MODELS FOR VARIOUS ASPECTS OF DWARF NOVAE AND NOVA-LIKE STARS

I. HISTORY OF MODELING

A. HISTORICAL OVERVIEW

RELEVANT OBSERVATIONS: For about the first 100 years of research in dwarf novae, almost only outburst light curves were known. Neither photometric nor spectroscopic variabilities on time scales shorter than a couple of hours were resolved. Practically no nova-like systems (in the modern sense of the word) were known.

see: 15, 21, 95

ABSTRACT: The first attempts to explain the nature of dwarf novae were based on the assumption of single-star phenomena, in which emission lines were assumed to be caused by circumstellar gas shells. The outburst behavior was tentatively ascribed to the kind of (also not understood) mechanism leading to nova outbursts. The realization that some, and possibly all, dwarf novae and nova-like stars (and novae) are binaries eventually led to models which bore more and more similarities to the modern interpretation on the basis of the Roche model.

The first dwarf nova, U Gem, was detected in 1855; SS Cyg followed in 1896; and soon afterwards they were joined by ever more similar objects. Along with these detections, from the beginning of this century on, ever new attempts were undertaken to develop conceptual models of what a dwarf nova may consist of physically. The major problem, and the reason for the failure of all attempts to arrive at a suitable explanation until the 1950's, was that no strictly periodic pattern could be found in any of the dwarf nova light curves. Only the availability of ever larger telescopes and ever more sophisticated technology, finally, in 1955, revealed the searched-for periodicity in the nova-like star AE Aqr and in the old nova DQ Her and thus their binary nature (the strong similarity between novae, dwarf novae, and nova-like stars meanwhile had well been established). This quickly led to the formulation of the Roche model (Chapter 4.II.A) which is still today the only serious model for these objects and, in fact, in an ever more sophisticated form, seems to be able to account for at least the majority of the observed features.

A brief survey is given below of the development from initial tentative interpretations to the modern understanding. As we approach the present era in the historical summary, more and more features in the successive attempts to explain these stars emerge which resemble the modern explanation, and, in retrospect, the development towards the Roche model appears to have been rather direct, until, at one point, this model must have seemed an obvious next step.

For about the first century of observations of dwarf novae, photometric observations were mostly restricted to visual brightness estimates; spectroscopic measurements were, when possible, of very low wavelength and temporal resolution. Thus, attention was focused on the outburst behavior alone, while the existing periodicities, on the order of hours and less, remained undetected. This led to some immensely
valuable statistical investigations of the outburst behavior of the two brightest objects, SS Cyg and U Gem, which are discussed in Chapter 2.II.A. However, detailed and extensive as these studies were in revealing some obvious correlations, they provided but little aid for constructing models. Certain observational features were particularly puzzling, like the occasional standstills of Z Camelpardalis stars or the temporary suspension of activity in SS Cyg (the only known example at that time); but in first attempts at interpretation these were ignored.

Eventually the first observations of U Gem were almost entirely forgotten. Prejudices due to our modern ideas about its nature make it hard to believe that there is any basis at all to what was historically reported: Hind (1856), on December 15, 1855, was the first person to detect U Gem as a 9th mag star in the sky, in a field which, he claims, he had been familiar with for five years (the outburst period of U Gem has been between 60 and 150 days ever since, and the maximum brightness always has been on the order of 9 mag). Did it only in 1855 start its outburst activity? Some support is given to this conjecture by reports of very unusual behavior at somewhat later times (van der Bilt, 1908): Pogson, who was reputed to have been a very careful observer, reports having seen "the variable subject to strange fluctuations of intervals of 6 to 15 seconds, and quite to the extent of 4 mgs. The neighboring small stars were steady, not at all twitching like the variable. The phenomenon.... was watched for about half-an-hour...."; and according to van der Bilt, other observers also described the star as being very variable at that time; no similar reports can be found for later times. Similar behavior has never been reported for any other dwarf nova or nova-like star.

Van der Bilt (1908) reports a first interpretation by Nijland, who tentatively assumes that long and short eruptions are essentially identical, with only short outbursts due to a superposition of an eclipse at the end of the outburst. In fact, a subtraction of a short outburst from a long one (for SS Cyg and U Gem) leaves a curve like that known from Algol variables. It is not explained, however, how an eclipse, i.e., a geometrical phenomenon, can account for the clearly not constant time intervals between successive short maxima, and for the irregular sequence of short and long maxima.

More usual attempts at explanation were based on the appreciation of the features common to novae and dwarf novae: a quick and sudden rise to maximum, and a slower decline. In some cases, similarities in the spectra can be noted as well. Common, or at least similar, underlying causes for the brightness changes were suspected, supported by the realization that dwarf novae and recurrent novae seem, statistically, to follow the same relation between time spent in minimum state and amplitude of the outburst (Chapter 2.II.A.3). This relation would predict recurrence times of several thousand years for novae, in full agreement with their having been observed only once in historical times. And it also is in agreement with Gordeladse's (1938) finding that the mass ejected by dwarf novae (supposing the same mechanism is at work as in novae), integrated over outbursts during some 5000 years, amounts to approximately the same mass (he derives some 10^{29} g) as ejected by a typical nova during one outburst.
A suggestion by Vorontsov-Velyaminow (1934) was based on the apparent similarities of the spectra of several types of stars, all of which he classifies as "nova-like" in modern terms, these are recurrent novae, dwarf novae, nova-like stars, and also R CrB stars, P Cyg and others. He supposes that they all are related to novae in some way: when a nova undergoes an outburst, he assumes, its atmosphere expands considerably; this, however, is not a stable state because the star collapses again, and at some state of collapse it reaches a new equilibrium; according to the resemblance of their spectra to those of the outbursting novae, the "nova-like" stars are regarded as different final states of this collapse — recurrent novae, for instance, and symbiotic stars (using modern terms) did not collapse very far; since this state is not quite stable in the long run, one day these stars will undergo an outburst. Dwarf novae and modern nova-like stars have collapsed much further and thus are moderately stable. That this is no satisfactory explanation, in fact no explanation at all, for the photometric behavior of dwarf novae is obvious, since practically none of the then already known observational features are explained. No further work along this line has been pursued.

Gerasimovič and Payne (1932) investigated color changes in U Gem and SS Cyg as they changed from minimum to maximum state, respectively. Since they could find changes in the color index by only 0.3 mag, they concluded that almost all of the brightness changes must be ascribed to changes of the radii of the stars, and that temperature changes play only a subordinate part.

This opinion is strongly contradicted by Hinderer (1948), who gained extensive photometric and spectroscopic observations of SS Cyg during various stages of activity. He concluded that most of the changes in brightness are due to mere changes in temperature (from some 5650 to 9600 K), except during the brightest phases of an outburst when the star's radius increases by a small amount. From the strong hydrogen emission lines during maximum a Zanstra temperature of some 20000 to 50000 K (or in some modified version of 10000 to 15000 K) would have to be derived, values which to Hinderer seemed unacceptably high; so he concluded that the conditions for the application of this method are not met and that the star's shell may be more highly excited than what corresponds to the equilibrium value. He also mentioned the additional difficulty that He II 4686 Å was seen in emission during quiescence, which also required an ionization temperature of at least some 20000 K. During maximum state, only absorption lines were observed, with no trace of emissions, which led him to conclude that the shell producing the emission lines had largely disappeared or had at least become fainter with respect to the photospheric region. During intermediate stages of decline, a superposition of broad symmetric absorption lines and narrower, but also symmetric, emission lines was observed. Since the line centers of both components coincided within the accuracy of these measurements and no changes in radial velocity could be determined, Hinderer concluded that the whole outburst was mostly an effect of variable opacity in the atmosphere, not connected with any relevant motions of material. The very broad absorption lines which were observed (corresponding to Doppler velocities of up to almost ± 3000 km/s) do provide a problem: if they would result from rotational motions of the star, this star would be torn apart. Elvey and Babcock (1943) argued that such rotational velocities may still be possible for a short time when the star goes through a period of instability during rise. However, this would not explain how these velocities could be maintained for many days almost all the way down the decline. Hinderer rejected this possibility altogether. He also discarded the possibility of the widening originating from the Stark effect, since this would require high pressures and thus high densities which cannot be reconciled with the existence of emission lines. Why there could not be two independent regions of line emission from which the two
sorts of lines could be emitted, he did not com-
ment on. So he concluded with only the as-
sumption that the lines were broadened by
turbulent motions in the outer layers of the star,
without considering any further what turbulent
motions of the required strength might do to
the atmosphere, and what might cause them.

On the basis of these conclusions, Hinderer
devised a model for the physical processes oc-
curring in dwarf novae during their outburst cy-
cle. He emphasized that the observations he
made of SS Cyg and the conclusions based on
them were in excellent agreement with the spec-
troscopic observations of many other dwarf
novae by Elvey and Babcock (1943), so he con-
juctured that his model might be applicable to
the entire class. Noting that there was no con-
clusive evidence for dwarf novae to be white
dwarfs (as was suggested by Elvey and Babcock
and also by Miczaika and Becker (1948)),
Hinderer, in accordance with Joy (1948),
assumed that dwarf novae were main sequence
stars of type G3 to G4 corresponding to a color
temperature of about 5000 K. In his picture,
they are surrounded by a shell of thin gas ex-
tending outward to some 1.8 stellar radii; this
shell produces the observed emission lines; the
cause of the outburst is believed to be the libera-
tion of some inner energy of the star, as for
novae, which leads to an increase in
temperature and a moderate increase of the visi-
ble radius out to about the outer limit of the
emission line region. The importance of this
region to the emitted radiation is thus largely
reduced, leaving only the absorption spectrum;
only slowly, when the star's radius shrinks
during decline, does it gain back its old impor-
tance. The different shapes of the outburst light
curves are due to the energy not always being
liberated in exactly the same way; the similari-
yty of the brightness changes in all the decline
phases results from the system properties during
relaxation, after the energy supply is turned off.
During the time between outbursts new energy
is accumulated. The Z Camelopardalis stars are
regarded as a transition state to "normal"
stars.

The dwarf nova SW UMa was observed spec-
troscopically by Wellmann (1952) during a
maximum. Here also very broad absorption
lines are visible, and Wellmann is concerned
with their interpretation. He immediately
discards rotational broadening as a likely
possibility for the same reason as Hinderer,
namely, that a stellar rotation of 1600 km/s (in
the case of a system such as SW UMa) simply
would not be stable. Turbulent motions as sup-
posed by Hinderer also do not seem reasonable,
since, in order to be strong enough to account
for the observed line width, the turbulent
pressure would inflate the atmosphere up to 10
to 50 stellar radii, in which case the emission
line spectrum would have a distinctly different
appearance. As another possibility, Wellmann
suggests that very broad absorption lines which
already possess the full observed line width,
may partly be filled in by emission originating
in an extended shell of optically thin material
surrounding the star. No explanation is pro-
vided for the questions as to what process may
produce such strong absorption lines and why
the emissions almost always just fill in the line
center and appear as emission profiles of their
own only in less luminous stages.

From radial velocity measurements in 1954,
AE Aqr, classified in modern astronomy as a
dwarf nova or a nova-like star (Joy, 1954), and
DQ Her, a nova (Walker, 1954b), were detected
to be binary stars; and they were soon follow-
ed by others. This led to the speculation that
possibly all cataclysmic variables are binaries.
In fact, the subsequent detection of the binary
nature of eventually almost all well investigated
cataclysmic variables finally led to a quite pro-
mising model for these stars and represented to
many astronomers (at least until the present
time, and there is no real indication for this
period to be about to end) an end to a long
search. This new "Roche Model," first outlined
by Crawford and Kraft (1955; 1956), is prac-
tically unchallenged and forms the basis for all
modern interpretation of all the members of the
class of cataclysmic variable stars. It will be
presented and discussed in detail in Chapter 4.11. For its derivation, observations of the few known nova-like stars, in particular of UX UMa, were of great importance. This star and early explanations of its nature shall now be presented briefly.

UX UMa has been known since 1933 to be an eclipsing variable star (Beljawsky, 1933). It is not just an ordinary eclipsing star, but a very remarkable one (Figure 4-1): for a long time it was the one system with the shortest known orbital period, of only 0.1967 days, but it does not show the W UMa-type light curve which would be expected if normal stellar components are very close together; its light curve looks similar to that of an Algol star. Any attempts to derive orbital elements failed, however, since a hump appears shortly before and extends until shortly after eclipse, and the exact shape of both the hump and the eclipse are available from one cycle to the next. Thus it was clear that this system was not just two normal stars eclipsing each other. Linnell (1950) proposed prominences or gaseous streams at some distance from the eclipsed star, to account for both the hump before and after eclipse as well as for a hold seen at rise from minimum light; the additional gas was assumed to be hidden by one of the stars at times of normal brightness, and the most active region on one of the stars was supposed to be eclipsed at times of the hold; the observed variability of the hump might be a sign of either variable prominence activity or variable activity in the steam-like region. The fact that the hump was decidedly more pronounced before than after eclipse could be understood as either different cross sections of the stream being seen or rather as some persistent hot spot which was connected with the advancing hemisphere of the bright star. (Essentially this same scenario was later adopted by Walker and Herbig (1954) and further refined on the basis of more observational
material.) A gaseous cloud was proposed with inhomogeneous temperature distribution having a hot “head” and a cool “tail” to circulate around the bright component (Figure 4-2). As will be seen in the next chapter, this interpretation of UX UMa bares a striking similarity to the modern explanation via the Roche model: the gas cloud will be replaced by a gaseous disc, the bright “head” by a “hot spot,” but otherwise it remains about the same.

Before turning entirely to the Roche Model, one final, though unsuccessful, attempt to explain the outburst behavior of dwarf novae should be mentioned. Also inspired by the discovery that possibly all cataclysmic variables are binary stars and by his own new theory of nova outbursts, Schatzmann (1959) tried to explain the outbursts of the dwarf nova SS Cyg in terms of a resonance phenomenon between binary motion and non-radial oscillations of the surface of one of the stars, leading to ignition of He burning in locally constrained areas of the star’s surface. Based on this theory, Zuckermann (1961) deduced a more elaborate model of SS Cyg. She assumes that the system consists of a G5 main-sequence star and a blue subdwarf — the explosive star — which is embedded in a spherical envelope of ionized gas; all the system may be surrounded by a cloud of circumstellar material. By this explanation the colors of the minimum spectra can be accounted for. During an outburst, she supposes, the outer layers of the shell which surrounds the blue dwarf are accelerated outward by the hot events underneath, but in the course of the process of expansion the shell cools appreciably.

Neither of these concepts on the cause of a dwarf nova outburst or on how the system is structured and evolves otherwise could stand up to further confrontation with observations. The light curves of eclipsing dwarf novae during the outburst clearly demonstrate that the seat of the outburst cannot be a restricted area on the surface of one of the stars; and an outburst is accompanied by a very appreciable heating in the system and clearly not by any sort of cooling.

Eventually the Roche model became the currently only widely accepted model for a cataclysmic variable. This is not to say that it provides the complete solution, and it cannot be excluded that one day new discoveries will force us to give up this model and to replace it with another one; but so far it mostly works. Thus a gross model seems to apply to all cataclysmic objects (with the possible exception of symbiotic stars), subject to refinements and modifications which may account for characteristics of sub-classes and peculiarities of single objects. At the same time, it cannot be denied that there are observations which either cannot be explained at all within the framework of this model, or which can only be explained by introducing additional, questionable assumptions.

II. MODERN INTERPRETATION

The history of the first 100 years of observations of dwarf novae and nova-like stars was reviewed in the previous chapter. Attempts to understand the physical nature of these systems invariably failed until it was determined that all of these objects, including novae, are probably binary stars. Now, more than 30 years later, there is no observational evidence to contradict this hypothesis, and for theoretical modeling it has proven to be extremely useful.

Not all cataclysmic variables are known binaries. In fact, with respect to the entire number of known objects, the proven binaries are still the minority, but all the brightest variables are in fact known to be binaries. Not a single system is known which exhibits the usual characteristics of a cataclysmic variable and at the same time can be declared with certainty to be a single star. Two systems are known, the dwarf nova EY Cyg and the recurrent nova V1017 Sgr, in which, in spite of intensive search, no radial velocity variations
have been found; but they still exhibit composite spectra consisting of a bright continuum, an emission spectrum, and a cool absorption spectrum. If the Roche model is correct, it is to be expected that a small percentage of objects is viewed pole-on, so orbital motions do not make themselves felt as Doppler shifts of spectral lines. So even these two systems support the hypothesis that all cataclysmic variables (with the possible exception of symbiotic stars) are binaries.

II.A. THE ROCHE MODEL

RELEVANT OBSERVATIONS: All well-studied cataclysmic variables turn out to be binary systems with orbital periods of typically 90 minutes to four hours. One component usually is a cool main sequence star (if the orbital period is not larger than some 10 hours), and the other is usually a hot object with a geometry and flux distribution much unlike normal stars.

see Chapters 2 and 3

ABSTRACT: The canonical model of a cataclysmic variable is a Roche lobe-filling cool main sequence star which loses matter into the Roche lobe of the white dwarf. The transferred material has too much angular momentum to fall onto the surface of the white dwarf, but builds out an accretion disc in which it slowly spirals towards the white dwarf to eventually be accreted.

Observational evidence for the various components has been given in the theoretical abstracts following major sections in Chapters 2 and 3. Here, a more detailed description of the basic Roche model is given. For every binary system it applies that from the combined effect of the gravitational potentials of the two stars and their motion around each other, for each component there is a maximum geometrical volume, the Roche volume (or, in a two-dimensional picture, the Roche lobe), matter contained in which is bound gravitationally to the respective star; outside of this, matter either is bound to the system as a whole or, even further outside, it is hardly affected by the system at all (Figure 4-3). If it is further assumed that, for the special case of a cataclysmic variable binary system, one component is a white dwarf and the two components are close enough to each other so that one component (normally, but not necessarily, a main sequence star) fills its Roche lobe, then the model follows immediately.

Based on the evidence that all cataclysmic variables are binaries (as was first expressed by Struve, 1955), and making use of promising aspects of former models, the Roche model for cataclysmic variables was first developed by Crawford and Kraft (1955; 1956) and was quickly established, mainly due to extensive research by Kraft, Krzeminski, Mumford, Walker and co-workers (Kraft, 1962b; Krzeminski, Kraft, 1964; Mumford, 1964; Kraft, 1965; Mumford, 1966; Mumford, 1967; Mumford, 1971). The model became more refined and sophisticated in order to be able to explain ever more observational details and properties of cataclysmic systems, but in essence it never was changed. In particular, the hypothesis gained increasing support that the binary nature is a necessary condition for a system to become a cataclysmic variable.

THE ROCHE MODEL FOR CATACLYSMIC VARIABLES:

One component of the system, the primary star, is a white dwarf much smaller than its Roche lobe, the other, the secondary star, fills its Roche lobe (Figure 4-4). At the inner Lagrangian point L, the secondary star slightly overfills this volume, and thus matter is spilled over into the Roche Lobe of the white dwarf. Since angular momentum has to be conserved, and because of the tiny dimensions of the primary, this material flying along its trajectory inside the white dwarf's Roche volume would not hit the star's surface, but rather would meet again the stream of injected matter, thus forming a ring around the central object. As viscous forces are at work, the matter gradually loses angular momentum, and this
Figure 4-3. Roche surfaces. Material contained in one of the critical surfaces clearly belongs to the respective star, the two volumes touch each other at the inner Lagrangian point $L_1$ (Kopal, 1978).

Figure 4-4. The Roche model for cataclysmic variables.

ring eventually spreads out to form a disc which lies in the rotational plane of the system, eventually extending down to the white dwarf. At its outer rim, where newly incoming material hits the disc, the so-called hot spot, or bright spot, is formed. When matter is approaching the white dwarf it has to get rid of excess gravitational energy, half of which — according to the Virial Theorem — is converted into kinetic energy of the disc material, while the other half is transformed into radiative energy, causing the disc to shine as a luminous object. The white dwarf need not rotate at Keplerian velocity; thus, at the interface between the innermost disc area and the white dwarf, the disc material will have to be braked down to the velocity of the white dwarf, in the process of which additional radiative energy will be liberated and the boundary layer will be formed.

In this picture, a dwarf nova outburst is attributed to a sudden collapse of much of the accretion disc, induced either by some process intrinsic to the disc itself, or by a sudden mass outburst from the secondary star. A nova-like star is thought to be a dwarf nova in permanent outburst; and a nova-explosion itself is thought to be due to a thermonuclear runaway on the surface of the white dwarf. A special sort
of cataclysmic variable originates if the primary's magnetic field is strong enough to entirely prevent the formation of an accretion disc (as postulated in the AM Herculis-type nova-like variables). In this case, the matter which falls in from the secondary star is funneled directly onto the white dwarf.

II.B. CONCEPTUAL PROBLEMS AND BASIC EQUATIONS

ABSTRACT: The Roche model appears to determine the basic physics of cataclysmic variable systems. Concerning details, however, the situation is not yet clear, but it is probably far more complicated than can reasonably be handled theoretically at this stage.

It is one thing to devise a conceptual model of a cataclysmic variables system and to develop plausible explanations for certain observed features. It is quite another matter to try to reproduce theoretically observed light curves or spectra. Allowing the Roche model to become "alive" and to consist of physically reasonable objects rather than two neatly defined balls of gas (the stars) and a smooth disc, it is obvious that the physical processes in a cataclysmic system are very likely to be rather intricate. This somehow has to be accounted for if computations of the emitted spectrum and the outburst light curves are to have any relevance. It will turn out, however, that most of the physical effects one can imagine likely to be present cannot now be dealt with theoretically in a satisfactory way, either for lack of adequate physical understanding or numerical skills, or due to limited computer capacity.

Observations indicate that the secondary stars are main sequence stars or at least lie close to the main sequence. In some cases their temperature and radius can be determined. From observations the spectral types are determined to be approximately G to M. In analogy to what is known about such stars (when they are single), magnetic activity brings about spots on the surface, surges, prominences, magnetic loops, and coronae and mass outflow, all of which probably does have some effect on other parts of the system and in turn is influenced by them. It is not known what effect it has on the structure and evolution of the star to be confined to the non-spherical shape of the Roche volume and what influence the dramatic lowering of the gravitational acceleration in the vicinity of the Lagrangian point L₁ may have on the outer layers of the star, and what the enforced co-rotation (at a speed much higher than normal for this kind of stars) may do to the secondary.

The masses, and thus the radii, of the white dwarfs in cataclysmic variables are known only with large uncertainty, which probably is of only little importance for the dynamics of the system, in particular the disc; but this is of great importance for the emitted spectrum (Chapter 4.IV). Furthermore, the white dwarf may possess a magnetic field, like many single white dwarfs do. Fields with a strength of up to some $10^6$ G normally cannot be measured directly in these stars, but such fields are able to push the inner disc radius far away from the surface of the white dwarf because, at the distance of the Alfvén radius, the influence of the stellar magnetic field on the matter circulating in the disc becomes strong enough to force the matter to co-rotate with the star's field lines rather than follow the motion of the disc, thus disrupting the disc at this distance from the stellar surface. For a strongly magnetic white dwarf which creates a bi-polar field, the distance of the Alfvén radius from the stellar surface and the gross structure of the magnetic field can be calculated easily (Lamb et al, 1973). If, however, the magnetic field is weak enough so that the disc is disrupted relatively close to the surface of the white dwarf, the field is no longer bi-polar and conditions become rather complicated.

Problems are also severe in the seemingly simpler case of a non-magnetic white dwarf. For the disc to be stable, matter in the disc is
expected to rotate at Keplerian velocity. It is fairly unlikely, however, that the white dwarf also rotates with this velocity, which for it would be at the limit of disruption; indeed, it may not be rotating at all. The probable case is somewhere between these two extremes. If the white dwarf is rotating at a velocity slower than its Keplerian velocity, the matter arriving from the disc must be slowed down in the boundary layer before accretion onto the white dwarf is possible. According to the Virial Theorem the kinetic energy of material close to the white dwarf is of the same order as the total energy radiated away by the entire disc. Thus in the case of a non-rotating white dwarf all this energy is liberated in the boundary layer. The faster the white dwarf is rotating, the less kinetic energy has to be converted into radiative energy, and the less spectacular the boundary layer and its radiation will be.

If no frictional forces were at work, the material transferred into the accretion disc from the secondary would circulate infinitely around the white dwarf without ever being accreted. It is only the viscosity which causes the particles to lose energy and thus to slowly spiral onto the white dwarf while the remaining angular momentum, carried by a small fraction of the material, is being carried away from the white dwarf. Thus, the physical parameter which very critically determines the appearance and temporal behavior of the accretion disc is the viscosity, which, however, is largely unknown.*

The viscosity determines how much energy is liberated at any point in the disc (i.e., the temperature); it governs the geometrical and, together with the temperature, the optical thickness at any distance from the white dwarf; and it also probably governs the cause and development, the presence or absence, of any outburst behavior. Whatever computations involving the viscosity are carried out, some essential, but still somewhat arbitrary, assumptions are needed; and the constraints imposed by available observations are not as tight as one would like.

There is one particular assumption — although it is certainly somewhat arbitrary — which has proved very useful for spectrum computations in particular (Chapter 4.IV), since it provides a simple analytical formula for the radiation emitted by the accretion disc. The assumption is that the whole disc is stationary, i.e., matter transferred from the secondary star is transported at a constant rate all the way down onto the white dwarf — in other words, the mass transfer rate $\dot{M}$ is constant throughout the disc. In this case the entire energy emitted by the disc is

$$L_d = \frac{G M_{\text{wd}} \dot{M}}{2 R_{\text{wd}}} \tag{4.1}$$

and if a non-rotating white dwarf is assumed, that same amount is emitted once more in the boundary layer (see Chapter 4.IV.F). Equally, under this assumption the effective temperature $T_{\text{eff}}(r)$ at each distance $r$ from the white dwarf can be shown to be

$$T_{\text{eff}}^4 (r) = \frac{3 G M_{\text{wd}} \dot{M}}{8 \pi \sigma r^3} \left(1 - \sqrt{\frac{R_{\text{wd}}}{r}}\right) \tag{4.2}$$

(all symbols have their usual meaning; for details of the derivation see Verbunt, 1982). The term in brackets accounts for the transfer of

*It should be realized that “viscosity” is just a different word to describe the mechanisms of probably turbulent angular momentum transport in the disc which are suspected to be present. Its parameterization by, for instance, a quantity $\alpha$, defined as the ratio of the turbulent velocity to the sound velocity, including the actions of probably present magnetic fields (Shakura and Sunyaev, 1973), at any point in the disc (thus the often used expression “$\alpha$-disc” which refers to just this way of parameterization) has no real physical meaning other than that of a basically free parameter which is likely to have different values at different points in the disc. As for its size, considerations about disk stability require it to be on the order of $10^{16}$ cm/sec (e.g., Hensler, 1982b), which is decidedly larger than molecular viscosity.
angular momentum between the disc and the white dwarf and imposes a certain, though in practice probably unimportant, uncertainty on the value of the effective temperature. It needs to be stressed, however, that the radial temperature distribution is not known theoretically if accretion occurs in some non-stationary fashion: besides irradiation which is not considered here, the only energy source in the disc is gravitational energy, set free by viscous interaction; if material circulates in the disc at a constant orbit, no energy is gained. In Chapter 4.IV.E, ways will be discussed how temperature distributions can possibly be derived in an empirical way. No matter what the viscosity is (unless it is unreasonably large), the material in the disc is rotating at Keplerian velocity \( v_\varphi \), corresponding to its distances from the white dwarf:

\[
v_\varphi^2 = \frac{G M_{WD}}{r}
\]  

The approximate size of the disc can be estimated by, on one hand, the observational as well as theoretical result that is smaller than some 2/3 of the entire Roche radius and, on the other hand, by the total size of the Roche radius which is estimated from the orbital period and the masses of the two stars as approximately

\[
\frac{R_2}{R_\odot} = 0.959 \frac{M_2}{M_\odot}
\]  

(for details and inherent assumptions see Chapter 4.II.C.1). The inner disc radius is set by the white dwarf's radius in the case of a non-magnetic, or weakly magnetic, star; and in the case of a magnetic white dwarf, the exact value of it at some distance from the surface depends on the mass of the star, the field strength, and the mass accretion rate.

The hot spot is the place at which the stream of infalling matter from the secondary star hits the accretion disc. From observations, its azimuthal angle with respect to the line connecting the centers of the two stars is approximately known; not so, however, its radial distance from the white dwarf, nor its exact shape. Depending on how wide the stream is with respect to the geometrical thickness of the outer disc, the stream can be imagined to either swamp this part of the disc or, if it is much narrower, to deeply penetrate into it. Clearly both the radiation (through the radiation characteristics of the spot) and possibly the outburst behavior (through the place and amount of energy deposited in the disc — see Chapter 4.III.C.2) of a cataclysmic system depend on the nature of the hot spot.

One final point is that in a cataclysmic system radiation emitted by different components and, to the extent they are present, magnetic fields will interact with each other and with other parts of the system changing conditions there. One can imagine that irradiation is particularly important if strong X-rays are emitted from the boundary layer and illuminate, and thus heat, the inner disc close to the white dwarf. A sort of corona may be formed, which probably accounts for the very strong emission lines of heavy elements seen in dwarf novae during quiescence, as well as strong P Cygni profiles seen during outburst in dwarf novae and in many nova-like stars.

II.C SYSTEM PARAMETERS

Absolute values of parameters of many systems have been quoted in the literature (e.g., Córdoval and Mason, 1983; Smak, 1983, Patterson, 1984; Patterson and Raymond, 1985a; Ritter, 1984, 1987; Warner, 1987). Since it is important to have as reliable values of physical quantities as possible, and since it is even more important to know how reliable these are, different methods for determining various of these quantities will be described and discussed in the following sections. This discussion is not intended to be totally comprehensive; its aim is mainly to caution, and to point out where in the determination of these parameters problems are likely to occur.
II.C.1 STELLAR MASSES, STELLAR RADII, AND INCLINATION ANGLES

ABSTRACT: Great care must be taken in the determination of radial velocity curves that inhomogeneities in the disc and irradiation of the secondary star are properly taken into account. Constraints due to the Roche geometry have proved to be important in determining system parameters.

In a normal detached binary system, absolute masses and radii as well as the inclination angle can be derived from radial velocity curves and light curves, applying well-established standard procedures. Prejudice-free (i.e., model independent) application of this procedure to cataclysmic variables caused much confusion in the past before the Roche model was applied to cataclysmic variables; and it led to practically no useful results. Only two eclipsing double-lined spectroscopic binaries are known among these systems, so strictly speaking it should only be possible to determine masses for these two systems. Furthermore, from the shape of the light curves it is clear that the system geometry is not really that of two detached stellar components, so it is highly questionable whether the inclination angles — and radii — of the stars could be determined in a meaningful way using the standard procedure for binary stars.

The establishment of the Roche model brought about a considerable improvement in this situation. Geometrical considerations and the assumption that the secondary star fills its Roche lobe impose enough constraints so that system parameters can now be derived for a large number of cataclysmic variable systems. At the same time, however, new problems arise.

The function $M \sin^3 i$ can be derived from the radial velocity curve of each component, as in other binary systems. There are, however, major problems in the determination of the radial velocity curve itself. In the case of the primary component (white dwarf), the spectrum emitted by the star itself is only very rarely visible, and if so, it is still: heavily contaminated by radiative contributions from the disc. In most cases the lines which are observable (in dwarf novae during the quiescent state these are mostly emission lines) are those originating in the accretion disc surrounding the white dwarf. If the disc were perfectly rotationally symmetric in its radiation pattern, its presence should not pose a problem, and the orbital motion of, in effect, the white dwarf would still be measurable. But it is not. Certainly the hot spot brings a decidedly inhomogeneous element to the radiation pattern (which clearly is visible in the line profiles). However, since neither its temperature nor its radiation characteristics nor its position in the disc are known for sure, it is not clear what part of the line profile is due to the hot spot and what is due to the disc in any single case. So clearly the normal way of fitting a Gaussian profile to the wings of a line (at least in systems which exhibit a hump) is bound to include all the distortions caused by the hot spot, and thus will not yield reliable results. (Incidentally, Stover (1981b) suggests that the phase shift of some 5-10° between emission and absorption components in double-lined spectroscopic binaries (Chapter 2.III.B.1.d) is due to just the effect). The often found result that the line centers move with different velocities than the (extreme) line wings is merely a result of inhomogeneities like the hot spot. In many, in particular high-inclination, systems, the Balmer lines exhibit a double-peaked profile which is explained as being due to the rotation of the disc; the peaks are ascribed to the outer edges of the disc and often radial velocities are derived from them. It is not clear, however, what this outer edge is, physically, or whether it is circular — thanks to the hot spot it probably is not — so the same problem will be encountered as described above.

The most reliable parts of a line profile for deriving the radial velocity curve are the extreme wings. Radiation producing them in a Keplerian rotating disc originates in the innermost disc close to the white dwarf, in an area
where at least the influence of the hot spot should be negligible, even though it cannot be excluded that inhomogeneities exist also there. At any rate, this seems to be the safest way to measure the radial velocity of the primary star. Shafter (1985) compared radial velocity curves of T Leo derived from different portions of the line wings (Figure 4-5) and found that the results depend dramatically on how the measurement was carried out. He suggests that the most reliable results can be obtained when several radial velocity curves are derived from different points in the line wings, so the point can be determined where the line wings just begin to merge into the continuum, believing that this then probably reflects most closely the motion of the white dwarf.

As for the secondary star, at first glance there should not be any problems; as long as the cool absorption spectrum is measurable it probably represents the cool star. That things are not that simple was first pointed out by Robinson et al (1986) who realized that the seven (!) measurements of the radial velocity amplitudes $K_1$ and $K_2$ for SS Cyg (which really should be the best measured radial velocity curve of all) all yield largely discordant results which do not even agree within the indicated errors (Table 4-1). A closer investigation of this case revealed that heating of the secondary star (of that half facing the disc) by the accretion disc changed the amplitude of the radial velocity curve ($K_2$) and also distorted it into a non-sinusoidal shape (Figure 4-6).

If both radial velocity curves, that of the emission spectrum (disc/white dwarf) and that of the absorption spectrum (secondary star) can be measured with some confidence, the orbital period $P$ and the ephemeris $T_0$ can be determined; furthermore, it is possible to obtain the "system velocity" $\gamma$, the mass ratio $q$, the distance between the centers of gravity, and the individual masses, all as some function of the angle of inclination.

In a double-eclipsing system the angle of inclination can be derived with fairly high accuracy from geometrical considerations, using contact times and system dimensions (Figure 4-7a), if the eclipses are identified as those of the white dwarf in first ingress and first egress and of the hot spot in second ingress and second egress (Smak, 1979; Ritter, 1980; Patterson, 1981; Vogt et al, 1981; Berriman, 1984; Priedhorski et al, 1987). In the case of a single eclipse the system geometry can be invoked only in combination with the mass ratio to derive some upper and lower limits for the inclination angle (Figure 4-7b). If no eclipse can be observed, guesses of the inclination angle $i$ are much more complicated and mostly rely on the presence or absence of a hump in the optical light curve.

As already noted, the constraints imposed by the Roche geometry have proved to be a big advantage in determining system parameters. The radiation of a spherical star filling the same volume as a Roche lobe-filling star has been computed by Paczynski (1971) to be

$$R_2/a = 0.38 + 0.20 \log q \quad (4.5a)$$

for $0.3 < q < 20$

$$R_2/a = 0.462 \left(\frac{q}{1 + q} \right)^{1/4} \quad (4.5b)$$

for $0 < q < 0.8$

with

$$q = \frac{M_2}{M_1},$$

i.e., $R_2$ only depends on the mass ratio and the distance between the two stars. Furthermore, there is plenty of evidence that the secondary

*Observations of several dwarf novae during outburst (e.g., SS Cyg, SU Ursae Majoris stars — see Chapter 2.III.B.2) demonstrate, however, that $\gamma$ can be a convolution of the system velocity and something else, since it is observed to be variable.
stars in systems with orbital periods below some 6 hours are indistinguishable from main sequence stars; for the latter there exists a relation between mass and radius which, to a fairly good approximation, is given by

$$R_2/R_\odot = 0.959 \frac{M_2}{M_\odot} \quad (4.6)$$

(Warner, 1972; 1976) for the range of interest for cataclysmic variables. Combining this with equation 4.4 and Kepler’s third law yields

$$M_2/M_\odot = 3.18 \times 10^{-5} \, P[s] \quad (4.7)$$

(Warner, 1976). So the knowledge of $M_1$, $M_2$, $\sin^3 i$, $q$, and the orbital period immediately provides a reasonably reliable inclination and thus individual masses.

The spectrum of the secondary star normally can only be measured in longer period cataclysmic variable systems, the more normal case being that only the emission spectrum, originating in the accretion disc, is visible. However, in this case too the Roche geometry helps to determine more system parameters than one normally would expect, although results become less reliable as more assumptions about system components and dimensions have to be introduced. Warner (1973) shows that the ratio between the radial velocity amplitude $K_1$ and $\nu \sin i$ for the equilibrium radius for a particle which is newly spilled into the Roche lobe is a function of only the mass ratio:

$$\frac{K_1}{\nu(r_d) \sin i} = \frac{f^2(q) \, q}{(1 + q)^2} \quad (4.8a)$$

with
Figure 4-6. Theoretical radial velocity curves for a K5V star in SS Cyg for different mass-transfer rates. The distortions of the shape and amplitude are due to heating of the secondary star by the disc and the white dwarf (Robinson et al, 1986).

Table 4-1. Determinations of amplitudes of the radial velocities, $K_1$ and $K_2$, in SS Cyg (Robinson et al, 1986).

\[
\begin{array}{ccc}
\text{K5 V Star} & \text{White Dwarf} & \text{Reference} \\
11^5 & 122 & \text{Joy} 1956 \\
11^5 & 122^* & \text{Walker and Chincarini} 1968 \\
10^* & 93 \pm 4 & \text{Kippenherr} 1979^* \\
13^* & 90 \pm 2 & \text{Stover et al} 1980 \\
120 \pm 8 & 114 \pm 3 & \text{Crawley et al} 1980 \\
117 \pm 2 & 107 \pm 3 & \text{Walker} 1991 \\
156 \pm 3 & 90 \pm 3 & \text{Herzog et al} 1984
\end{array}
\]

* Walker and Chincarini adopted Joy's $K$ velocities.
** Kippenherr rejected this velocity because it disagreed with Joy's and because the absorption lines in his spectra were sometimes doubled.

Figure 4-7b. In the case of a single eclipse, only upper and lower limits of the inclination angle can be obtained, from considerations about the system geometry (Robinson, 1974).

Figure 4-7a. In a double eclipsing system, the angle of inclination can be derived with high accuracy from the contact times of the eclipses of the white dwarf and hot spot, respectively (Ritter, 1980).

\[
f(q) = 0.500 - 0.227 \log q, \quad (4.8b)
\]

for $0.1 \leq q \leq 10$

where $f$ is the distance between the center of the white dwarf and $L_1$ (Warner, 1976). Furthermore, he shows that

\[
\frac{R_2}{M_2^{1/6}} = \left( \frac{G}{4\pi} \right)^{1/6} \left( \frac{1 + q}{q} \right)^{1/6} \times
\]

\[
\times P^{1/3} (0.38 + 0.20 \log q), \quad (4.9)
\]

which is only weakly dependent on $q$ provided $0.1 < q < 1$ (Warner, 1973). So in combination with the mass-radius relation (equation 4.4), the
mass of a cataclysmic variable system can be
determined for every system for which the
radial velocity curve of the emission lines can
be measured.

Difficulties with this method are ob-
vious: how can $v(r_d) \sin i$ be measured, and
what part of the line profile represents radia-
tion emitted from this hypothetical equilibrium
radium in the disc? Warner (1976) tentatively
identifies the half-width of the emission pro-
files with this value, but the physical
significance of this is not at all clear. Schoembs
and Vogt (1980) suggest a modification of this
method of mass determination which is based
on measurements of the far line wings, which
originate in an area adjacent to the white dwarf.
Their method probably yields much more
reliable results since it is based on more reliable
measurements. Further complications arise, of
course, from measurements of $K_1$, as discus-
sed above, and from the not very well determin-
ed mass-radius relation.

Robinson (1976) suggests an alternative to
determine the masses if either the mass ratio $q$
is known (for all double-lined systems) or, in
the case of single-lined systems, if the inclina-
tion angle is known. In the first case, use is
made of the relation between the secondary’s
mass and the orbital period, similar to equa-
tion 4.5:

$$M_2^2 = \left(0.996 \times 10^{-4} P^2 (1 + q) \right. \times (0.38 - 0.2 \log q)^3)/0.804 \quad (4.10a)$$

with $q = M_1/M_2$ \quad (!) \quad (4.10b)

(Robinson, 1976). In the second case, the above
equation is solved simultaneously with the
equation for the mass function for cataclysmic
variables:

$$\frac{M_2 \sin^3 i}{(1 + q)^3} = \frac{P K_1^3}{2\pi G}, \quad (4.11)$$

the right-hand side of which only contains
known values (in the case of a single-lined
system). The advantage of this method over the
method suggested by Warner (1976, see above)
is that all the necessary input values can be
determined reliably (while the basic assumption
about the nature of the system and in particular
about the secondary are the same). The disad-
vantage is that the difficulty of measuring the
radial velocity curve of some ill-defined radius
at the outer edge of the disc is replaced by the
not necessarily easier task of guessing the in-
clination angle.

Another procedure for determining the
parameters of the secondary star relies on con-
siderations of, in particular, IR colors (Wade,
1982). If the contribution of the disc to the IR
flux can be estimated — e.g., from the assump-
tion of a stationary disc with some outer radius
(a maximum of course is the radius of the
Roche lobe), or from the optical and IR flux
distribution, which in some objects shows a
distinct rise to the IR that can be ascribed to
the secondary star — then any excess flux in
the IR is ascribed to the secondary and thus its
spectral type can be determined. Assuming that
it is a main sequence star, the mass and radius
follow immediately. The obvious shortcoming
of this method is that the IR flux of the disc
is extremely poorly known. A more reasonable
estimate would go the other way, namely, if
from the size of the secondary’s Roche lobe
(i.e., from the orbital period — see equation
4.5) its IR flux is estimated, the contribution
of the disc at IR wavelengths can be
determined.

An unconventional method of determining
the inclination angle was devised by Horne et
al (1986). They made use of the constraints that
the mass of the white dwarf must not exceed
the Chandrasekhar limit and that the secondary
is a main sequence star, and then used Monte
Carlo techniques to determine that value of i
which with the highest probability produced the observed values of $K_1$ and $K_2$.

II.C.2 ABSOLUTE MAGNITUDES AND DISTANCES

**ABSTRACT:** In principle the spectral type of the secondary star can be used to determine absolute magnitudes. Relatively well-defined observational relations exist between the $V$-magnitude of a dwarf nova in outburst and the orbital period, and between the $V$-magnitude and the equivalent width of the $H\beta$ emission line, the UV absorption feature at 2200 Å, if present, should be indicative of the distance.

A search in the literature for distances and absolute magnitudes of dwarf novae and nova-like stars yields a wealth of values for each system which, even within the error limits, disagree. Since these values were derived under various assumptions, this very clearly reflects the serious difficulties which are encountered in this task, and leads one to suspect the large errors reflect a strong model dependence.

The first determinations of absolute magnitudes and/or distances of dwarf novae were based on parallax determinations. For U Gem and SS Cyg, trigonometric parallaxes were derived (Strand 1948; van Maanen, 1938; Becker and Mazaika 1948). A mean absolute brightness during minimum was derived for U Gem, SS Cyg, RU Peg, and EM Cyg from radial velocities and proper motions (Kraft, 1962b). This latter method is particularly dubious since, as has been discussed above, measurements as well as interpretations of the radial velocity curves are very difficult, and it is not at all clear what is being measured.

In many systems, in particular in those with longer orbital periods, the spectrum of the secondary component is visible in the optical, or at least in the IR. Under the (justified — see Chapter 4.IV.C) assumption that it is a main sequence star, it is in principle possible to determine its absolute magnitude — thus that of the entire system — from estimating its spectral type. This method has been applied to many systems (e.g., Wade, 1979, 1982; Berriman et al, 1985). If radial velocity measurements of the late spectrum are available, a determination of the spectral type should be straightforward and fairly accurate. If only low-resolution spectra or colors are available, however, it is usually not clear what contribution to the observed flux can be ascribed to the outer cool areas of the accretion disc. The constraint that the secondary fills its Roche lobe helps to provide a fairly reliable guess in cases where the mass ratio and the orbital period are known.

Bailey (1981, see also Warner, 1987) calibrated the surface brightness of the cataclysmic variable secondaries in the K-band, arriving at the relation

$$S_K = K + 5 \log d + 5 \log \left(\frac{R}{R_\odot}\right)$$

with

$$S_K = 2.56 + 0.508 (V-K) \quad \text{for } (V-K) < 3.5$$

$$S_K = 4.26 + 0.58 (V-K) \quad \text{for } (V-K) > 3.5$$

(Figure 4-8). He points out that the interstellar absorption is no problem at these wavelengths, and that the K magnitude is relatively insensitive to the temperature of the cool star. If the period and mass ratio of a system are known, the secondary's radius can be derived (see Chapter II.C). So the main difficulty again is to estimate the contribution of the star in the V band — which is a very serious problem for most systems, since here the flux from the star is probably weak while that from the disc can be fairly strong. Unless other information about the secondary is also available to constrain possible values, the errors in distances derived by this method are expected to be considerable.
Figure 4-8. The surface brightness of the secondary stars in the K band (22000 Å) is a good measure of their effective temperature (Bailey, 1981).

Warner (1987) found that there exists a relatively well defined relation between the absolute V magnitude of dwarf novae, in particular during outburst, and the orbital period (Figure 2-14; the theoretically inferred relation between the absolute magnitudes and mass transfer rates in cataclysmic variables is indicated as well)

\[ M_v(\text{max}) = 5.64 - 0.259 P[\text{hr}], \]
\[ \pm 0.13 \pm 0.024 \] (4.13)

which then provides absolute magnitudes for all dwarf novae for which the orbital period has been determined with an accuracy of ± 0.23 mag.

Finally, the interstellar absorption feature at 2200 Å has been used frequently to derive distances since the availability of UV spectra. Reddening curves have been published by, e.g., Nandy et al (1975) and Seaton (1979). In order to derive the amount of reddening, the observed spectra are corrected for interstellar absorption by using successively different values for E(B-V) until the absorption feature disappears from the spectrum. Normally results derived by this method are not in striking disagreement with values determined by other means, although error limits are fairly large — since the "best fit" is to some degree a question of taste. Furthermore, it is not certain that one single absorption law is applicable equally reliably to all directions in the Galaxy. Also experience has shown that great care must be taken to use well-exposed spectra for this procedure, since otherwise results are unreliable. Only an upper limit can be derived for most dwarf novae and many nova-like stars, however, since due to their proximity to the Earth no dip in the interstellar absorption is visible in the spectrum.

Patterson (1984) finds an empirical relation between the equivalent width of Hβ and the visual magnitude of cataclysmic variables (Figure 4-9):

\[ \text{EW}(\text{H}\beta) = 0.3 M_v^2 + e^{0.55 (M_v - 4)}, \] (4.14)

which allows an estimate of the absolute visual magnitude within some ± 1.5 mag for systems with strong emission lines (Wλ > 15 Å).
II.C.3 MASS TRANSFER RATES

ABSTRACT: It is necessary to distinguish between the mass transfer rate from the secondary star into the Roche lobe of the primary, the mass throughput through the disc, and the mass accretion rate onto the white dwarf. All determinations are strongly model dependent. Determinations based on eclipse-mapping techniques seem to be the most reliable.

A distinction must be made between the mass transfer rate from the secondary star into the Roche lobe of the white dwarf, the mass throughput through the accretion disc — which need not be uniform throughout the disc but well may depend on the distance from the white dwarf, and, as observations show, is clearly variable with time — and the mass accretion rate onto the white dwarf, although this is often not done carefully in the literature on cataclysmic variables. Only in the special (though theoretically most often considered) case of a stationary accretion disc are all these values by definition identical. In all other cases they are more likely to be different from each other, while relative relations can easily change with time: e.g., assuming that the mass transfer rate from the secondary into the primary's Roche lobe is not significantly changed during an outburst, then this mass transfer rate is probably the largest of the three during their quiescent state of dwarf novae, when material is preferentially stored in the outer disc rather than transported towards the white dwarf and thus the two other rates will be close to zero; during outburst, on the other hand, when the matter which was stored in the outer ring during quiescence is suddenly transferred through the disc to be accreted by the white dwarf, the mass throughput is probably largest, closely followed by the mass accretion rate*, while the mass transfer rate might temporarily be smallest. Although one can imagine that differences can be fairly large, normally no distinction is made when values for \( \dot{M} \) are determined, and implicitly either the mass transfer rate or the mass throughput are determined.

All known methods for determining the mass transfer rate are strongly model dependent, though all give results which agree within surprisingly narrow limits — as long as systems are considered which probably are close to the steady state (i.e., nova-like stars in the high state and dwarf novae during outburst). Fairly extensive discussions of various methods are given by (Patterson, 1984; Verbunt and Wade, 1984).

The theory of evolution of cataclysmic variables (Chapter 4.V.D) assumes that mass transfer in the ultra-short period systems, with periods shorter than 2 hours, is driven solely by gravitational radiation, while above the period gap magnetic braking (which results in strong mass loss) is at work. Computations for the amount of mass loss in short period systems yield qualitative agreement with mass transfer rates derived from observations (Patterson, 1984 — see also Figure 4-54), demonstrating that these mass transfer rates should be on the order of \( 10^{-10} \) to \( 10^{-11} \) \( M_\odot \)/yr. For long period systems above the period gap, larger mass transfer rates are expected. Thus, in general, mass transfer rates significantly lower than \( 10^{-11} \) \( M_\odot \)/yr can be excluded for cataclysmic variables. For an upper limit, values of \( \dot{M} \) both from observations and from computations (Figure 4-54), as well as values by other authors, suggest that mass transfer rates significantly in excess of \( 10^{-7} \) \( M_\odot \)/yr are quite unlikely to occur in cataclysmic variables, since then these systems would be dynamically unstable.

There are several ways to derive a mass throughput rate for the accretion disc, most of which, however, implicitly assume that the disc is stationary.

If somehow the radiation flux from the disc can be estimated, the mass throughput rate can be obtained from

---

*Some mass is lost in stellar wind, observable in the P Cygni profiles, and some is carried outward in the disc, taking care of the excess angular momentum.
A determination of the disc luminosity is difficult, since the bulk of the radiation is probably emitted at unobservable EUV wavelengths. So even if the disc is really stationary (which often is rather questionable), and even if the entire accessible wavelength range from the IR to the Lyman edge is covered observationally, some (largely arbitrary) bolometric correction has to be adopted in order to determine the luminosity. Correspondingly worse is the situation if either no UV data are available for the system, or if the disc is obviously not in a stationary state, as is the case for quiescent dwarf novae, for instance.

Another way of estimating the mass throughput rate is to compare the observed continuous flux distribution in the optical and UV with theoretical models and then derive values for $M$ from this. However, as will be discussed in more detail in Chapter 4.IV, the difficulty here is that the currently available models for stationary, and even less for non-stationary, discs are far too unreliable and far too dependent on details of the assumptions and of the program codes with which they were constructed to inspire any confidence, other than they might be crude estimates at best.

In some cases attempts have been made to determine average mass transfer rates from observed changes in the orbital periods of cataclysmic variable systems (e.g., Smak, 1971; Warner and Nather, 1971). These authors assume conservation of mass and angular momentum in the system, which, as Patterson (1984) points out, is inconsistent with evolutionary theories about these systems (Chapters 4.V.C and 4.V.D). Furthermore, since observations show that period changes do change their signs occasionally, it is clear that they cannot be due to mass transfer in the system.

In other cases attempts have been made to determine the mass transfer rate from the luminosity of the hot spot (Krzeminski and Smak, 1971; Paczynski and Schwarzenberg-Czerny, 1980). The obvious difficulties with this approach are that neither the shape nor the geometrical radiative characteristics of the hot spot are known, nor is its flux distribution (from which some kind of sensible bolometric correction could be derived), nor is the location of the hot spot in the disc which determines the amount of energy that is liberated.

A possibly reliable method to determine the mass throughput was derived by Horne and co-workers (Horne and Cook, 1985; Horne and Stiening, 1985; Wood et al, 1986) using the eclipse mapping method (Chapter 4.IV.E): the image of the accretion disc is reconstructed from photometric observations during the eclipse of a cataclysmic variable, preferentially in several colors; and from the colors at each point the temperature distribution and thus the local (!) values of $M$ can be determined.

III. THE ACCRETION DISC

Accretion discs have become very fashionable in astrophysics for explaining a wealth of phenomena which cannot be understood in terms of processes occurring, for instance, in normal stellar atmospheres. They have been applied to quasars, to active galactic nuclei, to X-ray and cataclysmic binaries, to symbiotic stars, to protostars, and to many more areas. Only in cataclysmic variables, however, does it seem that most astronomers are reasonably confident that the brightness changes observed in dwarf novae and nova-like stars in the optical and the UV are due directly to changes in the discs. Thus, the study and understanding of accretion discs in these systems can, and hopefully will, bear potentially valuable consequences for many other fields in astronomy.

It should be noted, however, that the unsolved or poorly understood theoretical problems concerning accretion discs are many and
difficult. A single particle added inside the Roche lobe of the white dwarf revolves around the central start in a circular orbit at a distance which is determined by the equilibrium between its own angular momentum and the gravitational attraction it experiences. When many such particles circulate in this way at the same distance from the star, they also interact with each other, and the distance-dependent Keplerian velocity ensures that shear stresses will develop between neighboring regions, probably causing many kinds of turbulence, while magnetic fields might also be present to further complicate the situation. The net effect is that mass and angular momentum become separated, with the bulk of the mass moving in orbits ever closer to the center of gravity as this material is deprived of ever more of its angular momentum, while most of the angular momentum, tied to very little mass, is transported outwards in the disc. (At the outer edge it either leaves the system, or, more likely, through interaction with the secondary star, is fed back into orbit as angular momentum.) Thus the mass which originally all circulated at roughly the same distance spreads out into a disc, both increasing the outer radius and decreasing the inner radius. Details of the strength and nature of the particle interaction are not known, but commonly they are covered conveniently under the term “viscosity.” To understand the nature and the role of the viscosity is the harder part, since complex processes of plasma kinetics, (magneto-) hydrodynamics, and poorly known radiative interactions, and other system parameters are involved. Notwithstanding, the study of the outburst behavior of dwarf novae and the brightness changes occurring during times between outbursts, might still provide some useful constraints on various concepts about, and models of, accretion discs. At all outburst stages the viscosity determines the size and geometrical shape of the disc, its radial and vertical temperature and pressure stratification, the spectrum emitted, and — since the gravitational potential is the only available energy source, made available exclusively by the viscosity — all observable (and unobservable) temporal changes. The theoretical study of these quantities shall be the concern of this section.

The general assumptions in theoretical work in the area of cataclysmic variables are: the neglect of relativistic effects and of self-gravity (i.e., the mass contained in the disc is negligible compared to the mass of the white dwarf); a disc which is assumed to be geometrically thin, lying flat in the orbital plane of the system and rotationally symmetrical with gravitational energy as the only energy source, which is converted into radiation energy by viscous processes which also cause the separation of

---

**Figure 4-10.** Qualitative model of an accretion disc. Indicated are the flow of mass (dotted lines) and angular momentum (dashed lines) in the disc; energy is generated by viscous heating which is the reason for the finite thickness of, and the radiation from, the disc (Katz, 1985).
angular momentum and mass. A certain vertical thickness structure of the disc is due to thermal pressure originating from the transformed gravitational energy (Figure 4-10.)

III.A. 2 DIMENSIONAL HYDRODYNAMICS

RELEVANT OBSERVATIONS: Photometric observations point to the disc being an essentially flat object.

ABSTRACT: The approximate position of the hot spot can be reproduced from computations of particle trajectories in the restricted three-body approximation. Hydrodynamic computations provide information about brightness distributions, surface density distributions, and velocity fields in the disc.

The incentive to introduce accretion discs for the explanation of cataclysmic variables came from observations. The first attempts to justify this hypothesis by means of computations were undertaken in the early 1960's (for a bibliographic overview see Flannery (1975a) and Hensler (1982a)), and in the course of the following decades the theories became ever more elaborate. Still, many properties of the disc can be derived from basic physical principles without extensive use of computers.

The secondary star can safely be assumed to co-rotate with the binary orbit (the synchronization time is very short, see Chapter 4.V.C), and matter leaving its surface at the inner Lagrangian point L₁ is likely to have a velocity on the order of the local sound speed (e.g., Lin and Pringle, 1976), which is about an order of magnitude less than the orbital velocity of the secondary star around the primary. When material is spilled into the Roche lobe of the white dwarf, it thus has an angular momentum with respect to the central star which is high enough so that it will pass by the star's surface on its trajectory in the white dwarf's gravitational field. The velocity is not high enough, however, for the particles to leave the field altogether, so they are deflected into an orbit around the white dwarf. A "hot spot" is formed where the orbit and stream of newly falling matter collide. Each single particle, were it left alone, would settle down into an orbit around the white dwarf corresponding to the equilibrium between its centrifugal force keeping it away from the star's surface and the gravitational force attracting it. All velocity components perpendicular to the orbital plane of the binary system (which may well be present as the particle enters the Roche lobe) are quickly smoothed out by the primary's gravitational attraction, which increases away from the plane, and thus the orbit will lie entirely in the orbital plane (see equation 4.17). Eventually many particles will orbit around the white dwarf all at (about) the same distance, making collisions and other interactions unavoidable. Thus some viscous mechanism has to be at work which is then able to separate mass from angular momentum: most of the material moves closer to the white dwarf (having less angular momentum), and comparatively little mass — ever less as the distance from the white dwarf increases — carrying most of the angular momentum is transported outward. The original torus spreads out (Figure 4-11).

First attempts to verify these considerations numerically followed the motion of individual particles in the restricted three-body approximation (e.g., Flannery, 1975a). The predicted position of the hot spot was in reasonable agreement with the observations, but more precise predictions were hardly possible. As more powerful computers became available, two-dimensional hydrodynamic computations were carried out to various degrees of sophistication (e.g., Novick and Wolter, 1975; Lin and Pringle, 1976; Hensler, 1982a; and references therein). Basically, the motion of many individual particles is followed through a geometrical grid which covers the entire Roche lobe; at consecutive time steps each cell is checked for the number and dynamical states of particles in it, some adopted numerical "viscosity" allows for momentum exchange as particles pass close by to each other, and, as ever more mass is fed into the initially empty
Figure 4-11. Formation of an accretion disc due to mass overflow from the secondary star: initially only a ring is formed, which eventually spreads out and forms an accretion disc; (d) is a side-view of the system (Petterson, 1983).

Figure 4-12. Two-dimensional hydrodynamic computations provide information about (a) the column density and (b) the radiation intensity of the accretion disc (Hensler, 1982a).

Figure 4-13. Temporal evolution of an accretion disc from first formation to a stationary state at constant mass transfer; displayed are the radius of the disc, the time, and (a) the surface density, (b) the temperature, and (c) the semi-thickness (Bath and Pringle, 1981).
Roche lobe, the process is followed until a stationary state builds up. In this way information about the mass density, the velocity field, and the radiation intensity* are obtained at the same time. Examples of such computations are given in Figure 4-12 for the column density and the radiation field. Similarly, Bath and Pringle (1981) computed the development of the temperature and surface density ($\Sigma = \int_0^q dz$) and the geometrical (semi-)thickness of the disc (Figure 4-13). Details about these results strongly depend on the assumptions about the viscosity, but the gross features are represented as expected from the observations.

It is also possible to obtain some approximate information on how much of the mass in the disc is accreted by the white dwarf, and on what happens to the mass that is carrying outward angular momentum. In principle, the alternative for material in the outermost disc areas is to either leave the system carrying away all the angular momentum or, exposed to ever stronger tidal forces by the secondary stars as it moves away from the white dwarf, to feed the angular momentum back into the orbital motion (Lin and Pringle, 1976; Papaloizou and Lin, 1979; Hensler, 1982a). The action of tidal forces on the outer parts of the disc is to try to slow down the rotation and to thus feed angular momentum back into the orbit. Viscosity, on the other hand, tends to counteract this and to make the material leave the Roche lobe, or even the system. With vanishingly little viscosity all angular momentum would be given back to the system, with no tidal forces acting, and 30% to 50% of the originally transferred mass would be lost from the system (Papaloizou and Pringle, 1977). What really will happen depends on the balance of the two, and thus also on the mass ratio between the components, which determines the strength of the tidal forces. Papaloizou and Pringle guess that the amount of mass lost from a typical cataclysmic variable system is on the order of a few percent*.

A hot spot originates where the stream hits the disc and strong velocity gradients occur. This feature is clearly present in all two-dimensional hydrodynamic computations, and it is considered responsible for the hump structure present in the light curves of many cataclysmic variables. More controversial is the geometrical shape and structure of the hot spot and its position in the disc. If a hump is seen at all in a cataclysmic variable light curve, it normally is visible only for about half of the orbital period, i.e., when the system is viewed from behind the hot spot it must be veiled by some material; at any rate the characteristic emerging radiation pattern is highly anisotropic. Furthermore, it is not clear from observations at what distance from the white dwarf the hot spot is located. For clarification of the structure and location of the hot spot, it is essential to know the vertical structure of the accretion disc (perpendicular to the rotational plane), and theoretical details about the hot spot depend very crucially on assumptions made about this structure or, what is almost identical, about the viscosity in the disc.

It is quite possible that the incoming stream penetrates the outer areas of the disc and only is stopped comparatively close to the white dwarf, where the density of the disc material is high enough to stop further penetration. Thus it follows that the actual shock may occur well within the disc far below optical depth one, hidden from direct observations, and that only gradually will the excess radiation find its way to the surface (e.g., Bath et al, 1983a, and Chapter 4.III.C.2).

---

*At this level of numerical treatment there was still no vertical component included in the computations, and thus no information about the vertical structure could be obtained, but only about the integrated values.

*Observations also lead to the conclusion that not much matter can be lost from the system, since no traces of gas shells have been observed around any system except for novae, which represent a different situation.
Clearly, in any further investigation of the disc structure, and in particular of dwarf nova outburst behavior, it will be necessary to take into account the third dimension of the problem, the vertical stratification of the accretion disc — which will be the issue of the next section.

The viscosity in the disc usually is assumed to be due to shear forces between the differentially (Keplerian) rotating particles in the disc, producing turbulent and/or magnetic interaction. Both are likely to be present in an ionized, rapidly rotating medium — but although this viscosity determines practically everything happening in the disc, no clear concept about its physical nature exists so far.

A kind of "effective viscosity" — the only physical function of which is to allow matter to be accreted by the white dwarf — has been suggested by Sawada et al (1986a; 1986b; 1987), whose hydrodynamic computations show that, under the action of tidal forces, shock waves are formed in the outer areas of an inviscid disc and then travel inwards toward the white dwarf. The gas in the disc loses enough angular momentum in these shocks for accretion to take place without any viscosity due to particle interactions being at work in the disc. Since these ideas are currently quite new, this alternative has not yet been pursued more deeply, so it will not be considered any further in what follows. All other computations of dwarf nova outburst behavior, etc., are based on the assumption of a turbulent viscosity.

III.B THE THIRD DIMENSION

RELEVANT OBSERVATIONS: Common sense says that the disc must be extended to some degree in the vertical direction.

ABSTRACT: Some idea can be obtained about the vertical structure of an accretion disc, assuming hydrostatic equilibrium in the vertical direction and employing the basic equations for stellar atmospheres and interiors. The least understood and at the same time the most important, parameter in the computations is the viscosity.

The hydrodynamic computations of accretion discs described above can provide averaged information about the vertical structure, which is the assumption underlying the computations. In order to obtain a clearer picture about disc properties, it is essential to consider the vertical stratifications. Certain properties of flux densities and surface densities are known from two-dimensional considerations. If prescriptions for the vertical pressure and density stratifications and some viscosity (energy source) are assumed, the normal equations for the computation of stellar atmospheres and interiors provide a vertical structure for each point in the disc when, as a good approximation, a locally plane-parallel approximation is assumed, and the gravitational acceleration is the vertical component of the force exerted by the white dwarf (equation 4.17). The weakest point in this procedure is again the poorly understood viscosity, which is generally assumed to be due to turbulence or small-scale magnetic fields. At present almost any assumption on the viscosity’s temperature and pressure dependence, from a simple relation with, say, the pressure ("c_disc") to a very intricate dependence, is equally conceivable and justifiable.

Meyer and Meyer-Hofmeister (1982) carried out detailed computations of this sort. They assumed hydrostatic equilibrium in the vertical direction and the diffusion approximation to obtain the vertical temperature dependence; energy transport in the vertical direction was allowed to occur by means of radiation and convection, whichever was required. In the first models, they adopted a constant $\alpha$ throughout the disc, which was assumed to be proportional to the local pressure ($\alpha$-disc); alternatively, allowance is made for "magnetic viscosity."

For different values of the mass transfer rate they construct model discs with a constant value of $M_{\text{wd}} = 1 \, M_\odot$. The geometrical shape of the disc surface ($\tau = 1$) is displayed in Figure 4-14. At some radius at the outside of the disc these models become convective, which leads to a flattening of the disc surface; at smaller
radii the disc becomes exponentially thicker with increasing distance from the white dwarf. Figure 4-15 shows the vertical structures for one point in the radiative regime and for one point in the convective regime, respectively.

It turns out from these computations that neither the exact value of the viscosity (keeping the general prescription of the dependence on only the local pressure) nor the mass of the white dwarf have a significant influence on the results (Figure 4-14). The viscosity law however, (in another paper: Meyer and Meyer-Hofmeister, 1983b) turns out to be of decisive influence. In Figure 4-16 the surface density vs. the effective temperature (the so-called S-curve) is displayed for essentially identical models (approximately the same as in Figures 4-14 and 4-15), but computed with either constant alpha or with alpha dependent on the ratio of the pressure scale height to the radial distance from the star, demonstrating that the choice of viscosity is a very important factor in the computations. A similar investigation was carried out by Pojmanski (1986), who demonstrates that the chemical abundance and the opacities used have no dramatic influence on the shape of the S-curve — and thus the vertical struc-

Figure 4-14. Thickness of an accretion disc at the height of the photosphere for various accretion rates as a function of the distance from the central object. The solid and dashed line give results for different viscosities; the crossed line represents computations for the same viscosity as the dotted line, but for a higher central mass (Meyer and Meyer-Hofmeister, 1982).

Figure 4-15. Vertical structure of one of the discs in Figure 4-14 i.e., for $\dot{M} = 10^{-9} M_\odot/yr$; solid line, (a) at a distance $\log r = 10.5 \, \text{cm}$, in the outer convective region, (b) at a distance $\log r = 9.5 \, \text{cm}$ in the inner radiative region. The course of the temperature, density, radiation flux, and, only in (b), the sound speed and the convective velocity are given as a function of the vertical height in the disc (Meyer and Meyer-Hofmeister, 1982).
Observations have led to a general concept of dwarf nova outbursts in which, during the quiescence state, matter is stored in the outer areas of the disc and then released by some mechanism towards the white dwarf during outburst; the sudden liberation of gravitational energy leads to the observed brightening. For a long time no suitable physical mechanism for this behavior could be identified. Then Meyer and Meyer-Hofmeister (1981) pointed out that the double-valued function of the surface density in the log $T_{\text{eff}}$ -- $\Sigma$ diagram (the S-curve in Figure 4-16) might be a promising candidate: due to the interplay of viscous heating and radiative or convective cooling the viscosity might jump from being low during quiescence to being high during outburst, regulated through the surface density. This so-called limit-cycle instability is the subject of the next section.

One way to obtain more constraints on possible values and dependencies of $\alpha$ is to compare the predictions of theoretical models with observations. The observations to which the value of $\alpha$ is most sensitive are the outburst light curves of dwarf novae, in particular their time-scales, including the duration of the quiescent state and the outburst, respectively, and brightness changes during rise, decline, and quiescence.

### III.C. MODELING THE OUTBURSTS BEHAVIOR

**RELEVANT OBSERVATIONS:** Dwarf novae increase in brightness by two to five magnitudes, in semi-regular intervals of time. Characteristic color changes occur; the rise at optical wavelengths in some objects precedes the rise in the UV by several hours, while in other objects both rise simultaneously. The much slower decline occurs simultaneously in all wavelengths.

---

*Figure 4-16. Viscosity -- surface density relations (S-curves) at different distances from the central object, where models are very similar to those in Figures 4-14 and 4-15. (a): the viscosity parameter alpha is constant throughout the disc; (b): alpha is assumed to be variable as a function of the ratio between the pressure scale height and the distance from the central star (Meyer and Meyer-Hofmeister, 1983b).*

---

*see chapter: 21*
Two possible outburst mechanisms have been suggested: viscous instability, or disc instability, in the accretion disc itself (Bath et al., 1974b; Osaki, 1974), and a suddenly enhanced mass transfer onto the disc from the companion star, called transfer instability (Bath, 1973; Bath et al., 1974b). Both are discussed in this order in the following sections.

### III.C.1. THE DISC INSTABILITY MODEL

**ABSTRACT:** The general, and to some degree even the detailed outburst light curves can be reproduced, with the assumption that the viscosity follows a hysteresis curve as the disc material changes between the convective and the radiative state.

Bath et al., (1974b) and Osaki (1974) suggested that dwarf nova outbursts might be due to "intermittent accretion," in which times of comparatively low accretion onto the white dwarf, during which matter infalling from the secondary star is stored in the outer disc, alternate with times of enhanced accretion, triggered by some instability in the disc itself. Following this, Bath and Pringle (1982) then suggested that it might be the viscosity which, under certain physical conditions, might have two stable equilibrium values between which the state of the disc can alternate if conditions are right. Because the viscosity \( \nu \) is responsible for the efficiency of material transport through the disc, this implies that the major changes really are of the surface density (Figure 4-17). For a disc to be stable, the condition \( d\nu/d\Sigma \geq 0 \) must be met (Lightman and Eardley, 1974); if a condition like the one in Figure 4-17 exists, there is a range of viscosity values for which no corresponding stable value of the surface density exists; the consequence is a limit-cycle behavior. Suppose the hypothetical equilibrium value of the viscosity (which would lead to a steady state) were \( \nu_0 \). An originally low surface density increases up to a value \( \Sigma_1 \) (in Figure 4-17). A further increase in \( \Sigma_1 \) implies a jump in viscosity; since the viscosity \( \nu_1 \) (this is the condition for the limit-cycle to occur) is higher than \( \nu_0 \), more material is transported out of this particular region in the disc than is fed in, and thus the surface density decreases until it reaches a value \( \Sigma_2 \). Now a further decrease implies another jump in viscosity, this time, however, to a value \( \nu_2 \), lower than the equilibrium value \( \nu_0 \). As a consequence the surface density increases again and the cycle continues, always trying to establish, unsuccessfully, an equilibrium state. Since in accretion discs a higher throughput of mass through a particular area implies a higher temperature, the semi-regular brightness changes in dwarf novae can be understood if somehow this limit-cycle activity would involve the entire disc, or at least a considerable portion of it. If, on the other hand, the mass input into the disc always is high enough to ensure a surface density for which there exists an accessible value of the viscosity to maintain an equilibrium between mass input and mass output, no "outbursts" would have to occur. This case is assumed to apply to UX Ursae Majoris stars and possibly to quiescent novae, while Z Camelopardalis stars and anti-dwarf novae are believed to be boundary cases in which the mass input is mostly just below or just above, respectively, the lower limit for equilibrium, and slight changes in the mass transfer rate from the secondary star have a dramatic effect, by either leading to temporary attainment of an equilibrium or by temporarily pushing the disc out of it.

Meyer and Meyer-Hofmeister (1981) point out that the change in surface density due to ionization of hydrogen (Figure 4-16), i.e., due to the change from a convective to a radiative state of the disc, is a likely candidate for providing the physical background for the limit-cycle activity. Thanks to the opposite temperature dependence of the hydrogen absorption coefficient for neutral and (partly) ionized hydrogen of

\[
x_1 = 10^{-36} \theta^{\frac{51}{10}} T^{10} \quad (4.16a)
\]

for \( T \lesssim T_0 \)
$$\kappa_2 \approx 1.5 \times 10^{20} \ e \ T^{-2.5}$$
for \( T \gtrsim T_0 \)

with

$$T_0 \approx 1.2 \times (10^8 \ e)^{0.53} \times 10^4, \quad (4.16b)$$

\((\approx 10^4 \text{K})\)

(Faulkner et al, 1983) either heating or cooling prevails in the area.

More recently, several groups have carried out computations of the behavior of accretion discs under the assumption that hydrogen ionization is the driving mechanism for the disc instability, and thus for dwarf nova outburst behavior (e.g., Cannizzo et al, 1982; Faulkner et al, 1983; Meyer and Meyer-Hofmeister, 1983b; Mineshige and Osaki, 1983; Papaloizou et al, 1983; Meyer, 1984; Meyer and Meyer-Hofmeister, 1984; Cannizzo et al, 1985; Lin et al, 1985; Mineshige and Osaki, 1985; Cannizzo et al, 1986; Mineshige, 1986; 1988; Meyer-Hofmeister, 1987; Meyer-Hofmeister and Meyer, 1987). Mostly these differ in their assumptions about the viscosity — whether it is assumed to be single-valued all over the disc, whether there are two different but constant values for the radiative and the convective state, respectively, or whether the viscosity is radially or vertically variable in some way — and whether, and how, radial interactions between adjacent parts of the disc are taken into account in the computations. The reader is referred to the original articles for details. It should be stressed, however, that all of these efforts succeeded to some degree in reproducing the general features of the outburst light curves of dwarf novae, in some cases in remarkable detail. Some of these results are presented in what follows.

In general the disc is found to be optically thin in its convective (cool) part, which for "typical" dwarf novae comprises most of the disc. Depending on adopted numerical methods and assumed particular values of the viscosity, the temperature of the outer disc is found to be between 5000 and 6000K (e.g., Papaloizou et al, 1983) or between 2000 and 3000 K (e.g., Cannizzo et al, 1986) over large areas. During the quiescent state the outer disc slowly becomes optically thick, and the temperature rises until an outburst occurs (Figure 4-18).

A very detailed discussion of the physics of dwarf nova outburst cycles based on the disc instability is given in Papaloizou et al (1983). It is found that, depending on details of the physical conditions, the outburst can start in different regions of the disc. A change to the ionized state can start from the innermost disc (called "type 1" by Papaloizou et al, and "type B" by Smak (1984)) if during quiescence all the disc is optically thin and cool, and if the viscosity is large enough for the inner regions to become optically thick first. In this case, the increased temperature heats up neighboring areas to produce a wave of ionization which travels outwards through the disc until the whole disc is in "outburst." If on the other hand \( \alpha \) is very small, the material piles up in outer disc areas and cannot be transported inwards efficiently, in which case the outburst starts at intermediate regions, causing two ionizing waves, one traveling outwards (like in the example above) and the other inwards (this is called a "type 2" outburst by Papaloizou et al, and "type A" by Smak). A third type of outburst originates if a sizable fraction of the central disc is permanently fully ionized and transitions are constrained to occur in the outer part only; the outburst is then triggered at the interface between the temporarily cool and the permanently hot region with the transition wave traveling outwards from there on ("type 3").

Corresponding to these three possibilities for initiating an outburst, there are also three ways of stopping it. In "type 1," all of the disc is depleted during the outburst since the mass throughput is higher than the mass input, the outer areas will be depleted first, the
Figure 4.17. Conditions for occurrence of a limit-cycle activity: for the viscosity \((v_0, \Sigma_0)\) which would lead to a steady state, there exists no corresponding accessible value of the surface density; conditions in the disc will oscillate about the equilibrium value between \((v_1, \Sigma_1)\) and \((v_2, \Sigma_2)\) which leads to the observed outburst activity (Bath and Pringle, 1982).

Figure 4.18. Development of the luminosity of the accretion disc and the optical thickness of the disc during the outburst cycle: at the end of an outburst the disc is depleted; in the outer areas it is refilled by material coming from the secondary star and ever larger parts become optically thick, until at some point in the disc conditions for a jump in viscosity are given, which triggers the next outburst (see also Figure 4.23) (Meyer-Hofmeister and Meyer, 1987).

Temperature falls below the critical temperature, and the material becomes neutral again. Ongoing depletion in more central areas as well as contact with neighboring cool areas causes a “downward transition” front to travel inwards. If, on the other hand, initially during the rise phase, not all of the outer disc was fully ionized, this cooling front originates at the outer interface between hot and cool areas and travels inward from there (“type 2”). Finally, a “type 3” decay is one during which the cooling wave can travel through only part of the disc, until it meets areas which stay fully ionized during quiescence, because the mass input rate and the viscosity are high enough.

Lin et al. (1985) constructed series of outburst light curves and investigated the influence of various parameters on the appearance of the light curves. They found that the amount of energy which is deposited in the disc at the hot spot, rather than being dissipated into radiation, influences the brightness level during quiescence by changing the general temperature distribution in the disc, but that it has practically no effect on the outburst light curve. Likewise the precise radial location and extent of the region into which matter is fed by the incoming stream (supposing the stream does penetrate the disc for some distance before it is stopped and forced to deposit its mass and energy) has no appreciable effect, unless a part of the disc happens to be very close to the transition to the fully ionized state during quiescence (Figure 4.19a). It is obvious that the outburst light curve is determined to a large extent by the condition in which the disc was left after the previous outburst, so that outbursts of considerably different shape may follow each other, and only sequences of two or more
Figure 4-19. Synthetic dwarf nova outburst light curves on the basis of the disc instability model. (a) a-b: Various amounts of energy are inferred into the disc by the hot spot; c-e: effect of location and physical characteristics of the mass stream; f-h: varying the mass input rate from 2.8 10^6 to 100. (b): Effect of varying the viscosity. (c),(d): Effect of varying the mass input rate (in M_o/yr): a: \( \dot{M} = 9 \times 10^{-11} \); b: \( 1.8 \times 10^{-11} \); c: \( 3.6 \times 10^{-11} \); d: \( 7.2 \times 10^{-11} \); e: \( 1.8 \times 10^{-10} \); f: \( 3.6 \times 10^{-10} \); g: \( 7.2 \times 10^{-10} \); h: \( 1.44 \times 10^{-9} \); i: \( 2.16 \times 10^{-9} \); j: \( 3.24 \times 10^{-9} \); k: \( 3.60 \times 10^{-9} \); l: \( 3.60 \times 10^{-9} \); m: \( 4.36 \times 10^{-9} \); n: \( 5.76 \times 10^{-9} \); o: \( 7.2 \times 10^{-9} \); p: \( 8.64 \times 10^{-9} \). (See text for description.) (d): Effect of increasing the mass input during decline from outburst (Lin et al, 1985).
Figure 4-20. If the outburst starts in the outer disc, the rise is fast and it starts later at shorter wavelengths than at longer; if the outburst starts in the inner disc the rise is slow and simultaneous at all wavelengths (Smak, 1984).

repeat. The shape of the outburst light curve depends very critically on both the adopted value of the viscosity and the mass input rate. In Figure 4-19b, the effect of increasing (each time-constant) viscosity is demonstrated. Very low values of the viscosity result in occasional large, long-lasting outbursts which affect all the disc, followed by a long quiescent time during which the disc has to be refilled; the next large outburst is preceded by a sequence of short low-amplitude bursts, between which the disc does not return entirely to its quiescent brightness level but during which the central area of the disc remains fully ionized. At increased values of the viscosity parameter $\alpha$ the long outbursts become shorter, the short ones become longer and more evenly distributed between the long ones, and the central disc area remains ionized all the time. If the viscosity becomes even larger all proper outburst behavior disappears, first leaving small amplitude brightness oscillations of the outer disc areas, until, at a further increase in viscosity, they disappear altogether.

The effect of increasing the mass input rate, but keeping it constant in each single case, is illustrated in Figure 4-19c. The characteristics of the light curve remain essentially unchanged, for larger ranges of the mass transfer rate while only the amplitude and duration increase, and the frequency decreases, with increasing mass transfer rate. Then suddenly the whole character changes at a certain critical value, as more areas of the disc become only marginally unstable or, finally, remain fully ionized all the time. It is to be emphasized that the appearance of the light curve can change from being very simple and regular to having the most complicated appearance, simply by varying the rate of mass input. Eventually, if all the disc becomes permanently ionized, all activity ceases.

Lin et al, suggest that the basic difference between old novae and dwarf novae may be a different mass transfer rate, a view which is supported by observationally deduced mass transfer rates and absolute magnitudes — see Chapter 4.II.C.3. In their picture, the mass transfer rate in novae is high enough to permanently prevent outburst activity and thus gives the system a larger intrinsic brightness than dwarf novae have. Some other old novae like GK Per might then be border cases, being at just the lower end of stability, so that they can undergo small-scale outburst activity at slight variations of the mass transfer rate.

Figure 4-21. Observed and computed color changes during the outburst cycle in the disc instability model (Smak, 1984).
Figure 4-22. Use of molecular opacities mainly changes the appearance of the S-curve in the cool areas of the accretion disc (compare with Figure 4-17) (Meyer-Hofmeister and Meyer, 1987).

Similarly, nova-like variables can be assumed to lie just above the critical mass transfer rate for outburst activity to occur.

Figure 4-23. Development of the temperature in the outer disc before outburst, corresponding to Figure 4-18. The temperature in the disc rises continuously in ever larger areas of the disc until conditions for an outburst are met at some distance from the white dwarf (Meyer-Hofmeister and Meyer, 1987).

Figure 4-24. The observed vague relation (solid line) between the outburst period and amplitude can be reproduced with the disc instability model; open circles denote models which take into account only radiation from the accretion disc, filled circles take into account contributions from the hot spot as well (Smak, 1984).

Figure 4-25. The observed relation between the orbital period and the rate of decline from an outburst can be reproduced with the disc instability model; the left panel shows the observed relation, while the right panel gives the results from models (Smak, 1984).

If certain values of the viscosity parameter $\alpha$ or of the mass transfer rate alone can bring about very complicated patterns of the light
curve, it is clear that changes of the mass transfer rate at some stage during the outburst cycle can also produce patterns of any degree of intricacy (Figure 4-19e). Temporary standstills entered from either a high or a low brightness state, as observed in Z Camelopardalis variables, can also easily be produced.

In all the above examples, $\alpha$ was kept constant throughout the activity cycle. In fact the calculated light curves only resemble the observed light curves in their general appearance. The observed feature of a much steeper rise than decline cannot be fitted in this way. Most other authors assumed a small value of $\alpha$ during quiescence and a much larger one during outburst (ionization). Under this assumption it is possible to fit either a fast or a slow rise (as compared to decline), if the outburst starts either at the outer (type A or 2) or at the inner (type B or 1) disc (Figure 4-20). Even the observed delayed rise at shorter wavelengths in the case of fast rises and the simultaneous rise at slow rises are reproduced with the correct time scales (see also Cannizzo et al, 1986). Similarly, optical color curves derived for the outburst behavior clearly do resemble the observations (Figure 4-21).

From the above results it appears that the observations probably best suited for distinguishing between different $\alpha$ prescriptions is the early rise phase. Meyer-Hofmeister (1987) and Mineshige (1988) carried out investigations of this phase, particularly with respect to the influence of molecular opacities on the surface density. Meyer-Hofmeister (1987) found that the characteristic curve at very low temperatures might indeed be more complicated than assumed so far (Figure 4-22) in exhibiting an additional halt in the limit-cycle $T_{\text{eff}} - \Sigma$ curve. The effect is that shortly before the actual rise to outburst the transition to the intermediate point occurs, resulting in a slow brightness increase during which a considerable area of the disc will be heated up to some 6000 - 7000 K. Only after this (Meyer-Hofmeister’s results show a steady rise for some 5 to 10 dayz; Mineshige’s results show a halt for about 1 day) does the actual rise occur (Figure 4-23, see also Figure 4-18) — in the case of the outburst starting from the outer disc.

Smak (1984) compared two observed relations concerning dwarf nova outburst behavior with his models: a vague relation between outburst period and amplitude (Figure 4-24), and the relation between the orbital period and the rate of decline from an outburst (Figure 4-25 — see Chapter 2.II.A.3. Both theoretical relations are in reasonably satisfactory agreement with the observations.

Similar disc-instability computations have been carried out for the hydrogen-deficient AM Canum Venaticorum variables as for “normal” hydrogen rich cataclysmic variables (Cannizzo, 1984; Smak, 1984). They find that dwarf nova-like outburst behavior due to He-ionization instability is quite possible in these objects as well. The disc temperature during quiescence is found to be of the order of 10000 K, and the disc is marginally optically thin. At outburst, Cannizzo (1984) found temperatures between 20000 and 50000 K in the disc. The system brightness rises by about 2 – 3 mag with an outburst period between 16 hours and some 10 days (in his computations), a value which is very critically dependent on the assumed mass transfer rate.

III.C.2. THE TRANSFER INSTABILITY MODEL

ABSTRACT: This alternative model assumes that instabilities in the atmosphere of the secondary star lead to temporarily increased mass transfer into the disc, which then causes the outburst. Modeling the observed brightness changes also yields satisfactory results.

In the above section, the cause of a dwarf nova outburst was assumed to be variable conditions in the disc itself, while the mass transfer rate was usually assumed to be constant with time. At any rate it influenced the outburst
behavior only rather indirectly. Alternatively, it has been suggested (Bath, 1973; Bath et al, 1974b) that the outburst might be due to an instability in the Roche lobe-filling companion star which leads to a temporarily increased mass transfer into the disc. This latter possibility has been pursued by Bath and co-workers (Bath, 1975; Papaloizou and Bath, 1975; Bath, 1976; 1977; Bath and Pringle, 1981; Bath et al, 1983b; Mantle and Bath, 1983; Bath, 1984; Bath et al, 1986).

The basis of this theory is the realization that the dynamic equilibrium of a star which is confined to a Roche lobe is quite different from that of a single isolated star, once the star's surface comes close to the confinement of the critical Roche surface (Bath, 1975; Papaloizou and Bath, 1975). The important difference is that the energy required for material to escape the surface of a single star must be sufficient for the material to reach infinity, while in the case of a Roche lobe it is sufficient to lift it only above the critical Roche surface, which requires a lot less energy. Papaloizou and Bath point out that there are two possible destabilizing effects in the confined atmospheres of a cool star: convection in the envelopes, and ionization zones in the vicinity of L1 at which point the net gravitational acceleration is zero.

The general idea of the transfer instability is that once the atmosphere of the secondary star becomes unstable in the vicinity of the Lagrangian point L1, enhanced mass overflow into (mainly) the Roche lobe of the white dwarf occurs, leading to an outburst and at the same time depleting the atmosphere. It stops when the energy provided by the instability is no longer sufficient to lift further material into the neighboring Roche lobe (Bath, 1976). Since this process has brought the star out of thermal equilibrium, the star contracts and detaches from the Roche surface, while during its "quiescent" state mass overflow is maintained only on a low level through normal stellar wind. Eventually, with the support of energy supplied from deeper stellar layers, thermal equilibrium is gained again, and the star expands and is ready for the next instability to occur.

Computed outburst periods (i.e., relaxation times of the destabilized atmospheres) are in the range of 10 to 200 days, mass transfer rates during outburst maximum are on the order of $10^{17}$ to $10^{19}$ g/sec and amplitudes are of several magnitudes — all in accordance with observed values.

Since a lot of material with low angular momentum is transferred into the disc, during the initial phases of an outburst the disc radius is expected to shrink as this new material mixes with material that is already contained in the disc, and only later does it expand again as angular momentum from the inner disc is transferred outward (e.g., Bath and Pringle, 1981). This prediction is quite in contrast to what the disc instability model predicts, namely, an increase in the disc's radius from the very beginning of an outburst if the outburst starts at the outer disc, or no change at all if it starts at the inner disc. Thus, in principle, it should be possible to decide between the two models from investigation of the outer disc radius, but since the outermost disc may be optically thin, determinations of the disk radii are not reliable.

Equally, an increase in the intensity of the orbital hump would be expected from enhanced mass transfer, while no effect should occur in the case of the disc instability. Here again the problem arises that neither the shape and position nor the geometrical radiation pattern of the hot spot are known. It is possible that the enhanced mass flux disappears entirely in the disc at the beginning of an outburst, while the excess radiation only eventually diffuses outward; and if the radiation pattern which originates then is rather isotropic, no enhanced hump would be seen at all — or, if this radiation is anisotropic, the enhanced hump may become visible with considerable delay (Bath et al, 1983b).
Figure 4-26. Temporal evolution of the accretion disc in response to a mass pulse: (a): evolution of the surface density; (b): evolution of the central disc temperature (at $z = 0$); (c): corresponding synthetic light curve (Bath and Pringle, 1981).

Figure 4-27. The typical shapes of dwarf nova outburst profiles (solid lines) can roughly be reproduced with the transfer instability model; different profiles can be obtained when the strength and duration of the mass pulse are changed. Dotted lines represent the accretion rates on the white dwarf. Occurrence of disc instabilities is numerically prevented (Bath et al, 1986).

The general course of an outburst induced by a transfer instability is shown in Figure 4-26: The mass pulse leads to a brief contraction of the radius, then the disc spreads out while its temperature increases, and the disc slowly relaxes to its equilibrium state; the corresponding outburst light curve is shown in Figure 4-26c.

A systematic investigation of theoretical outburst light curves has been carried out by Bath et al (1986). In their computations the mass burst profile is a free parameter, since knowledge of the atmospheres of the cool companions is not sufficiently detailed to make meaningful predictions. By appropriate choices of the pulse profiles they are able to reproduce gross features of the outburst light curves (Figure 4-27). In general they find that the rise is fast, following essentially the rise time of the mass flux, if the time-scale for the rise in mass transfer is shorter than the viscous time scale of the disc, while the whole behavior follows only the mass flux variation if the variations are slower than the viscous time scale of the disc. In particular, very long-lasting outbursts can be produced if the mass transfer goes on for a long time. For all these computations a viscosity parameter of $\alpha = 2$ was assumed and numerically any ionization instability was
Figure 4-28. When disc instabilities are allowed for, they lead to a flickering-like activity during quiescence but not during outburst (Bath et al., 1986). When compared with Figure 4-19, this might be due to the large alpha used in these models.

Figure 4-29. As in the case of the disc instability model, the observed relation between decay time and orbital period can be reproduced with the transfer instability model (Mantle and Bath, 1983).

III.C.3. CONCLUSIONS FROM OBSERVATIONS

ABSTRACT: From comparison with observations no clear distinction between the two proposed outburst models is possible so far.

Since the disc instability model and the transfer instability model do not produce the same light curves, it should be possible to distinguish between them by comparing theoretical results with observations. Pringle et al. (1986), Verbunt (1986), and Cannizzo and Kenyon (1987) tried to do this in a systematic way. In particular, Pringle et al. compare observed optical and UV flux changes which occurred during rise to an outburst in VW Hyi and CN Ori with predictions by various models.
In VW Hyi, a delay of the rise in UV with respect to the optical of about half a day is observed, while in CN Ori the rise proceeds simultaneously in all wavelengths. The decline in both systems is simultaneous in the optical and in the UV.

Adopting parameters suited for these two systems, Pringle et al find that the rise phase of VW Hyi cannot be modeled satisfactorily with a disc instability, whether it is starting from the inner or from the outer disc. In light of results presented in Chapter III.C.1, it would be expected that an instability in the outer disc would produce such a delay, but in their computations the temperature in the outer areas rises far too quickly to produce a delay of the order required. In the case of CN Ori, an instability in the inner disc can produce reasonably acceptable agreement with the observations. For both systems, the observed flux development can be reproduced by suitable choice of the mass flux profile which, given the uncertainties in that theory, is no surprise. In any case the decline depends only on the relaxation of the disc, so it is fitted about equally well by all three models, though the decline proceeds somewhat too fast in the transfer instability computations.

Cannizzo and Kenyon, on a more general basis, arrive at similar conclusions. All authors note, however, that these results do not imply a failure of the disc instability model, given the large uncertainties in the prescription of the viscosity and its dependence on other physical parameters. The only reliable conclusion one can arrive at, then, as Pringle et al point out, is that the dependence of $\alpha$ on local physical conditions certainly is neither simple nor universal if the disc instability is the mechanism for driving dwarf nova outbursts.

Since the decline process is more or less the same for all possible types of outburst, the rise phase to the outburst, if anything, should provide distinctive evidence for one model or the other.

The outer disc radius is predicted to shrink slightly at the very beginning of an outburst in the case of a transfer instability, while in the case of a disc instability it either increases (if the outburst starts in the outer disc) or remains unchanged (if it starts in the inner disc). In all cases it is expected to increase around the phase of outburst maximum and slowly shrinks as decline proceeds, so excess angular momentum can be carried away efficiently. It is questionable whether this predicted shrinking is observable. It may not occur at all if the stream material penetrates deeply into the disc, or it may occur only for a very short time and not be very pronounced, so it clearly could escape the observations. If shrinking could be observed without any doubt in a system, it would exclude the disc instability as the cause of the outburst, at least for this particular (hypothetical) system.

An indication in favor of the disc instability would probably be the observation of the predicted halt in brightness at about 6000 K for up to half a day at a level slightly above the quiescent brightness level before the real rise to the outburst occurs, as predicted by Meyer-Hofmeister (1987) and Mineshige (1988).

As discussed above, the behavior of the orbital hump cannot currently provide the basis for distinguishing between models, since it is not known how deeply the mass stream penetrates into the disc and how isotropic or anisotropic the characteristics of the outgoing radiation are. In particular, the hump size is not a useful discriminant, since it is quite possible that, as soon as the outburst starts, triggered by a disc instability, the secondary star is heated by irradiation and itself reacts with increased mass transfer, leading to a delayed increased hump (as is also expected from the deeply penetrating original mass transfer pulse). Conversely, it also is perceivable that a transfer instability triggers a disc instability, so again both effects become interwoven. The fact that in some objects all kinds of different shapes of outburst light curves are observed certainly indicates that the underlying physics is not trivial or straightforward.
Figure 4.30. Comparison between observed and various computed spectral changes during rise to an outburst (see text for discussion) (Verbunt, 1986).
Figure 4-31. Schematic model for a precessing elliptic disc. This model for a superoutburst combines characteristics of both the disc instability and the transfer instability model; a precessing eccentric disc can explain the photometric phenomena observed during superoutburst (Osaki, 1985).

III.C.4. SUPEROUTBURST MODELS

RELEVANT OBSERVATIONS: In some short-period dwarf novae, besides the normal outbursts, so-called superoutbursts are observed which are about one magnitude brighter and last considerably longer than normal outbursts. A number of characteristic phenomena are observed only during superoutbursts.

ABSTRACT: No single generally accepted model exists so far. The most promising of the current models combines features of the disc instability and the transfer instability which lead to a temporary deformation of the accretion disc.

Since the detection of superhumps occurring during superoutbursts in certain dwarf novae, the SU Ursae Majoris stars (Vogt, 1974; Warner, 1975), almost a dozen models have been developed to explain this phenomenon and other related phenomena, ranging from strongly magnetic rotating white dwarfs (Papaloizou and Pringle, 1979; Patterson, 1979b; Vogt, 1979) to flux tubes emerging from the surface of the secondary star and then connecting with the disc (Meyer, 1979), spots on the surface of the cool star (Vogt, 1977; Haefner et al, 1979), eccentric discs surrounding an inner concentric disc (Vogt, 1982b), outer eccentric discs (Gilliland and Kemper, 1980) and the response of a moderately strong magnetic white dwarf to a variable mass transfer rate (Warner, 1985). Vogt (1982b) and Warner (1985) present and discuss most of these models and show why almost all of them are unable to account for all, or even the most important, of the observed phenomena. There is no need to repeat this here. We will concentrate on only those few models which have survived to date.

Currently there are only two such models which seem plausible. The first was proposed by Papaloizou and Pringle (1979). They suggest that the orbits of SU Ursae Majoris stars are slightly eccentric (e = 10^{-4}). Due to the then variable distance between the two stars, the size of the secondary's Roche lobe is also slightly variable, which leads to a variable mass transfer rate with the precession period of the line of absides. If dwarf nova outbursts are caused by a disc instability, the mass transfer rate normally should not be influenced by this, so no major effects are expected on the hump amplitude either. Only if an outburst induces an increase in the mass transfer rate as well (e.g., due to a periodic instability of the secondary's atmosphere), do the mass transfer rate, and thus the hump amplitude, strongly vary with the precession period. Papaloizou and Pringle show that such an eccentricity is expected to be driven and maintained by tidal forces, particularly in systems with a large mass ratio (which is the case for all SU Ursae Majoris stars). The difficulties with this model are that neither the variable γ-velocity nor the late superhump find a satisfactory explanation.
A somewhat related model has since been suggested by Osaki (1985). He assumes a circular orbit and an accretion disc which undergoes outbursts due to a disc instability. The secondary star normally transfers material at a relatively low rate into the Roche lobe of the white dwarf. If the secondary is very cool (as is the case in the ultra-short period systems with periods less than 2 hours, since the Roche lobe is too small to accommodate a larger star), the star's atmosphere reaches a certain degree of instability semi-periodically. When a normal outburst occurs at about the same time this instability is reached, then during the outburst the hot disc and boundary layer bring the cool atmosphere out of equilibrium by irradiation, and for some time mass transfer proceeds at an enhanced rate. The first sudden pulse of mass causes a non-axisymmetric perturbation of the disc which deforms it into an eccentric shape (Figure 4-31). For some time, the stream impact and energy release always are much stronger when the stream hits the disc near periastron (closer to the white dwarf), thus maintaining the elliptical shape and producing a stronger hump as long as the enhanced mass transfer lasts. Since the disc precesses with some period, the observed superhump period is slightly longer than the orbital period, and the observed decrease in superhump period can be explained as due to a slight continuous contraction of the disc which leads to a somewhat enhanced precession period. Once the mass reservoir of the secondary is exhausted, the mass transfer rate returns to normal (causing the fast decline from superoutburst) and the disc relaxes to a circular shape, during which the late superhump (with the correct phase shift or about 180° in Osaki's computations) is visible.

III.D. RAPID OSCILLATIONS

RELEVANT OBSERVATIONS: During the optically high state (outburst) of some dwarf novae and some nova-like stars, low-amplitude brightness fluctuations with periods of typically some seconds to some minutes and coherence times between a few cycles and several hundred cycles can at times be observed.

ABSTRACT: Due to the short periods and low stability, the origin of the oscillations has to be placed either on the outermost surface of the white dwarf or in the innermost areas of the accretion disc. No satisfactory model is available as yet.

Rapid brightness variations on time scales of some 10 to 100 seconds have occasionally been observed in many dwarf novae during outburst, as well as in several nova-like systems during the high brightness state (Chapters 2.II.D.2., 3). The emerging picture is very confusing. Conventionally, a distinction is made between coherent and quasi-periodic oscillations, but it is by no means clear whether this reflects any real physical differences underlying the observed phenomena.

Likewise, the theoretical understanding is not very advanced. Patterson (1981) summarizes and comments on most of the suggested scenarios. The very short observed periods place the origin somewhere on, or in the vicinity of, the white dwarf (i.e., clearly the secondary star can be discarded as a possible source). Furthermore, the short coherence times and the period changes point to only very little mass being involved in the process.

The latter constraint excludes rotation of the white dwarf as the source, since there is no way to change its period on time scales of hours, and the oscillations clearly cannot be due to a spotty surface of the white dwarf due to magnetic fields. If the surface were rotationally decoupled from the white dwarf's interior (Paczynski, 1978) the hypothesis could possibly be saved from a dynamical point of view, but the question remains why no pulsations are seen during quiescence in those systems (e.g., HT Cas, Z Cha) which during quiescence seem to be dominated by the radiation from the white dwarf.

Non-radial pulsations (g-modes) of the white dwarf have been suggested as a possible source by several authors (e.g., Faulkner et al, 1972;
Osaki and Hausen, 1974). The problem with these pulsations is, as Papaloizou and Pringle (1978b) point out, that the observed changes in period and amplitude require energies and time-scales that are far too high for a cataclysmic system. Papaloizou and Pringle conclude, however, that when the rotation of the star is properly taken into account, another kind of non-radial oscillations, which they call r-mode oscillations, are confined to only the outermost surface layers of the star and could be responsible for the observed oscillations. Even so, the resulting phase coherence may still be orders of magnitude larger than observed (Córdova et al, 1980).

The latter authors tentatively place the origin of radial pulsations in the boundary layer between the disc and the white dwarf. Variations of the thickness and structure of the boundary layer would lead to changes in the period and amplitude of the pulsations.

Bath (1973) suggests inhomogeneities in the inner accretion disc as they might originate during outburst as possible source of the observed oscillations. They are periodically eclipsed by the white dwarf as they rotate; after a while they would dissolve and others would originate at slightly different radii. Radius changes would bring about period changes, as the inhomogeneities presumably rotate with Keplerian velocity. The large amplitudes and the long time scales of some of the observed oscillations seem problematic, as well as why always only one such blob should be visible.

Similarly, Sparks and Kutter (1980) invoke waves of turbulent condensed material produced in the process of accretion onto the white dwarf as the source for oscillations in dwarf novae. Again the question arises why only one such wave should be present at a time, and how its (often) long lifetime can be explained, and why these waves are not observable during all outbursts of all dwarf novae and in all nova-like systems with high mass transfer rates.

The most recent model by Tajima and Gilden (1987) supposes the reconnection of small-scale magnetic fields generated in the inner disc to be the physical cause of the oscillations. The problems here are identical with Sparks' and Kutter's model.

In general it seems safe to conclude that the radiation source of the observed short-period oscillations and pulsations is to be sought somewhere in the area of the inner disc, boundary layer, and/or white dwarf. Since this probably is the very region in a cataclysmic system where the physics and structure are least understood, there seems to be no real hope of improving our knowledge of the oscillations in dwarf novae until a better physical understanding of that part of the system is gained. The hope that observations of the oscillations might set limits on possible models of, in particular, the boundary is clearly greatly diminished by the totally confused picture the observations currently present.

III.E. SECULAR VARIATIONS OF THE ORBITAL PERIOD

RELEVANT OBSERVATIONS: Secular changes of the orbital period have been measured in several dwarf novae and nova-like stars. In some objects these changes seem to be semi-periodic with time scales on the order of 10 years or more.

see 45, 98, 111, 115

ABSTRACT: Several possible explanations have been suggested. None of these provide a satisfactory explanation of the observed phenomena. It is possible that a good deal of the changes are artifacts of numerical manipulation.

Observed secular changes in the orbital period of some cataclysmic variable systems have been reported. In principle, several mechanisms can be imagined to cause a change in the observed orbital period: the presence of a third (unseen) body in the system, rotation of the line of absides, loss of angular momentum and/or mass from the system, redistribu-
tion of mass and/or angular momentum within the system, motion of the relative position of the hot spot with respect to the other system components, or any combination of these mechanisms. In addition, it ought to be kept in mind that the determination of the orbital period is by no means trivial, and results must be scrutinized very carefully before it can be concluded that the period is variable at all. As the example of the photometric observations of CN Ori (Chapter 2.II.B.3) demonstrates, it is not necessarily possible to derive a reliable orbital period from the light curve of a non-eclipsing system; and if an eclipse is present, it is not obvious what is being measured, say, at the deepest point or at mid-eclipse, since the shapes of eclipses are known to be variable, and slight changes can be misinterpreted as period changes. Only the moment half-way between white dwarf/boundary layer ingress and egress of double eclipsing systems seems to give a stable reference point; the radius of the white dwarf is a stable feature, and, if a boundary layer is present, it is likely to be both thin (compared to the dimensions of the white dwarf) and symmetric in the rotational plane, which is what is observed in double-eclipsing variables. Finally, care must be taken of light travel times and leap seconds. Spectroscopically determined orbital periods (from radial velocities) which are sufficiently accurate have been available only for some 15 years and thus do not provide a sufficiently long time basis. Also some care must be taken in combining photometric and spectroscopic data, since for some objects (e.g., TT Ari, possibly CN Ori) these are different, for as yet unknown reasons. In view of all this, none of the published results of orbital period changes seem particularly reliable, for one reason or another. However, for the sake of theoretical considerations, published values are taken at face value in what follows.

To most of the published O-C data a parabolic as well as a higher order function, or even a part of a sine curve, can all be fitted about equally well, because too few points are known. In cases where observations are clustered about few epochs with no single observations in between, occasional occurrences of sudden changes rather than continuous changes cannot be excluded.

In three of those objects which have been observed for a long time (U Gem, UX UMa, DQ Her), the change in the orbital period has been seen to invert its direction; time scales are on the order of 15 years for U Gem and DQ Her, and 29 years for UX UMa. Time bases are still too short to allow for a decision on whether the changes are really periodic or merely semi-periodic. The data in the O-C diagram of U Gem (Figure 2-45) significantly deviate from a tentatively fitted sine curve. In the case of UX UMa there is no longer any support for a 29-year periodicity to be present in O-C. So far, no such conclusions can be drawn for any other object for lack of observational data.

Several mechanisms have been suggested to explain period changes in cataclysmic variables. To date, however, not a single one can be identified that seems likely to work. Among those mechanisms suggested is a third body circulating the cataclysmic system (Nather and Robinson, 1974; Patterson et al., 1978b) by means of which strictly periodic period changes could be explained in principle. However, besides the fact that this kind of change is not what is observed, constraints on the mass of this hypothetical third body are unreasonably high for it to be a likely explanation for many systems. Rotation of the line of apsides (Patterson et al., 1978b) again would produce strictly periodic period changes, and, in addition, the predicted time scales are on the order of months rather than decades. Redistribution of mass within the system, or loss of mass and/or angular momentum from the system (Pringle, 1975) all lead to only either a period increase or a period decrease, but clearly not to any kind of cyclic or even periodic changes. The only mechanism considered so far which might be able to explain non-monotonic changes of the
orbital periods is changes of the relative position of the hot spot in the system, though no explanation has been offered why this should happen on time scales of tens of years.

III.F. MAGNETIC ACCRETION

III.F.1. STRONGLY MAGNETIC SYSTEMS — AM HERCULIS STARS

RELEVANT OBSERVATIONS: In this class of objects photometric (at all observable wavelengths), spectroscopic, and polarimetric changes all occur with the same period.

ABSTRACT: The current model envisions a strongly magnetic, synchronously rotating white dwarf which prevents the formation of an accretion disc, so accretion occurs through magnetic funnels along the field lines.

There is a class of cataclysmic variables which exhibits very strong linearly and circularly polarized, Zeeman splitted lines during its photometrically low state, and in which all temporal variations occur with exactly the same period that is typical for the orbital period of cataclysmic variable systems.

The general interpretation which has been suggested for these AM Herculis stars is that the rotation of a very strongly magnetic white dwarf (field strengths of several times $10^7$ G have been determined from both polarization observations and Zeeman splitting) is phase-locked with the binary motion, and the strong field and correspondingly large Alfvén radius prevent the formation of an accretion disc in these systems. (For a recent review see Liebert and Stockman (1985).)

The gross picture is that material leaves the companion star at the Lagrangian point $L_1$ and enters the white dwarf's Roche lobe. For the first part of its journey to the white dwarf the trajectory of the material is hardly different from that in other cataclysmic variables, but eventually the magnetic field increasingly governs the flow and channels the matter onto one or both of its poles where it is braked and then accreted in a strong standing shock (Figure 4-32). The usual case is that one pole, the one closer to the secondary star (the white dwarf's rotational and magnetic axis can have any orientation with respect to each other) accretes most of the material, while the second, depending again on its position, in the extreme case may receive no material at all.

The observed hard and soft X-ray radiation, roughly having the shape of a bremsstrahlungs spectrum and a black body, respectively, are thought to be emitted from the accreting pole, whereby soft X-rays result from hard X-rays which have been degraded in the white dwarf's photosphere (e.g., Chanmugam, 1986). The third continuum component, a cyclotron spectrum which dominates the optical and the IR, is conjectured to originate higher up in the shock front (Figure 4-33). The observable photometric variations are brought about by aspect variations and/or temporary eclipses of the accreting pole(s), and differences between various systems are mostly due to differences in the inclination angle, the different orientation of the white dwarf's magnetic axis, and — related to this — whether and how strongly one or both poles accrete material. The observed sharp strong emission lines are believed to originate from the heated surface of the secondary star (e.g., Mukai et al, 1986), while the board bases of these lines seem to be emitted by the gas stream close to the shock front as it approaches the white dwarf. The very broad emission features in the optical and IR continuum (Figure 4-34a) are understood to be cyclotron emission lines (e.g., Wickramasinghe and Meggitt, 1982).

At irregular intervals of time, some AM Herculis systems are seen to drop in brightness by three to five magnitudes, an effect which is ascribed to temporary cessation of the mass overflow like that in other nova-like stars. At
Figure 4-32a. Phase diagram of events around the orbit of the AM Herculis object VV Pup (Patterson et al, 1984).

Figure 4-32b. Schematic model of an AM Herculis system (for description see text) (King, 1983).

Figure 4-33. (a) Components of radiation from an AM Herculis system and (b) loci of origin above the accretion pole (Lamb, 1985).

Figure 4-34. Model fits to observed spectra (a) and photometric and polarimetric variations (b) of the AM Herculis system VV Pup (Chanmugam, 1986; Osborne et al, 1986).
these states the X-ray influx is reduced, the strong narrow emission lines are strongly reduced or disappear altogether (the secondary star is not heated any longer), their broad bases disappear (since the accretion column is more or less non-existent), and in some systems the spectrum of the secondary component becomes visible in the IR (e.g., Bailey et al, 1985). At this stage several lines in the optical appear to exhibit Zeeman absorption components corresponding to the strength of the white dwarf’s magnetic field as derived from polarization measurements (e.g., Mitrofanov, 1980; Schmidt et al, 1983; Bailey et al, 1985). The polarization stays unaltered during low states — supporting the view that it originates from the white dwarf and that mass transfer from the secondary star is the cause for the observed changes.

Theoretically a very difficult and controversial point about AM Herculis stars is the actual physical structure and the geometrical size and shape of the accretion shock. A wealth of literature has been published on this subject during recent years, but many points have yet to be clarified (e.g., King, 1983; Langer et al, 1983; Frank and King, 1984; Meggitt and Wickramasinghe, 1984; Wickramasinghe and Meggitt, 1985a; 1985b; Chanmugam, 1986). An extensive review of the current state of the theory of AM Herculis stars, discussing the difficulties and prospects, has been given by Lamb (1985).

For conditions found in these systems, the main cooling process is optically thin cyclotron emission. Model computations based on this assumption fit the observed flux and polarization fairly well, in spite of the obviously still serious theoretical problems (Figure 4-34). From such fits the strength of the magnetic field, the geometrical size and shape of the accretion column, the various inclination angles, temperatures of the emitting region, and accretion rates can be obtained. The still obvious disagreements with the observations seem to become smaller if a reasonable temperature structure for the shock region, other than a single temperature, and other cooling effects than just cyclotron radiation are considered.

III.F.2. WEAKLY MAGNETIC SYSTEMS — DQ HERCULIS STARS

RELEVANT OBSERVATIONS: These systems exhibit more than one highly stable photometric period. Otherwise their appearance is like that of other (non-magnetic) nova-like stars.

see 112

ABSTRACT: The magnetic field of the white dwarf is of intermediate strength and disrupts the disc at some distance from the star; final accretion occurs along the field lines onto the magnetic poles. The white dwarf rotates asynchronously with the binary orbit. Illumination of system components by the hot accreting poles, or radiation from the poles themselves, produces the additional photometric periodicities.

The appearance and temporal variability of most cataclysmic variables can be understood satisfactorily without invoking magnetic fields; but, as was discussed in the previous section, there are others in which strong magnetic fields are directly measurable, and these fields determine the nature of these systems to a very considerable extent. If very weakly magnetic, or non-magnetic, systems and very strongly magnetic systems exist, it seems reasonable to assume the existence of systems with moderately strong magnetic fields.

A couple of systems have been observed, the DQ Herculis stars, which exhibit two or more extremely stable photometric periods. These have been tentatively identified with moderately strong magnetic systems (Bath et al, 1974a; Lamb, 1974; Patterson et al, 1978a; Chester, 1979; Patterson, 1979b; Patterson, 1980; Hassall et al, 1981; Patterson and Prince, 1981; Warner, 1983; 1985a; Cordova et al, 1985; Warner, 1985b; 1986a. For recent reviews see Warner, 1983; 1985a.) Neither polarization nor
Zeeman splitting in the lines is observable in DQ Herculis systems, but if magnetic fields are assumed to be one to two orders of magnitude smaller than those observed in AM Herculis stars, one would not expect to observe any direct evidence for these fields with current techniques.

The basic physics of these systems is identical to that of non-magnetic cataclysmic variables: a white dwarf accretes material from a Roche lobe filling secondary star through an accretion disc. However, here the magnetic field of the white dwarf is strong enough (estimates — mostly based on observed spin-up rates of the white dwarfs — predict field strengths on the order of $10^5$ to $10^6$ G, i.e., 1–2 orders of magnitude weaker than fields in AM Herculis stars (Lamb, 1985)) to disrupt the inner edge of the disc at some distance from the white dwarf. Interior to this Alfvén radius all material is forced to co-rotate with the field lines (i.e., with the white dwarf), thus accretion in this area occurs through accretion funnels which end in two hot regions around the magnetic poles on the surface of the white dwarf. Since the white dwarf does not rotate synchronously with the binary orbit, and since its magnetic axis is not in general aligned with the rotational axis, a beam of light from each magnetic pole sweeps over the system with the rotational period of the white dwarf. Thus, what is observable from these systems first of all are the normal brightness variations of a cataclysmic variable system, like humps, eclipses, etc. In addition, beams from one or both magnetic poles might be seen. Furthermore, the secondary star and the inner edge and/or the surface of the accretions disc or the hot spot are irradiated by the magnetic poles which produces radiation that is modulated with the rotational period of the white dwarf and/or some beat period with the orbital motion. Which part contributes how much to the observed radiation strongly depends on the system geometry, the orientation of the magnetic axis, and the position relative to the observer. Furthermore, changes in the disc geometry in particular, as for instance due to changes in the mass transfer rate, can account for more or less irregular brightness (amplitude) variations, or even time delays, of the degraded radiation. X-ray radiation is expected to be emitted from the accretion poles, which may or may not be directly observable depending again on the inclination angle; but certainly the X-rays are a powerful heat source in the system.

Warner (1986) carried out a systematic investigation of theoretically predicted and observed Fourier spectra of photometric variabilities in DQ Herculis variables, and was able to explain in a consistent way the optical and X-ray idiosynchrasies of as different systems as V1223 Sgr, FO Aqr, and AO Psc, and several others by simply assuming different inclination angles of the magnetic axis (Figure 4-35).

Phase shifts of the oscillations (71 sec period) of about 180° have been observed in DQ Her during the eclipse (Patterson et al, 1978a, see
Figure 3-26. Assuming the reflection of a rotating beam at the inner edge of the disrupted disc, a progradely rotating white dwarf with a rotational period of 71 sec, and a system inclination of about 90°, Patterson (1980) was able to theoretically reproduce the observed behavior. A similar shift seen in the occasionally present short-period oscillations in the very similar, but probably not magnetic, system UX UMa can be explained in the same way, if a somewhat larger inclination angle is assumed, and if bright blobs circulating in the inner disc are assumed to be the source for the brightness variations.

IV. MODELING THE OBSERVED SPECTRA

The observed spectra of dwarf novae and nova-like stars have been presented in the previous chapters. They comprise a fairly large range: pure emission spectra, pure absorption spectra, a mixture of both, asymmetric line profiles, very different slopes of the continuous flux distribution — and one single system may exhibit all of these features at different times. Changes from one sort of spectrum to another have been observed to occur as quickly as within one hour or less. Usually, however, they occur within a couple of hours or days.

From considering many of the observed properties, a conceptual model has been developed of what a cataclysmic variable may look like, the so-called Roche model, or a canonical model (see Chapter 4.II.A). This model makes predictions as to what parts of the system emit what kind of radiation. So in principle it should be possible to synthesize a theoretical spectrum of a cataclysmic variable by means of — probably slightly modified — conventional spectrum computations. Agreement and disagreement between computed and observed spectra should show whether or not the Roche model is applicable and where it probably will have to be modified and improved.

During the past ten years a couple of such attempts have been undertaken which are discussed in this section. It will be seen immediately that a reasonably comprehensive approach which accounts for all the observed features is currently well beyond our physical understanding of these systems, not to mention the numerical problems to be encountered along the way. The problem of computing spectra for cataclysmic variables is split into basically three separate problems: that of specifying the physical model to serve as a basis for the equations and the computations; that of computing the continuous spectrum; and that of computing the line spectrum.

IV.A. THE BLACK BODY APPROACH

RELEVANT OBSERVATIONS: The observed flux distribution of the continuous radiation does not resemble that of normal stars.

see 65

ABSTRACT: An approximate fit of the observed flux distribution is possible if it is assumed that only the accretion disc contributes to the radiation, and that at each distance from the white dwarf it radiates like a black body with a radial temperature distribution derived for a stationary disc.

No attempts have ever been made to compute spectra of cataclysmic variables based on a conceptual model other than the Roche model, since, in spite of some obvious shortcomings, no reasonable alternative exists. In fact, computations based on the Roche model do yield results which bear definite similarities to many of the observed features.

The assumption that the Roche model is valid grossly determines which parts and properties of the system — i.e., the masses of the stars, the orbital period, the inner and outer disc radius, and the angle of inclination of the system with respect to the line of sight of the observer — primarily determine the spectrum. Observations impose some further limitations
on possible parameter combinations, such as minimum and maximum orbital periods of some 80 minutes and some 10 to 15 hours, respectively, and also the observation that the secondary stars are usually close to the main sequence. Furthermore, from observed energies and velocities it is certainly justifiable to neglect all relativistic effects.

The easiest way to test the suitability of the Roche model from the spectroscopic point of view, and also to gain some experience and understanding of the radiation emitted by an accretion disc, is to assume that every point of its surface radiates like a black body. This first has been done by Tylenda (1977). He assumes a stationary accretion disc and thus can make use of the theoretical radial temperature distribution given by equation 4.2. For numerical purposes, the disc is divided into concentric "rings" of homogeneous physical conditions (Figure 4-36). The radiation of the entire disc is obtained from integrating over the contributions of all these "rings." He demonstrates that it is possible to qualitatively reproduce the observed spectra of the dwarf nova SS Cyg in outburst and of the old novae RR Pic and V603 Aql in a rather satisfactorily way.

This kind of computation proved to be a very valuable and straightforward tool for investigating the major influences of various parts of the system on the integrated radiation (la Dous, 1986; 1989). In Figure 4-37, the contribution functions of the "rings" of which the synthetic disc consists are displayed together with the resulting integrated spectrum. It is evident that in the example given, the UV radiation is entirely due to the very innermost 1 - 5% of the disc area, whereas the optical radiation is produced by both the inner and the middle areas, and the IR comes from the middle and the outer areas. It is particularly important to realize that very dramatic changes of the size of the cool outer disc hardly affect the optical radiation and do not at all affect the UV radiation. Furthermore, the radius of the white dwarf (which is critically dependent on the mass (Nauenberg, 1972)) has a very determining influence on the UV radiation. Likewise, the rate of mass throughput through the disc is by definition identical to the mass transfer rate from the secondary star (in a stationary accretion disc) is of crucial importance to the radiation emitted by the disc.

The influence of the boundary layer between the disc and the white dwarf on the spectrum,
is hard to estimate, since the amount and wavelength distribution of that radiation depends critically on the rotational velocity of the white dwarf as well as on the geometrical size and physical structure of the boundary layer, none of which are known. Computations demonstrate that this radiating component may or may not be of any importance in the UV, depending on local conditions. Clearly, however, it can always be neglected at optical and IR wavelengths.

Considering the contribution the secondary star provides to the integrated radiation from the system, its immediate influence is restricted to long wavelengths in the red and IR regions, since temperatures typically are on the order of 4000 K or less, and the projected surface area is comparable to the size of the disc. The secondary's contribution can become important, however, if the angle of inclination of the system is such that the disc is seen almost edge-on, so that its hot central parts do not contribute to the observed radiation.

For a very large accretion disc which comprises a very large temperature range, a spectral index of $\alpha_2 = 1/3$ ($\alpha_3 = -7/3$, respectively) is predicted analytically by just integrating the radiation of a black body disc (Lynden-Bell, 1969). It has often been claimed in the literature that the observation of such a continuum slope is an indication that the radiation is emitted by a stationary disc. From adopting maximum reasonable system parameters for the computations, however, it becomes immediately clear that accretion discs in cataclysmic variables are both too small and too cool for the spectral index to be around $\alpha_2 = 1/3$ for more than a very small spectral range. Thus, the fact that such an index has been observed on occasions has to be taken as an indication that the disc emitting that spectrum is not stationary but rather possesses some different radial temperature distribution.

**IV.B. CONTINUOUS AND ABSORPTION LINE SPECTRA**

**RELEVANT OBSERVATIONS:** In dwarf novae during outburst, one observes mainly broad absorption lines in the optical and UV.

**ABSTRACT:** A reasonable first-order fit of the spectra of dwarf novae during outburst is possible, adopting current methods of spectrum computations.

When optical and UV continuous and absorption line spectra of cataclysmic variables are computed, to a rather good approximation it is sufficient to concentrate just on the radiation emitted by the disc and to ignore contributions from all other components of the system. Furthermore, as in the case of black body discs (see above), it is quite reasonable to numerically divide the disc into concentric rings of homogeneous physical conditions and compute the spectrum emitted by each such ring, in close analogy to stellar (absorption) spectra. The basic difference between a star and a disc is the energy source, nuclear in one case, gravitational energy in the other, but as long as the central plane of the disc is optically thick, so that it is safe to assume that all the energy has been set free well below the photosphere, it does not matter for the emitted spectrum what the nature of the energy source is.

Spectrum computation requires the knowledge of the chemical abundance, the effective temperature, and the gravitational acceleration in the atmosphere. As to the chemical abundance, to adopt solar composition seems reasonable and in agreement with considerations about the evolutionary status of cataclysmic variables. Furthermore, slight changes in the chemical abundances show hardly any effect on the emitted spectra of discs. Again, under the assumption that the disc is stationary (which has been assumed so far in all published computations), equation 4.2 provides the radial temperature distribution. The gravitational acceleration in the vertical direction is determined by the geometrical thickness
Figure 4-38. (below) Spectrum emitted by the same system, seen under inclinations of 23°, 60°, and 83° (la Dous, 1989). The inclination angle is of great importance for the received radiation.

The inclination angle is of great importance for the received radiation, and increases with increasing distance from the central plane:

\[
g(z,r) = \frac{G M_{\text{WD}}}{\sqrt{r^2+z^2}} \frac{z}{\sqrt{r^2+z^2}^2} = \frac{G M_{\text{WD}}}{r^3} z \quad \text{for } z \ll r
\]  

(4.17)

(all the symbols have their usual meaning). The disc’s photosphere is defined to be at the geometrical height \(z_0\). From observations it can be concluded that \(z_0\) is small compared with the distance from the white dwarf, but its actual value is unknown. Slightly different assumptions were made by different authors about the value of \(z_0\), but any of these is essentially as arbitrary as any other. A comparison of synthetic disc spectra with various values for \(z_0\) in the range between 0.05° ≤ \(\theta\) ≤ 15° (\(\theta = \tan^{-1} \frac{z_0}{r}\)), as well as values which resulted from hydrodynamic computations of optically thick discs by Meyer and Meyer-Hofmeister (1982), revealed that differences in the computed spectra are too small to be seen in actual observed spectra unless the inclination becomes very large and the limb darkening law makes itself felt, in particular since effects of other theoretical parameters on the spectrum are much more serious (la Dous, 1986). Thus it can be concluded that for geometrically thin, optically thick accretion discs any value of \(z_0(r)\) is reasonable as long as \(z_0(r) < \ll r\) or \(\theta < 10°\).

Also, in many of the published spectrum computations it is assumed that the gravitational acceleration does not vary within the atmosphere. Whether or not this is a justified assumption depends on the relative thickness of the atmosphere with respect to the underlying optically thick disc. Again, test computations revealed that as long as the disc is really (and not just marginally) optically thick in the central plane, the assumption that \(g(z) = \text{constant}\) is quite justified (la Dous, 1986; 1989). The situation of course changes entirely either if energy production in the disc atmosphere is taken into account, by means of which the atmosphere can become very extended with its outer parts then called a "corona," or if the disc is optically thin in the central plane.
Figure 4.40. Contribution functions from an accretion disc model: (a) lines have not yet been broadened by the disc rotation; (b) lines broadened by the Doppler velocity.
Figure 4-40. (Continued) Integrated disc radiation (c) in the UV and (d) in the optical (la Dous, 1989).
One further principle difference between atmospheres of stars and atmospheres of discs is that, unlike stars, discs are essentially two-dimensional objects and thus the angle of inclination \( \text{i} \) between the normal on the rotational plane and the line of sight of the observer is important for the observed radiation. This mostly affects the limb-darkening in that mainly the radiation emitted by the cooler layers of the atmosphere is received at the higher inclinations (Figure 4-38). With respect to computing a spectrum this means that, in a disc, it is the angle-dependent flux emerging from each part of the atmosphere which has to be integrated over all the disc rather than the angle-averaged flux. Since published stellar spectra (computed or observed) all yield only the angle-averaged flux, this implies serious limitations to comparing computed stellar spectra with observations when used as a basis for synthesizing relevant disc spectra.

The parameter ranges covered by accretion discs in cataclysmic variables suggest that, with the possible exception of the very central parts close to the white dwarf (where the boundary layer probably has a large influence), LTE is a good approximation for atmosphere computations (Figure 4-39). Radiation pressure can well be neglected. Electron scattering, however,
can become very important in some areas of the disc. Convection does become important at larger distances from the central object. However, irrespective of their large geometrical size, the contribution of these cool parts to the total radiation of a stationary, optically thick accretion disc is negligible in the optical and UV, so proper treatment of convection is not of crucial importance.

When line radiation is included, the effect of the Keplerian rotation of the disc must be taken into account, which leads to both additional line broadening as well as to double-peaked absorption profiles (in analogy to emission profiles, see Figure 4-45, Chapter 4.IV.C). The main effect on the Balmer lines is that they appear double-peaked, particularly at high inclination angles, while the UV lines of heavy elements can be broadened by many times their pressure-broadened value, and a lot of detailed information can be washed out... an effect which is strongly enhanced by the integration of radiation from all parts of the disc (Figure 4-40).

Spectrum synthesis for stationary, geometrically thin, optically thick accretion discs has been performed by several authors (Herter et al, 1979; Kiplinger, 1979; 1980; Mayo et al, 1980; Pacharintanakul and Katz, 1980; Tylenda, 1981a; Wade, 1984; la Dous, 1986, 1989), using either published sets of, or program codes for, stellar atmospheres and including, or in some cases not including, absorption lines in the computations. The effects of parameter changes on the theoretical spectra have been investigated (quoted references should be checked for details). The major results which emerged are that, of the possible free parameters investigated which might influence the character of the computed spectrum (the mass $M_{\text{WD}}$ of the white dwarf, the mass transfer rate $M$, the inner and outer disc radii $r_i$ and $r_o$, respectively, and the angle of inclination $i$), the mass of the white dwarf — not considering its influence on the star's radius — and the outer disc radius are of practically no importance. The drastic effect of the inclination angle $i$ on lines as well as on the continuum flux already has been shown in Figure 4-38. The mass transfer rate $M$ enters the radial temperature profile and thus is a very determining factor for the appearance of the disc spectrum (Figure 4-41), and variations of $r_i$ strongly alter the temperature range. If due to a magnetic field which the white dwarf may possess, the inner disc radius $r_i$ also is a variable parameter, this is even more important for the disc spectrum than $M$ (Figure 4-42), since its variation means including or neglecting the very hottest areas of the disc which entirely determine the UV and also partly the optical radiation.

In general, almost any parameter variations have much more pronounced effects on the UV than on the optical radiation, so that no useful information about the system as a whole can generally be obtained from the optical colors of cataclysmic variables alone. Including also UV colors is of limited use. Meaningful system parameters may possibly be derivable from high-quality, high-resolution spectroscopic data of the optical plus UV range or from continuum observations alone, if the mass of the white dwarf is known with some confidence (la Dous, 1989), but this has not been tried yet.

Irradiation of the disc by hot central areas and/or the boundary layer, or generation of energy in the atmosphere, lead to the formation of extended chromospheres and coronae above the accretion disc which change considerably the nature of the emitted radiation, even if the accretion disc itself is optically thick (Schwarzenberg-Czerny, 1981; Kříž and Hubeny, 1986; Shaviv and Wehrse, 1986). Results so far have not led to more than the demonstration that such thin extended regions are present.
IV.C. EMISSION LINE SPECTRA

RELEVANT OBSERVATIONS: During the quiescent state, most dwarf novae are dominated by strong emission lines of hydrogen in the optical and by strong metal resonance emission lines in the UV. see 73

ABSTRACT: The line emission is attributed to optically thin areas in the inner and/or outer accretion disc, or even to the entire disc. Modeling is considerably more difficult than for outburst spectra since there is evidence that the disc is not stationary (i.e., no radial temperature law is available a priori), and non-LTE effects might become important, in particular in the presence of energy generation in the atmosphere.

The spectra of quiescent dwarf novae, some nova-like systems, and old novae, usually exhibit strong emission lines of H, HeI, HeII, and, particularly in the UV, of highly ionized elements such as C IV, Si IV, N V. Different lines in one spectrum can, and usually do, exhibit vastly different profiles; in particular, the Balmer and He lines often are double-peaked if the system is seen under a large inclination angle, while normally all other lines exhibit single-peaked profiles. The shapes of the radial velocity curves are different for different lines and species, normally lines are eclipsed if the continuum undergoes an eclipse, but different lines are eclipsed in different ways, and all are affected differently than the continuous radiation. In general, it can be concluded that emission lines originate from the Roche lobe of the primary stars (white dwarf); that most emission lines originate near the orbital plane, i.e., in or near the accretion disc; that not all lines are formed in the same region of the disc; and that some lines, in particular those with the highest ionization potential, are formed far away from the orbital plane.

From lines and line ratios seen in the spectrum of the old nova V603 Aql, Ferland et al (1982a) conclude that the line emission probably originates in an extended corona filling most of the white dwarf's Roche lobe which is heated by irradiation from the disc. Similarly, Jameson et al (1980) conclude from the UV spectra of the somewhat peculiar cataclysmic variable AE Aqr that the cool emission lines, Mg II, Ca II, and about half of Lyα are likely to originate in optically thin parts of the accretion disc itself, while the lines of the highly ionized elements, in particular N V, Si IV, C IV, and He II are likely to originate in a chromospheric region outside the disc, which again is heated by the hot central parts of the disc.

In general, emission lines in cataclysmic variables can originate either in a corona above or below the accretion disc, or in optically thin parts of the disc itself. Williams (1980) included the possibility for the disc to be optically thin in the continuum in the outer cool disc areas (far away from the white dwarf), while it is still optically thick in the lines. The particular features of his results are highly dependent on the choice of the viscosity and on the assumption that the entire disc is stationary, but the general features are of wider significance. He finds that larger areas of the outer disc become optically thin in the continuum with decreasing mass transfer rate while the inner disc is still optically thick in the lines. The particular features of his results are highly dependent on the choice of the viscosity and on the assumption that the entire disc is stationary, but the general features are of wider significance. He finds that larger areas of the outer disc become optically thin in the continuum with decreasing mass transfer rate while the inner disc is still optically thick (Figure 4-43a); at very small transfer rates (of the order of $M < 10^{-12}$ M⊙/yr in his models) the entire disc becomes optically thin in the continuum. The Balmer lines remain optically thick throughout almost all of his model discs. The intensity ratio he obtains for the emissions of Hα, Hβ, and Hγ are in reasonable agreement with observed ratios; however, he does not take into account contributions from absorption lines produced by the optically thick parts of the disc. None of his models produce any He lines of appreciable strength, from which Williams concludes that — for discs of the kind investigated — the lines, which are observed to be fairly strong at times, originate in some other, hotter, optically thin regions of the system. The disc in Bailey and Axon (1981) is assumed to be stationary, thus the optically thin outer disc has to heat up considerably over the respective black body
temperature in order to radiate away all of the locally created energy. This results in the temperature in these outer regions staying more or less constant (at about 6500 K for the viscosity chosen in William's example) over geometrically large areas (Figure 4-43b).

Tylenda (1981a) arrives at very similar results from only slightly different computations. He stresses the point how increasingly poorly, particularly at optical and UV wavelengths, black bodies represent the continuous radiation emitted by accretion discs as ever larger optically thin areas are involved, increasing in size with decreasing $M$ and increasing outer disc radius (Figure 4-44a), but also how sensitive the actual temperature reached in the outer disc is to the particular choice of the viscosity (Figure 4-44b). He also finds that the emission line profiles are critically dependent on $M$ and viscosity, and also on the inclination angle, as one would expect.
When it comes to modeling actually observed emission line profiles, it is known that they depend on a wealth of parameters in addition to the influences already discussed above in the case of absorption lines; on the radial and azimuthal brightness distribution in the accretion disc, on the inner and outer radius of the region where the particular line under investigation originates, on gas densities in the region, on the amount of turbulent motions present, on the rotational velocities in the discs and their relative size in different regions, on the shape and geometrical origin of the continuous spectrum as well as of the absorption lines, and on even more parameters, depending on the degree of sophistication of the computations. Several increasingly promising attempts to model emission line profiles have been undertaken so far, in particular of the Balmer lines of hydrogen; but clearly a lot of research remains to be done. The double-peaked lines of high-inclination systems are particularly well suited for testing of models, since these profiles have sufficient structure to impose considerable constraints on model parameters.

Double-peaked profiles are believed to be due to the combined effects of the disc being a two-dimensional object and Keplerian rotation, in the same way double-peaked absorption profiles arise in the computations of optically thick accretion discs (Chapter 4.IV.B). In principle, the peak separation should be indicative of the rotational velocity of the outermost radius in the disc, where a considerable contribution to the line under investigation originates (Figure 4-45). Computations indicate that such a peak separation should already be visible at inclinations as low as 15°. There are objects on the other hand, like BT Mon, LX Ser, and SW Sex, which exhibit clear single-peaked hydrogen emission profiles, yet they show eclipses of the continuous as well as of the line radiation which point to fairly high inclinations (Williams et al, 1988). One suggestion for solving this problem was to assume the presence of an additional component in the core of the emission lines, the geometrical origin of which was not specified further (Smak, 1981). Recently however, Marsh (1987) and Williams et al (1988) pointed at the importance of Stark broadening in smoothing out the peaks of theoretical emission line profiles in accretion discs, an effect which is also able to explain the presence of very broad wings of the Balmer emission lines in some systems which are seen nearly pole-on. The actual importance of the Stark effect depends largely on the surface density of electrons and thus on the temperature. Due to different ways in which the Stark effect acts on H and He, a situation is conceivable wherein double-peaked H lines co-exist with single-peaked He lines (Williams et al, 1987) — a situation which is actually observed at times (see Chapter 2.II.B.1, Figure 2-93). Marsh (1987) stresses the point that while Stark broadening certainly has a very important influence on line profiles in the inner disc, it may not be very important for the formation of double-peaked profiles of the H lines if the bulk of emission originates in the outer cool disc areas where the electron density is low. Furthermore, there is an observational indication against Stark broadening to be the cause of the single-peaked profiles, at least in the nova-like system LX Ser, since the Paschen lines should still have double structure if Balmer lines appear single-peaked due to Stark effect, which in LX Ser they do not, when observed at sufficiently high wavelength resolution (Young et al 1981a).

A Keplerian rotating accretion disc implies the presence of significant velocity gradients in the atmosphere. This does not pose a problem for the continuum radiation, nor does it probably pose one for absorption lines according to Rybicki and Hummer (1983). They elaborate on the importance of this additional asymmetry when dealing with accretion discs which are optically thin in the continuum. They discuss in particular conditions in the outer cool parts of the discs where the strong H emission lines are formed. They try to apply the Sobolev theory (a simplified version of the general escape probability method, applicable in the presence of high velocity gradients), but conclude that
physical conditions in accretion discs in cataclysmic variables at best marginally allow for applicability of this particular method.

Horne and Marsh (1986) deal further with this problem, again assuming a condition where the disc area investigated is optically thin in the continuum and either optically thin or optically thick in the lines. In order to show the general influence of some physical effects on the emission line profiles they still adopt the Sobolev approximation for sake of simplicity, fully aware, however, of its limitations. They find that the local anisotropy can become a very important factor for the profiles of optically thick emission lines, particularly when the disc is seen under large inclination angles, while it is of only minor importance for optically thin lines, which by definition radiate isotropically. The effect on optically thick lines is illustrated in Figure 4-46a: for inclination angles larger than about 60° the minimum between the two peaks of the emission lines becomes considerably deeper if shear broadening is taken into account, compared to the case when it is not. Comparison of these computed profiles with an observed profile (Figure 4-46b) of Hα in Z Cha demonstrates that the fit is decidedly better.
Figure 4-47. The principle of the line deconvolution: when a disc with asymmetric brightness distribution is seen under different angles, different line profiles result; the deconvolution method tries to reconstruct the information contained in the line profiles (Marsh, 1986).

Figure 4-48. Reconstructed emission of Z Cha during quiescence in four Balmer lines: (a) total emissivity.
Figure 4-48(b) Fractional deviations from symmetry.

When the effect of this velocity gradient is taken into account, a similar situation accounts for observed profiles of U Gem (Stover, 1981a).

In order to compute the profile of a line emitted by an entire accretion disc some radial dependence of the line emissivity has to be adopted (like assuming the disc is rotationally symmetric). The assumption of a power-law dependence:

\[ f(r) \sim r^{-b} \]  

for the radial variation of the emissivity proved successful for modeling observed eclipses of H and He lines in DQ Her, with \( b = 0 \) for the
Balmer lines and \( b = 2 \) for He II 4686 Å. Applying the same hypothesis and analyzing observed line profiles of different cataclysmic variables, values in the range \( 1 < b < 2.2 \) provided the best fit (Smak, 1981, and references therein). Horne and Marsh demonstrate how various choices of \( b \) mostly affect the wings of the theoretical profiles (Figure 4-46b).

In basically the same way in which Horne et al (Horne, 1985; Horne and Cook, 1985; Horne and Stiening, 1985; Wood et al, 1986 see also Chapter V.IV.E) tried to reconstruct the continuum brightness distribution of accretion discs from UBV observations during eclipses, Marsh (1986) attempted to reconstruct the brightness distribution within the discs for line emission, by relating profiles observed at different orbital phases (Figure 4-47). The basic assumptions for this computation were Roche geometry and that the line radiation be emitted by the stationary, Keplerian rotating accretion disc. As a first approximation an azimuthally homogeneous disc was assumed. The resulting radial and azimuthal distribution of the emissivity in \( \text{H}_\alpha \), \( \text{H}_\beta \), \( \text{H}_\gamma \), and \( \text{H}_\delta \) in the eclipsing dwarf nova Z Cha during quiescent state (as reconstructed from phase-resolved echelle spectra) is displayed in Figure 4-48. Approximately the expected pattern is obtained: the emissivity basically is azimuthally (rotationally) symmetric, with the single exception of \( \text{H}_\gamma \) which for unknown reasons departs considerably from symmetry; fractional deviations from symmetry indicate the presence of the hot spot at about the place where it is expected according to theoretical considerations, which, since it only becomes visible when deviations from rotational symmetry are considered in turn indicates that the contribution of the hot spot to the integrated radiation cannot be more than a few percent; and, finally, the radial distribution of the emissivity clearly does not follow a simple power law (which would be a straight line in Figure 4-48c) but a more complicated pattern which is unique for each of the investigated lines.

Marsh attempts to reconstruct theoretically the observed spectrum with this probably reasonable distribution of the line emissivity, taking into account Stark broadening and shear broadening and using both a fairly accurate brightness temperature distribution derived from eclipse mapping methods and also a fairly reliable angle of inclination of the system (derived in Wood et al, 1986). He concludes that the disc is probably not in a steady state, since the mass transfer rate estimated from the luminosity of the hot spot is much larger than permissible for an optically thin accretion disc; an optically thick disc, on the other hand, matches the observations decidedly worse than an optically thin one; secondly, non-LTE conditions have to be assumed in order to match the high Balmer emissivity observed from the center of the disc.

4.IV.D. P CYGNI PROFILES, WINDS, AND CORONAE

RELEVANT OBSERVATIONS: Strong P Cygni profiles or blue-shifted absorptions are seen in C IV, and occasionally also in Si IV and N V, during the outburst of dwarf novae and in many nova-like stars. These profiles are considerably different from those of normal hot stars. Systems with very high inclination angles exhibit only strong emission profiles in these lines. All other lines are seen in absorption as usual.

see 84, 99, 107

ABSTRACT: The observations can be reproduced rather well assuming an optically thick accretion disc and a wind driven out from the disc center and/or boundary layer.

In many nova-like stars and in dwarf novae during the outburst state the Balmer lines of H in the optical are seen in absorption, with profiles that are symmetric about the line center. In the UV, however, many systems exhibit absorptions in C IV (1549 Å), N V (1240 Å), and in Si IV (1400 Å), which are asymmetrically blue shifted by 3000 - 5000 km/sec or more. In addition, the C IV line also has an appreciable red-shifted emission component.
which is often seen at the same time, making it a classical P Cygni profile; while no more than a slight indication of an emission can be detected in the Si IV and N V lines. Other systems exhibit spectra with C IV in emission during the photometrically bright state, while the other lines are either in absorption or in emission; but all lines are roughly symmetric about line center (Chapters 2.II.B.2.3). When compared with system parameters it turns out that P Cygni profiles seem to appear exclusively in low-inclination systems, while very high-inclination systems show at least C IV in emission during the photometrically bright state. Eclipse observations of RW Tri and UX UMa show that Si IV and N V are decidedly less affected than the continuum at the time of strong attenuation of the continuous flux, while C IV is hardly eclipsed at all (Drew and Verbunt, 1985). From these observations in these two stars, it can be concluded that all of these lines are formed in an extended region centered about the white dwarf, and that C IV also originates in a much larger volume which thus is practically insensitive to the eclipse. Furthermore, from the observation that C IV does exhibit P Cygni profiles while Si IV and N V do not, it can be concluded that the abundance of the latter two ions is much less than that of C IV.

In analogy to P Cygni profiles observed in O, B, and Wolf-Rayet stars, it was inferred from their mere presence that mass loss occurs from dwarf novae and nova-like stars. Several authors (Krautter et al., 1981b; Cordova and Mason, 1982; Greenstein and Oke, 1982; Klare et al., 1982) tried to derive mass-loss rates for cataclysmic systems from comparing synthetic P Cygni profiles for O-stars by Olson (1978) or Castor and Lamers (1979) with the observed profiles. They arrived at some computed values, but most of their computed profiles can hardly be matched by any of the observed ones, and also, as pointed out by Drew and Verbunt (1984; 1985), it is by no means obvious that their inherent assumption that the physics underlying mass outflow from hot single stars is basically identical to that working in cataclysmic variables applies. In other words, is the geometry of stars and discs different enough to be likely to have effect on observable line profiles? Drew and Verbunt attempted to gain more insight into the problem of mass loss from cataclysmic variables by actual modeling of line profiles and comparing them with observations (Drew and Verbunt, 1984; 1985; Drew, 1986; 1987). Their findings are reported in what follows.

They start by investigating the properties of a wind which is driven away from the vicinity of the white dwarf. Such a wind cannot be driven thermally by conditions in the inner disc or boundary layer, since temperatures of around $10^8$ K would be required to produce a wind in that way.* Such temperatures could not, however, drop down sufficiently on relevant time-scales to account for the observed ionization states. Consequently, the authors assume that the wind is probably driven by radiation pressure due to radiation from the boundary layer, i.e., by conversion of (E)UV photons into kinetic energy of the ions, which requires a much lower temperature in the vicinity of the white dwarf.

Assuming that the wind is radiatively driven, and assuming that the ions were to escape from near the surface of the white dwarf to infinity — which requires an escape velocity of order 5000 km/sec, a velocity of the order observed in the blue-shifted absorptions of the UV resonance lines — an estimate of the mass-loss rate from the system can be obtained:

$$\dot{M}_w < \frac{v_{esc}^2}{v_\infty c} \approx \frac{v_{esc}}{c}, \quad (4.19)$$

where $\dot{M}_w$ is the mass loss rate, $\dot{M}$ is the mass accretion rate, $v_{esc}$ and $v_\infty$ are the escape

---

*In the boundary layer of a non-rotating white dwarf, a considerable amount of energy, equal to the luminosity of the entire disc, could be released in principle but there is no evidence for $10^8$ K temperatures there — see Chapter 4.B.2.
velocity and the terminal velocity in the wind, respectively, and c denotes the speed of light. This result implies that the mass-loss rate should not amount to more than a few percent of the mass accretion rate.

Based on these findings and considerations of the physical structure of winds in cataclysmic variables, Drew (1986; 1987) carried out a systematic investigation of theoretically expected resonance line profiles (C IV). She
assumed an optically thick, geometrically thin accretion disc in which the temperature is assumed to decrease with increasing distance from the white dwarf, according to equation 4.2 — i.e., no hot boundary layer is required for the production of these profiles. In this case the wind originates preferably from the disc center and/or boundary layer. The outflow was assumed to be spherically symmetric, unless mentioned otherwise below. Computations were carried out in the Sobolev approximation, which as Drew points out, will yield results which may not be accurate in detail, but will show gross properties in a reliable way.

As shown in Figure 4-49, there is a very marked difference in P Cygni profiles originating from single stars and from extended discs, respectively. Due to the much larger extension of the geometrical size of the continuum source, a mass loss rate higher by an order of magnitude is needed in cataclysmic variables in order to produce any observable effect in the line profiles. The shapes of both the absorption and the emission components in P Cygni profiles originating from discs look totally different from those of stars: the emission component is much shallower and the absorption component much weaker than for a comparable situation in stars. One consequence of this is that, somewhat dependent on the radial velocity law in the wind, conditions could exist under which the absorption minimum does not reach below half of the continuum intensity even in totally optically thick lines. It is in accordance with the observations that the absorption minimum lies close to the central wavelength in the case of a disc, whereas it can be blue-shifted by a considerable amount in the case of stars. Furthermore, the amount of scattering in the winds from discs is found to be fairly unimportant for the appearance of the line profiles, much unlike the case in stars, where it is very important. The effect of the radial velocity of the wind on the line profiles is demonstrated in Figure 4-49b, which shows that at high M slow wind acceleration causes the absorption to develop a sharper, deeper minimum, while the emission is strongly enhanced. Comparison with observations thus leads to the conclusion that acceleration in discs is likely to occur much more slowly than in single stars. The dependence of the profile on the inclination angle is considerable (Figure 4-49c), as expected from observations where, as pointed out above, P Cygni profiles are seen only in systems where the inclinations are not too high; for a small angle in the computations, when the disc is seen more or less face-on, only blue-shifted absorption components can be seen, and only for discs which are seen almost edge-on (i.e., at inclination angles which most likely would produce eclipses of the continuous radiation) does the emission become very strong. If some limb-darkening of the disc continuum is assumed (the exact characteristics of which are unknown), or when a bi-polar wind outflow is assumed rather than a spherically symmetric one, the absorption disappears completely for high inclinations, while the profiles for low inclinations are hardly affected. Mauche and Raymond (1987) arrive at very similar results from similar investigations.

In conclusion, of greatest importance for determining the actual shape of the resulting line profile according to these calculations are the mass loss rate, the velocity profile of the wind, and the temperature and the temperature profile of the accretion disc.

4.IV.E. ECLIPSE MAPPING

RELEVANT OBSERVATIONS: Photometric eclipses observed in dwarf novae and nova-like stars contain a lot of information about which parts of the system are being eclipsed.

see 35, 96, 103

ABSTRACT: The surface brightness and temperature distribution of the accretion discs can be reconstructed from eclipse observations by means of image processing techniques. The results agree well with the concepts of the Roche model.
When trying to compute the spectra of accretion discs, the most serious problem encountered is that the temperature distribution in the disc is in general not known. However, a useful method for the empirical determination of the surface-brightness distribution in accretion discs has been developed by Horne (1985). He applies methods of image-processing to light curves of eclipsing cataclysmic variables. The general idea is that an eclipse light curve is determined by, and thus contains information on, the brightness distribution in the accretion disc. Being a one-dimensional set of data, it does not contain all the information of an essentially two-dimensional object, but, provided some additional information is given about the system investigated, the range of possible solutions and parameters can be narrowed considerably. The assumptions made are that the eclipse is one of an axi-symmetric two-dimensional object and that a default image ensures an optimum reconstruction of the radial brightness distribution on the expense of azimuthal information (i.e., the hot spot will cause a bright ring at the outer radius of the reconstructed image, and it will appear as a hot peak in the derived radial temperature law — see Figure 4-50b, d). Out of the several solutions still possible for image deconvolution, that image is chosen which maximizes the entropy for each element (Horne, 1985, equation 5). The data shown in Figure 4-50 demonstrate that the reconstruction (Figure 4-50b) of the input disc image (Figure 4-50a) from the “ob-

![Figure 4-50. Test images for the eclipse mapping method: (a) input image is a steady state black body accretion disc with a hot spot; (b) the reconstructed image; (c) the synthetic eclipse light curve used for the reconstruction; (d) the reconstructed brightness temperature distribution (dots) as compared with the input distribution (solid line); the hot spot appears as peak in temperature at the other disc radius (Horne, 1985).]
Figure 4-51. Surface brightness and radial temperature distributions of Z Cha in outburst (a, b) and quiescence (c, d), reconstructed by using the eclipse mapping method (Horne and Cook, 1985; Wood et al., 1986).

served" light curve (Figure 4-50c) is convincing. One other input into this test image was a radial temperature law as it is assumed to hold for stationary accretion discs, $T(r) \sim r^{-3/4}$, which is reconstructed with 20% accuracy (Figure 4-50d). In terms of mass transfer rates for stationary accretion discs this translates into an accuracy of within a factor of two.

Possible applications of this method are obvious for the derivation of any radial temperature profile — thus also allowing for the localization of geometrical origin of the various line radiations, for the distinction between optically thick and optically thin disc areas from the analysis of light curves obtained in different colors, for tracing the development of the disc through the outburst cycle, and, as already mentioned, for the determination of the mass transfer rates, and finally of distances.

Some of these possibilities have not been exploited yet, but this eclipse mapping method has been applied to two objects already, to
UBV observations of the dwarf nova Z Cha during outburst (Horne and Cook, 1985) and in white light during quiescence (Wood et al, 1986), and to the nova-like system RW Tri. The reconstructed images and radial temperature profiles of Z Cha are shown in Figure 4-51. From these it is clear that during outburst the radial temperature distribution agrees with that of a stationary accretion disc, while this is not at all the case during quiescence. Furthermore, all the disc is probably optically thick during the bright state, while in the low state the inner disc is probably optically thin (no conclusive results can be obtained from measurements in merely white light) and the outer disc is optically thick, having a much higher mass transfer rate than the inner disc. Both results are in agreement with the concept that during quiescence matter is accreted only in the outer disc, while it is accreted onto the white dwarf during outburst. Inferred mass transfer rates are $10^{8.9 \pm 0.3} M_\odot/yr$ during the observed outburst, and somewhere between $10^{-10.23}$ and $10^{-10.08} M_\odot/yr$ for the outer parts of the quiescent disc, down to some $10^{-11.7} M_\odot/yr$ at disc center. During the quiescent state the disc temperature increases only very slowly from some 4000 K at the outer rim to some 7000 K in the center; during outburst it ranges from some 7000 K at the outer edge to some 35000 K in the center (it should be kept in mind here that optical colors are quite insensitive to UV flux, so that the temperature of 35000 K should be regarded as somewhat approximate). The radial colors for the disc in outburst all fall between the theoretical lines for black bodies and for main sequence stars for the hot central parts of the disc ($r \gtrsim 0.3$ of the radius of the Roche lobe), indicating that the vertical temperature gradient in the disc is flatter than that of main sequence stars. Finally, a distance of $105 \pm 20$ pc is derived from the angular diameter of the disc, assuming a radial velocity amplitude of the red companion star of 400 km/sec.

A similar analysis was carried out for U, B, R observations of RW Tri (Horne and Stiening, 1985). This disc is obviously optically thick*, and it follows roughly a radial temperature dependence as expected for a stationary disc. The temperature ranges from 10000 K to 40000 K, and a mass transfer rate of $M = 10^{-7.9 \pm 0.4} M_\odot/yr$ and a distance of some 500 pc are derived.

4.IV.F. THE BOUNDARY LAYER

RELEVANT OBSERVATIONS: Hard and soft X-ray radiation which cannot be ascribed to the accretion disc is observed from many dwarf novae and nova-like stars.

see 60, 69

ABSTRACT: A boundary layer between the disc and the white dwarf is assumed to form where the Keplerian rotating disc material is braked down to the rotational velocity of the white dwarf in order to be accreted. The detailed structure of this layer is not yet clear.

A final weak point in current spectrum computations is that only very little is known about the physical structure of the radiation emitted from the boundary layer. The current state of knowledge is summarized in this section. For other recent reviews see e.g., Shaviv (1987) and Stanley and Papaloizou (1987).

Material in the accretion discs of cataclysmic variables is assumed to circulate with Keplerian velocity at any distance from the white dwarf. The white dwarfs, on the other hand, are not likely to rotate at the Keplerian velocity corresponding to their surface (which would mean just short of brake-up). So, at the boundary where the inner disc meets the surface of the star, a velocity gradient is likely to be present. By some mechanism the excess kinetic energy of the disc material has to be converted into some other form of energy so that at least a significant fraction of this material can eventually settle onto the surface of the white dwarf. This region is referred to as the boundary layer.

---

*This combination of colors permits a clear distinction to be made between optically thick and optically thin areas.
The amount of energy liberated depends on the velocity gradient. An upper limit can be obtained from assuming a non-rotating star in which case no boundary layer would be formed. In general the luminosity of the boundary layer is

\[ L_{\text{BL}} \lesssim \frac{G M_{\text{WD}} M}{R_{\text{WD}}} \quad (4.20) \]

where all the symbols have their usual meaning; i.e., once again about as much energy as that radiated away by the entire disc can be liberated in the boundary layer, an energy of the order of \(10^{33}\) erg/sec. Given the small geometrical size of the boundary layer, it is the probable place of origin of the observed X-ray radiation from cataclysmic variables. While other places of origin have been suggested, like the coronae of the secondary stars or the hot spot, these other sites lead to severe theoretical problems concerning the amounts of energy liberated (Patterson and Raymond, 1985a); thus the boundary layer seems the most likely site. Observations indicate that during the quiescent state the hard X-ray flux obtained from dwarf novae can be as high as the optical plus the UV flux from the disc, while no soft X-ray flux can be seen. During outburst the hard X-ray flux increases only slightly, if at all, while the soft X-ray radiation increases by one to two orders of magnitude (Chapter 2.III.A.2). The sample of observations in the latter case is not really statistically significant (three objects: U Gem, SS Cyg, VW Hyi), but negative observations of other sources do not contradict a generalization of this kind.

Theoretically, the physical structure and luminosity of the boundary layer are not yet known reliably. For making estimates, it is almost always assumed that the white dwarf is non-rotating (consistent with high observed X-ray luminosities), in which case hard X-rays are likely to be due to an optically thin boundary layer which has to heat up considerably in order to radiate away all of the energy produced. Soft X-rays occurring at high accretion rates, on the other hand, are attributed to optically thick boundary layers where the radiation is thermalized before being able to escape the system (Pringle, 1977; Pringle and Savonije, 1979; Tylenda, 1981b; Frank et al, 1985; Patterson and Raymond, 1985a; 1985b). Energy is assumed to be liberated either by very efficient viscous interaction (Tylenda, 1981b) or by shocks, as the fast-rotating material approaches the surface of the white dwarf (Pringle, 1977; Pringle and Savonije, 1979). One problem with this latter approach is the necessity to produce radial shocks in a basically azimuthally moving material, where shear flows and consequent turbulence reduce the efficiency of the shocks (Frank et al, 1985). King and co-workers (King and Shaviv, 1984; Frank et al, 1985) propose that during quiescence, when the turbulence is expected to occur, the boundary layer heats the gas to temperatures up to \(10^8\) K and also causes it to expand out of the boundary layer, thus forming a hard X-ray emitting corona around the central object. At high accretion rates — a value typically considered, in agreement with the observations, is \(M \gtrsim 10^{16}\) g/sec — the expansion is suppressed, thus reducing the hard X-ray flux and at the same time increasing the soft X-ray radiation. Elaborating on an idea suggested by Icke (1976) that shocks created by turbulent activity in the disc form a corona above the accretion disc, Jensen et al (1983) suggest that a dynamo coupling differential rotation and convection generates a magnetic field, which leads to the formation of a hard X-ray emitting corona, in analogy to models suggested for the formation of the solar corona. By this concept the authors are able to explain qualitatively the observed relation between oscillations seen in hard X-rays and in optical wavelengths in particular. Such a corona could also be responsible, at least partly, for the UV resonance lines observed in emission in most quiescent dwarf novae; it cannot be responsible, however, for the P Cygni pro-
files seen during outburst, since then the observed X-ray radiation is softer. Any of the scenarios proposed above can be reconciled with the tentatively established relationship between the angle of inclination and the observable X-ray flux (Patterson and Raymond, 1985a) which suggests an obscuration of the source of hard X-rays, thus implying that the emitting region must be confined to an area fairly close to the white dwarf.

How much radiation in X-ray energies is to be expected from cataclysmic variables is subject to a wealth of assumptions and hypotheses. A so-called "mystery of the missing boundary layer" has been mentioned a couple of times in the literature (Ferland et al., 1982a; Burkert and Hensler, 1985; Kallman and Jensen, 1985). What is meant by this term is that the observed soft X-ray flux from many cataclysmic variables is lower by a factor of 10 to 100 than what has been predicted from theoretical considerations. A careful examination reveals that all the quoted estimates of expected X-ray flux are upper limits based on the assumptions of a non-rotating white dwarf plus a couple of other somewhat arbitrarily chosen parameters which will be considered below. The "mystery" seems to consist of still questionable assumptions, not paradoxical results. The amount of observable radiation depends on: the geometrical size of the boundary layer, which is normally assumed to be small, on the order of the thickness of the disc near the white dwarf, the rotational velocity of the white dwarf, which gives a range to the flux of anything between zero and the flux of the entire disc; the mass and thus the radius of the white dwarf which can make a difference of up to one order of magnitude in the predicted flux; the mass accretion rate, which is not known at all for quiescent dwarf novae in particular, only that it clearly is different from the mass transfer rate from the secondary star, and the distance of the object from the observer, which usually also only is known within fairly wide limits.

Considering that observations, in particular of X-rays, are only carried out in certain energy bands and do not cover the entire spectrum, the wavelength dependence of the emitted spectrum also has to be considered when discussing expected fluxes. Qualitative considerations clearly seem to support the view that a) the boundary layer is the origin of the observed X-rays, b) the boundary layer is optically thin for low accretion rates, i.e., for quiescent dwarf novae, and c) it becomes optically thick for high accretion rates (Patterson and Raymond, 1985a; 1985b). In no case, however, have self-consistent computations of the physical structure of the boundary layer been carried out predicting any spectrum. Program codes for two-dimensional hydrodynamic boundary layer computations are being developed by Robertson and Frank (1986), Kley and Hensler (1986), and Kley (1989). The first results from these do not contradict the qualitative concepts reported above, but they do indicate that the boundary layer seems to be much more extended and expansive than assumed so far. Papaloizou and Stanley (1986) carried out computations of the temporal development of the boundary layer and found that quasi-periodic oscillations can originate due to small-scale viscosity instabilities in there — as observed occasionally in nova-like stars and in dwarf novae during outburst (Chapters 2.II.D.2, III.D).

4.V. THE EVOLUTIONARY STATE OF CATACLYSMIC VARIABLES*

4.V.A. SPACE DISTRIBUTION AND SELECTION EFFECTS

ABSTRACT: Various determinations of the space densities of sub-classes of cataclysmic variables depend critically on assumptions about outburst time scales (for novae and recurrent novae) and about observational selection effects.

In order to understand the context of cataclysmic variables in the evolutionary big picture of all stars, it is of great importance to

---

*Cataclysmic variables in this section are restricted to novae, nova-like stars, dwarf novae, and possibly recurrent novae; symbiotic stars are not included in the discussion.
Table 4-2. Determinations of space densities of cataclysmic variables (Downes, 1986).

<table>
<thead>
<tr>
<th>Class</th>
<th>This Work</th>
<th>Previous</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dwarf novae</td>
<td>2.4</td>
<td>3</td>
<td>Patterson 1984</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>Warner 1974</td>
</tr>
<tr>
<td>Novalike</td>
<td>1.5</td>
<td>5</td>
<td>Warner 1974</td>
</tr>
<tr>
<td>(confirmed only)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novalike</td>
<td>4.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(with candidates)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novae</td>
<td>1.4</td>
<td>4</td>
<td>Patterson 1984</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13</td>
<td>Duerbeck 1984</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>Bath and Shaviv 1978</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>Warner 1974</td>
</tr>
</tbody>
</table>

Several attempts have been undertaken to derive the space density of cataclysmic variables (e.g., Kraft, 1965; Warner, 1974b; Bath and Shaviv, 1978; Duerbeck, 1984; Patterson, 1984; Downes, 1986; Ritter and Burkert, 1986) with different results, depending on what assumption is made.

know the space distribution and space density of these objects. Whatever the evolutionary model for cataclysmic variables, their space density and distribution must be compatible with those of supposed progenitors and descendants.
tions were made about the absolute magnitudes of cataclysmic variables and about the completeness of observed samples. An extensive investigation of the space distribution of different sorts of cataclysmic variables and possible progenitors has been carried out by Duerbeck (1984). Some of the resulting distributions are displayed in Figure 4-52. One of the conclusions from this study is that novae, dwarf novae, and possibly supernovae of type I have very similar galactic distributions, which in turn are fairly similar to those of W Ursae Majoris stars and Algol systems. Recurrent novae and symbiotic stars have similar galactic distributions, but these are distinctly different from the former ones. Thus there is support for the view that either W Ursae Majoris stars or Algol systems, or both, might be progenitors of most cataclysmic variables on the basis of their space densities; however, since the angular momentum of W Ursae Majoris stars is too small for the formation of a white dwarf, they can be discarded (Ritter, 1976). Space distributions of dwarf novae are consistent with the hypothesis that they might develop into type I supernovae (see also Chapter 4.V.E).

The shape of the galactic distribution of dwarf novae and novae (see also Figure 1-1) points at their being population I objects. Some novae (T Sco, N Oph 1938) and possibly some dwarf novae are members of globular clusters, however, so they are population II objects (Webbink, 1980).

According to Duerbeck’s study, novae and dwarf novae seem to have almost equal space distributions, but dwarf novae are some 10 times as abundant as novae. On the other hand, Bath and Shaviv (1978) determined the space density of dwarf novae to be 200 times less than that of classical novae; Patterson (1984) derived the density of dwarf novae and nova-like stars together (without AM Herculis stars) to be about as high as that of novae; Warner (1974b) derived about equal densities for both novae and dwarf novae; and Downes (1986) arrived at approximately the same distributions for novae, dwarf novae, and nova-like stars from a survey of UV-excess objects! (Table 4-2). The point of disagreement is the completeness of the observed sample. On one hand this is a reflection of the problem of accurate distance determination (see Chapter 4.II.C.2), which, however, for statistical purposes is probably not too severe in the case of dwarf novae, since all those which are observable are relatively close by. A much more important question is the assumed outburst period of classical novae about which nothing is known at all.

For dwarf novae and nova-like stars alone, Ritter and co-workers (Ritter, 1986a; Ritter and Burkert, 1986; Ritter and Özkan, 1986) investigated the influence of various system parameters on the observable, magnitude-limited, sample. They found that the selection is rather insensitive to assumptions about the assumed galactic distribution function and the interstellar absorption, as well as to the bolometric correction. By far the strongest selection effect was due to the mass-radius relation of white dwarfs, according to which the star’s radius decreases considerably with increasing mass, yielding a larger gravitational potential and thus a brighter disc. The consequence is a strong statistical overrepresentation of the brighter systems containing massive white dwarfs. Finally they found that, since different effects act in different ways on the system’s brightness, selection effects become less severe as the limiting apparent brightness decreases. Owing to these selection effects, they concluded that only a very minor fraction, about 1 out of 200, of the existing cataclysmic variables are actually observed.

4.V.B. THE WHITE DWARFS

White dwarfs in cataclysmic variables seem to have statistically higher masses than single white dwarfs. This difference can be fully accounted for by selection effects.
The determination of the masses of the two stellar components in a cataclysmic system is very difficult and not particularly reliable (Chapter 4.II.C.1). In particular, there are only two relatively highly reliable determinations for the masses of the white dwarfs, based on the analysis of double-lined spectroscopic eclipsing systems: EM Cyg and U Gem. The masses obtained are $0.57 \pm 0.08 \, M_\odot$ for EM Cyg (Stover et al, 1981) and $1.18 \pm 0.15 \, M_\odot$ for U Gem (Stover, 1981a). Less reliable masses have been obtained for several other cataclysmic systems (Ritter, 1987). With the necessary caution, however, this sample may be regarded as statistically representative of the white dwarf masses in cataclysmic variable systems. For the mass distribution of the white dwarfs in sub-classes of cataclysmic variables, see Figure 1-5. There is no evidence for any striking systematic difference in the masses of different sub-classes of cataclysmic variables. The mean white dwarf mass in cataclysmic variables is $0.90 \pm 0.06 \, M_\odot$, whereas that of single white dwarfs is only $0.62 \pm 0.08 \, M_\odot$ (Ritter, 1987). Even if the masses derived for white dwarfs in cataclysmic variables are not very reliable, it seems fairly unlikely that such a severe distortion can be due to poor data, particularly since the two reliable values, for EM Cyg and U Gem, fit into the general pattern.

The white dwarfs in cataclysmic variables, as white dwarfs in general, are believed to be formed from red giants or supergiants which are eventually stripped off their envelopes as will be discussed later. The white dwarf is simply the He or CO core of the evolved star. Since the core mass keeps increasing as long as the star still possesses its shell, the mass of the remaining white dwarf depends on the time when the surrounding shell is lost. Thus, in the geometrically constrained case of the evolution taking place within the Roche lobe, statistically, the masses of the white dwarfs in cataclysmic variables are expected to be even smaller than those of single white dwarfs, since the shell was lost when it exceeded the Roche limit, no matter how little advanced the giant's evolution was at that stage.

Another possible explanation for this large discrepancy is to assume that it is in the nature of cataclysmic variables, or rather their progenitors, that they are formed preferentially with massive white dwarfs. Computations have been carried out by, e.g., Law and Ritter (1983; see also Ritter, 1983) and Livio and Soker (1984a) in order to investigate this question. Both these computations must be seen in the context of the general concept of binary evolution leading to the formation of cataclysmic variables which will be discussed in a later Chapter 4.V.E. Here only the basics, relevant to the present question, will be outlined. Law and Ritter suggest that white dwarfs in close binaries cannot only be formed through what they call case B,* leading to low-mass white dwarfs, and case C,* leading to massive white dwarfs, events but also through what they call case BB. In their case B mass transfer event and common envelope evolution, the primary of an initially wide binary system loses the binary period and thus the size of the Roche lobe has decreased considerably. If conditions are then right, i.e., if the mass of the helium star is not too high ($M_{\text{He}} < 3.4 \, M_\odot$, as a larger mass would eventually lead to a supernova explosion) and the Roche lobe is not too large, a case B mass transfer event may occur, leaving a white dwarf near the Chandrasekhar limit of $1.4 \, M_\odot$. However, estimates of the abundances of possible progenitors for this sort of white dwarf leads to the conclusion that, while they probably do exist, their probability of occurrence is much too small to provide an explanation for the observed discrepancies in the mass spectra of white dwarfs.

---

*Catastrophic mass transfer in a binary system is referred to as case B when it sets in before, and is referred to as case C when it sets in after, core helium burning in the more massive component.
Livio and Soker (1984a) investigate the question whether details of the common envelope phase of binary evolution favor massive white dwarfs in some way. They show that the two stellar components are much more likely to survive the spiralling-in process during common envelope evolution if the envelope is a relatively less dense supergiant envelope — and thus the core is a massive white dwarf — rather than a denser giant envelope. However, Ritter and Burkert (1986) hold against this hypothesis observations of detached post-common-envelope systems whose mean white dwarf mass \((0.62 \pm 0.08 \, M_\odot)\) is not any higher than that of single white dwarfs.

Another possibility that has been suggested is that the mass of the white dwarf might grow secularly due to continued accretion of material from the disc. Ritter (1985; see also Ritter and Burkert, 1986) discusses this possibility in more detail. He rejects it with two arguments: first, a much larger number of type I supernovae would be expected if the accumulation of a significant amount of mass were possible whenever the Chandrasekhar limit for the white dwarf mass was exceeded (see also Nomoto and Sugimoto, 1977); second, analysis of nova ejecta* suggests that the white dwarf loses mass during the course of a nova outburst.

Yet another possibility to explain the two different mass spectra of the white dwarfs is the action of selection effects on the observed sample. This aspect has been dealt with by Livio and Soker (1984b) and Ritter and co-workers (Ritter 1986a; Ritter and Burkert, 1986; Ritter and Özkân, 1986). Livio and Soker estimate that there are \(6.86 \times 10^3\) more outbursts for a \(1.3 \, M_\odot\) white dwarf than for a \(0.6 \, M_\odot\) white dwarf with otherwise identical conditions. Considering this theoretical dependence of outburst time-scales of novae on the white dwarf masses (on which no observational data are available) they conclude that these relatively frequent outbursts in novae containing a massive white dwarf could account for the excess average mass in novae. A little discouraging in this respect are the derived masses of novae which, while mostly higher than the average of \(0.6 \, M_\odot\) for single white dwarfs, are somewhat at the lower end of the distribution in cataclysmic variables (see Figure 2-3a).

A selection effect favoring dwarf novae with high-mass white dwarfs is the mass-radius relation for white dwarfs according to which, as the material a white dwarf consists of is degenerate, the radius shrinks considerably as the mass increases, thus making the respective accretion discs much brighter due to the higher gravitational potential in the hot inner disc. Ritter and co-workers (see above) carried out extensive computations in order to find out the strength of this selection effect, and found that this mass-radius relation for white dwarfs can fully account for the observed mass spectrum of white dwarfs in cataclysmic variables as compared to that of single white dwarfs. Ritter (1986b) investigated a sample of precataclysmic objects which is probably not strongly biased observationally and found that the mean white dwarf mass for these systems is \(0.6 \pm 0.08 \, M_\odot\), in perfect agreement with the results for single white dwarfs.*

---

*One cautioning remark about white dwarf masses in cataclysmic variables should be added. A lot of the discrepancy between single white dwarf masses and those of white dwarfs in cataclysmic variables seems to be due to difficulties in deriving the masses from observations, akin to the problem of determining the radial velocity curve (see Chapter 4.11.C.1). When masses given in the catalogues by Ritter from 1984 and 1987 are compared, the average has shifted down from \(1.04 \pm 0.3 \, M_\odot\) (1984) to \(0.90 \pm 0.06 \, M_\odot\) (1987), largely as a result of a redetermination of white dwarf masses using more elaborate techniques.
4.V.C. THE SECONDARY STARS

RELEVANT OBSERVATIONS: The secondary components are seen in the optical range in systems with orbital periods in excess of some 6 hours; often they are also seen in shorter period systems at infrared wavelengths.

see 78

ABSTRACT: Observationally, the secondary stars cannot be distinguished from main sequence stars. On theoretical grounds, they must originally have been considerably more massive, but lost most of their mass during the common envelope phase and later evolution. After the common envelope phase the stars are drawn together by magnetic braking of the secondary component until this star comes into contact with its Roche lobe. Thereafter, mass transfer is driven primarily by magnetic braking above the period gap, and by gravitational radiation below it. Black dwarf secondaries are expected to exist below the period gap but have not been detected so far; selection effects can account for this.

As in the case of white dwarfs, masses of the secondary stars can only be determined reliably for the two eclipsing double-lined spectroscopic binaries, U Gem and EM Cyg (numerical values for these and other systems are given in Chapter 4.II.C.1). Furthermore, in all those systems which have a known orbital period, the radius of the secondary can be derived from the assumption that the star just fills its critical Roche volume. In this case the mass and the radius of the star are linked only by an almost constant factor the value of which depends mainly on the orbital period and only very weakly on the mass ratio (Faulkner et al, 1972; Patterson, 1984). Furthermore, the spectral types of the secondary stars are known for a couple of systems. The assumption that the radius and spectral type of the secondary ought to correspond to each other puts further limits on the possible mass of the companion star. Finally, dynamical stability of the system requires that the mass of the secondary star must not exceed significantly that of the white dwarf (although at least the case of EM Cyg, with q = 0.77, demonstrates that nature is not quite as restrictive as our theories). From this constraint it immediately follows — and is observationally confirmed — that the secondaries in systems with $P_{\text{orb}} \gtrsim 10$ h must be evolved stars.

Ritter (1983b; 1985; 1986a) derived physical parameters for 13 secondary components using just the above criteria, in order to test the frequently made assumption that the secondary stars in cataclysmic variables are main sequence stars. He compares these results with theoretical and observed (from detached visual and spectroscopic main sequence binaries) mass-radius relations, as well as with the spectral types and mean densities of main sequence stars (Figure 4.53). From these investigations it becomes apparent that, observationally, the secondaries in cataclysmic variables with orbital periods smaller than about 10 hours are very close to the main sequence and may not even be distinguishable from main sequence stars, which in turn implies that they are essentially unevolved. As Ritter (1983b) stresses, it cannot be concluded from this that they are zero-age-main-sequence stars, a point which will be considered more closely in a moment. It might be useful to emphasize, however, that for cataclysmic variables with orbital periods smaller than 10 hours the assumption that the secondaries look like main sequence stars provides an easy, though not highly reliable, method to obtain the masses of both stars. (See Chapter 4.II.B.1).

If the masses of the secondary stars, which are typically smaller and may be much smaller than $1 M_\odot$, are taken at face value, it is not likely that they have evolved significantly during their lifetimes. On the other hand, the progenitors of cataclysmic variables are not yet known with certainty, so not much is known about the initial mass of these stars. It probably can be taken for granted from general evolutionary theory that the mass of the primary star was originally the larger one of the two, and that it evolved faster than its companion and
eventually became a white dwarf. If the secondary star originally was only slightly less massive than the white dwarf, it also should have evolved significantly by the time the common envelope phase of the evolution was initiated by the primary (see below). The mass of the white dwarf primary after this stage has to be less than the Chandrasekhar limit of 1.4 \( M_\odot \), so, if stability against mass transfer* is to be attained, the secondary star also has to lose a certain possibly large fraction of its initial mass. If this happens after the star has had a chance to evolve, the remaining secondary is expected to exhibit decided traces of its evolution. The effect is expected to be the stronger the closer to unity the initial mass ratio is. In fact most secondaries, in particular those in systems with orbital periods smaller than 10 hours, do not seem to be significantly evolved, since, if they resembled normally evolved stars they should be found above the main sequence, and if they were the chemically homogeneous cores of formerly fairly massive stars they should lie below the main sequence (Ritter, 1983b). This situation can only be understood, in terms of the typical critical mass ratios having originally been larger than two, in which case, theoretically, no significant trace of evolution is expected to be present in the remaining secondary star.

The secondary star obviously loses mass continuously to the primary at a rate on the order of \( 10^{-8} \) to \( 10^{-11} \) \( M_\odot/yr \) (see Chapter 4.II.C.4). If the star has a comparatively large mass, or the system a long orbital period, it hardly is affected by this. As the star loses mass, its radius shrinks due to magnetic braking or gravitational radiation (Chapter 4.V.D), and the Roche lobe shrinks as well, so the star loses

---

*Mass loss from the secondary star results in a shrinking of the star's radius, rather than an expansion. The latter would lead to an uncontrolled increase of mass transfer.
still more mass. Mass transfer continues without dramatic consequences, since the thermal time scale for the secondary to adjust to mass loss is much smaller than the time-scale at which gravitational radiation proceeds. If this process continues for long enough, however, eventually the star is driven out of thermal equilibrium and becomes a degenerate black dwarf. As will be discussed in the next section, at the time when the star is out of thermal equilibrium the system reaches a minimum orbital period, which at later stages of the evolution increases again as the secondary continues losing mass.

Rappaport et al (1982) point out that close to the minimum orbital period the secondaries cannot be assumed to follow the mass-radius relation of main-sequence stars any longer, but they are expected to have systematically smaller masses by eventually as much as 20%, after they are driven out of thermal equilibrium. No such effect is apparent in observations (Figure 4.53). However, this might be due to the still relatively long orbital periods of the systems for which the secondaries have been investigated.

For systems containing a degenerate black dwarf secondary, i.e., systems which have evolved through the minimum period, a secular increase of the orbital period is predicted. Such an increase has been observed in Z Cha (Cook and Warner, 1981; for observations see Chapters 2.II.B.5, 3.II.A, 3.III.A, 3.IV.A, and also Chapter 4.III.E). Faulkner and Ritter (1982) investigate whether the assumption of a black dwarf secondary is in agreement with observational properties of Z Cha and find that it is in clear contradiction: the predicted amplitude of the radial velocity is smaller than observed; the radii of both the accretion disc and the white dwarf are smaller than observed (Z Cha is a double-eclipsing system which allows for fairly accurate determination of these parameters); the mass transfer rate of $10^{-13} \ M_\odot/\text{yr}$ inferred from the system's apparent brightness, would bring it close enough to the Sun for it to show a large parallax and a high proper motion, neither of which are observed; and, finally, although the sign of the observed period change is what is to be expected in the case of a black dwarf, the predicted period change on a time-scale of some 10 years is in strict contradiction to the value of $9.6 \pm 10^6$ years claimed by Warner and Cook. By analogous arguments Faulkner and Ritter also exclude the presence of black dwarf secondaries in two other ultra-short-period systems, HT Cas and OY Car.

Rappaport et al (1982) estimated that up to 20% of the ultra-short-period cataclysmic variables (systems with orbital periods shorter than 2 hours) could contain a black dwarf secondary. With such a high postulated percentage it is surprising that not a single such system could be found so far, although selection effects may act strongly against them. When the secondary component is a normal undegenerate main sequence star, the hot spot is eclipsed during the eclipse of the white dwarf at very high inclination angles, rendering the system with a double eclipse. Since the time interval between the two parts of a double eclipse depends on the geometrical size of the secondary star, becoming larger as the secondary becomes smaller, the two parts of the double eclipse should occur with a much larger time interval between them in the case of a black dwarf secondary, or even be separated into two independent eclipses, one immediately following the other (unpublished computations on this matter are reported in Ritter (1983a)). This means that in principle one should be able to identify black dwarf systems from observations of eclipse light curves. Nevertheless, the expected fraction of double eclipsing systems is very low, even among cataclysmic variables with main sequence secondaries. It is even lower by a factor of up to three in systems with a black dwarf, since the probability of occurrence of such an eclipse decreases with increasing mass ratio between the primary and secondary.
star. Furthermore, it is the mass transfer rate which determines the luminosity of an accretion disc; just this, however, is lower by one or two orders of magnitude from a degenerate secondary than from a main sequence star, which renders black dwarf systems intrinsically a lot fainter than others. Taking all these effects into account, only between 0.1 and 3.0 percent of the observed systems in a magnitude-limited sample are expected to contain a black dwarf. Finally, it is possible that the outer disc as well as the hot spot are optically thin, if the low predicted accretion rates are correct. In this case, even if the geometry were right for a double eclipse, only the eclipse of the white dwarf would appear in the light curve, and thus the system would be indistinguishable from other cataclysmic variables.

4.V.D. THE PERIOD GAP, THE MINIMUM PERIOD, AND THE SECULAR EVOLUTION OF CATACLYSMIC VARIABLES

RELEVANT OBSERVATIONS: In the distribution of orbital periods of cataclysmic variable systems, the period gap between some two and three hours and the minimum orbital period of about 80 minutes are outstanding features.

ABSTRACT: The upper end of the period gap is understood to be due to the cessation of magnetic braking which detaches the secondary star from its Roche lobe so that mass overflow stops. Gravitational radiation then brings the star and its Roche lobe back into contact, by which time the period has decreased to some 2 hours. The period keeps decreasing due to gravitational radiation until, at the minimum orbital period, the secondary mass has decreased enough so the star cannot maintain nuclear energy generation any longer. At this point, the star rapidly becomes degenerate (a black dwarf) and with further mass loss the orbital period increases.

Figures 1-2 and 2-2, featuring the distribution of orbital periods of cataclysmic variables, reveal a couple of interesting properties of the systems. The orbital period is the physical property of a cataclysmic system which, if measurable at all, usually is by far the most accurately known of all.

There are mavericks at both ends of the distribution: at the short period end they are the AM Canum Venaticorum stars. They all fail to show any trace of hydrogen in their spectra, and thus are assumed to be degenerate. They will not be dealt with for the moment, but only later in this section.

At the other end there is GK Per, a somewhat peculiar old nova which, in its quiescent state as a nova, exhibits brightness fluctuations reminiscent of dwarf nova outburst behavior. This system is dealt with in detail in Chapter 8. Since the secondary in this system certainly is an evolved star, GK Per also will not be included in the following discussion.

The bulk of cataclysmic variables have orbital periods shorter than 10 hours, and their number rises appreciably toward shorter periods. There is a very characteristic gap between about 2 hours 50 minutes (TU Men) and 2 hours 15 minutes (AR And) in which no object has been found, although at both ends many objects are known. There is also a very sharp cut-off at about 1 hour 16 minutes (AF Cam) below which only objects have been found which do not possess any hydrogen.

The sharpness of the two cut-offs, the period gap and its width, and the minimum orbital period, suggest that they are of real physical significance. Furthermore, the tendency for certain sub-types of cataclysmic variables to appear preferentially at one or another side of the gap (see Chapters 1, 2.1.C) suggests an evolutionary cause of the observed distribution.

*No distinction will be made in the following discussion between the different sub-types of cataclysmic variables.
The most obvious agent of secular evolution in cataclysmic variables is the mass loss from the secondary star, which is bound to cause secular changes in the orbital periods of these systems. An extensive literature has been published on computations and general theoretical aspects of this issue (e.g., Faulkner, 1971; 1976; Taam et al, 1980; Paczynski, 1981; Paczynski and Sienkiewicz, 1981; Verbunt and Zwaan, 1981; D’Antona and Mazzitelli, 1982; Rappaport et al, 1982; Joss and Rappaport, 1983; Paczynski and Sienkiewicz, 1983; Ritter, 1983b; Spruit and Ritter, 1983; Taam, 1983; Patterson, 1984; Verbunt, 1984; Nelson et al, 1985). The main results shall be reviewed briefly in the following.

For Roche-lobe overflow to occur, either the secondary has to expand beyond its Roche lobe or the Roche lobe has to shrink. Expansion of the secondary can occur either through nuclear evolution or through dynamical instabilities in the star’s outer layers. For stars on the main
sequence, mass transfer rates due to nuclear evolution are less than $10^{-12} \, \text{M}_\odot/\text{yr}$ and are thus smaller by at least an order of magnitude than the mass transfer rates inferred from observations (see Figure 4-54 and Patterson, 1948)*. Dynamical instabilities, on the other hand, produce pulse-like events which might explain the outburst behavior of dwarf novae (Chapter 4.III.C), but not the continuous mass transfer at rates of $10^{-11}$ to $10^{-8} \, \text{M}_\odot/\text{yr}$ occurring at quiescent state.

Thus, the more likely cause for Roche-lobe overflow is shrinking of the volume available to the secondary star, which also implies a secular period decrease of the system. Simple conservative mass transfer from the (lower mass) secondary to the (higher mass) primary as well as mass loss from the system both result in a period increase, and thus a growth of the secondary's Roche lobe. The only known mechanism to counteract the above tendency for the period to increase, even leading to a period decrease (i.e., shrinkage of the primary's Roche lobe), is loss of angular momentum from the system (Patterson, 1984). Here again two possible mechanisms are known: gravitational radiation and magnetic braking.

The concept of gravitational radiation is based on the field equations for gravitational forces in general relativity which predict the existence of gravitational waves and corresponding energy loss from every object. It was not clear for a long time whether this radiation could be of any astrophysical importance; estimates for single star evolution suggested that it was not. The situation is different for stars whose geometrical size is determined by factors external to the star itself, like the Roche volume in close binary systems. In this case gravitational radiation can become important enough to control the evolution of the system.

Energy loss from a binary system depends on the entire mass of the stars involved as well as on the orbital period of the system (Faulkner, 1976). The time scale $\tau$ for significant changes in the system due to this form of energy loss is on the order of

$$\tau = 3 \times 10^8 \, \text{p}^{8/3},$$

if the two masses are of roughly comparable size. This means that the typical time scale for changes due to gravitational radiation is shorter than $10^{10}$ years for cataclysmic systems with orbital periods smaller than 8 hours, it becomes considerably shorter than this for very short period systems, and eventually it becomes even shorter than nuclear time scales (Faulkner, 1976). When the secondary's Roche lobe shrinks due to gravitational radiation, the star must lose the excess mass which lies outside the critical Roche surface, preferentially through the zero-gravity point $L_1$, which means into the Roche lobe of the white dwarf. The theoretically inferred mass-loss rate is of the order of $10^{-10}$ to $10^{-11} \, \text{M}_\odot/\text{yr}$ for low-mass main sequence stars, and is nearly independent of the orbital period (Rappaport et al, 1982; Patterson, 1984; see also Faulkner, 1976). If this value is compared with observationally inferred mass-transfer rates, it is clear that for systems with orbital periods in excess of 3 hours, i.e., for systems above the period gap, this mechanism, though probably present, is not strong enough to explain the mass-transfer rates which are one to two orders of magnitude higher. For systems with periods below the gap on the other hand, it provides just about the correct mass-loss rate (see Figure 4-54).

Another mechanism which has been suggested for causing a secular period decrease is magnetic braking. This mechanism is at work in cool main sequence stars ($M \lesssim 1.5 \, \text{M}_\odot$) which possess a convective envelope and a

*For numerical values of mass transfer rates, possibilities or their determination, and their reliability, see Chapter 4.II.C. In the context of this chapter it should be kept in mind that the numerical values are not very accurate, but they are not likely to be seriously in error either.
radiative core; due to solar-like activity these stars lose considerable amounts of angular momentum in their stellar winds which decreases the stars' rotational velocity. In a close binary system, time scales of tidal forces working to synchronize the rotation of the secondary star are significantly shorter than any other time scale of evolutionary significance; and thus, in practice, the secondary always rotates synchronously with the orbital motion no matter how much angular momentum is carried away by stellar wind. This in turn leads to an increase of the stellar wind and drains even more angular momentum from the system, thus forcing the two stars closer together and the Roche lobe of the secondary to shrink (Huang, 1966; Mestel, 1968; Eggleton, 1976; Verbunt and Zwaan, 1981; Patterson, 1984; see also Paczynski, 1985). To estimate mass-loss rates to the primary's Roche lobe from this scenario is very difficult since the secondaries in cataclysmic variables are rotating much faster (at an almost constant velocity of some 130 km/sec for all systems according to Patterson (1984)) than any known single stars of similar spectral type due to forced synchronism. Assuming that gravitational radiation plus magnetic braking are at work above the period gap and that only gravitational radiation is at work below, since these stars are fully convective, Patterson tries to derive mass transfer rates and arrives at values well in agreement with the observationally derived values (Figure 4-55).

A different approach has been undertaken by Spruit and Ritter (1983). They point out that just at the stage when the stars are assumed to become fully convective, a strong decrease in indicators of magnetic activity (Ca II and H α emission, and X-rays) is observed, pointing to break down of the magnetic field — which also is expected on theoretical grounds when the interface between the radiative core and the convective envelope no longer exists as the star becomes fully convective at an orbital period of about three hours. Since the star is significantly out of thermal equilibrium due to former excessive loss of angular momentum (see above), it detaches from its Roche lobe and shrinks until it reaches its equilibrium radius. Even if no magnetic braking is acting any longer, the Roche lobe keeps decreasing due to gravitational radiation until, eventually, it again meets the star's surface (Figure 4-56). After this, mass transfer starts again, proceeding at the rate of some 10^{-10} to 10^{-11} M_☉/yr due to gravitational radiation. Choosing appropriate parameters, Spruit and Ritter are able to

Figure 4-56. Evolution of a cataclysmic variable into and out of the period gap: (a) the relative angular momentum loss drops as soon as magnetic braking ceases to be at work; (b) the mass transfer rate is high above the gap, drops to zero in the gap as the secondary becomes detached from the Roche surface, and is resumed at a small rate below that when gravitational radiation has brought the star into contact with the Roche surface again; (c) corresponding changes in the star's radius (solid line), its equilibrium radius (dashed lines, and the size of the Roche surface (dash-dotted line) (Spruit, Ritter, 1983).
Figure 4-57. Evolution of the secondary star below the period gap under the influence of gravitational radiation: (a) A minimum orbital period of the system is reached as the star becomes degenerate (Ritter, 1983 after Paczynski and Sienkiewicz, 1981). The exact value of the minimum orbital period depends on the total mass of the system (as indicated in the figure); (b) for solar composition it always is on the order of 80 minutes (Paczynski and Sienkiewicz, 1981); (c) when the chemical abundance is changed, the minimum orbital period changes as well (Sienkiewicz, 1985).

magnetically active stars possess many star spots which disappear together with the magnetic activity; this alone can produce a gap like the one observed, even if the star remains in thermal equilibrium all the time.

Whichever of these above two scenarios is applied to explain the period gap, results still depend strongly on the assumed initial conditions and parameter values, and clearly considerable theoretical work remains to be done.

For systems below the period gap there is theoretical agreement that mass transfer is driven by gravitational radiation alone. The orbital period keeps decreasing as the secondary loses ever more mass. When the star’s mass approaches some 0.1 \( M_\odot \), corresponding to an orbital period of some 120 min, the thermal time-scale becomes comparable to the time-scale of gravitational radiation. The star cannot adjust to the permanent mass loss any longer and is driven out of thermal equilibrium. This occurs when its radius becomes distinctly larger than the radius of a main sequence star having the same mass (see above). The continued mass loss causes the nuclear burning in the core to be extinguished. Since there is no longer any support by nuclear generated thermal pressure against gravitation contracting the star, the star cools off and collapses. As a consequence it becomes degenerate (Rappaport et al, 1982).

Both the period gap and the strong cut-off at a minimum orbital period still need to be explained. In principle there are three possible ways for producing a period gap: either, a) for some reason systems are not formed in the gap and also do not eventually evolve into it, or b) they move through the gap very quickly so the detection probability decreases strongly, or c) for a while the system is a detached binary, mass transfer stops, and thus they are not recognizable as cataclysmic variables. Patterson (1984) and Verbunt (1984) summarize and discuss previous, unsuccessful, attempts to explain the gap; these are not reviewed here.
Currently, two possible additional scenarios are under consideration. The basic idea of one of these concepts is (D'Antona and Mazzitelli, 1982; Joss and Rappaport, 1983) that, as the secondary star loses matter and its internal structure tries to adjust to the ever new conditions (staying close to the main sequence, nevertheless), the convection zone penetrates deeper and deeper into the star's interior while the star's central temperature decreases. Since new He 3 is mixed into the core from the outer, chemically unconsumed layers of the star, the energy production increases conspicuously just as the star becomes fully convective at an orbital period of some three hours. The increased energy production leads to a temporarily enhanced mass transfer, hence stronger mass loss from the star and also a fall of its internal temperature which in turn means nuclear energy generation becomes less effective. The system becomes detached, as nuclear energy generation becomes too inefficient and the stellar radius shrinks below the Roche radius. Only when gravitational radiation has reduced the Roche lobe strongly enough for it to reach once more the surface of the secondary star does mass transfer begin anew.

The orbital period of a binary system depends to first order only on the radius and thus on the mass of the secondary star. For a non-degenerate star the radius shrinks as it loses mass, whereas the situation is opposite in the case of degenerate material. Due to this the binary period reaches a minimum — the exact value of which depends on the total mass of the system and its chemical composition — just at the time when the secondary star becomes degenerate, after which it increases as the secondary loses even more mass (Figure 4-57a). Paczynski and Sienkiewicz (1981) and Sienkiewicz (1985) investigated the theoretical value of the minimum orbital period as a function of the total mass of the system ($M_1 + M_2$) as well as of the chemical composition. They find that, quite independent of the entire mass, the minimum orbital period for hydrogen-rich systems for which they assume $X = 0.70$, $Y = 0.27$, $Z = 0.03$ is $81 \pm 6$ minutes (Figure 4-57b), while for hydrogen-poor systems with mass fractions down to 0.01 the minimum period can be as short as 40 minutes (Figure 4-57c). Thus the period of GP Com, 46.5 minutes, can be explained in this way, while the periods of PG1346 + 182 (24.8 min) and AM CVn (17.5 min, but somewhat questionable) require another, as yet unknown, mechanism. Eventually the mass will become too low for the star to be degenerate. From this moment on the period will decrease again as the star loses mass, and only gravitational radiation will be at work until the mass of the star will be almost entirely lost. What little mass will be left is probably just a rapidly spinning white dwarf (Nather, 1985).

4.V.E. PROGENITORS AND DESCENDANTS

Abstract: In general, the progenitors of cataclysmic variables must be wide binaries with orbital periods of months or years. No known class of stars has been identified with them so far. Alternatively, a small fraction may be formed by extensive mass accretion onto a planet from a star during a common envelope phase. At the end of the lifetime of a cataclysmic variable, mass overflow from the secondary star continues until this mass is entirely transferred. A single white dwarf then remains.

All cataclysmic variables contain a white dwarf primary which can only have evolved from the core of a red giant or supergiant, thus from a progenitor whose radius was much larger than the dimensions of a cataclysmic system. Since most observed white dwarfs in cataclysmic variables are rather massive ($M_{\text{WD}} \gtrsim 0.45 \, M_\odot$), they must have originated through case C mass transfer which implies a large angular momentum of the original system, most of which (up to 95%) must have been lost during the evolution toward a cataclysmic variable (Ritter, 1976).

This is about all that can be concluded about the progenitors from inspection of existing
observed systems. Several attempts have been made to at least qualitatively model stellar evolution before cataclysmic systems emerge (e.g., Paczynski, 1976; Ritter, 1976; Webbink, 1979; Law and Ritter, 1983; Paczynski, 1985). Although the details are by no means well understood theoretically, there is general agreement about the following overall scenario:

The progenitors must be fairly wide binaries with orbital periods of many months or years, in order to later be able to accommodate a red giant or supergiant in the primary’s Roche lobe, which in turn is needed in order to eventually produce a white dwarf. W Ursae Majoris stars can be excluded as possible candidates, even though their space density and distribution well match those of cataclysmic variables (Duerbeck, 1984, and see Chapter 4.V.A), since their angular momentum is so small that no white dwarf could be formed (Ritter, 1976). No other group of stars has been identified as a possible class of progenitors.

Whatever the progenitors are, the two stars must evolve independently for most of their lives on time-scales dictated by their masses. The more massive star eventually becomes a red giant, harboring a future white dwarf in its core. It will expand and fill its critical volume as either a giant or a supergiant, depending on the size of the Roche lobe, and, correspondingly, case B or case C Roche lobe overflow will take place onto the still rather unevolved secondary star, unless $q = 1$ (Kippenhahn and Weigert, 1967). Initially, the mass transfer onto the secondary star is very violent with a mass-loss rate as high as 0.1 $M_\odot$/yr (Webbink, 1979). The secondary star clearly is not able to adjust to this flood of material and possibly as soon as within a few orbital periods (Paczynski, 1976), blows up to become a red giant-like object. Thus both stars together find themselves wrapped in a common envelope. Angular momentum is transferred to the envelope, and the two cores are drawn ever closer together while the envelope expands. It is not clear theoretically how this binary system and the envelope finally become detached from each other. If this scenario is qualitatively correct, one would expect to observe a close binary system consisting of a white dwarf and a main sequence star with an orbital period of less than 2 days to occur inside a planetary nebula.

Observational support for this scenario came from the detection of several close binary systems consisting of a white dwarf and a main sequence star, ten of which are central stars of planetary nebulae (Ritter, 1986b). In particular the planetary nebulae must be very young since they become invisible some $10^4$ years after their ejection. These so-called pre-cataclysmic binaries (or V471 Tauri stars) are considered the immediate progenitors of cataclysmic variables (Patterson, 1984; Bond, 1985; Ritter, 1986b).

They are not quite cataclysmic systems yet, however, since the orbital periods are still too long for the secondary components to fill their Roche lobes. Again, the known mechanisms for angular momentum loss which are efficient enough to act on relevant time-scales are gravitational radiation and magnetic braking or expansion of the secondary on nuclear time-scales. Gravitational radiation is at work in all binary systems. It has been mentioned above, however, that contraction of these systems due to gravitational radiation takes a very long time for long orbital periods, so that systems for which only this mechanism applies are not likely to become cataclysmic variables within the lifetime of the Galaxy. Magnetic braking, on the other hand, only works within fairly narrow limits of physical conditions, namely, a differentially rotating secondary which possesses a convective envelope, implying that it must be a main-sequence star with a mass between roughly 0.3 and 1.3 solar masses in a sufficiently close binary system which ensures tidal interactions to be able to force the stars to co-rotate. Bond (1985) points out that, given all these constraints, cataclysmic variables may well be only a tiny fraction of possible end-products of common envelope evolution.
Evolution toward and through the life of a cataclysmic variable was already outlined above (Chapter 4.V.D): magnetic braking eventually brings the period down to some three hours; when this mechanism brakes down, the secondary detaches from its Roche lobe and mass transfer stops. All further evolution is dominated by gravitational radiation bringing the stars ever closer together until mass overflow resumes again. The orbital period keeps decreasing until a minimum value is reached when the secondary becomes degenerate. After this it increases again until so much of the secondary’s mass is transferred that it is not even degenerate any longer. The period then decreases again as further mass is lost, until all of this star has finally been cannibalized by its white dwarf neighbor.

On the way to this stage, however, further obstacles can be met. If the white dwarf should succeed in crossing the Chandrasekhar limit by accreting enough mass (that this can happen theoretically is by no means obvious — see Chapter 4.V.B), the white dwarf will undergo a supernova explosion of type I which may or may not destroy the secondary; at any rate such an event would seriously influence the further fate of the system. Also, it is not clear theoretically what repeated nova explosions, if they occur at all, might do to a system in terms of angular momentum balance and white dwarf masses.

As mentioned earlier (Chapter 4.V.D), AM Canum Venaticorum stars, i.e., systems containing two degenerate stars orbiting around each other with periods of less than 40 minutes, cannot be explained by the above scenario, since the minimum orbital period from gravitational radiation is larger than the orbital period of two of these systems. A possible explanation for the existence of such objects was given by Livio and Soker (1983, 1984b; Livio, 1983 --- see also Rappaport et al, 1982). They start out from a star-planet system. The star evolves and becomes a red giant. The planet is exposed first to a strong stellar wind, later it is embedded in the envelope and accretes matter, and may, if conditions are right, finally become a low-mass star of order 0.14 M⊙. Spiralling-in diminishes the distance between the planet/star and the white dwarf core of the giant. It depends on the initial conditions what the end product of this evolution will be. If the initial separation is too small (smaller than some 500 R⊙), the planet will spiral into the star before the star had lost its envelope and only the star is left, if it is too large (larger than some 2000 R⊙) the planet will not be able to accrete a significant amount of mass and the separation between the two cores will not be reduced. Concerning masses, if the initial mass was too small (smaller than 0.01 M⊙) the planet will evaporate during the spiralling-in phase; only if the mass is larger than some 0.0125 M⊙ will a low-mass star result, of some 0.14 M⊙, whose mass will be determined almost entirely by the mass contained in the giant’s envelope.

4.V.F. NOVAE — DWARF NOVAE — AND NOVA-LIKE STARS

RELEVANT OBSERVATIONS: Except for their outburst behavior and its immediate consequences, novae, dwarf novae, and nova-like stars cannot be physically distinguished from each other.

ABSTRACT: It is suggested that all systems cyclically pass through all these stages of activity.

In all of the above discussion no reference was made to the type of cataclysmic variable (novae, dwarf novae, or nova-like star) considered. In fact, observationally, no striking statistical differences can be found between the members of the various sub-classes of cataclysmic variables except, it seems, for the outburst activity. This may be due partly to difficulties in determining system parameters reasonably accurately (see Chapter 4.II.C). It is certainly still an open question why some objects appear as novae, others as dwarf novae, and even others as different types of nova-like
systems. Probably the only clear cases are AM Canum Venaticorum systems which contain two degenerate stars, and AM Herculis stars which are governed by the magnetic field of the white dwarfs (these two kinds of systems are not considered here).

Since system parameters of essentially all cataclysmic variables seem to be approximately identical, the general feeling is that all these systems may really be effectively identical, and just be seen at different stages of a cyclic evolution. A nova explosion is expected to occur as soon as a sufficient amount of hydrogen-rich material has accumulated on the surface of the degenerate white dwarf. The necessary material is assumed to come from the secondary star via the accretion disc. Furthermore, novae by definition only have been observed to erupt once, and so far there has been no observationally imposed reason to change this definition.* Of course this does not mean that all novae erupt only once, but only that the outburst interval is in excess of a couple of hundred years. Since a nova outburst obviously is not very traumatic for a cataclysmic system, such an event well may happen to the system many times.

There are a few old novae which occasionally exhibit small-scale brightness fluctuations fairly reminiscent of dwarf nova outbursts (see, e.g., Livio, 1987): The most prominent example is GK Per, discussed extensively elsewhere in this book. In a tentative scenario, U Geminorum variables, Z Camelopardalis stars, anti-dwarf novae, and finally UX Ursae Majoris stars all can be linked comfortably to stages in an evolutionary sequence with an increasing fraction of time spent in the high/outburst state.

Following this general idea, Vogt (1982b, see also, e.g., Shara et al, 1986) tentatively suggested a cyclic behavior in which a nova explosion blows away the entire accretion disc which then eventually is built up again: first the system starts to exhibit dwarf nova-like behavior of an old nova; then it becomes a U Geminorum star; then, as the torus/disc mass increases it will be a Z Camelopardalis star, and later an UX Ursae Majoris star, until enough hydrogen-rich material has accumulated on the surface of the white dwarf for it to undergo another nova explosion. There are observational as well as theoretical problems with this scenario. Observationally, there is no evidence for quiescent novae to be physically significantly different from other cataclysmic variables at least for the first couple of decades after eruption. In particular there is every indication that they all possess accretion discs. Theoretically, the mass-transfer rates of typically some $10^{-9}$ $M_\odot/yr$ in UX Ursae Majoris stars are about an order of magnitude too high for thermonuclear reactions to occur on the surface of the white dwarfs (Paczynski, 1985).

The mass-transfer rates derived for quiescent novae are of about the same order, some $10^{-9}$ $M_\odot/yr$ or even higher than those derived for UX Ursae Majoris stars. This poses the problem that in novae new nova outbursts are not expected to occur. On the other hand the general feeling is that the nova phenomenon is likely to be a recurrent one. To propose a way out, Shara et al (1986) and Livio (1987) suggested that after the nova outburst high mass-transfer rates are maintained for some 50 to 300 years by irradiation of the secondary by the still hot white dwarf. The two stars are driven somewhat apart by the outburst. (A period decrease after, as compared to before, the outburst has been claimed for the old nova BT Mon, observed in 1939 (Shaefer and Patterson, 1983; Livio, 1987). Thus as the white dwarf cools and is ever less able to heat the secondary,

---

*Some recurrent novae have been observed. There is plenty of support, however, for the assumption that they all contain evolved secondaries which might have played a part in their evolution. At any rate, they can comfortably be excluded from considerations concerning the bulk of cataclysmic variables by stating that they are not typical representatives of this class, and in fact some of them, like T CrB, seem almost indistinguishable from symbiotic systems — see Chapters 9 and 11.
this star will shrink to its normal dimensions which, due to the now somewhat enlarged Roche lobe, will underfill its critical volume and thus mass transfer will either be strongly reduced or be stopped altogether (Livio calls this state "hibernation"), until, by magnetic braking or by gravitational radiation, the system is brought into contact again and dwarf nova activity is resumed until the next nova outburst occurs. Observational support for this is an observed decrease in luminosity in the old nova RR Pic (Nova Pic 1925) as well as very low mass-transfer rates for the two very old novae CK Vul (1670) and WY Sge (1783).

At any rate, concerning this aspect of understanding cataclysmic variables, considerable theoretical work remains to be done.
5

SUMMARY

In the preceding chapters an attempt has been made to convey a picture of the current state of research in dwarf novae and nova-like stars. Chapters 2 and 3 summarize the observational appearance of dwarf novae and nova-like stars, respectively. The aim in these chapters was to present the data independent of any interpretation, while at the same time to point out implications these observations have for any model. In order to facilitate understanding, in particular for newcomers to the field, a "general interpretation" of the main observed features within the framework of the Roche model was given at the end of each major section. Finally in Chapter 4 the Roche model for cataclysmic variables and the main streams of current theoretical work in dwarf novae and nova-like stars were presented with abstracts of the relevant observations for the models preceding major theoretical sections. In addition, a system of cross-references should enable easy comparison between similar features in different sub-classes of dwarf novae and nova-like stars.

The emerging picture is that, in gross features and in most respects, dwarf novae and nova-like stars, as well as quiescent novae are almost indistinguishable. Nevertheless, in addition to their different outburst behaviors, there appear to be some further differences between dwarf novae and nova-like stars such as:

- a tendency for the Balmer emission lines of hydrogen to have larger equivalent widths in dwarf novae than in nova-like stars (see Chapter 2.III.B.1.a);
- in dwarf novae Hβ is often of comparable strength to, or stronger than, Hα, while in nova-like stars Hα is normally the stronger line (see Chapter 2.III.B.1.a);
- in some nova-like systems the changes between high and low brightness states are considerably more pronounced at optical wavelengths and in the IR than in the UV, while in dwarf novae it always is the UV that is most strongly affected; other nova-like stars behave in the same way as dwarf novae (see Chapters 2.III.A, 3.III.B, 3.V.C);
- in several eclipsing nova-like stars the orbital hump appears at some times before, at others after eclipse, while in dwarf novae it always appears before eclipse (see Chapters 2.II.B, 3.II.A.1, 3.IV.B);
- in dwarf novae the hump amplitude always seems to be strongest in the optical, decreases in intensity toward shorter wavelengths, and is practically absent in the UV; in some nova-like stars, however, it was seen to be strong in the optical, decreases around the U filter, and increases again shortward of this (see Chapters 2.II.B, 3.II.B, 3.III.A.2).

The reasons for these differences are not yet understood. In fact, hardly any attention has been paid to them so far. Commonly the understanding is that nova-like stars can be regarded as dwarf novae in a permanent state of outburst. There is evidence, though, that the differences between these two classes of stars go beyond this view.
In general it is difficult to make statements about "the" behavior of nova-like stars since this class is far from being a homogeneous one. The definition of a nova-like star is that, except for the outburst behavior, it exhibits all the photometric and spectroscopic characteristics of a dwarf nova; and then, as was discussed in Chapter 3, the whole class is divided again into various sub-classes according to certain observed properties.

Some comments about this classification should be made. As was pointed out in earlier chapters, the distinction between dwarf novae, novae, and nova-like stars is by no means as clear as it may seem at first glance. As examples of this, systems like EX Hya and AE Aqr, which are classified primarily as nova-like stars of sub-type DQ Herculis, occasionally are regarded instead as dwarf novae; and the system WZ Sge clearly is to be placed somewhere intermediate between dwarf novae of sub-type SU Ursae Majoris and recurrent novae. Furthermore WZ Sge clearly exhibits (or rather exhibited, since no investigations of this have been published for times after the last outburst in 1978) strictly coherent, highly stable oscillations which would qualify it as a DQ Herculis stars, and thus a nova-like variable. Finally, DQ Her itself is regarded as both an old nova as well as a nova-like system. This list can be extended, and for many objects, it turns out, classification is a matter of taste.

It also becomes apparent that problems of classification are encountered most often with the nova-like stars of sub-type DQ Herculis. This group literally comprises objects of all other classes of cataclysmic variables (see also Ritter, 1987). The defining characteristic of DQ Herculis variables is the occurrence of oscillations which, in any object, are present with the same periods at all times. From the problems over classification, it seems likely that these highly coherent oscillations are a rather common property of cataclysmic variables as a whole, quite independent of outburst properties; consequently their observation is probably a very unfortunate basis for classification... as long as the classification scheme is based primarily on the properties of the long-term light curves.

Another difficulty with the currently used classification scheme exists in the distinction between UX Ursae Majoris stars and anti-dwarf novae, since this requires a long and complete record of observations. Most of the time the objects in both classes cannot be distinguished from each other. However, at unpredictable times, anti-dwarf novae suffer a drop in brightness by 2 to 5 magnitudes, from which, after weeks or months, they recover to the normal "high" state. The first indication for such behavior was found in 1980 in the object TT Ari (Krautter et al., 1981a). Since then, a considerable number of (former) UX Ursae Majoris variables were reclassified into antidwarf novae, based on either direct observations or from inspection of old photographic plates. It may well be that all UX Ursae Majoris stars suffer the fate of an anti-dwarf nova from time to time, and thus both classes are really identical. The observational proof of this might be merely a matter of time.

Finally there exists a problem in the classification of dwarf novae in that the distinction between U Geminorum stars, on one hand, and SU Ursae Majoris and Z Camelopardalis stars, on the other, also requires rather extended, and in the case of SU Ursae Majoris stars also rather thorough, observations. By definition, all systems that are not clearly SU Ursae Majoris or Z Camelopardalis stars are referred to as U Geminorum stars. Z Camelopardalis stars are defined by occasionally undergoing more or less extended standstills, a behavior which can be detected simply from long-term monitoring. The occurrence of standstills seems to be restricted to systems with short outburst periods, so it becomes clear which U Geminorum stars are candidates for being thus far unrecognized Z Camelopardalis stars, and these can be monitored more closely. Certainly the classification should be clear by now for the...
long-known systems. Amazingly the exact classification of the dwarf nova CN Ori, which first was detected in 1906 (Wolf and Wolf, 1906), is still a matter of debate: it is commonly referred to as a Z Camelopardalis star, although except for one hand-drawn light-curve in a notoriously unreliable source (Glasby, 1968, see also Pringle and Verbunt, 1986) there is no evidence that CN Ori was ever observed to undergo a standstill.

The occurrence of superoutbursts seems to be restricted primarily to dwarf novae with orbital periods below 2 hours (i.e., below the period gap), although TU Men is proof that SU Ursae Majoris systems can also be found at least just above the gap. In any event, short period dwarf novae are clearly candidates for being SU Ursae Majoris systems, and, again, the extensive migration from being classified as U Geminorum to SU Ursae Majoris stars indicates that possibly all short-period dwarf novae are SU Ursae Majoris stars. Accordingly, thorough investigation of the (so far) U Geminorum stars below the period gap eventually might be able to tell whether or not this is really the case.

In light of all these problems with classification, clearly the question arises whether the outburst behavior, which currently is the basis of almost all classification, is really a suitable criterion for sorting cataclysmic variable into physically related groups. The two remaining classes of nova-like systems, the definitions of which are based on something other than outburst activity, are the AM Herculis stars with their strong polarization, and the AM Canum Venaticorum stars in the spectra of which no hydrogen can be found. These two classes are clearly and unambiguously defined. Possibly, from a physical point of view, DQ Herculis stars are a well-defined class. . . . or maybe the occurrence of stable oscillation is such a common property that again it is not suitable. Nevertheless, it seems that, as more detailed observations of cataclysmic variables become available, the more pressing becomes the need to revise their classification scheme, a scheme which, after all, was designed when nothing but the outbursts of cataclysmic variables could be observed.

In spite of all these difficulties, dwarf novae and nova-like systems are statistically almost identical in most of their properties, and theoretically they can be, and have been treated together, without — at least at the current level of theoretical understanding — the need for a more sophisticated distinction arising. To a large extent this also applies to the modeling of novae in the quiescent state (i.e., a long time after outburst).

All dwarf novae, novae, and nova-like stars that were investigated thoroughly turned out to be short-period binary systems, and there is no contradiction to the hypothesis that all of these objects are indeed binaries. The theoretical foundation of all modeling these days is the Roche model (Chapter 4.II.A), and indeed, within this framework, a surprisingly large fraction of the observations can be explained. In particular, all of the gross features can be accounted for. Thus clearly it makes sense to retain this general model for the time being. However, agreement between theory and observations is so far still restricted to these gross features, and our current level of understanding is far from inspiring hope that we will soon understand the nature of cataclysmic variables in detail.
REFERENCES


PRECEEDING PAGE BLANK NOT FILMED


Beljawsky, S. 1933, *Pulk. Circ.*, 9, 34


Brunt, C. C. 1982, Ph.D. thes., Cambridge


Campbell, L. 1934, *Harv. Ann.*, 90, 93


Chanmugam, G. 1986, in *Plasma Penetration into Magnetospheres*, N. Vyiylatis, J. Papamastorakis, J. Ventura (eds), Crete Univ. Press, P. 153


Kholopov, P.N., and Efremov, Y.N. 1976, *Var. Stars*, 20, 277


Kraft, R.P. 1956, *Carnegy Yearbook*

Kraft, R.P. 1962a, *Adv. in Astr. and Astroph.*, 2, 43


Livio, M. 1987, *Comments in Astrophys.*, 12, 87


Mantel, K.-H. 1986, *Diplomarbeit*, Univ. Muenchen


Mayall, M.W. 1972, AAVSO Report, 29


Meyer-Hofmeister, E. 1987, Astron. Astrophys., 175, 113


Miczaika, G.R. 1934, Zeitschr. f. Astr., 8, 292


Mineshige, S. 1988, Astron. Astrophys., 190, 72


Müller, G., and Hartwig, E. 1918, Geschichte und Literatur des Lichtwechsels der bis Ende 1915 als sicher veraenderlich erkannten Sterne nebst einem Katalog der Elemente ihres Lichtwechsels, Poeschel and Trepte, Leipzig


Mumford, G.S. 1962, *Sky and Tel.*, 23, 71


Parenago, P., and Kukarkin, B. 1934, *Variable Stars*, 4, 249


Petit, M. 1958, *Ciel et Terre*, 74, 195


250


Sienkiewicz, R. 1984, Acta Astr., 34, 325


Smak, J. 1971, Acta Astr., 21, 15


Smak, J. 1979, Acta Astr., 29, 325


Smak, J. 1984, Acta Astr., 34, 161


Szkody, P. 1985b, in Proc. ESA Workshop: Recent Results on Cataclysmic Variables, Bamberg, FRG, 17–19 April, 1985, ESA SP-236, p. 39


van der Bilt, J. 1908, *Recherches Astronomiques de l'Observatoire d'Utrecht III*


Vogt, N. 1981, Habilitationsschrift, Univ. Bochum
Vogt, N. 1983c, Sterne und Weltraum, 22, 588
Voikhanskaya, N.F. 1980, Sov. Astr., 24, 300
Voikhanskaya, N.F. 1983a, Sov. Astr., 27, 5
Voikhanskaya, N.F. 1983b, Sov. Astr., 27, 670
Wade, R.A. 1979, Astronom. J., 84, 562
Walker, M.F. 1957, in IAU Symp. 3, Non-Stable Stars, held in Dublin, Ireland, 1 Sept., 1955, G.H. Herbig (ed.), Cambridge Univ. Press, p. 46


Warner, B., and van Citters, G.W. 1974, Obs, 94, 116


PART II

CLASSICAL NOVAE AND RECURRENT NOVAE

Written by
Antonio Bianchini
Hilmar Duerbeck
Michael Friedjung
Margherita Hack
Pierluigi Selvelli