The viscous circle: the first steps toward modeling Galactic binary X-ray sources

S. N. Shore

Dipartimento di Fisica “Enrico Fermi”, Università di Pisa, largo B. Pontecorvo 3, Pisa 56127, Italy
e-mail: shore@df.unipi.it

Although Galactic X-ray sources were first identified (Giacconi et al. 1971) in the decade preceding the paper by Pringle & Rees (1972, hereafter PR72), at the time of its publication there was a vigorous debate about their explanation. The models divided into two broad categories, those that drew analogies with active galactic nuclei and those based on cataclysmic variables. The confusion in this early period was remarkable. Energetic phenomena have generally been associated with strong radio emission so most attempts to explain the X-ray emission were linked to synchrotron emission and the presence of strong magnetic fields (e.g. Manley 1966); no suggestion was too extreme, e.g. the pulsar-inspired cocoon model proposed by Davidson et al. (1971). Classical novae were already understood to be extremely short-period (≤ 1 day) binary systems containing a compact object (generally thought to be a white dwarf, Kraft 1964). Inspired by such observations, Shklovsky (1967) proposed an “old nova” model, founded on known cataclysmics in which a stream from a normal, though low-mass, companion filling its Roche surface supplies the material to a compact object, in this case a neutron star (to reach the appropriate temperature regime for the emission). He did not treat the transport or accretion process in any detail, focusing instead on the emissivity and temperature derived from different emission lines observed in Sco X-1 and citing the conditions in WZ Sge as observational support for the system parameters. The connection with X-ray systems was hinted at by Mumford (1967) in his general review of binary models for cataclysmic variables.

The discovery of rapid flickering in several sources – Sco X-1 (Sandage et al. 1966; Lawrence et al. 1967) and Cyg X-2 (Kristian et al. 1967) – with timescales of only a few hundred seconds or less suggested gravitational powering on a compact scale, much smaller than the size of the binary system, and appeared to rule out any simple comparison with quasars or other active galactic nuclei. But the fundamental clue came with the observation of radial velocity variations, although not periodic motion, for one of the brightest sources, Cyg X-2, by Burbidge et al. (1967). Prendergast & Burbidge (1968, hereafter PB68), citing Cyg X-2 as their exemplar X-ray binary system, proposed that a viscous stress (using an unspecified mixing length model and explicitly neglecting compressive viscosity) in a Keplerian disk provides the torque and heating in the midplane that is then radiatively transferred to the surface of the disk for which the scale height is determined by local hydrostatic support with the projection of the gravitational acceleration from the central star (including both scattering and absorption opacities). They obtained power-law radial dependences for the disk temperature and the flux emitted but did not compute a spectrum. The models proposed for Galactic X-ray sources by Shvartzman (1971) included disk accretion almost as an afterthought; the picture he discussed was accretion from the interstellar medium, but the last part of the paper discusses both accretion disks and possible supply of matter from a companion instead of a diffuse environment.

Then within a single year – 1972-3 – three complementary papers proposed the same solution, that the high energy emission can be explained by a viscous accretion disk around a compact object, a neutron star or black hole in a close binary system: PR72 in A&A, Novikov & Thorne (1973, in the Les Houches proceedings from the previous year and specifically addressing the relativistic problem in the context of accreting black holes), and Shakura & Sunyaev (1973, SS73) in A&A. To place PR72 in context, it is important to note the discussions of binary systems that had developed during the previous decade. Although the role of the Roche surface had been cited as early as 1943 by Kuiper as an explanation for the Algol paradox (that the most evolved star in the β Per system is the least massive), not until the mid-1960s was it possible to start including such effects in evolutionary codes. But that was for normal stars since the main cases being considered were those leading to Algol-like systems in which the evolved star is a Roche lobe-filling giant with a main sequence companion. Instead, at the stage of white dwarf evolution, little was known. Although core collapse supernovae produce neutron stars, arguments had been presented

1 Notably, he states that “The preliminary evidence would seem to suggest that the optical counterparts of Sco X-1 and Cyg X-2, while somewhat similar to novalike variables, may actually represent a new class of objects”.

2 They also noted that such systems can be sources of gravitational radiation, already suggested for WZ Sge by Kraft et al. (1962), which is not cited.

3 This paper is mentioned in a note added in proof in PR72 as a preprint that reaches the same conclusions, see Andrew King’s commentary and the SS73 paper, this volume.

4 The binary scenarios were outlined in the first comprehensive review by Paczynski (1971) that included the effects of accretion on compact objects.
that binaries would disrupt if more than half the mass were lost in less than an orbital period and, for low-mass companions or close systems, this seemed virtually certain. Nonetheless, white dwarfs might accrete enough mass to undergo accretion-induced collapse, massive stars might form black holes, and there were X-ray emitting neutron stars in compact binary systems (see Schreier et al. (1972) for Cen X-3, especially eclipses).

There are several important simplifying assumptions in the PR72 model. They used a thin disk approximation with the scale height $b$ derived assuming Keplerian motion despite possible viscous and magnetic effects. The radial drift rate was estimated to be a fraction of the local Keplerian velocity, $v_{\text{K}}$, following the prescription of Lynden-Bell (1967, LB67), a scenario for the role of induced magnetic stresses in a differentially rotating, thin, non-self-gravitating flow to feed Schwarzschild black holes in AGNs where the material was supplied by the environment via cloud accretion under rather general conditions. In contrast, and this was a particularly important feature of PR72, the flow was a stream lost from a "normal" companion that fills its Roche surface (although the Roche radius and system parameters are left unspecified). Classical novae are not treated. Instead, the central object is assumed to be either a neutron star or black hole, and they cite the eclipses in Cen X-3 and the proposed black hole in Cyg X-1 (e.g. Bolton 1972). Thus the nastiest problem of the AGN-like scenario was alleviated: the compact object can be supplied at whatever relevant rate is required with appropriate adjustment of the system masses and orbital parameters. The viscous stress in the nonmagnetic case is not specified (this remained to be adduced by SS73), the vertical structure is integrated by assuming mean values for the temperature gradient and flux, the disk is treated in steady state, and for the magnetic case the ratio $B_{\text{L}}/B_{\text{K}}$ is kept constant. The rate of accretion is actually a poorly constrained parameter in the paper; but since the radial velocity is a constant fraction, $v_r / v_{\text{K}}$, of the Keplerian velocity, $v_{\text{K}} \sim \sqrt{GM/R}$, the energy generation scales as $p(r) \sim (M/R)F$, where $F$ is the mass flux, $F \sim \rho v_{\text{K}}$ and the density, $\rho$, comes from the requirement that the mass transfer through any cross section of the disk $Fbr$, where $b$ is the disk scale height, balances the mass transfer, $M$. They stressed, however, the importance of the Eddington luminosity, which provides the upper limit on the mass accretion rate depending on the luminosity generated at the gain (Shvartzman also mentions this). Because the luminosity cannot exceed $L_{\text{Edd}}$, the X-ray emission is necessarily limited, which provides a test of the model if the distances and luminosities can be determined.

Interpretations of, for example, the spectra of Sco X-1 and Cyg X-2 were complicated by the contradictions in the physical conditions that seemed required for nebular analyses, but the paradoxes of co-spatial temperatures differing by an order of magnitude, extreme luminosities, and apparently contradictory structures did not lead to a shift in the picture. Here PR72 made a fundamental contribution, that power-law spectra are not necessarily the result of nonthermal processes but of nonbarotropic emitting surfaces. Since the temperature is a power-law in radius, the disk emits a power-law spectrum. The viscous energy generation is locally balanced by radiative diffusion, thus determining the radial temperature structure of the disk self-consistently within the constraints of a model for the viscosity. The disk emits with a locally Planck source function such that when integrated over the disk surface, PR72 obtained a power-law flux distribution $S_{\nu} \sim \nu^{1/3}$ form, although they showed that the exponent depends on the physical processes included for the heating. The problem of magnetospheric accretion was also tackled in this paper, possible effects from accretion onto a rotating (Kerr) black hole were also briefly discussed, and the possibility of an X-ray pulsar (due to channeling of material to the poles alone) is proposed (e.g. Cen X-3). At the end PR72 speculate that the disk may become thermally unstable for some mass accretion rates, an important insight. The later innovation, by SS73, of introducing turbulent stresses (and at about the same time the assumption by NT73) results in a parameterization that is more flexible than the prescription for the radial drift in LB67 and PR72. The time-dependent problem was soon tackled by Lightman & Eardley (1974) and Sunyaev & Shakura (1975). Mineshige & Osaki (1983) and others treated thermal stability and viscous heating, self-gravitation was included in the $\alpha$-prescription, and the magnetorotational instability was proposed as a source of disk turbulence (Balbus & Hawley 1991), all with roots in the results of PR72.

Acknowledgements. My thanks go to Claude Bertout for his valuable comments on a draft of this paper, and to Steve Balbus, Tom Bolton, John Cannizzo, Ivan Hubeny, Mirek Plavec, John Scalo, and Robert Wilson for discussions. I also wish to record my indebtedness to Kevin Prendergast for discussions (in 1979) about accretion disk processes that first sparked my interest in the problem.

References
Paczynski, B. 1971, ARA&A, 9, 183
Shvatzman, V. F. 1971, SvAL, 15, 377
Sunyaev, R. A., & Shakura, N. I. 1975, SvAL, 1, 158

The magnetic stresses are produced by shearing of the infalling mass that distorts a pre-existing field. The tangential stress, $B_{L}B_{K}/4\pi r$, was cited as the source of a strong torque that, also allowing for a complex perhaps chaotic topology, would both transfer angular momentum and possibly accelerate high-energy particles. This is a forerunner of many of the current pictures of Poynting flux jets and X-ray-luminous structures in protostellar disks. Similar speculations, far less well developed, were advanced by Hoyle (1960, QJRAS, 1, 28) for the angular momentum transfer required in the protosolar disk, attributing the idea of magnetic torquing to Alfvén in his 1954 monograph on the origin of the solar system. Thus the idea had been around for some time before being applied to accretion disks around compact objects.