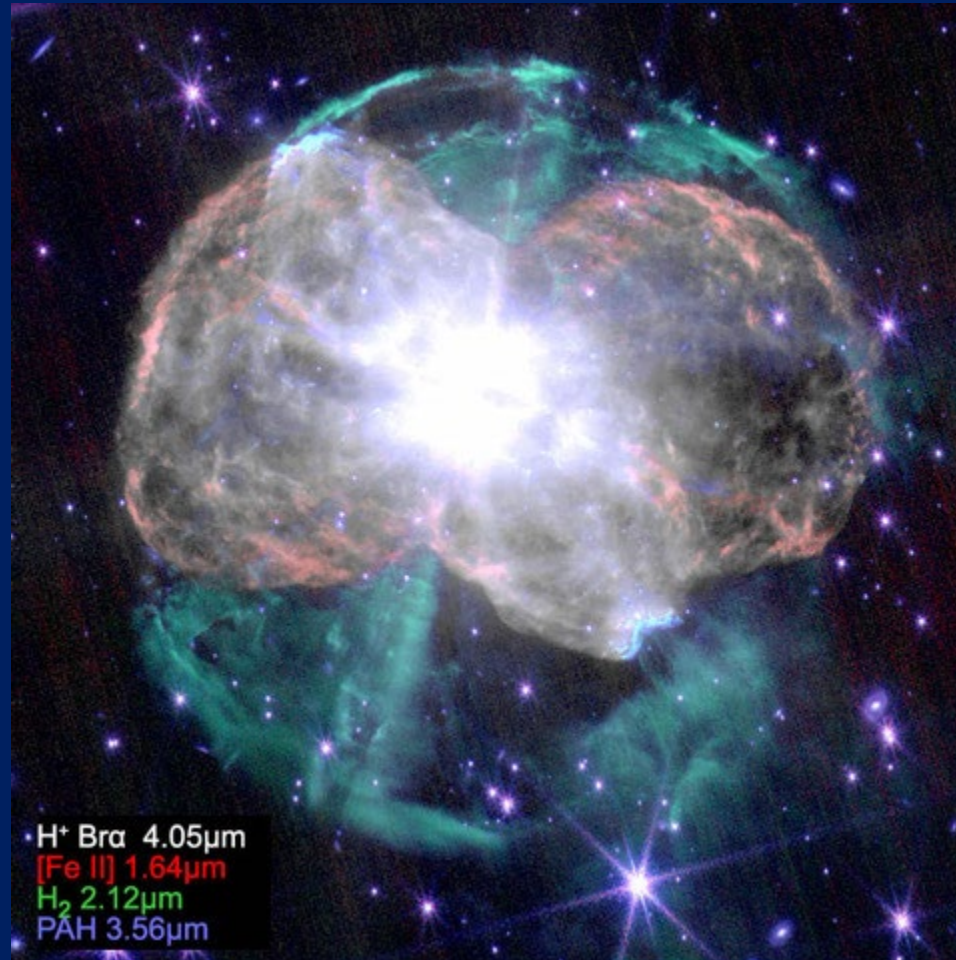


30 sec, unprocessed

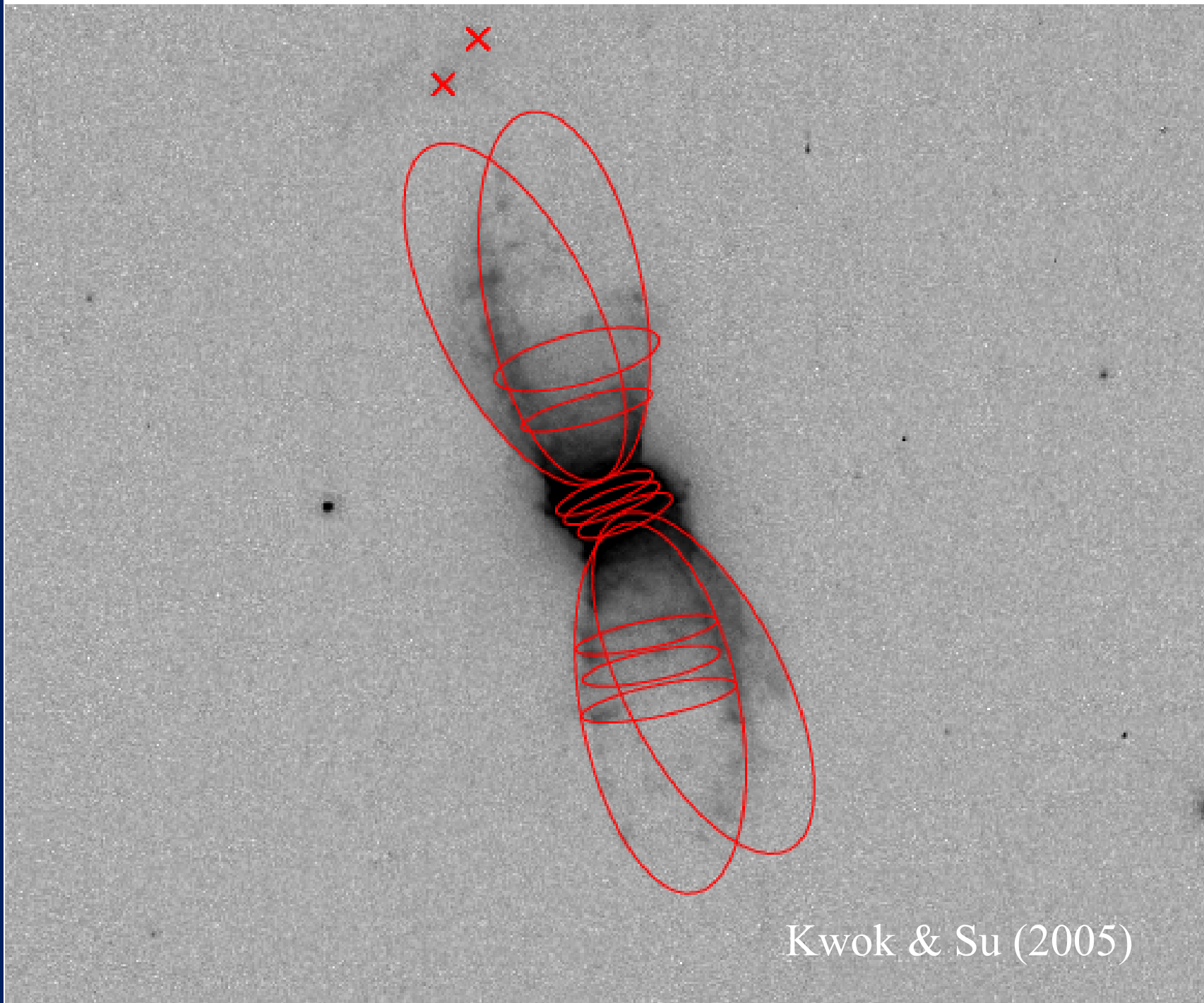


1200 sec, unprocessed

JWST NIRCam image of NGC2440

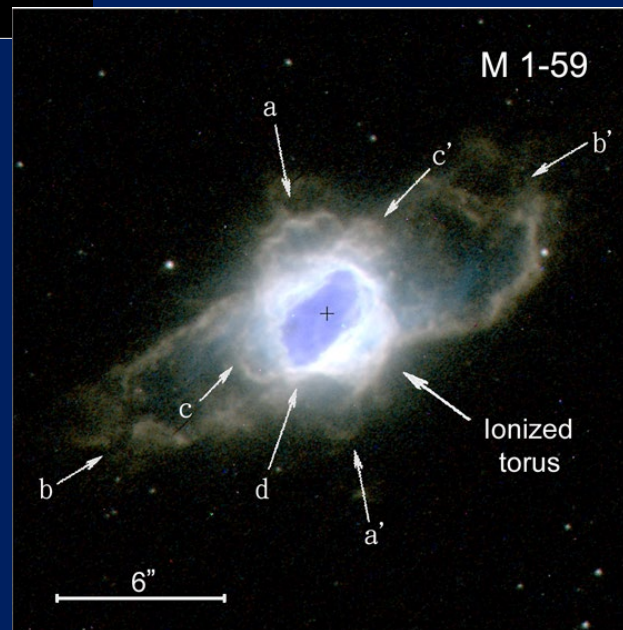
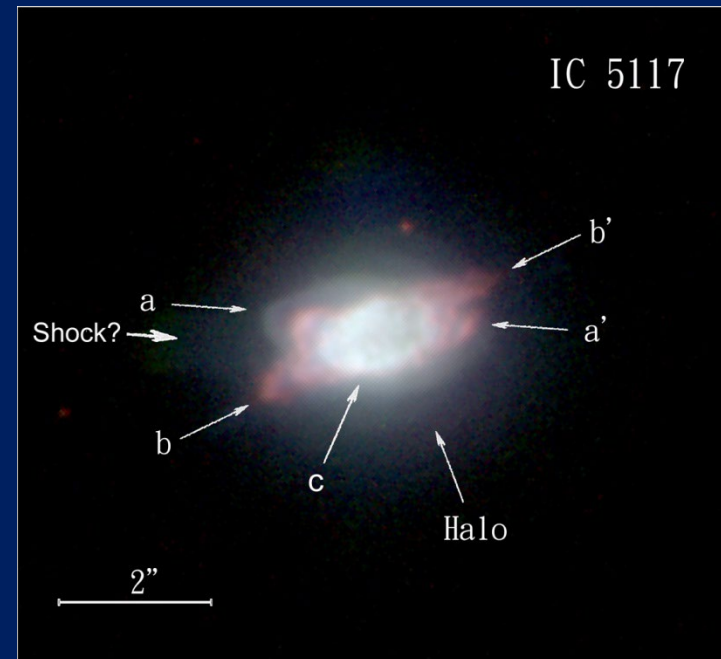
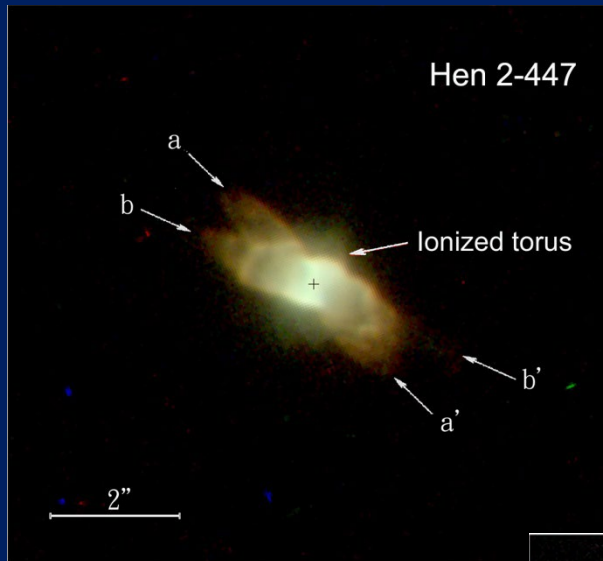


Bruce Balick, private
communication



Kwok & Su (2005)

Multipolar nebulae



Hsia et al. 2014



NGC 6072, IRAC 4-band image

Kwok et al. 2010

Model of PN

- Most of the mass of a PN is in the remnant AGB (spherical) envelope
- Post-AGB collimated fast outflows create cavities in the envelope (channeled by a torus?)
- PPN: dust scattering of visible photons from the central star lead to bipolar nebulosity
- PN: UV photons from the now hot central star ionize the low-density cavity, illuminating the bipolar region

Mass not in the bright regions

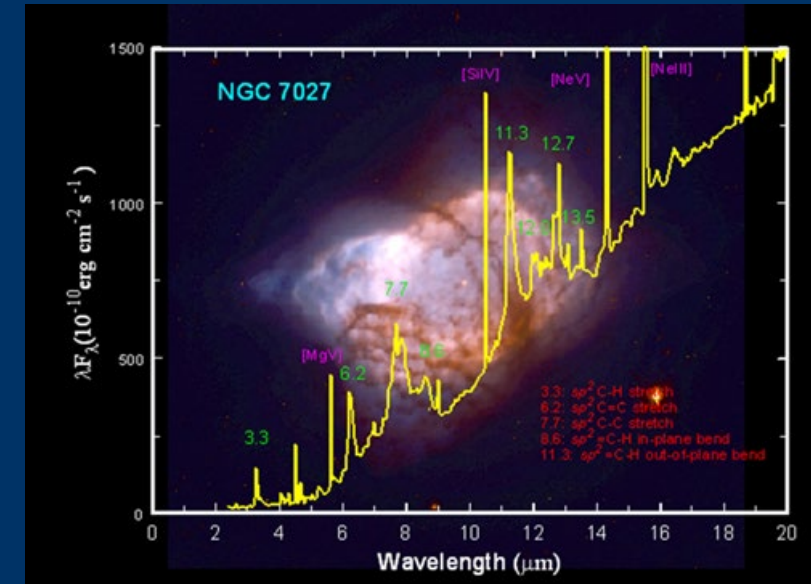
Optical morphology of PN is not defined by regions of matter ejection, but are defined by holes in the matter distribution where densities are low enough for UV to ionize

Morphological structures

- Bipolar nebulae revealed by higher dynamic range images
- Bipolars found by IR imaging
- True 3-D structures only found by proper modeling
- Dust in the torus, lobes, and haloes
- Model of PN must take into account not just the visible components
- *The optical lobes of bipolar/multipolar PN are not indicators of mass distribution but cavities of low density*

Things to be done

- High angular resolution mappings of the ionized, molecular, and dust components as the ionized component does not contain most of the mass
- Detailed time sequence of organic synthesis: carrier of the unidentified infrared emission bands
- Physical mechanisms for various wind components leading to the shaping of PN

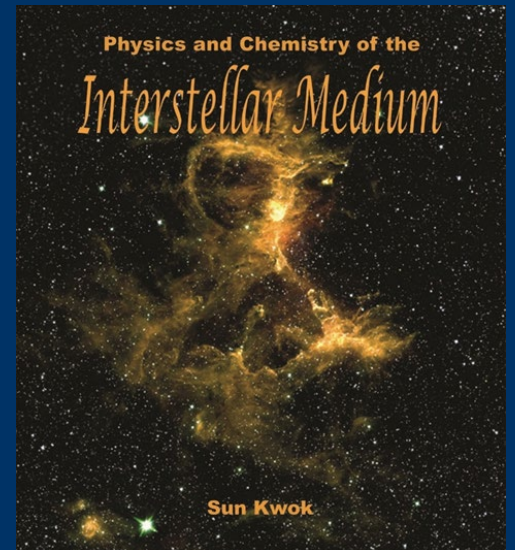
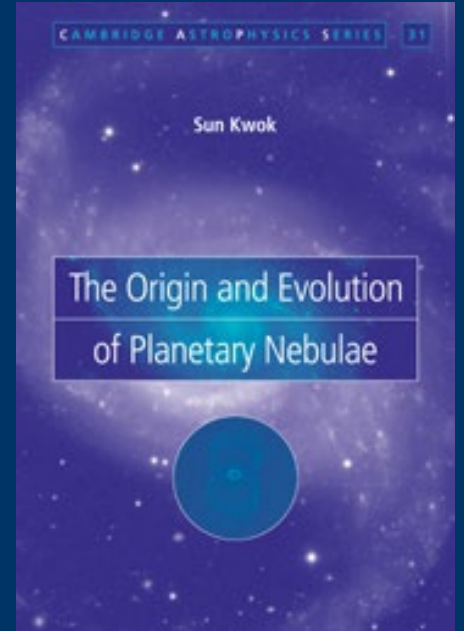


Kwok, S. The Mystery of Unidentified Infrared Emission Bands, *Astrophys. Space Science*, **367**, 16 (2022)

Kwok, S. Complex Organics in Space: A Changing View of the Cosmos. *Galaxies*, **11**, 104 (2023)

Summary (cont.)

- nucleosynthesis (origins, chemical environment)
- galactic evolution (chemical, SF history)
- cosmology (H_0 , dark matter, baryonic mass)
- origin of life



Kwok, S. 2024, Planetary Nebulae Research: Past, Present, and Future, *Galaxies* 12(4), 39

What is the future of PN research?

- Stay relevant as a tool and laboratory
- Advances help other fields
- Make use of new facilities
- There are still many unsolved problems
- Intrinsic scientific merit of the field vs. fashion

*Recognition of PN as a post-AGB phenomenon only happened after 1970, 200 years after the discovery of PN.
Paradigm shift did happen.*

To challenge this paradigm, one has to present a self consistent alternative explaining morphology, abundances, kinematics, statistics, etc.

Kwok, S., Balick, B., Chu, Y.-H., Hrivnak, B., López, A., Parker, Q., Sahai, R., Zijlstra, A.
Current Unsolved Problems in Planetary Nebulae Research, Galaxies, **14**, 30 (2026)