RECENT PROGRESS IN UNDERSTANDING THE ERUPTIONS OF CLASSICAL NOVAE

Michael M. Shara

October 1988

Operated by the Association of Universities for Research in Astronomy, Inc., for the National Aeronautics and Space Administration
RECENT PROGRESS IN UNDERSTANDING THE ERUPTIONS OF CLASSICAL NOVAE

Michael M. Shara
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218

To be published in the December 1988 issue of the
Publication of the Astronomical Society of the Pacific
Received 21 September 1988; Accepted 21 September 1988
Abstract

Dramatic progress has occurred in the last two decades in understanding the physical processes and events leading up to, and transpiring during the eruption of a classical nova. The mechanism whereby a white dwarf accreting hydrogen-rich matter from a low-mass main-sequence companion produces a nova eruption has been understood since 1970. The mass-transferring binary stellar configuration leads inexorably to thermonuclear runaways on the white dwarf surfaces, mass ejection, and photon emission so luminous it can be detected at distances of megaparsecs. This review summarizes the efforts of many researchers in understanding the physical processes which generate nova eruptions; the effects upon nova eruptions of different binary-system parameters (e.g. chemical composition or mass of the white dwarf, different mass accretion rates); the possible metamorphosis from dwarf to classical novae and back again; and observational diagnostics of novae, including X-ray and γ-ray emission, and the characteristics and distributions of novae in globular clusters and in extragalactic systems. While the thermonuclear-runaway model remains the successful cornerstone of nova simulations, it is now clear that a wide variety of physical processes, and three-dimensional hydrodynamic simulations, will be needed to explain the rich spectrum of behavior observed in erupting novae.
**TABLE OF CONTENTS**

I. Introduction and Scope of the Review ..................................................... 1

II. The Pre-Outburst Stage ............................................................................. 2
   II.1 Radial Accretion of Hydrogen-Rich Material onto a White Dwarf ............ 2
   II.2 Diffusion During Pre-Nova Envelope Accretion ....................................... 3
   II.3 Shear Instability Mixing of Accreting Material Possessing Angular
       Momentum .................................................................................................. 4
   II.4 The Effects of Heating from the Boundary Layer on Accretion Models
       for Novae .................................................................................................... 7

III. The Outburst Stage .................................................................................... 8
    III.1 Localized Thermonuclear Runaways (LTNR) .......................................... 8
    III.2 The Evolution of a Classical Nova Model Through a Complete Cycle ....... 10
    III.3 The Helium/Hydrogen Ratio in the Ejecta of Classical Novae ............... 11
    III.4 Thermonuclear Runaways on the Surfaces of Oxygen-Neon-Magnesium
        White Dwarf ............................................................................................. 12
    III.5 Thermonuclear Runaways on the Surfaces of Iron White Dwarfs .......... 14
    III.6 Thermonuclear Runaways on the Surfaces of White Dwarfs near the
        Chandrasekhar Limit ................................................................................. 15

IV. The Ejection Phase .................................................................................... 16
    IV.1 The Steady State, Constant Luminosity, Continuous Wind Ejection
        Model of Novae ......................................................................................... 16
    IV.2 Dust in Nova Systems ............................................................................. 17
    IV.3 X-Rays from Outbursting Novae ............................................................ 18
    IV.4 Positron Annihilation Gamma Rays from Novae .................................... 18

V. Hibernation and Ancient Novae ................................................................. 19
    V.1 The Hibernation Scenario of Cataclysmic Binaries .................................... 19
    V.2 Hunting for Novae Seen by Ancient Astronomers .................................... 24

VI. Recurrent and Magnetic Novae ................................................................. 25
    VI.1 Recurrent Novae .................................................................................... 25
    VI.2 Novae on Strongly Magnetic White Dwarfs ............................................ 26

VII. Novae in Clusters and Galaxies ............................................................... 28
    VII.1 Novae in Globular Clusters .................................................................... 28
    VII.2 Extragalactic Novae and the Distance Scale ......................................... 30

VIII. Summary and Prospects for the Future .................................................. 31
    References .................................................................................................... 32
    Figure Captions ............................................................................................ 36
I. INTRODUCTION AND SCOPE OF THE REVIEW

The first systematic observations of erupting novae were made by Chinese imperial astronomers two thousand years ago. Their underlying reason for observing novae was the same as that of modern astronomers. Anything that brightens overnight by fifteen magnitudes is interesting; however, the use to which sky watchers put novae has changed in the last two millenia. Instead of being regarded as astrological omens, novae are now acknowledged to be powerful cosmological tools. Novae are excellent standard candles; they are extremely luminous at maximum, and their rates of decline are tightly correlated with their absolute magnitudes at maximum (Arp, 1956). The better a standard candle is understood, the less chance there is for misusing or misinterpreting it. This alone justifies the detailed study of classical novae. For many workers in the field, however, the rich potpourri of theoretical physics, numerical simulation, and range of observational phenomena involving novae keeps research going at a brisk pace regardless of these systems' cosmological value.

It is ironic that extragalactic novae were at first part of a serious stumbling block to cosmology. The supernova of 1885 in the nucleus of M31 was incorrectly argued to be a classical nova by Shapley in his famous debate with Curtis. Only with the resolution of Cepheids in M31 by Hubble were the relative brightnesses and distinct natures of classical novae and supernovae finally understood by astronomers. Had the supernova of 1885 not thrown astronomers off the track, novae might have helped settle the nature of the nebulae years earlier.

A wealth of observational data on classical novae was summarized by Payne-Gaposchkin (1957) and by McLaughlin (1960). It was clear to these authors that classical novae ejected significant amounts of matter at velocities of hundreds or thousands of kilometers per second into space. Faint, hot blue stars remained after nova eruptions, and the stars appeared about as bright after eruption as before. The intrinsically most luminous novae were observed to belong to the fastest "speed class" i.e., to fade most rapidly. Very fast nova decline by \( \sim 3 \) magnitudes in \( \sim 10 \) days; very slow novae may require a year to dim by the same amount.

The crucial breakthrough in understanding why novae erupt was the observation by Walker (1954) and by Kraft (1959, 1964) that novae occur in white dwarf-main sequence binary systems where hydrogen-rich matter is being accreted onto the white dwarf. Accretion occurs via a disk which is generated by the red dwarf overflowing its Roche Lobe. Considerable energy and angular momentum losses allow the disk to feed matter onto the white dwarf's surface.

Starrfield, Truran, Sparks and Kutter (1972) and Starrfield, Sparks and Truran (1974a, 1974b) demonstrated convincingly that hydrogen-rich envelopes on the surfaces of white dwarfs undergo thermonuclear runaways (TNRs). These explosions are violent enough to eject matter at high velocity if the envelopes are enriched in CNO nuclei relative to the Sun. Bath and Shaviv (1976) showed that radiation pressure can drive winds for extensive periods from novae. Subsequent observations in the infrared, ultraviolet, and X-ray wave bands have demonstrated that the bolometric luminosities of novae remain remarkably constant at the Eddington luminosity for months after the visual luminosity decreases. The physical processes leading to this observed behaviour include large bolometric corrections created by the thinning of the wind (Bath and Shaviv 1976), the contraction and heating of the remnant envelope (Prialnik, Shara and Shaviv 1978) and dust formation and heating (Bode 1983).

This appealingly straightforward scenario of nova eruptions was summarized a decade ago in an extensive review by Gallagher and Starrfield (1978). Since then, many researchers have been exploring the wide range of physical processes involved in nova eruptions, and the
dramatic effects these have on the subsequent eruption. The aim of this review is to describe
the efforts of these workers, over the last ten years, to progress from a simple understanding
of the basic nova mechanism to a detailed picture of the three-dimensional, time-dependent,
highly nonlinear processes that govern nova eruptions.

Several valuable reviews have been published in recent years on subjects related to cata-
clysmic binaries. An extraordinary compilation of finder charts, accurate coordinates, and
outburst properties of Galactic novae has been assembled and published by Duerbeck (1987).
This remarkable atlas of ~200 classical novae is of enormous value to all researchers in the
field. A further, important compilation of cataclysmic binary data and interpretation is
that of Patterson (1984). Specialized reviews have been published on the subjects of cata-
clysmic binaries with moderate magnetic fields attached to the white dwarf (the DQ Her
stars: Warner 1985); the strongly magnetic AM Herculis stars (Liebert and Stockman 1985);
the evolution of magnetic cataclysmics (Lamb and Melia 1987); IUE ultraviolet studies of
galactic novae (Starrfield 1986); dust in nova systems (Bode and Evans 1985); and nucle-
leosynthesis in novae and the effects of $\beta^+-$unstable nuclei and their effects on nova eruptions
(Truran 1985). I will touch only lightly on these topics.

The $\sim$2–5 magnitude outbursts of dwarf novae, repeating on time scales of weeks to years,
are seen only in low mass transfer-rate binaries. They probably occur when a recurring disk
instability rapidly causes much of the accretion disk to be dumped onto the white dwarf.
The significant release of gravitational potential energy manifests itself as an erupting dwarf
nova. Dwarf novae are briefly considered as metamorphic forms of classical nova systems
within the context of the hibernation theory ($\S$VI.1). Finally, the evolution of binary stars
to and from the cataclysmic stage, and the reason for the total lack of cataclysms in the
2–3 hour period range will be discussed in a forthcoming review by Livio (1989).

II. THE PRE-OUTBURST STAGE

II.1 RADIAL ACCRETION OF HYDROGEN-RICH MATERIAL

ONTO A WHITE DWARF

The first generation of TNR models (Starrfield, Sparks and Truran 1974a; 1974b; Pri-
alnik, Shara and Shaviv 1978, 1979) were based upon hydrogen-rich envelopes already in
place on the white dwarf. While these models succeeded in reproducing the energetics and
early light curves observed in classical novae, it was clear that the assumed envelope masses
and thermal structures were not necessarily self consistent. Numerical simulations of the
envelope accretion and evolution were first carried out by Kutter and Sparks (1980), Narai,
Nomoto, and Sugimoto (1980), and by Prialnik et al. (1982).

The most important parameter in determining whether a TNR occurs was found to be the
mass accretion rate $M$. Rapid mass accretion ($M > 10^{-9} M_\odot \text{yr}^{-1}$) produces significant
compressional heating at the hydrogen-rich envelope base. A mild TNR occurs because the
accreted mass is not very degenerate, and the envelope puffs up into a red giant-like
configuration. Only for $M < 10^{-9} M_\odot \text{yr}^{-1}$ is there sufficient degeneracy built up at the
envelope base prior to ignition to result in a strong TNR. The recent inclusion of boundary
layer heating in accretion calculations (Shaviv and Starrfield 1988; Prialnik, Kovets and
Shara 1988) implies even a lower value, $M < 10^{-10} M_\odot \text{yr}^{-1}$ for a strong TNR (see $\S$II.4).
At first sight, these upper limits conflict with observations of “old” novae (systems 10–100 years after eruption). Patterson (1984) has compiled observations of old-nova brightnesses and distances. He finds that $M \sim 10^{-6} M_\odot \text{yr}^{-1}$ is implied by the observed brightnesses of most old novae. Yet if the white dwarfs in old novae continue accreting matter from their companions at $10^{-6} M_\odot \text{yr}^{-1}$, simulations indicate that they will not again erupt as classical novae. This is contrary to the findings of Ford (1978) who showed conclusively that classical novae must suffer hundreds or thousands of TNRs during their lifetimes. This is because $\sim 80$ novae/year are observed to erupt in M31 (see §VII.2). Over a Hubble time ($\sim 1.5 \times 10^{10}$ years) roughly $7.5 \times 10^{11}$ nova eruptions occur in M31. This is comparable to the number of stars in that galaxy. Close white dwarf-red dwarf binaries are only a small fraction of all stars in the Galaxy. If this is also true in M31, then each nova must erupt $\sim \frac{1}{3}$ times, where $f$ is the ratio (number of nova binaries/total number of stars). A precise value of $f$ is not known, but it is probably in the range $10^{-3} < f < 10^{-6}$.

The finding of all the numerical studies cited above, that $\dot{M} \lesssim 10^{-9} - 10^{-10} M_\odot \text{yr}^{-1}$ is crucial to permit nova recurrence, is central to the hibernation scenario of cataclysmic binaries described in §V.

II.2 DIFFUSION DURING PRE-NOVA ENVELOPE ACCRETION

Many theoretical studies have shown that CNO enrichment of the accreted hydrogen-rich envelope is required to produce a fast nova (Starrfield, Truran and Sparks 1978; Fujimoto 1982). There is, however, no observed correlation between CNO abundances in the ejecta and the nova speed class. Shara, Pirslnik and Shaviv (1980) found that at least five important physical parameters affect the speed class of a nova. These are the white dwarf mass $M_{\text{wd}}$, the hydrogen-rich envelope mass $M_{\text{env}}$, the envelope metallicity $Z_{\text{env}}$, the mass accretion rate onto the white dwarf $\dot{M}$, and the underlying white dwarf luminosity $L_{\text{wd}}$. These workers and MacDonald (1983) suggested that novae of different speed classes lie at different points in the $(Z_{\text{env}}, M_{\text{env}})$-plane. More recent work has shown that neither $Z_{\text{env}}$ nor $M_{\text{env}}$ is a free parameter that may be chosen arbitrarily, or even independently. Both are determined by the accretion phase preceding the outburst. In current thinking, the fundamental parameters which determine the speed class of a nova are $M_{\text{wd}}, L_{\text{wd}},$ and $\dot{M}$. Novae begin to fade (in the optical part of the spectrum) only after they’ve ejected most of their envelopes (Bath and Shaviv 1976). $M_{\text{wd}}, L_{\text{wd}}$ and $\dot{M}$ determine the white dwarf envelope mass at the time of eruption, and the strength of the eruption. This determines how quickly the envelope is ejected, and hence how quickly the optical nova fades (i.e., the speed class).

$M_{\text{wd}}, \dot{M}$ and $L_{\text{wd}}$ change as the cataclysmic binary evolves. Thus the character and frequency of nova outbursts should change in a given system over thousands of outbursts spanning many millenia. A self-consistent, nova binary evolution simulation over many outbursts has not yet been attempted (but see III.2 for the evolution through two outbursts).

Prialnik and Kovetz (1984) and Kovetz and Prialnik (1985) have proposed a simple mechanism by which CNO enhancement of the accreting hydrogen-rich envelope is achieved. This mechanism is diffusion of hydrogen into the outer core of the C-O white dwarf and C-O into the hydrogen-rich envelope from the white dwarf core. Ignition first occurs in a very hydrogen-poor, but not hydrogen-free region in the outer white dwarf core. The large amount of energy liberated as soon as ignition occurs causes convection to mix all the matter from the ignition zone ($Z \lesssim 1$) up to the surface ($Z \sim Z_{\text{red dwarf}} \sim Z_\odot$).
The profiles of heavy elements at the onset of TNRs for various models are extremely instructive. Figure 1 (from Kovetz and Prialnik 1985) shows that the sharp boundary between hydrogen-rich envelope and hydrogen-absent core is smeared out by the diffusion process. In all diffusion models, traces of hydrogen diffuse into an outer core layer comparable in mass to the accreted hydrogen-rich envelope. This results in CNO abundances in the convective envelope at the onset of the simulated nova outburst from \( Z = 0.08 \) to \( 0.40 \).

The diffusion calculations carried out by Kovetz and Prialnik (1985) coupled the diffusion equations, which include concentration, pressure, and thermal terms to the evolution equations for each timestep and mass shell. By using a coarse model grid in \( M, M_{\text{wd}}, \) and \( L_{\text{wd}} \) Kovetz and Prialnik (1985) were able to determine roughly the simultaneous dependence of \( M_{\text{env}} \) and \( Z_{\text{env}} \) on these parameters, as shown in Figure 2.

The white dwarf mass \( M_{\text{wd}} \) was found to be the single most important parameter in determining the accreted envelope mass \( \Delta m_{\text{acc}} \). The depth reached by hydrogen diffusing into the core is proportional to the accretion interval \( \Delta t_{\text{acc}} = \Delta m_{\text{acc}} M^{-1} \). Thus the total amount of mass \( \Delta m_H \) for which \( X > 0.005 \) is also strongly correlated with \( M_{\text{wd}} \). The ratio \( \Delta m_H/\Delta m_{\text{acc}} \), which directly determines \( Z_{\text{env}} \) is almost independent of \( M_{\text{wd}} \), but is strongly dependent on \( M \).

\( Z_{\text{env}} \) is mostly determined by the accretion rate \( \dot{M} \) in the sense that higher \( Z_{\text{env}} \) values result from lower accretion rates. Low \( \dot{M} \) permits hydrogen to diffuse for a longer time, and hence to a greater depth into the white dwarf core.

The accreted envelope mass \( \Delta m_{\text{acc}} \) increases with decreasing \( L_{\text{wd}} \), because low luminosity corresponds to low envelope temperature and correspondingly longer time to reach TNR. Contrary to previous simulations, Kovetz and Prialnik (1985) found that \( \Delta m_{\text{acc}} \) decreases with decreasing \( M \). This is because diffusion provides enough enrichment for a strong TNR in hotter regions than were possible without diffusion. Thus very fast novae should occur for a combination of high \( Z_{\text{env}} \) (low \( \dot{M} \)) and low \( M_{\text{env}} \).

The diffusion scenario predicts that all novae show some degree of CNO enhancement as a result of diffusion. It also explains, in a natural way, how CNO nuclei can be enhanced by very different factors in novae of different speed classes, as shown in Figure 2.

In summary, white dwarfs of different masses and luminosities, undergoing accretion at different rates require very different time intervals to reach TNR. Diffusion can produce the wide variety of observed \( Z_{\text{env}} \) in a wide range of \( M_{\text{env}} \). It is \( M_{\text{wd}}, L_{\text{wd}} \) and \( \dot{M} \) that determine \( Z_{\text{env}} \) and \( M_{\text{env}} \) at the beginning of a TNR, and hence the speed class of an erupting nova.

### II.3 SHEAR INSTABILITY MIXING OF ACCRETING MATERIAL

**POSSESSING ANGULAR MOMENTUM**

The hydrogen-rich material accreted by a white dwarf in a cataclysmic binary possesses very high specific angular momentum. There must exist an interface region between the outer part of the white dwarf core and the inner part of the accretion disk where material is transferred from the disk to the core. Important and difficult questions concerning this interface are:

1) How effective is shear mixing at transporting accreting material into the white dwarf core?

2) Are there processes that can transport matter away from the equatorial region of the white dwarf on a timescale short in comparison with that required to build up an envelope sufficiently massive to generate a TNR?
3) What, if any, are the effects of rotational energy dissipation on the thermal equilibrium of the white dwarf envelope, and of the resulting TNR?

4) Are there any observed or predicted consequences of shear mixing of accreted and core material?

Durisen's (1977) pioneering study of the accretion of a rotating fluid by a barytropic model of a white dwarf outlined many of the important physical principles and effects that must be considered. In particular, the effects of degenerate electron viscosity and various fluid instabilities that could lead to turbulence and enhanced transport of angular momentum were considered. Two particularly promising mechanisms outlined were the formation of an Ekman layer, and oscillating, non-equilibrium meridional circulations. The Ekman spin-up theory is particularly attractive and has been treated in some detail by Livio and Truran (1987) and is discussed at the end of this section.

Durisen (1977) concluded that the overall structure of an accreted envelope depends critically on the turbulent redistribution of angular momentum. A schematic model for the white dwarf in a cataclysmic binary is adopted from Durisen (1977) and shown in Figure 3. The star consists of a degenerate core in slow, nearly rigid rotation; a mostly degenerate, relatively narrow, turbulent, high-shear region; and a nearly Keplerian envelope which may or may not have a substantial degenerate subregion.

Durisen (1977) suggested that the nonsphericity of old nova shells (McLaughlin 1960; Warner 1972; Hutchings 1972) could be due to TNRs in nonspherical envelopes of white dwarfs. This remains a distinct possibility, though interaction of the expanding shell with the accretion disk and/or the binary companion, magnetic fields, or localized TNRs are also possible. Durisen (1977) also pointed out that the high-shear region could provide core-envelope mixing and the CNO enhancements required by the Starrfield, Sparks, and Truran (1974) nova models.

Kippenhahn and Thomas (1978) (KT) claimed that accretion belts do form on accreting white dwarfs. They stated that the Richardson criterion, in which the effects of temperature, angular velocity and molecular-weight gradients are taken into account, is the condition that determines the onset of shear-mixing.

In the presence of a gravitational field, turbulence is hindered by a density distribution increasing towards the center of gravity. To maintain turbulence caused by shear motion, the kinetic energy of turbulent elements must be larger than the work done by those elements against gravity and centrifugal force. This is quantified by the Richardson number $R_i$ (for details see Prandtl 1942): $R_i = \text{work done against buoyant and centrifugal forces/kinetic energy of turbulence.}$ Marginal stability occurs in the case $R_i = 0.25$. KT noted that if the timescale of accretion is long compared to the timescale during which the instability can mix material, then the involved layer is nearly marginal. Freshly accreted material causes further exchange of angular momentum and of hydrogen between the disk and the white dwarf. This makes the last accreted layer only slightly unstable. The bottom of the accreted layer moves deeper into the white dwarf and nuclear burning eventually starts.

An important simplification used by KT was that the condition for marginal stability was satisfied separately in the radial and horizontal directions. Furthermore, the existence of the belt was assumed not to perturb the structure of the white dwarf. The main numerical result of Kippenhahn and Thomas (1978) was that the accretion belt was found to extend to latitudes of $\pm 21^\circ$, roughly independent of depth.

Macdonald (1983) studied the effects of instabilities due to differential rotation and chemical inhomogeneity in the envelopes of accreting stars. By considering growth rates of
nonaxisymmetric perturbations MacDonald (1983b) claimed that freshly accreted material rapidly suffers dynamical instabilities and covers the entire white dwarf. Mixing inwards, however, towards the white dwarf core occurs on a much longer (thermal) timescale.

The ratio of soft X-ray flux to visual flux for many old novae is several orders of magnitude less than predicted by Pringle's (1977) study of the boundary layer between the accretion disk and white dwarf (Verland et al. 1982). Perhaps the simplest solution to this problem (suggested by Verland et al. 1982) is that of a boundary layer comparable to the size of the white dwarf rather than a thin belt at the white dwarf's equator. A larger surface area permits a lower effective temperature in the boundary layer for the dissipation of rotational energy. Unobservable EUV radiation, rather than soft X-rays, would then be emitted. MacDonald (1983b) suggested that shear-induced turbulence spreads the thermal energy (produced by the dissipation of rotational energy) evenly over surfaces of constant effective gravity on a dynamical timescale. "The missing boundary layer" becomes, in this interpretation, evidence that shear-induced turbulence mixes accreted matter on a dynamical timescale over the entire white dwarf.

Sparks and Kutter (1987) and Kutter and Sparks (1987) carried out extremely detailed numerical simulations of accretion and thermonuclear runaways on white dwarfs. They too assumed that the angular momentum distribution of accreted material on a white dwarf was in marginal stability and was determined by the Richardson criterion. The distribution of mass they considered had a half width of 27.5° at half intensity, covering 42.3% of the white dwarf's surface. This is more of a saucer than a belt. In all cases studied, Sparks and Kutter (1987) failed to obtain a sufficiently strong nuclear runaway for nova-like mass ejection. In their radial accretion models gravitational compression heating triggered the TNR too soon. When they included the effects of angular momentum on the accreted matter, the support due to the centrifugal force lowered the mechanical confinement to the burning regions too much to permit a strong TNR and mass ejection.

Sparks and Kutter (1987) considered several possibilities to increase the pressure at the base of the accreted material during TNRs. One suggestion was that only white dwarfs significantly more massive than 1 M⊙ can generate classical novae. However, the estimated masses of white dwarfs in nova binaries is generally 1 M⊙ or less (Warner 1976, Ritter and Burkert 1986; but see Livio and Soker 1984). Another possibility is that the angular momentum of the accreted matter might be removed by tidal torques' braking effects exerted by the disk or by the secondary star, or possibly by mechanical coupling between the accreted matter and the disk. This could permit mass to penetrate the core without carrying in angular momentum. The third possibility is the interaction (during a TNR) of convection with layers of material whose horizontal velocities have a gradient.

Livio and Truran (1987) considered the generation of a very thin turbulent layer on a white dwarf surface dynamical time scale (seconds) as soon as the accretion disk first comes into contact with the white dwarf. The conclusion that on a turbulent, viscous time scale an almost rigidly rotating thin layer spreads over the entire white dwarf surface. As the thin rotational shear layer (the "Ekman layer") becomes turbulent it produces another shear layer, with the process repeating itself. This mechanism transfers angular momentum towards the poles and to the body of the white dwarf much more efficiently than the Kippenhahn-Thomas model. Spreading the angular momentum over the white dwarf's surface and into the core reduces the centrifugal force in the outer layers and precludes nova-like TNRs in the Sparks and Kutter scenario.
Researchers in the field are still debating whether diffusion or shear mixing is the dominant mechanism in producing CNO and heavy element enrichments in nova ejecta. A simple and elegant test has been proposed by Livio and Truran (1987). The test is based on the conclusion of Webbink et al. (1987) that the outbursts of the recurrent novae T Pyx and U Sco (and by implication V394 CrA) result from TNRs on the surfaces of very massive white dwarfs. Because of the short recurrence times of these objects (1–2 decades) diffusion (which requires centuries for significant mixing) cannot be operative in them. Any-heavy element abundances seen would have to be due to shear mixing. Conversely, a complete absence of heavy elements in the ejecta spectra of all three recurrent novae would imply that shear mixing is rather unimportant on all nova white dwarfs. Diffusion would remain the contender as the enrichment mechanism in non-recurrent novae.

T Pyx is due for an eruption in the next few years and should be carefully monitored spectroscopically for heavy element enrichments. U Sco is not enriched in CNO abundances relative to the Sun (Williams et al. 1981). However, it may be heavily enriched in helium. Shear mixing might have occurred with a helium-rich layer left behind from the previous outburst. The observed high ejection velocities remain unexplained.

A photometric and spectrographic study of the second known eruption of V394 CrA has been reported by Sekiguchi (1988). The amplitude of the outburst (ΔV ≈ 11 magnitudes) is comparable to that of classical nova. The decline rate (3 magnitudes in 5 days), high ejection velocities and spectra resemble those of U Sco.

II.4 THE EFFECTS OF HEATING FROM THE BOUNDARY LAYER
ON ACCRETION MODELS FOR NOVAE

Once matter in the accretion disk of a cataclysmic binary reaches the environs of the white dwarf its dynamical state must change quickly. While in the accretion disk the material orbits the white dwarf with Keplarian velocity. There must be strong thermal and dynamical interaction between the outer part of the white dwarf and inner part of the accretion disk in order for accreting material to transfer from the latter to the former. The structure of this boundary layer has not yet been solved. However, Shaviv and Starrfield (1987) have made the point that boundary layer heating of the outer white dwarf layers must be important, and must be included in accretion models of cataclysmics.

The virial theorem implies that at least half the boundary layer luminosity of material accreting onto a white dwarf is radiated, while up to half is transported into the surface layers of the white dwarf by the accreted material. The effects of this accreted energy significantly affect the evolution to a TNR and nova outburst. In particular, Shaviv and Starrfield (1987) found that inclusion of boundary layer energy makes it more difficult to produce a nova outburst for a given set of parameters (M, M_{white dwarf}, L_{white dwarf}) usually chosen in studying novae. This is because boundary layer heating was found to produce a nuclear burning region extending through all or much of the accreted envelope. This burning forced the entire accreted envelope to become convective. Because the degeneracy of material in the envelope was significantly lowered by the heating, the strengths of TNRs were also dramatically reduced.

Shaviv and Starrfield (1987) found their results to be strongly dependent on the degree of boundary layer heating and on the ratio of mixing length to scale height. In the words of the authors... “we have complicated an already cloudy situation”. Nevertheless, it is clear that heating from the boundary layer must be included in realistic accretion models.
The reader should not be misled into thinking that TNRs are impossible to produce when the effects of boundary layer heating are included. We note that both the $1.35M_\odot$ accreting models of Starrfield, Sparks and Shaviv (1988) and the $1.1M_\odot$ iron white dwarf models of Prihnik, Kovetz, and Shara (1988) produce strong TNRs even when boundary layer heating was included. The latter authors concluded that the maximum $\dot{M}$ that could be tolerated, and still produce a strong TNR on a white dwarf of given mass, had to be decreased by about an order of magnitude because of boundary layer heating. Thus accretion rates $\lesssim 10^{-10}M_\odot\text{yr}^{-1}$ are theoretically necessary for TNRs when boundary layer heating is considered.

III. THE OUTBURST STAGE

III.1 LOCALIZED THERMONUCLEAR RUNAWAYS (LTNR)

Most studies of thermonuclear runaways on the surfaces of white dwarf stars have made the simplifying assumption that the runaways are completely spherically symmetric. This is a reasonable first approximation in trying to understand what the dominant physical processes are in the runaway. However, even simple estimates (described below) show that TNRs usually start at a point somewhere in the surface layers of an accreting white dwarf star. How likely it is for the thermonuclear runaway to spread from the point, and how quickly spreading can occur are much more involved problems.

The first detailed study of localized thermonuclear runaways was conducted by Mitrofanov (1980), who was concerned with TNRs on magnetic white dwarfs. Mitrofanov suggested that the outbursts of dwarf novae result from the unstable thermonuclear burning of hydrogen on the surfaces of magnetic degenerate dwarfs. He noted that with a magnetic field of $10^6-8$ gauss a white dwarf would be able to confine hydrogen-rich plasma to accretion columns. A temperature inhomogeneity on an accreting white dwarf's surface is then generated as accretion heats only a small part of the surface. Furthermore, a strong magnetic field leads to heterogeneities in the heat conductivity. A typical area exposed to accretion was crudely estimated to be about $10^{14-15}$ cm$^2$, or a few percent of the surface area of the white dwarf.

Mitrofanov pointed out that a competition takes place between the rate of local accumulation of material at a magnetic pole and the rate at which material spreads away from the pole over the surface of the white dwarf. If local accumulation is rapid then ignition might occur in a small area on the surface of the white dwarf. He estimated the mass of burning hydrogen to be approximately $10^{22}$ to $10^{24}$ grams. Thus if $10^{15}$ ergs gm$^{-1}$ are released during the TNR (an amount of energy per gram similar to that released in a nova eruption) then the total dwarf nova outburst energy is about $10^{39}$ to $10^{40}$ ergs. This agrees with the energetics of dwarf nova outbursts.

The recent demonstration that Nova Cygni 1975 is a strongly magnetic white dwarf (see §VI.2) has cast serious doubt on the picture of Mitrofanov. The magnetic field of Nova Cygni 1975 did not confine most of the hydrogen to small areas at the polar cap of the star. The total energetics of Nova Cygni demand that about $10^{-5} M_\odot$ of hydrogen was accumulated on the white-dwarf surface without being burnt to helium ($10^{-6} M_\odot$ at a time) in successive dwarf nova eruptions. While Mitrofanov's hypothesis of dwarf nova eruptions being caused by thermonuclear runaways is probably incorrect, his paper was the first to point out that
localised thermonuclear runaways should occur on white dwarf stars under certain special circumstances.

The next studies of localised runaways were carried out by Shara (1982) and by Fryxell and Woosley (1982). Shara (1982) considered the energy transport inside a degenerate hydrogen rich envelope on a white dwarf. He demonstrated that only when thermal conductivity is the dominant mode of energy transport inside a hydrogen rich envelope can a local TNR occur. The thermalisation time scale between adjacent mass elements $\Delta R$ in size is

$$t_{TH} \sim 10^4 \left[ \frac{\Delta R}{10^8} \right]^2 \left[ 1 + 0.45 \left( \frac{T}{10^9} \right)^2 \left( \frac{10^5}{\rho} \right)^2 \right]^{-1} \left( \frac{10^8}{T_F} \right) \text{years,}$$

where $T$ is the temperature, $\rho$ is the density and $T_F$ is the Fermi temperature. For typical white dwarf envelope values ($\rho \sim 10^4 \text{ g cm}^{-3}$ and $T \sim 15-20 \times 10^6 \text{ K}$) a localised TNR is able to occur long before the envelope can reach thermal equilibrium.

A TNR should start in the initially hottest part of the envelope covering only a small part of the white dwarf surface area. In some cases neither conductive, convective, radiative, advective, nor acoustic transport of energy from the site of the LTNR seem capable of heating all or much of the surrounding hydrogen-rich envelope on a time scale comparable with the eruptive phase of the TNR.

Temperature inhomogeneities are crucial in making a localized rather than a global TNR. Four mechanisms for setting up temperature gradients were suggested by Shara (1982). These are the degenerate dwarf's magnetic field, rotation, non-spherically symmetric accretion of matter, and the red-dwarf tidal force.

As discussed in §II.3 the accretion of matter by the white dwarf is certainly not spherically symmetric. Matter is accreted via an accretion disk, hence the deposition of matter and turbulent energy on the white dwarf outer layers occurs in a toroidal region around the equator of the degenerate dwarf. More compressional (PdV) work is done on material in this toroidal region than on material at higher latitudes on the white dwarf surface. Temperature differences of a few percent or even larger over regions $10^{-8}$ cm apart are straightforward to produce. Hydrogen-rich envelopes must be massive for local TNRs to occur because only then does thermal conductivity dominate radiative energy transport.

If a localised runaway does begin in the envelope of a degenerate dwarf, what are the logical consequences? In particular, will a localised runaway remain localized, or will it spread in a short time to engulf the entire dwarf envelope?

Once a TNR (local or global) starts on a white dwarf, radial transport of energy is convective. Determining the method of meridional transport of energy is much more difficult. Because the dominant energy transport mechanism is not certain, neither is the timescale of energy transport. As a result, it's not certain how long localized runaways remain localized. In the absence of turbulent mixing or horizontal convection, Shara (1982) claimed that it would take years for a diffusively propagated burning wave to cover the surface of a white dwarf.

A localised TNR has a built in "safety valve" not available to a spherically symmetric TNR. In the latter case, the only direction in which matter can expand is radially. In the localised case, horizontal expansion is also possible. In particular, as material expands quasi-hydrostatically out of a cylindrically shaped LTNR, that material must spread horizontally, away from the site of the localized runaway. This is sketched in Figure 4. A volcanic outpouring of hot, hydrogen-helium "lava" onto the white dwarf surface is the obvious analogy. Convective entrainment and turbulence were not included in these qualitative scenarios, and
any estimate of the spreading time scale would depend on these effects. Fryxell and Woosley (1982) suggested that these processes spread the runaway over the entire dwarf surface in days or weeks.

Orio and Shaviv (1988) have recently extended the analytic estimates made by Shara (1982) and by Fryxell and Woosley (1982). They included the accretion process of the hydrogen-rich material onto the white dwarf and they extended the calculations to areas comparable to the white dwarf surface. The most important physics introduced by Orio and Shaviv (1988) is consideration of the fraction of the radiative flux $\eta$ distributed in the meridional direction. By taking into account the curvature of the white dwarf's surface, Orio and Shaviv (1988) also allowed for temperature inhomogeneities on scales of the order of the radius of the white dwarf.

The general trends found confirm many of Shara's (1982) results. In particular, the result that TNRs which start at a point can remain local is confirmed and extended by Orio and Shaviv's (1988) work. They find that for a minimum temperature differential $\Delta T$ a very small TNR will spread into a somewhat larger one, but not quickly cover the entire white dwarf surface.

Numerical simulations indicate that Li'NRs are most likely to occur on white dwarfs accreting at high rates, on hot massive white dwarfs, or on cool intermediate mass white dwarfs. Meridional transport of energy lengthens the propagation time of a local TNR. A months or years-long brightening before the rapid, dramatic rise to maximum light is frequently observed in classical novae. Orio and Shaviv (1988) suggest that this rise to maximum light might be due to meridional transport of energy slowing down the TNR. The pre-nove EUV radiation from the white dwarf could cause weeks or months of heating and brightening of the red dwarf (seen in visible light).

Orio and Shaviv (1988) have demonstrated how complex LTNRs can be; until convection and turbulence are included in realistic, three dimensional simulations of local TNRs, the effects and behavior of LTNRs on white dwarfs will remain rather poorly understood.

III.2 THE EVOLUTION OF A CLASSICAL NOVA MODEL THROUGH A COMPLETE CYCLE

The simulated evolution of even a single thermonuclear runaway on a white dwarf surface can take hours of time on a large computer. The coupled equations of hydrodynamics and nuclear energy generation and transport, diffusion, and mass accretion make any attempts at accurate analytic solutions hopeless.

Until very recently, the complexities of the numerical simulations have prevented even the simulation of two successive runaways on the same white dwarf. Prialnik (1986) has now followed the evolution of a classical nova through a complete cycle. The accretion, outburst, mass loss, decline, and resumed accretion, leading to a second outburst, were all calculated with a hydrodynamic code, which included diffusion and an extensive nuclear reaction network. The model's luminosity history is shown in Figure 5, and that of the envelope is shown in Figure 6. The white dwarf chosen was a 1.25 M$_\odot$ C-O star, while the accretion rate was taken to be $M = 10^{-11}M_\odot yr^{-1}$. Matter with close to solar composition was accreted onto the white dwarf. The exact parameters of the model are less important than the result of the simulation which is that the second accretion phase resembled the first one very closely, and the second outburst was almost identical to the first. While this has been suspected by many workers in the field for many years, this work was the first
clear demonstration of the correctness of this statement. The implications for the long term evolution of classical novae, and of successive nova eruptions, are extremely important.

Many of the results of Prialnik (1986) of the first stage of accretion through outburst and decline have already been seen in previous TNR simulations. The low accretion rate chosen permitted significant amounts of diffusion of hydrogen-rich material into the C-O white dwarf and the resulting enrichment over solar (by an order of magnitude) in Z lead to a fast nova. The composition of the ejecta and the light curve and ejection velocities were in good agreement with e.g. Nova Cyg 1975 and Nova Cyg 1978. Within a few centuries, the white dwarf cooled to its preoutburst structure, with one very significant difference. A low-mass but helium-rich envelope was left on the white dwarf. It was this layer which raised the helium abundance of the ejected matter of the next nova outburst (see the next section).

Because of diffusion and the resulting enrichment in CNO nuclei of the accreted hydrogen-rich envelope, the white dwarf mass was found to decrease slightly. In addition, the red dwarf mass decreased secularly because of mass transfer to the white dwarf. Thus the orbital period, which depends only on the secondary mass (Patterson 1984) should decrease. In the event of very low mass-rate accretion, like that considered in Prialnik’s work, the white dwarf’s intrinsic luminosity decreases steadily and the occurrence of flashes does not prevent the normal cooling of the white dwarf. However, for higher mass accretion rates, where heating from the boundary layer must be included, the opposite may be true. Whether the white dwarf luminosity increases or decreases is critical to understanding the long term evolution of a nova binary. As Prialnik (1986) notes “a significant decrease in the white dwarf temperature would result in an appreciable increase of the accretion time prior to outburst and, consequently, larger envelopes and ejected masses.” The opposite statement is also true. Thus, despite this important first step in understanding the long term evolution of classical novae, a simulation of several successive outbursts which include a self-consistent determination of the mass transfer rate throughout the outburst cycles is still needed to determine the long-term trends in nova eruption behaviour.

III.9 THE HELIUM/HYDROGEN RATIO IN THE EJECTA OF CLASSICAL NOVAE

Highly enhanced CNO abundances are not always required to produce a TNR and a classical nova (Prialnik, Shara, Shaviv 1978, 1979; Sparks, Starrfield and Truran 1978; Narai, Nomoto and Sugimoto 1980). Observations of the spectra of novae ejecta, and many TNR simulations have shown that a variety of nova speed classes can be produced by a wide variety of CNO isotope abundances. We note that various combinations of large overabundances, relative to the Sun, of helium, carbon, nitrogen, oxygen, neon, magnesium and perhaps iron have all been seen in various novae (see Truran and Livio 1986 for a recent summary).

There seem to be only three reasonable ways to create this overabundance (Prialnik 1988). These are:
1) the accreted matter must itself be extremely rich in helium, CNO, and perhaps heavier isotopes; or
2) during the outburst, the heavy elements are created by nucleosynthesis; or
3) the accreted and core materials are mixed by diffusion and/or shear mixing during the accretion process.

The mass transferring red dwarfs in cataclysmic binaries in the solar neighborhood should be composed of solar-like material. Furthermore, no indication of abundance peculiarities in
the accretion disk spectra of nova-like variables (widely believed to be nova progenitors) has ever been reported. Thus the first possibility can be discounted. Nucleosynthesis of large amounts of heavy isotopes can also be discounted. Less than 1% of the hydrogen envelope need be burnt in order to provide the energy of a nova outburst.

If mixing is the only viable explanation for high metallicity and the violent mass ejection seen in novae, and heavy isotope production during thermonuclear burning is not important, then the ratio of He/H should be that of the accreted material, i.e., solar. The observations, recently summarized by Prialnik (1988) show that enhanced He/H ratios are very common. Three outstanding examples are RR Pic (nova 1925), DK Lac (nova 1950), and HR Del (nova 1967). In each of these three objects the ratio Y/X is approximately 1.0, while the solar value is 0.34. An inescapable conclusion seems to be that the outer layers of the white dwarf are very rich in helium.

A significant complication to this simple picture, however, is that large CNO excesses usually accompany the helium excesses. Thus mixing must be able to bring hydrogen-rich material down through the top helion layer into the white dwarf C-O core. The helium layer thickness must, therefore, be comparable to the hydrogen-rich envelope mass, i.e., of order $10^{-6} - 10^{-4}M_\odot$. A thin primordial shell of helium would be quickly eroded by just a few nova outbursts, and would almost never be seen as enriched helium in the ejecta of most novae.

An elegant alternative mechanism, proposed by Prialnik (1988) is that a thin helium-rich mantle forms just after each nova outburst phase. This occurs because only a fraction of the hydrogen-burning envelope is ejected during each nova eruption. Some fraction of the mass, $\lesssim 30\%$, is left behind and much of this material burns to helium. Such behaviour has been demonstrated in TNR simulations which eject mass. The mass accreted before the next eruption constitutes between 40–80% of the total envelope mass. Diffusion, and/or shear mixing enrich the envelope both in helium (left over by the last eruption) and in C-O or O-Ne-Mg from the white dwarf.

In the case of a C-O white dwarf, most of the carbon dredged up from the core is burned into nitrogen by the eruption. The TNR temperatures are generally too low for oxygen to be affected (Starrfield et al. 1978). The C/O and high He/H ratios observed in novae ejecta provide strong support of the scenario of the depletion and regeneration of a thin helium shell during each nova cycle.

III.4 THERMONUCLEAR RUNAWAYS ON THE SURFACES OF OXYGEN-NEON-MAGNESIUM WHITE DWARF

As long as the temperature in the hydrogen-rich envelope of a white dwarf, which is undergoing a thermonuclear runaway, does not exceed $4 \times 10^8 \, ^\circ K$ “breakout” from the CNO cycle does not occur (Truran 1982; Wiescher et al. 1986). The C, N, and O isotopes present act only as catalysts; they are neither created nor destroyed in the CNO cycle burning. Only when $T > 4 \times 10^8 \, ^\circ K$ can the Ne-Mg-Al-Si group of nuclei be created (i.e., breakout). No single published TNR model to date, including those dealing with white dwarfs approaching the Chandrasekhar mass, reaches or exceeds $4 \times 10^8 \, ^\circ K$. Thus, any element heavier than oxygen, seen in nova ejecta, must either have been transferred into the white dwarf envelope from the red dwarf companion, or somehow have been dredged up from the underlying white dwarf.
One of the first quantitative reports of high element abundance was that by Ferland and Shields (1978) of a twenty-fold enhancement of the abundance of neon relative to solar in the ejecta of nova Cygni 1975. Since then, even more dramatic enhancements have been reported by Snijders et al. (1984) for nova V1370 AQL 1982, and by Williams et al. (1985) for nova V693 CrA 1981 (see Figure 7). Andriailat and Houziaux (1985) observed very strong Mg II emission both in IUE and optical spectra, for the nova Vul 2 1984.

Ferland and Shields (1978), Law and Ritter (1983), and Williams et al. (1985) have all suggested that the high neon abundances seen in these novae imply the existence of underlying O-Ne-Mg white dwarfs. Just as the high abundances of C, N, and O in many nova ejecta imply the ablation of the underlying white dwarf, so does the presence of oxygen-neon-magnesium enriched material imply the erosion by the nova process of the underlying white dwarf.

Stellar evolution calculations predict that stars with initial masses in the range 1–8 $M_\odot$ will evolve to yield C-O white dwarfs. Those of initial mass 8–12 $M_\odot$ will evolve further to yield ONeMg white dwarfs (Barkat, Reiss and Rakavy 1974). A simple Salpeter (1955) initial mass spectrum predicts the eventual formation of 35 C-O white dwarfs for every ONeMg white dwarf. Iben and Tutukov (1984, 1985) arrived at a similar ratio, after including possible effects of binary evolution and mass transfer onto the white dwarf.

Because at least four of the dozen recent brightest novae have exhibited strong neon and/or magnesium enhancements, the frequency of oxygen-neon-magnesium white dwarfs appears significantly larger than that predicted by the above theoretical considerations. However, Truran and Livio (1986) have shown that selection effects are very important in determining which kinds of white dwarfs are seen in nova eruptions. ONeMg white dwarfs may be more massive on average than C-O white dwarfs, and, therefore, are predicted to erupt with lower mass envelopes. If this assumption is correct then ONeMg white dwarfs accumulate critical mass envelopes and suffer nova eruptions at shorter recurrence intervals. Thus, despite the fact that ONeMg white dwarfs occur in only ~ 3% of all cataclysmic binaries, their large masses ensure that they account for ~ 1/3 of all observed classical novae.

The first hydrodynamic simulations of oxygen-neon-magnesium novae have been carried out by Starrfield, Sparks and Truran (1986). These authors studied thermonuclear runaways on the surface of a 1.25 $M_\odot$ white dwarf. Accretion rates of $10^{-8}$–$10^{-10} M_\odot$ yr$^{-1}$, as well as different chemical compositions in the accreted matter were tested. Mixtures with $O^{16}$ abundances of 0.25 and 0.50 were used to simulate accretion onto oxygen-neon-magnesium white dwarfs. The results of these simulations were extremely violent TNRs and nova models which exceeded the Eddington luminosity $L_{\text{Edd}}$ at maximum. Ejection velocities of over $4 \times 10^3$ km s$^{-1}$ were achieved. While the ejection velocities are higher than those in C-O white dwarf models, the ejected masses are significantly lower.

More realistic accretion scenarios, including diffusion and shear mixing, should be attempted to try to answer some of the remaining difficulties with the oxygen-neon-magnesium white dwarf scenario. In particular, is it possible to make a nova eject roughly comparable fractions of hydrogen, helium, CNO and neon as seen in V693 CrA (nova 1981)? At least as puzzling is the appearance of significant enhancements of iron relative to solar in several recent novae. As Starrfield (1987) has emphasized, high iron abundances are not predicted by TNR scenarios on either C-O or ONeMg white dwarfs. A first attempt at explaining these high-iron observations has been made by Prialnik, Kovetz and Shara (1988). The results are described in the next section.
III.5 THERMONUCLEAR RUNAWAYS ON THE SURFACES OF
IRON WHITE DWARFS

As described above, the large enrichments seen in several novae of oxygen, neon and
magnesium, relative to solar, have suggested a population of ONeMg white dwarfs and
cataclysmic binaries. Significant overabundances of iron have also been seen (in N Muscae
1983 and N Aquilae 1982). Conventional thermonuclear runaway theory does not predict
the formation of large amounts of iron during nova eruptions. This has prompted Prialnik,
Kovetz, and Shara (1988) to study accretion onto, and nova events on the surfaces of Fe
white dwarfs. Half the free-fall energy of the accreted matter was assumed to be transmitted
to the accreting star via the accretion disk boundary layer (a heating process ignored in most
previous simulations). Diffusion between H, He, CNO, and Fe was taken into account.

The material which was accreted onto the iron white dwarf was assumed to be of solar
composition. During the accretion phase hydrogen was found to diffuse deeper into the Fe
core than its CNO catalysts. Thus a TNR did not begin at the deepest levels penetrated
by the hydrogen, but further out where the CNO abundance was highest. This natural
enhancement of the CNO abundances does serve to trigger TNRs in every case studied, but
these overabundances were smeared out by convection as soon as the TNR began. Furthermore,
accretion rates \( \dot{M} \) of \( 10^{-9}, 10^{-10}, \) and \( 10^{-11} \, M_\odot/\text{yr} \) produced qualitatively different
runaways because of different envelope masses and thermal structures at the time of TNR.

Large iron enhancements (over solar) were obtained, particularly at \( \dot{M} = 10^{-10} \, M_\odot/\text{yr}^{-1} \). Generally, a high degree of contamination of nova ejecta by core material is obtained either
from rapid accretion onto a non C-O white dwarf or from slow accretion onto a C-O white
dwarf (if diffusion followed by convection is the mixing mechanism). In the former case, this
is because the admixed core material does not participate in the thermonuclear burning,
and because hydrogen will only diffuse to a Lagrangean mass depth in the white dwarf core
of a few times the mass of the hydrogen-rich accreted shell. A long period of slow mass
accretion yields a more massive envelope, but not one more enriched in iron. In the case of
a C-O white dwarf, the admixed carbon and oxygen enhance the TNR so slow accretion is
preferable to obtain a high degree of C-O enhancement.

Mass loss from the iron white dwarf was found to occur in two distinct stages. The
episodes of mass loss occurred several days or weeks apart as the erupting star expanded,
ejected matter, contracted and then expanded again, shedding more mass. RR Pic (nova
1925) showed multiple mass loss episodes, and probably was iron-rich (Payne-Gaposchkin,
1957).

The maximum temperature obtained—\( 2.85 \times 10^8 \, \circ K \) — was not high enough for heavy
elements (CNO, Ne) to be produced. Thus there was no breakout from the CNO cycle. The
authors concluded that, in order to account for the large iron abundances detected in the
ejecta of some classical novae, within the framework of conventional TNR theory, the white
dwarf progenitor should be rich in iron as well as CNO and Ne.
III.6 THERMONUCLEAR RUNAWAYS ON THE SURFACES OF
WHITE DWARFS NEAR THE CHANDRASEKHAR LIMIT

Recurrent novae have been exhaustively reviewed by Webbink, Livio, Truran and Orio (1987). These authors concluded that the only two near-certain thermonuclear runaway powered recurrent novae were TY Pyx and U Sco. The extraordinarily short recurrence times of these two novae, (eight years in the case of U Sco) imply that the underlying white dwarfs must be more massive than \( \sim 1.3 M_\odot \) (Nariai and Nomoto 1979; Fujimoto 1982; MacDonald 1983) if it really is a TNR that powers these objects. This is because the high surface gravity of a massive white dwarf can produce a TNR with a much lower envelope mass (which can be accreted quickly) than that on a low mass white dwarf.

The first simulations of TNRs on the surfaces of white dwarfs near the Chandrasekhar limit were carried out by Shara (1981) to see if “breakout” from the CNO cycle could be obtained (i.e., if Ne or Mg nuclei could be created during hot hydrogen burning). Temperatures of \( 3 - 3.5 \times 10^8 K \) were the maximum found in these simulations, and “breakout” (which requires \( T \sim 4 \times 10^8 K \)) did not occur.

Starrfield, Sparks, and Truran (1985) studied a \( 1.38 M_\odot \) white dwarf accreting matter with solar composition at \( 1.7 \times 10^{-9} M_\odot \text{yr}^{-1} \) and \( 1.7 \times 10^{-9} M_\odot \text{yr}^{-1} \). In the former case, only \( \sim 33 \) years were required to reach the peak of a TNR which ejected 6\% of the accreted envelope. The remaining material burned to helium in \( \sim 2 \) years and settled back onto the white dwarf. The light curve, accretion time required to reach outburst, and quantity of mass ejected during this simulation agreed well with the observed outburst properties of nova U Sco 1979.

The lower accretion rate simulation took over \( 10^3 \) years to reach the burst phase of the evolution. Again, only a small fraction (13\%) of the accreted envelope was ejected at slow velocities. Thus in both cases the mass of the white dwarf was found to grow toward the Chandrasekhar limit.

To try to explain the extraordinary eight year recurrence time of U Sco (eruptions observed both in 1979 and in 1987), Starrfield, Sparks and Shaviv (1988) simulated the accretion of \( 1.1 \times 10^{-6} M_\odot \text{yr}^{-1} \) onto a \( 1.35 M_\odot \) white dwarf. Incredibly, a TNR was achieved in only \( \sim 2.6 \) years, and it was able to eject \( 4 \times 10^{-7} M_\odot \) moving in excess of \( \sim 400 \) km s\(^{-1}\). The same important conclusion as before was reached as was the ultimate fate of the white dwarf. Its mass is inexorably increasing, as only a small fraction of the accreted envelope is ejected during each eruption. Thus, the authors predict that U Sco will absorb enough mass in a rather short length of time (\( \sim 10^5 \) years) to exceed the Chandrasekhar limit, and will erupt as a type I supernova.
IV. THE EJECTION PHASE

IV.1 THE STEADY STATE, CONSTANT LUMINOSITY, CONTINUOUS

WIND EJECTION MODEL OF NOVAE

It is a remarkable fact that all erupting novae display the same sequence of spectral behavior as the light declines (McLaughlin 1960). While it is certainly true that individual novae display unique characteristics in their outburst properties (perhaps due to varying binary separation, masses, abundances, etc.) the overall similarities are so strong that some universally applicable physical conditions must hold in the ejecta of novae. This basic philosophy has been adopted by Bath and Shaviv (1976) in their study of the outflowing photosphere of a classical nova.

All novae near peak brightness display the approximately Eddington luminosity of a \( \sim 1-2M_\odot \) object. This led Bath and Shaviv (1976) to suggest the following set of conditions. Radiation pressure in the optically thick continuum of the envelope was assumed to continuously accelerate matter away from the underlying white dwarf. Electron scattering was taken to be the dominant opacity mechanism. The total bolometric luminosities of erupting novae were assumed to be constant in time throughout the optical outburst. (UV, infrared and X-ray observations have since confirmed this suggestion). Finally, the A-F supergiant-like spectra of novae near maximum light (with \( L \sim 10^4L_\odot \)) imply photospheric temperatures \( \sim 10,000^\circ K \) and thus radii of \( \sim 10^{12} \) cm. Such radii are \( \sim \) ten times larger than the binary separation between the white and red dwarfs. Thus frictional dissipation must result from binary motion within the common outflowing envelope. The optical decline of a nova was found by Bath and Shaviv (1976) to be due to an increasing bolometric correction as the wind thinned out, and the photosphere receded to hotter, deeper layers.

On the basis of their model Bath and Shaviv (1976) were able to determine the mass outflow rates from classical novae as \( M \sim 10^{21}-10^{22} gm s^{-1} \) at maximum light. Most important of all, they showed that the fractional decline from maximum light in the optical depends only on the physical conditions in the outflowing photosphere. This accounts for the observed strong correlation between “decline stage” (i.e., brightness decrease from maximum light) and spectral type observed in classical novae.

Optically thick winds in classical novae were examined in more detail by Bath (1978) and by Ruggles and Bath (1979). These studies showed that as the wind thins during the optical decline of classical novae, the photospheric location and hence temperatures rises as shown in Figure 8. Ions of increasing ionization potential are then photoionized by the increasingly hotter emergent spectrum. Bath’s (1978) sequence of models with progressively thinner wind and thus higher effective temperatures yielded the decline stage at which oxygen and nitrogen could be photoionized by the underlying continuum. The predicted stages at which various ionized species are produced by photoionization by the increasing ultraviolet continuum are in good agreement with those observed in average novae and are shown in Figure 9. Ruggles and Bath (1979) further showed that the outflowing wind must be adiabatic deep within the photosphere and in the region near the critical (sonic) point, while small deviations from adiabatic flow determine the transonic structure in the outer envelope. Only flows which become supersonic close to the white dwarf surface can produce the ejection velocities observed in novae. While radiation pressure is dominant throughout the envelope, most of the luminosity of the eruption is at first carried away from the white dwarf by advection. In the outer part of the envelope most of the luminosity is transported outwards by radiation.
Friedjung (1987) and Taylor et al. (1987) have emphasized the importance of the collision of different velocity materials in nova ejecta. Friedjung pointed out that the principal absorption system, which appears to represent most of the mass ejected by a classical nova, could be explained in the following way. Relatively slowly moving material, ejected during the premaximum phase of a nova outburst, is overtaken by more rapidly moving ejecta in the form of a high velocity wind seen after maximum and which is responsible for the emission lines seen in the "diffuse enhanced" and "Orion" spectrum (McLaughlin, 1960). This could lead to the production of a relatively dense shell, which Friedjung suggests is the matter responsible for the "principal system".

Taylor et al. (1987) suggest the same model to explain their radio observations of Nova Vulpeculae 1984 number 2. The general features of the radio light curve were explained by a wind with a mass loss rate $M \sim 10^{-5} \, M_\odot \, yr^{-1}$, continuing for a year after optical outburst. The early radio emission, which has a brightness temperature in excess of $10^5 \, K$, was modeled as a strong shock propagating outward through the principal ejecta of the nova. The source of the shock is interaction of the slower, earlier ejected material with a later, high velocity wind from the white dwarf.

Detailed models to account for the spectra of nova during outburst have not yet been attempted. The wealth of high quality ultraviolet, visible and infrared spectra of novae in outburst are just beginning to be qualitatively understood. A detailed quantitative understanding and simulation of outburst spectra may require an effort comparable to that invested by theorists over the last two decades in simulating thermonuclear-runaways.

**IV.2 DUST IN NOVA SYSTEMS**

A comprehensive review of observations and theory of dust in nova systems has been given by Bode and Evans (1988). I will discuss only a few of that review's highlights and mention more recent IRAS results.

The seminal observations were those of FH Serpentis (1970) from 1 to 22 $\mu$m by Geisel, Kleinmann and Low (1970). These authors observed the visible and infrared fluxes to decline together for seven weeks before unexpectedly starting to differ dramatically. The visual light curve underwent a steep decline, while the IR flux rose steadily until it dominated the nova light output. Similar behavior in Nova Vul 1976, and the subsequent IR decline is shown in Figure 10. Circumstellar dust was the obvious IR source. The lack of 10 $\mu$m spectral features ruled out silicates, while the $\sim 1000^\circ K$ dust temperatures implied refractory material. Geisel et al. (1976) concluded that the dust was graphite grains, though Lewis and Ney (1979) have argued for metallic iron grains. Clayton and Wickramasinghe (1976) have shown that micron-sized grains can form in the time available.

Observations since then have shown a correlation between nova speed class and IR luminosity: the faster the nova, the less prominent its IR excess. This was explained by Gallagher (1977) as a Strömgren sphere ionization process. The faster the nova (and the higher its bolometric luminosity) the larger its Strömgren sphere and the less able it is to produce dust. Bode and Evans (1982) expanded this model's scope by considering the rise in nova effective temperature as the wind starts to thin. They showed that the distance from the star at which grains can condense increases with time. Expanding ejecta must catch up to the receding condensation distance before its declining density becomes too low to permit significant grain formation. Bode and Evans (1982) determined that novae with an initial rate of decline slower than 0.1 mag/day could make grains (in the context of their model) while faster novae could not.
Dinerstein (1986) found four correspondences between the positions of IRAS sources and those of 67 classical novae. The slow nova Sgr 1982 showed a warm dust shell, while the fast nova Mus 1983 did not. Dinerstein claimed that the detected infrared emission from HR Del (1967), FH Ser (1970) and DQ Her (1934) is inconsistent with a simple dust condensation plus cooling model. She suggested that strong IR lines might be responsible for most of the IR flux from old, metal-rich nova remnants.

**IV.3 X-RAYS FROM OUTBURSTING NOVAE**

The visual brightness of a classical nova begins to decline once mass ejection subsides, after most (\(\gtrsim 70\%\)) of the envelope has been ejected. As the envelope remnant collapses back onto the white dwarf, a large bolometric correction develops and most of the energy radiated by the white dwarf is in the EUV to X-ray band (Prialnik, Shara and Shahiv 1978, 1979). A blackbody temperature of \(\sim 3 \times 10^5\) °K was predicted by these authors to be expected from classical novae for a few years after optical maximum.

The first detections of erupting classical novae in the X-ray band have recently been reported by Ögelmann, Beuermann, and Krautter (1984) and by Ögelmann, Krautter, and Beuermann (1987). The three brightest erupting novae during the EXOSAT mission were all detected as soft (0.5 to 2.5 keV) X-ray sources. The observations of nova Vul 84 No. 1 and nova Vul 84 No. 2 were particularly illuminating. Both were observed approximately four months and ten months, respectively, after outburst, and both brightened by factors of \(\gtrsim 2\) during this time. During this same period of time the optical light curves of both novae were declining significantly. In addition, nova Vul 84 No. 1 was observed by EXOSAT eight days after optical maximum and the nova was too faint to be detected at this time.

Nova Muscae 1983 was clearly seen to be fading between 2–3 years after eruption. This short time scale conflicts with the shocked circumstellar gas model initially proposed by Brecher et al. (1977) in order to explain the transient soft X-ray sources. This cooling time is in good agreement with the constant bolometric luminosity model of a hot white dwarf remnant. In this model, the source remains at constant bolometric luminosity until the envelope hydrogen is all processed to helium.

From the X-rays’ fluxes and energies Ögelmann, Krautter and Beuermann (1987) attempted to determine the white dwarf mass, and the mass of the nuclear processed material for Nova Muscae 1983 (see Figure 11). \(M_{\text{wd}} \sim 0.8-0.9M_\odot\) and \(M_{\text{burn}} \sim 10^{-6}M_\odot\) were estimated, though significant uncertainties remained due to the likely presence of an optically thick wind from the white dwarf remnant. A wind with \(M \gtrsim 10^{-7}M_\odot\,\text{yr}^{-1}\) would decrease the photospheric effective temperatures in the energy range observed by EXOSAT and allow more massive white dwarfs, \(M_{\text{wd}} > 1M_\odot\) (Ruggles and Bath, 1979; Kato, 1984).

Effective temperatures \(T_e\) and luminosities \(L\) of novae will, in future, provide stringent tests of the constant bolometric luminosity model. Future soft X-Ray missions with spectral capability will be able to measure changing \(T_e\) and \(L\) as mass ejection slows and the remnant envelopes collapse back onto the white dwarf.

**IV.4 POSITRON ANNIHILATION GAMMA RAYS FROM NOVAE**

Positron annihilation gamma rays are potentially powerful diagnostics of novae (Clayton and Hoyle 1974, Leising and Clayton 1987). The latter authors have included improved estimates of the nucleosynthesis expected in novae, and were particularly motivated by the
approaching launch of the Gamma-Ray-Observatory as a potential probe of thermonuclear-powered hydrodynamics. It is on the observational predictions and results of this latter paper that I concentrate.

Leising and Clayton (1987) adopted the standard thermonuclear scenario of classical novae. Their numerical simulations were based on simplified models of the physics of the nova TNR which, nonetheless, included a sophisticated study of the generation and transfer of gamma ray photons out of the erupting, expanding envelope. The temperature and density in a nova envelope must be accurately calculated to determine the mode of annihilation of the positrons produced in the weak decay of proton-rich nuclei created during a TNR. These physical conditions determine what fraction of the positron annihilations result in line photons, and what fraction appear as orthopositronium continuum. A Monte Carlo simulation was used to study the photon propagation through the atmosphere, yielding the surface emissivity and energy spectrum.

The two most important sources of annihilation line photons from positrons early in a nova outburst are Nitrogen-13 and Fluorine-18. The emerging intensity of gamma rays depends on the transfer and escape of the emitted photons. Most of the escaping gamma rays emerge from the outer few scattering lengths because the atmosphere is extremely thick to Compton scattering.

Unfortunately, the most crucial assumption in the Leising and Clayton (1987) simulation was the distribution of the positron emitters at the beginning of the evolution. The emitted gamma ray fluxes are extremely sensitive to whether convection can carry positron emitters to within a few gamma ray scattering lengths of the surface. Assuming that convection does, indeed, bring the positron emitters up to the surface, Leising and Clayton (1987) compared the calculated fluxes of annihilation lines, integrated over ~4-40 minutes for various isotopes, with the sensitivities of the Gamma-Ray Spectrometer on the Solar Maximum Mission (SMM GRS) and the Oriented Scintillation Spectrometer Experiment on the Gamma-Ray Observatory (GRO OSSE). Nova envelope models of $10^{-3} M_\odot$ produced enough $^{13}N$ and $^{18}F$ annihilation photons to be detected up to distances of several kiloparsecs during the first few hours of a nova outburst.

Clearly the detection of gamma rays from outbursting novae would be an exciting confirmation of existing theoretical scenarios. This is particularly true for nuclear decay lines, which would be powerful diagnostics of the TNR process and of the early stages of the outburst. Novae should be considered very high priority targets of opportunity for the Gamma-Ray Observatory.

V. HIBERNATION AND ANCIENT NOVAE

V.1 THE HIBERNATION SCENARIO OF CATAclySMIC BINARIES

Dozens of novae erupt in M31 every year (Arp 1956, Ciardullo et al. 1987). Because only one stellar system in thousands is the close white dwarf–red dwarf binary necessary to produce a nova, all novae must therefore recur thousands of times (Ford, 1978). It’s therefore of interest to ask the question: what do novae look like during the millenia between eruptions? The first in-depth historical studies were those done by McLaughlin (1960). McLaughlin concluded that there is no essential difference in the appearance of a nova before and after eruption. He also determined that the initial rise to maximum light must be extremely rapid, usually one or two days.
A much more extensive study was carried out by Robinson (1975). Robinson examined the pre-eruption light curves of 33 novae which erupted from 1899 to 1971. Robinson’s results supported McLaughlin’s first conclusion, namely that there is no apparent difference in the brightness of a nova before and after its eruption. McLaughlin’s second conclusion clearly had to be modified. Most novae do show significant increases in luminosity between one to fifteen years before eruption. The pre- and post-nova eruption brightness of novae has been shown to be \( \sim L_\odot \) (Warner, 1987), mostly emitted by the disk and due to accretion energy. However, the canonical picture of old novae has recently been cast into serious doubt by the recoveries of two much older novae: CK Vulpeculae (nova 1670) (Shara, Moffat and Webbink, 1985) and WY Sagittae (nova 1783) (Shara et al. 1984).

Once a white dwarf ejects most of its hydrogen envelope in a TNR, the remaining envelope settles back onto the white dwarf and begins to cool on a time scale of order 100 years (Prialnik, Shara and Shaviv, 1978). Thus CK Vul and WY Sge, as the two oldest recovered novae, are much more likely to be typical of “between outburst” systems than any 19th or 20th century novae. Both objects are amongst the intrinsically faintest known cataclysmic variables. WY Sagittae, at B \( \sim 19 \) has M \( \sim 7-8 \) (Shara et al. 1984, Kenyon and Berriman 1988). CK Vul (nova 1670) has M \( \sim +10 \), i.e. a luminosity only \( 10^{-2} L_\odot \). Certainly in the latter case, mass transfer is diminished by at least two orders of magnitude relative to that seen in most old novae. In addition, WY Sagittae (nova 1783) and several other old novae have shown at least one dwarf nova outburst (Shara et al. 1984; Livio 1987). If WY Sge and CK Vul are typical of most old novae, then these observations suggest that

1) \textit{old novae decline in brightness and in mass transfer rate on a time scale of a few centuries after eruption, and}

2) \textit{the long term evolution of cataclysmic binaries is going to be dramatically affected by this variable M “hibernation scenario”}.

Two other sharp disagreements between observation and theory are neatly resolved by the hibernation scenario. The space density of cataclysmic variables found in galactic surveys, of order \( 10^{-6} \text{pc}^{-3} \) (summarized by Patterson, 1984) is 2-3 orders of magnitude lower than that deduced from simple nova theory and from the eruption frequency of novae in M31 (Bath and Shaviv 1976). The only way cataclysmic binary space densities of \( 10^{-6} \text{pc}^{-3} \) can be compatible with the observed number density of erupting novae in the Galaxy is if the recurrence time of classical novae is \( < 10^3 \) years. This in turn demands high accretion rates, with average M \( > 10^{-9} M_\odot \text{yr}^{-1} \) onto the white dwarfs in cataclysms. Numerical simulations of mass accretion onto white dwarfs demonstrate that M \( > 10^{-10} M_\odot \text{yr}^{-1} \) keeps the accreted envelope too hot for strong degeneracy to develop (Shaviv and Starrfield 1988, Prialnik et al. 1982), and, therefore, \textit{never leads to nova eruptions}. Instead, continuous burning or mild TNRs puff the envelopes of the mass accreting white dwarfs out into red giants.

A hibernation scenario resolves these apparent conflicts by hypothesizing that cataclysms hide from most surveys during most of the millenia between eruptions. They do this by 1) shutting off mass transfer a few centuries after eruption, 2) keeping mass transfer off for centuries or millenia, and then 3) turning mass transfer back on again for a few centuries before eruption. While this might seem very contrived there are sound physical reasons for a cataclysmic binary to behave this way.

Before explaining why mass transfer rates should vary so dramatically in cataclysmic binaries, let’s step through the scenario and describe what happens at every stage of a nova outburst and the subsequent evolution of the underlying binary.
1) When a nova erupts, the underlying white dwarf's hydrogen rich envelope suffers a TNR. In a very short period of time the Eddington luminosity is reached, and mass ejection continues for weeks or months until most of the envelope is ejected.

2) The visual luminosity of the star then starts to decrease as the bolometric correction increases. This is because the mass ejection rate from the white dwarf decreases and observers see progressively deeper into the atmosphere of the material still surrounding the white dwarf.

3) Eventually the white dwarf stops ejecting mass, though its envelope remains hot and luminous for about a century after eruption. The very hot white dwarf irradiates its red dwarf companion, and this irradiation helps keep mass transfer high for one to two centuries after the nova outburst (Kovetz, Prialnik, and Shara 1988).

4) After a few centuries the rate $M$ of mass transfer from the red dwarf to the degenerate dwarf decreases. As $M$ drops below the critical value required for dwarf nova eruptions, such eruptions begin. In particular, when $M < 10^{-9} M_{\odot} \text{yr}^{-1}$, accretion disk instabilities should occur and old classical novae should be able to undergo dwarf nova eruptions. At least eight such systems are known and Livio (1988) has recently tabulated them. It is not clear just how low the mass transfer rates in cataclysmic binaries become. They may decrease to the point that dwarf nova eruptions become infrequent or stop altogether.

5) After a long interval of very low $M$ (that is, hibernation) the mass transfer rate starts to increase; dwarf nova eruptions begin again; and if the mass transfer rate keeps on rising, dwarf nova eruptions will eventually be suppressed when $M > 10^{-9} M_{\odot} \text{yr}^{-1}$. Above this rate the system would appear as a nova-like variable, as it would have the spectroscopic properties of an old nova with no sign of a visible shell and no history of a recent TNR.

6) Eventually a nova-like variable (that has gone through high $M$, then low $M$, then high again $M$) must erupt again as a classical nova when the pressure and temperature at the hydrogen-rich envelope base becomes large enough to trigger a TNR. In this way, a cataclysmic binary can re-cycle itself thousands of times, becoming in turn a dwarf nova, a hibernating system, a born-again dwarf nova, a nova-like variable, and finally a classical nova.

The detailed questions we now seek to answer are:

1. Why are post-novae bright?
2. Why does $M$ decrease in the centuries following eruption?
3. How low can $M$ in a cataclysmic binary drop?
4. Why do the systems stay in hibernation for thousands of years before $M$ rises again?
5. Why are novae as bright before eruption as after?
6. Do all cataclysmic binaries and nova systems undergo a hibernation phase?

The answers to these 6 questions are as follows:

1. During nova eruptions, and for decades afterward, the red dwarfs in cataclysmic binaries are irradiated with hundreds of times more luminosity than they themselves produce. The simulations of Kovetz, Prialnik and Shara (1988) followed the structure and evolution of red dwarfs in cataclysmic binaries blasted by tremendous amounts of luminosity from their erupting and subsequently cooling companions. Simulations of the time dependent irradiation of three red dwarf models of 0.25, 0.5 and 0.75 $M_{\odot}$ indicate that the red dwarfs swell $\sim 1\%$ in radius because of the irradiation. This swelling is maintained for several decades after outburst. As a result, the mass transfer rates forced by irradiation after nova eruption are found to be enhanced by up to two orders of magnitude due
to the irradiation. As the white dwarf cools on a time scale of a century, the enhanced mass transfer declines.

2. M will decrease in a cataclysmic binary if the mass-transferring red dwarf loses contact with its Roche lobe. This can occur in the following natural way after a nova eruption.

During the weeks or months of nova eruption and mass ejection from the white dwarf, the mass of the underlying binary system is, of course, decreasing. Of order $10^{-4}-10^{-5}M_\odot$ is lost. This results in a weaker gravitational interaction between the two stars in the binary. Assuming spherically symmetric ejection from the white dwarf, Shara et al. (1986) investigated the effects of mass and angular momentum losses during a mass ejection phase. In most cases mass loss dominates, and the separation a of the binary stars increases during eruption. The relationship between the separation increase $\Delta a$ and amount of mass loss $\Delta M$ is given by

$$\frac{\Delta a}{a} \simeq \frac{\Delta M}{M}.$$

The key point is that the red and white dwarf's separation increases as a result of the mass loss. The theory also predicts that the orbital period of the binary should fractionally increase by about $10^{-4}$ to $10^{-5}$. Such an increase has recently been observationally confirmed by Schaefer and Patterson (1983). BT Mon (nova 1939) suffered a fractional period increase $\Delta P/P \simeq +3 \times 10^{-5}$ when it underwent eruption. This is exactly in the sense and size predicted by Shara et al. (1986).

Another possibility to detach the red dwarf from its Roche lobe is the injection via synchronization torques of spin angular momentum of the white dwarf into the binary orbit (Lamb and Melia 1987). This can occur if the timescale for synchronizing the white dwarf is much shorter than that for the radius of the secondary to change appreciably.

An increase in separation $\Delta a/a \simeq 10^{-4}-10^{-5}$ is much less than the increase in the size of the red dwarf atmosphere due to irradiation by its hot white dwarf companion. As a result, the red dwarf atmosphere completely fills its Roche lobe and mass transfer proceeds at an enhanced rate. The essential point to note is that the irradiated red dwarf swells enough to continue to fill its Roche lobe and to transfer mass at a high rate for a few white dwarf cooling times, i.e., a few centuries after the nova eruption.

During the next phase of evolution of the binary, M decreases as the irradiation luminosity from the cooling white dwarf decreases. The red dwarf atmosphere starts to contract back to its pre-eruption value. Furthermore, the increased separation of the red and white dwarfs in the binary helps to break contact of the Roche lobe with the red dwarf atmosphere. Although the computed values and times dependence of M are very model dependent, the long term trend does seem to be clear: M decreases on a time scale of one or two centuries after eruption. Once M does decrease to the critical value $10^{-9}M_\odot\text{yr}^{-1}$, nova dwarf eruptions can begin. If the mass transfer rate observed in old nova systems, $10^{-9}M_\odot\text{yr}^{-1}$, acts for only a century or two after eruption then the scenario predicts that only a few $\times 10^{-6}M_\odot$ is accreted during this phase. During the subsequent hibernation this initial envelope will become rather cold and degenerate, and it will have ample time to diffuse into the carbon-oxygen or oxygen-neon-magnesium white dwarf. It is this rather small amount of mass, located deep in the C-O or O-Ne-Mg outer layers, and accreted in the century or two after eruption that is crucial to triggering the next TNR millenia later.
3. How low can $M$ become in a cataclysmic binary? The intrinsically faintest cataclysmic known, CK Vul (nova 1670), must have $M < 10^{-11.5} M_\odot yr^{-1}$ (Shara, Moffat and Webbink 1986) to account for its extremely low luminosity. Lower values might lead to systems which appear as non-interacting, short-period white dwarf-red dwarf binaries.

4. In an attempt to understand the behavior of accreting novae with variable M, Prialnik and Shara (1986) studied the effects of rapid accretion and then hibernation in the preoutburst history of classical novae. As in previous simulations, no hibernation time and $M = 10^{-8} M_\odot yr^{-1}$ produced a mild TNR, yielding a red giant rather than a nova. The hibernation models produced dramatically different results. As the hibernation times increased, so did the amount of diffusion of hydrogen into the white dwarf carbon-oxygen cores. When mass transfer was turned on again at the rate of $10^{-8} M_\odot yr^{-1}$ after hibernation ($M = 0$) times of $10^3$, $10^4$ and $10^5$ yr, violent TNRs occurred in each of the three cases. This occurred despite the fact that all accreted mass was deposited at $10^{-8} M_\odot yr^{-1}$.

Much of the mass necessary to trigger the next nova eruption arrives at the white dwarf surface in the $\sim 10^4$ yr before eruption when $M$ again becomes high, i.e., about $10^{-8} M_\odot yr^{-1}$.

Livio, Shankar and Truran (1988) have extended the work of Prialnik and Shara (1986) in studying the case of mild hibernation. They found that a reduction in the accretion rate by a factor of about 100 over a period longer than a few thousand years is generally sufficient to ensure nova-type outbursts even in the presence of high pre-outburst accretion rates. Thus a total shutdown of mass transfer is not necessary to ensure strong TNRs once mass transfer begins again.

5. The emergence from hibernation to a dwarf nova and eventually a nova-like state, as $M$ increases, is due to the red dwarf resuming contact with its Roche lobe. This occurs as gravitational radiation and angular momentum losses through magnetic stellar and disk winds drive the red and white dwarf together. In most cases, magnetically driven winds are the dominant effect and they return the stars back to their original separation on a time scale of $\sim 10^4$ years. Once $M$ reaches the high pre-nova eruption state of about $10^{-8} M_\odot yr^{-1}$, only a thousand years or so are needed until the critical envelope mass is achieved and the next nova eruption occurs. Warner (1987) has suggested that there is a natural limit to the mass accretion and mass transfer rates in cataclysmic binaries of about $10^{-8} M_\odot yr^{-1}$ because so many systems have converged to this value. This might occur if $10^{-8} M_\odot yr^{-1}$ is the upper limit to mass transfer in an accretion disk, and could explain why $M$ is so similar before and after nova eruptions.

6. Livio and Shara (1987) studied which cataclysmic binaries should be most likely to undergo hibernation. These are the systems which suffer the largest reduction in mass transfer rate, after a nova event, caused by the increase $\Delta a$ in orbital separation. The assumptions used in the derivation were:

1) The secondary fills its Roche lobe;
2) The secondary obeys a lower main-sequence mass-radius relation;
3) The mass ejected in the nova outburst is equal to the mass necessary to trigger a TNR;
4) The Shara et al. (1986) assumptions of conservation of mass and angular momentum transfer during a nova outburst;
5) The Papaloizou and Bath (1975) red dwarf scale height $H$ (as a function of binary parameters).

These assumptions lead to a complex functional dependence of $\Delta a/H$ on the orbital period $P$ and the binary mass ratio $q$. It was found numerically that for a given orbital
period, systems with a higher mass ratio are more likely to enter hibernation. For a given mass ratio the systems with shorter orbital periods are more likely to enter hibernation.

Livio (1988) has stated that a change in the separation of the red and white dwarfs may not yield as large a mass transfer rate change as previously claimed. He suggested that the periods of high M before and after eruptions are due to irradiation of the red dwarf by the white dwarf. The hibernation scenario (in this viewpoint) is unchanged, but the mechanism to produce M is variable red dwarf irradiation rather than variable binary separation.

The hibernation scenario explains quantitatively the apparent discrepancy between mass accretion rate observations and the theory of mass accretion onto pre-novae; the surprising rarity of nova discovered in surveys; how metallicity can vary from nova to nova; and how moderate and fast novae can appear even in metal-poor globular clusters and galaxies. The latter effect, of course, is due to the diffusion of hydrogen into the degenerate dwarf core during the hibernation phase.

The greatest advantage of the hibernation scenario is that it offers a unifying picture of dwarf novae, nova-like variables and classical novae. In the hibernation scenario all three are the same system; we simply see metamorphosis between one cataclysmic type and another as the mass transfer rate in the binary changes over the millennia.

V.2 HUNTING FOR NOVAE SEEN BY ANCIENT ASTRONOMERS

A clear prediction of the hibernation scenario is that very old novae (much older than CK Vul) should be intrinsically much fainter than ex-novae 10–100 years after eruption. Searching for novae which erupted before the invention of the telescope is a difficult task. European observations of novae prior to 1600 A.D. are almost non-existent. We must therefore rely on records kept by Japanese, Korean and Chinese astronomers.

For a nova to have been noted by Oriental astronomers, it must have reached at least third magnitude. With an amplitude of about 12 magnitudes, which is typical for most novae, very old novae should now be no fainter than V ~ 15 if the hibernation scenario is incorrect. If the hibernation scenario is correct, then these objects may be as faint as 20th magnitude.

Over the last several years, I and my colleagues (M. Bode, F. R. Stephenson, M. Potter and A. Moffat) have been conducting a search for ancient novae. We are attempting to recover a dozen objects observed over the last two thousand years by Oriental astronomers. Deep U and B Schmidt plates of those areas of the sky that the Oriental astronomers reported to contain tailless, stationary "guest" stars (to exclude comets) have been obtained. Scanning and processing of the plates is presently underway.

The first object we've searched for is the possible nova of September 19, 712 A.D. The astronomical treatus of the Chiu-t'ang-shu records the following: "At night the moon was eclipsed and it was complete, a star entered into the moon's spirit." The star was not a planet and there is no star brighter than magnitude 7.0 in the computed path of the moon that night (Stephenson 1986). Our U and B plates reveal no object which is brighter on the U plate than the B plate down to B = 18. If the hibernation theory is wrong, and our ancient colleagues did indeed see a nova being occulted, we would expect to have found a 15th magnitude nova-like variable. There clearly is not one in the 0.5 degree² of the sky where the Oriental astronomers may have recorded a nova in 712 A.D. This, of course, is not direct evidence that the hibernation theory is correct, but the result is consistent with the predictions of the hibernation scenario. If in any of the remaining survey fields we find bright nova-like variables, serious doubt would be cast on the hibernation scenario. If, on the
other hand, for a large number of fields with certain or near certain ancient novae, we can recover no nova-like variables down to 19th or 20th magnitude, support for the hibernation scenario of cataclysmic binaries will have been gained.

A system which might represent a case in which hibernation did not occur is 0623+71. This object is a blue, 12th magnitude nova-like variable surrounded by a nebula (Ellis, Grayson and Bond 1984). If the nebula is, in fact, a shell ejected by a nova many centuries ago, then 0623+71 could represent a binary system in which accretion at a relatively high rate persists. Alternately, the system may be a planetary nebula with a binary nucleus (Bond 1988). Unfortunately there is no historical record of the outburst of this object, and its age remains problematical.

Bruhweiler and Sion (1986) have suggested that V471 Tau may have erupted as a nova many centuries ago. If this suggestion is correct then this detached red dwarf-white dwarf Hyades binary is now in a state of deep hibernation (i.e., \( M = 0 \)).

An exciting recent discovery by Bragaglia et al. (1988) is that two white dwarfs, out of twenty examined for radial velocity variations, have close red dwarf companions. One of the two (WD0034-211) is probably a weakly mass-transferring system. This could be a system in deep hibernation long after an eruption. If so, the implied space density of hibernating novae is (very roughly) 0.05 that of white dwarfs, or \( 10^{-3} \sim 10^{-4} \) pc\(^{-3} \)—just that predicted by the hibernation scenario.

VI. RECURRENT AND MAGNETIC NOVAE

VI.1 RECURRENT NOVAE

The nature of the recurrent nova has been reviewed in great detail by Webbink, Livio, Truran and Orio (1987). This paper offers a broad overview for anyone interested in recurrent cataclysmic binaries. I will only touch on a few of the highlights discussed by Webbink et al. (1987).

Theorists and observers generally agree that dwarf novae are accretion-powered events (probably due to accretion-disk instabilities quickly dumping mass onto white dwarfs), while classical novae are powered by thermonuclear runaways on the surfaces of white dwarfs. The outburst mechanism for recurrent novae is less clear. Recurrent novae suffer outbursts on time scales of decades to ~ one century, intermediate between the dwarf novae and classical novae. Erupting recurrent novae brighten by \( \sim 7-11 \) magnitudes, a larger range than all dwarf novae but smaller than most classical novae. Finally, recurrent nova speed classes range from very slow (T Pyx) to extremely fast (U Sco, V394 CrA, T CrB, and RS Oph are amongst the fastest known novae). Webbink et al. (1987) have concluded that the "class" of the recurrent nova is in fact an extremely heterogeneous group of objects. This, they point out, is largely due to the lack of a rigorous definition as to what constitutes a recurrent nova. They offer the following criteria, which any object must satisfy in order to be classified as a recurrent nova:

i) at least two distinct recorded outbursts, with absolute magnitudes at maximum comparable to those of classical novae \( (M_v \lesssim -5.5) \), and

ii) ejection of a shell at outburst, with ejection velocity comparable to those of classical novae \( (V_{\text{ej}} \gtrsim 300 \text{ km s}^{-1}) \).

These criteria cleanly distinguish recurrent novae from classical and dwarf novae, and also from symbiotic stars.
By carefully reviewing the observations of eleven objects usually referred to as recurrent novae, Webbink et al. were able to eliminate all but five from the class. Only two were considered to be caused by TNRs on the surfaces of massive white dwarfs: U Sco and T Pyx. Accretion-powered events probably underly the eruptions of T CrB and RS Oph. These latter objects are suggested to be giants transferring bursts of mass to their main-sequence companions. Finally, V1017 Sgr appears to be a hybrid case which exhibits both accretion events and TNRs. Recently V394 CrA underwent a second outburst very similar to those of U Sco (Sekiguchi et al. 1988). It is likely a TNR-powered nova.

Webbink et al.'s (1987) main criterion for distinguishing thermonuclear from accretion events in yet to be discovered recurrent novae is the spectrum at optical maximum. TNR powered recurrent novae must occur on white dwarfs with masses close to the Chandrasekhar mass, and with envelope masses that must be extremely low \((10^{-6}-10^{-7} M_\odot)\). The density of such envelopes at optical maximum is so low that electron scattering opacity can exceed the true absorption opacity by several orders of magnitude. The continuum spectrum observed must be formed deep within the envelope, and is expected to depart significantly from a black body. Webbink et al. (1987) thus predict an emission-line spectrum, as observed in T Pyx, for TNR powered recurrent novae.

In the case of accretion powered recurrent novae, the resultant shock ejection should produce exceptionally high velocities, decreasing with time; and the appearance of high-excitation coronal line emission and X-ray emission during decline.

Finally, because of the very short accretion timescales involved in recurrent novae, T Pyx and U Sco offer an outstanding opportunity to differentiate between the diffusion and the shear-mixing mechanisms of CNO enrichment of novae ejecta. This was discussed in §II.3.

**VI.2 NOVAE ON STRONGLY MAGNETIC WHITE DWARFS**

The evolutionary relationship between magnetic and non-magnetic nova systems is unknown. The white dwarfs in the former can have strong magnetic fields. The magnetic cataclysmic variables divide naturally into two subclasses. The DQ Herculis stars (recently reviewed by Warner 1985) are inferred to have magnetic fields of \(10^5-10^7\) gauss. The AM Herculis stars, recently reviewed by Liebert and Stockman (1985), typically have magnetic fields of \(2-3 \times 10^7\) gauss. An extensive review of the evolution of magnetic cataclysmics has recently been published by Lamb and Melia (1987).

DQ Herculis, the prototype of the DQ Her systems, erupted as a classical nova in 1934. Novae were not known to be binary systems in 1934 and we are unaware of any attempts to measure periods in the system then. Nevertheless, DQ Her impressed upon theorists that a classical nova could indeed erupt on a white dwarf with a moderately strong magnetic field. Until 1987, however, it was not clear that a classical nova could erupt on an extremely magnetic white dwarf, i.e., in an AM Herculis binary (Livio 1983; but see Livio, Shankar and Truran 1988).

Kaluzny and Semeniuk (1987) recently carried out CCD-imaging photometry of V1500 Cygni. The light curve of this old nova, now at 17th magnitude, revealed sharp features reminiscent of AM Herculis sources. Following up on this discovery Stockman, Schmidt and Lamb (1988) (hereafter SSL) carried out a program of polarimetric observations. Periodic variations of optical circular polarization with semi-amplitudes of \(\pm 1.5\%\) were found by SSL. The period of polarimetric variation is 1.8% shorter than the stable photometric period, first obtained by Patterson (1979), and refined recently by Kaluzny and Semeniuk (1987). SSL ascribe the circular polarization to optical cyclotron emission from the accreting...
magnetic white dwarf primary. However, most of the system light and the photometric variability are due to the heated companion star.

Nova Cygni 1975 achieved a maximum absolute magnitude $M_v = -10.2$ as one of the most luminous novae ever seen (see the recent review by Lance, McCall, and Uomoto 1988). A 3.3 hour variation was seen in the early nova emission lines and then in the optical continuum. Patterson (1978, 1979) observed the period to decrease from 0.141 days to 0.138 days during the first year after eruption and then to increase and finally stabilize at 0.140 days towards the end of 1977. The main theoretical thrust of SSL's work has been to offer a simple, reasonable explanation for the observed period changes.

The stable 0.140 day photometric period observed since 1977 is probably due to viewing the heated face of the orbiting red dwarf at moderate inclination (Patterson 1979). This is energetically plausible if the central white dwarf luminosity was $L_{wd} \approx 10^4 L_\odot$ in 1978. Detailed models of the cooling of post-novae (Prialnik, 1986) and the resulting heating of the secondary (Kovetz, Prialnik and Shara 1988) support this high white dwarf luminosity, as well as the high resultant luminosity of the secondary a decade after eruption. A detailed comparison of the optical polarimetric and spectrographic observations with the theory is given by SSL. Here we concentrate on their explanation of the period changes due to asynchronous rotation set up in the V1500 Cyg system. The period evolution is sketched in Figure 13.

The absolute visual magnitude of V1500 Cyg was $> +8.3$ before eruption (Wade 1987, Lance, McCall and Uomoto 1988). AM Her and VV Pup have $M_v \sim 8.6$ in their bright states (Patterson 1984). These values correspond to mass transfer rates of $\dot{M} \sim 10^{-10} M_\odot$ per year. The angular velocity $\Omega$ of the strongly magnetic white dwarf was probably phase locked to the orbital angular velocity $\Omega_b$ of the binary before outburst (SSL). A massive white dwarf (Shara 1981, Prialnik 1986) and sufficient accretion time to permit hydrogen envelope enrichment in CNO isotopes are the other requirements which enabled Nova Cygni to become one of the fastest novae ever observed.

SSL have proposed the following scenario for the observed period changes in V1500 Cyg. As the nova erupted the envelope expanded to engulf the entire binary. This significantly increased the moment of inertia of the white dwarf. The relative change in the moment of inertia is given by

$$\frac{\Delta I}{I} \approx \frac{4\pi}{30} \left( \frac{\mu}{k} \right)^4 \left( \frac{a}{3} \right) \left( \frac{GM_1}{4} \right)^3 \frac{\beta^4}{1 - \beta} \left( \frac{R}{R_g} \right)^2 \left( \frac{1}{M_1} \right)$$

where $\mu$ is the mean molecular weight per baryon, $k$ is the gas constant, $a$ is the radiation constant, $R$ is the outer radius of the envelope and $R_g$ is the radius of gyration of the original white dwarf. Assuming that $\beta = 1 - L_{\text{nuc}}/L_{\text{Edd}} \approx 0.1$, where $L_{\text{nuc}}$ and $L_{\text{Edd}}$ are the nuclear and Eddington luminosities, a density $\rho(r) \approx 10^{-7}$ gm cm$^{-3}$ in the vicinity of the secondary, $R \approx$ the binary separation, $M_1 \approx 1.4 M_\odot$ and $R_g \approx 7 \times 10^{-4} R_\odot$, SSL find $\Delta I/I \sim 10^{-2}$. By assuming that convection, driven by the steep temperature gradient during the initial expansion phase dominates energy transport, then the angular velocity $\Omega$ will also decrease by $\Delta \Omega/\Omega \sim 10^{-2}$. Thus, immediately following the TNR and resultant envelope expansion the white dwarf rotates more slowly than the binary. SSL identify the 0.141 day period observed six days after outburst with the slower white dwarf rotation period. They also suggest that a hotspot, which illuminated the expanding nebula and which was seen photometrically, might have been caused by a local area of enhanced nuclear burning (Shara 1982), if the magnetic field or a convective plume could transport the luminosity to
the photosphere. The expanded envelope coupled the white dwarf to the orbiting companion and was quickly spun up from the 0.141 day period to the 0.140 day binary period. Some of the envelope was ejected, but some radiated enough thermal and potential energy to sink back onto the white dwarf; this caused the photometric period to decrease from 0.140 days to the polarimetric, white dwarf rotation period of 0.137 days. By 1977 the white dwarf had become faint enough in the visible part of the spectrum for the heated secondary to be the dominant system light source. At that point, the photometric variability stabilized at the orbital period of 0.140 days. SSL have suggested that it might be possible to phase the current polarimetric observations with the 1976 photometric data to test their model and to identify the dominant accretion/nuclear burning pole in V1500 Cyg.

Clearly much more work remains to be done on this fascinating object, both by theorists and by observers. It is particularly gratifying to see the strong overlap and interaction between theory (local thermonuclear burning, strong magnetic fields, expanding envelope interaction with the secondary, irradiation of the secondary) and the painstaking observations of subtle polarization and minute changes in orbital period. Invaluable information and understanding is clearly to be gained by looking in detail at novae not only during the months after eruption but years and decades afterwards.

VII. NOVAE IN CLUSTERS AND GALAXIES

VII.1 NOVAE IN GLOBULAR CLUSTERS

Two nearly certain classical novae have been seen in globular clusters: T Sco, which erupted in M80 in 1860, and the nova of 1938 in M14 (Hogg and Wehlau 1964). In addition, two dwarf novae are known in globulars: V101 in M5 (Margon, Downes and Gunn 1981; Shara, Moffat and Potter 1987) and V4 in M30 (Margon and Downes 1983). These objects may be hibernating novae (see the section on hibernation). Intriguing photometric evidence for the presence of cataclysmics in the inner regions of M71 is seen in the beautiful color-magnitude diagram of Richer and Fahlman (1988); but spectra now indicate that the candidates are subdwarf O stars (Richer 1988).

Novae which erupted in galactic globular clusters before ~1920–1930 (when systematic photographic surveys of globulars began), during the fall and winter months when globulars are less observable, and those novae which erupted near the centers of clusters with high central surface brightness would not have been detected. Trimble (1977) concluded that most globular cluster novae are in fact not observed, and that the nova rate per unit mass in globular clusters is significantly higher than the rate in the Galaxy.

Bright X-ray sources are known to be much more common in globular clusters than anywhere else in the galaxy (Katz 1975). This is believed to be brought about by the process of tidal capture in globular cluster cores (Fabian, Pringle, and Rees 1975; Sutantyo 1975). Two unbound stars passing sufficiently close to each other in the core of a dense globular cluster can raise significant tides on each other and divert enough orbital energy into dissipative stellar oscillations to bind the pair. Succeeding periastron passages repeat the dissipative process, and eventually the orbit circularizes with a semi-major axis of a few stellar radii. This mechanism works as well, of course, for white dwarfs capturing main sequence stars as for neutron stars (which presumably form the bright X-ray sources). This, then, is a mechanism for manufacturing nova-producing binaries in globular clusters. The similarity of the distributions of novae and globular clusters about the central bulge of M31
prompted Ciardullo et al. (1986) to suggest that many or most novae are created in globular clusters and then ejected into the field.

Very close binaries are believed to drive the evolution of their host globular clusters (Spitzer and Mathieu 1980). Because white dwarfs are much more numerous than neutron stars in globular clusters (and everywhere else in the Galaxy), there are probably tens or hundreds of times more cataclysmics in globulars than bright X-ray sources. Thus close, interacting binaries may control the structure and the evolution of all or most Galactic globular clusters. Finding these objects, and studying their properties is clearly important.

The first optical counterpart for a globular classical nova was recently identified by Shara et al. (1986). Nova 1938 in M14 was identified on the basis of positional coincidence, brightness and blue color to be a B=20.2 object within one arc second of the outbursting object photographed in 1938. Spectrographic confirmation of the candidate was difficult because of the star's faintness and because of the crowding of its field, but has now been accomplished (Shara, Moffat and Potter 1988). An even more challenging task will be the recovery of the nova of 1860 which erupted a few arc seconds from the core of M80. This object's spectrographic confirmation will certainly require the high spatial resolution of the Hubble Space Telescope, and will be observed by Dr. B. Margon in the first two years following HST launch.

Hertz and Grindlay (1983) identified low-luminosity Einstein X-ray sources in globulars, and suggested that they are short-period cataclysmic (white dwarf + red dwarf) binaries. Deep CCD multicolor photometry (Margon and Bolte 1987) and high time resolution photometry (Shara et al. 1988) in candidate and other fields have detected no optical cataclysmic binary candidates in ω Cen or in 47 Tuc. These data do not exclude Hertz and Grindlay's (1983) suggestion, but raise the following question: if the low-luminosity X-Ray sources in globulars are really like cataclysmic variables, then why aren't they blue, and why don't they flicker rapidly in brightness?

Novae recur on timescales of ~ 10^4–10^6 years. Even if only 2–3 novae erupt per century in all the globular clusters in the Galaxy, a few nova-producing binaries must still exist in most globular clusters in the Galaxy. Scott and Durisen (1978) pointed out that the eruptions of globular cluster novae would have a profound effect on the gas content of globular clusters.

Typical globular cluster stars should lose ~ 0.2M☉ during evolution from the main sequence to the white dwarf stage. One would expect ram pressure stripping to sweep accumulated gas out of a globular cluster during passages through the galactic plane. Between passages, however, ~ 100M☉ of gas should build up in a typical globular. Both 21cm and Hα searches of globular clusters have yielded stringent upper limits significantly below a hundred solar masses for neutral or ionized hydrogen (Taylor and Wood 1975; Smith, Hesser and Shawl 1976). The latter authors reported a conservative upper limit of 5.9M☉ of ionized gas in NGC 6388, and a "probable" upper limit of 1.8M☉. NGC 6388 has a very large escape velocity (~ 75 km s^-1) (Illingworth and Freeman 1974). Its inability to retain at least 15–20 times more gas than is observed suggests that another depletion process exists in globulars.

Scott and Durisen (1978) have calculated the heating effect of multiple nova eruptions on the interstellar gas in globular clusters. They demonstrated that even a few novae in a globular cluster will heat the interstellar gas to over 4 × 10^5 °K. Such temperatures will deplete globular clusters of their interstellar gas on a time scale of 10^4–10^5 years, much less than the 10^8 year galactic plane crossing time. This is probably why globulars are so gas-free.
VII.2 EXTRAGALACTIC NOVAE AND THE DISTANCE SCALE

Novae are extremely luminous at optical wavelengths, and plentiful and straightforward to find in external galaxies. Only supernovae are much more luminous, though they are much rarer than novae. The fastest novae are 3 magnitudes brighter than the brightest cepheids. Zwicky (1936) was the first to show that there is a tight relationship between the absolute magnitude of a nova at maximum light and its subsequent rate of decline. McLaughlin (1945) and Cohen (1985) have calibrated the absolute magnitude rate of decline relation for galactic novae. These observations have reaffirmed that classical novae are amongst the best standard candles available in astronomy (both because of their brightness and because of the small intrinsic dispersion in the maximum magnitude rate of decline relation).

Shara (1981) was able to show on physical grounds why the absolute magnitude-decline time relation works. The relevant physics used is:

1) the steady state, constant luminosity, continuous wind ejection model of novae first proposed by Bath and Shaviv (1976) (see §IV); (this physical model assumes that radiation pressure drives the nova-wind);
2) the assumption that nova visual shut off occurs because of envelope exhaustion;
3) the TNR hypothesis and the consequent pre-nova envelope structure; and
4) a simple white dwarf mass-radius relation.

The fastest novae occur on the most massive white dwarfs because these stars require the lowest mass envelopes to produce a TNR, and these envelopes can be ejected the most quickly. Thus bright novae fade rapidly, while fainter novae fade more slowly. Only one assumption in the derivation of the absolute magnitude-decline time relation may limit its validity to any particular type of galaxy, stellar population, or epoch.

The assumption is that metallicities in the ejected envelopes can be enriched even in metal poor systems. Moderate and fast novae (which demand \( Z > Z_{\odot} \)) have been observed in the globular cluster M80 and in the Large and Small Magellanic cloud. The diffusion and/or shear mixing processes should, and seem to work as well on carbon-oxygen white dwarfs in metal poor systems as in metal rich systems.

Another assumption is the implicit one that all novae are binaries, and that binaries do form in all stellar systems at all times, and that some eventually evolve to the close red dwarf + white dwarf pair needed for a nova. The red and white dwarfs are assumed to be like those in Galactic novae. This is supported by the fact that novae have been seen in the Galaxy and in M31 (both of type Sb), in M33 (Sc), the Large and Small Magellanic clouds (both irregulars) and even in M80 (a globular cluster) and in NGC 205 (dE). We don’t know enough about binary star formation or binary evolution to be sure that binaries form in galaxies at large redshifts. Despite these caveats it is important to note that novae in nearby galaxies obey the same absolute magnitude-decline time relationship found in the Milky Way, with no apparent dispersion due to metallicity or galaxy type. See Figure 14.

Liller and Mayer (1987) have been surveying the Galaxy for erupting novae since 1983. The deduce a rate of \( 73 \pm 24 \text{ yr}^{-1} \) in the Milky Way, larger than that observed in M31 by Arp (1956).

The pioneering observations of Arp (1956) and those of Rosino (1964, 1971) lead to a nova rate of \( 26 \pm 4 \text{ novae yr}^{-1} \) in M31. The density of novae was found to increase towards the center of the galaxy, but there was a curious “hole” found in the center of that galaxy. No novae were discovered by either author within 4 arc minutes of the center of M31. Arp’s observations also indicate that about 1/3 of the novae in M31 are fast, 1/3 are moderate and 1/3 are slow speed novae.
Recently, Ciardullo, Ford and Jacoby (1986) and Ciardullo et al. (1987) have searched for novae in M31 with charge coupled device cameras (CCDs) and Hα interference filters. The high quantum efficiency of the CCDs allow the searches to be carried out with 1 meter class telescopes. More important, the dynamic range of the CCDs permits the detection of novae even against bright backgrounds, particularly when using Hα narrowband interference filters. Erupting novae are 3 to 5 magnitudes brighter in Hα than in the continuum and so they can be seen much longer in Hα than in the continuum (Figure 15). Ciardullo et al. (1987) found that the number of novae per square arc minute increases monotonically to the center of M31 (Figure 16), and suggested that the “hole” seen by Arp and by Rosino is in fact an observational artifact. The Ciardullo et al. (1987) observations indicate that the nova rate in M31 is more than twice that previously claimed.

Ford and Ciardullo (1988) have shown that the time averaged B and Hα luminosity functions of novae are power laws. They thus have little utility as distance indicators. However Hα observations of galaxies remain a powerful tool for determining nova rates and parent populations.

A decade-long survey of the Large and Small Magellanic clouds has been carried out by J. Graham (1981). Patience is required as only ~1-2 novae per year erupt in the LMC, and only 1 every ~ 2 years in the SMC. Graham plotted the positions of all the known novae in the Large and Small Magellanic clouds (see Figure 17). He found that novae are not found in the bars of each galaxy, implying that the low mass stars there are not yet old enough to have formed novae. Both the speed class distribution and the spectra of a few erupting Magellanic cloud novae appear to be similar to those in M31 and in the Galaxy.

The most challenging detections of novae ever reported are those of Pritchett and van den Bergh (1986). During a fifteen night dark run at the Canada-France-Hawaii telescope they detected nine novae in the Virgo cluster. The brightest of these objects appeared at 21.75, and the faintest at B=25. The light curves of most of these novae are too poorly determined to permit an accurate distance measure to the Virgo cluster, but a rough estimate of the derived Hubble constant $H_0 = 69 \text{ km s}^{-1}\text{Mpc}^{-1}$ was given. These important observations demonstrate that novae can be detected out to cosmologically interesting distances. They will surely be used by Hubble Space Telescope observers as standard candles in the Virgo cluster and perhaps beyond.

**VIII. SUMMARY AND PROSPECTS FOR THE FUTURE**

An enormous amount of effort has gone into trying to understand the physical mechanisms and binary system parameters which govern nova eruptions. The accretion process, ignored in the early simulations of novae, is now understood to be extremely complex and crucial in determining the character of any outburst. Thermonuclear runaways are certainly not one dimensional, spherically symmetric events. Rotation, tides, magnetic fields, and anisotropic accretion all contribute to TNRs being local events in their early stages. If, and how quickly localized runaways spread over the white dwarf surface is not yet known.

The helium, CNO, and Ne abundances of nova ejecta are powerful clues both to the natures of the TNRs and the previous accretion history of the white dwarf. ONeMg and even iron-rich white dwarfs likely exist in some nova systems. White dwarf masses approaching the Chandrasekhar mass seem required to explain the extraordinarily rapid recurrences of U Sco, V394 CrA and T Pyx. X-rays are beginning to be useful diagnostics of post-eruption
white dwarfs, while γ-rays could be powerful probes of the nucleosynthetic products of novae in the next decade.

Dramatic changes in the orbital periods of V1500 Cygni have offered intriguing glimpses into the interplay between expanding envelope and underlying binary. The hibernation scenario of cataclysmic binaries offers a unifying picture of classical novae, dwarf novae and nova-like variables. These systems may undergo metamorphosis on timescales of hundreds to thousands of years as the mass transfer rate varies.

Studying novae formed by the tidal capture mechanism in globular clusters might provide valuable comparisons with field cataclysms formed by ordinary binary evolution. A particularly intriguing question is: Can tidal captures form cataclysmic binaries with orbital periods between 2–3 hours? Either binary stellar evolution does not make these binaries, or they are not active as mass transfer binaries when they evolve to 2–3 hour periods. Eruptive binaries in globular clusters have much to teach us about cataclysms in the field.

Continued surveys for erupting novae in M31, the Magellanic Clouds and more distant galaxies will give us a better understanding of the parent populations and hence evolutionary time scales of close accretion binaries. They may also contribute significantly to our knowledge of galaxy distances out to the Virgo cluster and beyond. As we learn more about novae, and understand which mechanisms and binary parameters control their outbursts, these powerful standard candles will be used with increasing frequency and confidence.

Bright, naked-eye novae appear about once a decade. Much has been learned about nova eruptions since the spectacular appearance of Nova Cygni 1975. Astronomers will be ready with a large battery of new observational tools and tests when the next bright guest star appears in the heavens.

Acknowledgements

It is a pleasure to thank Peter Stockman, Howard Bond, Dina Prialnik, Anthony Moffat and Lindsey Smith for reading and suggesting improvements to an earlier version of this paper.

REFERENCES

Andrillat, Y., and Houziaux, L. 1985, in Recent Results on Cataclysmic Variables, ESA Workshop (Bamberg) p. 197.


33
Richer, H. 1988, priv. communication.
Sekiguchi, K., Catchpole, R. M., Fairall, A. F., Feast, M. W., Kilkenny, D., Laney, C. D.,
796.
Snijders, M. A. J., Batt, T. J., Seaton, M. J., Blades, J. C., and Morton, D. C., 1984,
M.N.R.A.S., 211, 7P.
Starrfield, S., 1987, in New Insights in Astrophysics: Eight Years of Ultraviolet Astronomy
Truran, J. W., 1982, in Essays in Nuclear Astrophysics, C. A. Barnes, D. D. Clayton and D.
Schramm eds., Cambridge Univ. Press, Cambridge.
Truran, J. W., 1985, in Nucleosynthesis: Challenges and New Developments, eds. W. D.
Lamb and J. Patterson, Reidel, Dorchecht.

FIGURE CAPTIONS

FIG. 1. - Heavy element (Z) profiles at the onset of the thermonuclear runaway for models
with $M_{\text{sd}} = 1.25 \, M_\odot$. Solid lines correspond to $L_{\text{sd}} = 1.5 \times 10^{-2} \, L_\odot$, dashed lines to $L_{\text{sd}} = 7.5 \times 10^{-3} \, L_\odot$. The initial Z-profile is the dotted step function. The mass m
is measured in both directions away from the initial core envelope boundary. The labels
have the same meanings as in Figure 2. From Kovetz and Prialnik (1985).

FIG. 2. - The resulting accreted mass (in units of $10^6 \, M_\odot$) and $Z_{\text{env}}$ (%) in initial parameter
space. B is the massive $1.25 M_\odot$ white dwarf, H is the hotter white dwarf ($0.015 L_\odot$) while C
is the cooler object ($0.0075 L_\odot$). The suffixes 9 and 11 refer to mass accretion rates of $10^{-9}$
and $10^{-11} \, M_\odot\text{yr}^{-1}$, respectively. Thus the model BH9 is a $1.25 M_\odot$ white dwarf radiating
$0.015 L_\odot$ and accreting at the rate of $10^{-9} \, M_\odot\text{yr}^{-1}$. By the time its thermonuclear runaway
starts, it has accreted $6.71 \times 10^{-6} \, M_\odot$ and the envelope has $Z = .102$. From Kovetz and
FIG. 3. - Schematic model for the white dwarf in a cataclysmic variable, illustrated for the case where shear turbulence is not efficient at redistributing angular momentum over the entire degenerate region. If shear turbulence is efficient, the core will be more nearly in uniform rotation, and there will be no substantial degenerate subregion in the nearly Keplerian envelope. The solid line represents the photosphere of the white dwarf, and the dashed curve, a smooth extension into an optically thin inner disk. The dot-dash curve illustrates the schematic form of the angular velocity \( n(\omega) \). From Durisen (1977).

FIG. 4. - Qualitative sketches of the likely hydrodynamical behavior of the matter in a localized thermonuclear runaway on a white dwarf. From Shara (1982).

FIG. 5. - Evolution of a 1.25\( M_\odot \) nova model's bolometric luminosity (solid line), nuclear luminosity (dashed line), and visual magnitude (dot-dashed line) through a full cycle, as indicated and the accretion phase preceding it. The time axis is divided into intervals corresponding to phases of different evolutionary time scales. From Prialnik (1986).

FIG. 6. - History of a 1.25\( M_\odot \) nova model's envelope structure. The time axis is the same as in Fig. 5. The ordinate is the mass above the original WD core \((m > 0)\) and below it \((m < 0)\). The solid line labeled "M" shows the total stellar mass \((M_{\text{wd}} + M)\) as a function of time. The convective zone's boundaries are indicated by dot-dashed curves. Hatched areas within dashed boundaries denote burning shells (before the onset of convection), where the nuclear energy generation rate exceeds \(10^{37} \text{ ergs} \text{ g}^{-1} \text{ s}^{-1}\). The different stages of mass loss and the mass of the envelope remnant are pointed out by arrows. From Prialnik (1986).

FIG. 7. - Low-dispersion IUE ultraviolet spectrum of Nova CrA 1981 on 1981 April 12. Flux is in units of \(10^{-14} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ Å}^{-1}\). From Williams et al. (1985).

FIG. 8. - Variation of photospheric density, pressure, radius and temperature for outflowing winds as a function of \( \Delta M_Y \), the decline stage in visual magnitudes, for a blackbody continuum at constant luminosity. Three luminosities of \(0.5 L_{\text{ed}}, L_{\text{ed}}\) and \(2L_{\text{ed}}\) are considered. From Rath (1978).

FIG. 9. - Magnitude, \( M_B \), in a \( B \) filter as a function of the wind mass loss rate, \( \dot{M} \). An outflow velocity of \(10^6 \text{ cm/s} \), and luminosities of \(2L_{\text{ed}} \) (a) and \(0.5L_{\text{ed}} \) (b) are assumed. The stage at which oxygen and nitrogen could be photoionized by the underlying continuum, and thus possibly give rise to associated emission features are listed, and indicated on both curves. From Bath (1978).

FIG. 10. - The infrared development of Nova Vulpeculae 1976 (NQ Vul) is typical of that of moderate speed novae, showing the discontinuity in the visual light curve and the corresponding rise in the infrared (3.5\( \mu \text{m} \)). The temperature of the dust shell is also included (from Bode 1981, adapted from Ney and Hatfield 1978).
FIG. 11. - The range of experimentally allowed bolometric luminosity, blackbody combinations that are compatible with the average soft X-ray counting rate \((3.0 \pm 0.310^{-3} \text{cts}^{-1})\) of Nova Muscae 1983, 462 to 707 days after the optical maximum. The boundaries, indicated by solid lines, are determined by: a) limiting the luminosity to \(L_{\text{edd}}\) for a 1.4 \(M_\odot\) white dwarf as \(2.1 \times 10^{38} \text{erg s}^{-1}\) (top boundary); b) using the lower limits of counting rate, \(N_H\), and distance (lower left boundary); c) using the upper limits of counting rate, \(N_H\), and distance (upper right boundary); d) using a minimum surface radius of \(5.1 \times 10^8\) cm for the white dwarf (lower right boundary). The central line corresponds to the mean estimate of counting rates, \(N_H\), and distances. The dashed curves are evolutionary tracks of the central stars of planetary nebulae with core masses 0.8 and 0.9 \(M_\odot\), as adopted from Paczynski (1971). From Ogelman, Krautter and Beuermann (1987).

FIG. 12. - Emerging fluxes of annihilation line photons from positrons emitted by each unstable nucleus, versus time, for envelope of \(10^{-4} M_\odot\). From Leising and Clayton (1987).

FIG. 13. - A sketch of the orbital evolution of V1500 Cygni due to its nova eruption. From Stockman, Schmidt and Lamb (1988).

FIG. 14. - The absolute magnitude \(M_B\) of novae at maximum light vs. the logarithm of time (in days) required to decline by 3 magnitudes from maximum. From Shara (1981).

FIG. 15. - The B and H\(\alpha\) light curves for a typical nova in M31. This characteristic light curve shows that shortly after outburst the nova is much more easily detected in H\(\alpha\) than in B. From Ford and Ciardullo (1988).

FIG. 16a. - The positions of the 35 novae detected in the Ciardullo et al. (1987) H\(\alpha\) survey. The boxes represent their CCD fields; the ellipses reflect M31's isophotes as measured by de Vaucouleurs (1958). The position of M31's nucleus is assumed to be \(\sim (1975) = 0^h41^m22.2^s; \delta(1975) = 41^\circ\ 08'\ 05''\). Ten of the 35 novae are within 2' of the nucleus. From Ciardullo et al. (1987).

FIG. 16b. - The cumulative distribution of isophotal radii for our 35 novae in M31. The solid line is the total amount of \(B\) light contained in the survey fields vs. the corresponding isophotal radius. Near the center, the novae follow the light precisely; at larger distances, the nova density drops off faster than the light. From Ciardullo et al. (1987).

FIG. 17. - Distribution of Novae in the LMC. The vertical cross marks the centroid and its extent indicates the standard error of the mean value. The diagonal cross shows the center of rotation used by Feitzinger (1980). From Graham (1984).
Figure 1

Figure 2
Figure 3

Figure 4
Figure 5

Figure 6
Figure 7
Figure 8

Figure 9
Figure 13

a.) Synchronous Pre-Nova System

b.) Early Nova Outburst

V = 500 – 2000 km/s

P = 0.140d

P = 0.141d

c.) Frictional Angular Momentum Loss (FAML)

V = 100 – 200 km/s

P = 0.141d

Accretion Stream

d.) Envelope Cooling Phase

P = 0.137d

Expanding Photosphere

Contracting Photosphere

Convective Transfer

Figure 13
Figure 14

Figure 15
Figure 16a

M31 NOVA POSITIONS

Figure 16b

M31 NOVAE -> M31 LIGHT

NUMBER OF NOVAE

ISOPHOTAL RADIUS (arcmin)