A&A 496, 413–421 (2009)
DOI: 10.1051/0004-6361/200810928

Astronomy & Astrophysics

The thermal-viscous disk instability model in the AGN context

J.-M. Hameury[1], M. Viallet[1], and J.-P. Lasota[2,3]

[1] Observatoire de Strasbourg, CNRS/Université Louis Pasteur, 11 rue de l'Université, 67000 Strasbourg, France
e-mail: [hameury;viallet]@astro.u-strasbg.fr
[2] Institut d'Astrophysique de Paris, UMR 7095 CNRS, UPMC Univ. Paris 6, 98bis Bd Arago, 75014 Paris, France
e-mail: lasota@iap.fr
[3] Astronomical Observatory, Jagiellonian University, ul. Orla 171, 30-244 Kraków, Poland

Received 7 September 2008 / Accepted 16 December 2008

ABSTRACT

Context. Accretion disks in AGN should be subject to the same type of instability as in cataclysmic variables (CVs) or in low-mass X-ray binaries (LMXBs), which leads to dwarf nova and soft X-ray transient outbursts. It has been suggested that this thermal/viscous instability can account for the long-term variability of AGNs.
Aims. We test this assertion by systematically studying how the disk instability model (DIM) is applied to AGNs.
Methods. We use the adaptative grid numerical code we developed in the context of CVs, enabling us to fully resolve the radial structure of the disk.
Results. We show that, because the Mach numbers are very large in AGN disks, the heating and cooling fronts are so narrow that they cannot be resolved by the numerical codes that have been used until now. In addition, these fronts propagate on much shorter time scales than the viscous time. As a result, a sequence of heating and cooling fronts propagate back and forth in the disk, leading only to small variations in the accretion rate onto the black hole, with short quiescent states only occurring for very low mass-transfer rates. Truncation of the inner part of the disk by e.g. an ADAF does not alter this result, but enables longer quiescent states. Finally we discuss the effects of irradiation by the central X-ray source and show that, even for extremely high irradiation efficiencies, outbursts are not a natural outcome of the model.

Key words. accretion, accretion disks – instabilities – stars: dwarf novae – galaxies: active

1. Introduction

Accretion disks are found in a wide variety of astronomical objects, from young stars to active galactic nuclei (AGNs). Among these, close binaries have deserved special attention, because they are nearby, and they vary on short timescales that enable time-dependent studies of their light curves. In particular, a number of these systems show large outbursts, such as dwarf novae, which are a subclass of cataclysmic variables in which a low-mass companion transfers mass onto a white dwarf. These systems undergo outbursts lasting at least a few days, during which their brightness increases by several magnitudes (see e.g. Warner 1995, for a review). The outbursts are believed to be due to a thermal-viscous accretion disk instability (Meyer & Meyer-Hofmeister 1981) that arises when the disk effective temperature becomes $\lesssim 10^4$ K, enough for hydrogen to become partially ionized and opacities to depend strongly on temperature (see Lasota 2001, for a review of the model). Similarly, soft X-ray transients, which are a subclass of low-mass X-ray binaries in which the compact object is either a black hole or a neutron star also show outbursts, but their amplitude is greater and the time scales longer than for dwarf novae. The ionization instability of the accretion disk is also thought to be cause of the outbursts: the difference with dwarf novae stems from the difference in the mass of the compact object (and thus in the depth of the gravitational potential well) and from the effect of illumination of the disk, which is much more important in the case of X-ray binaries (see e.g. Dubus et al. 2001).

It was realized long ago (Lin & Shields 1986) that the same instability could be present in accretion disks around AGNs; it was found that, at radii $\sim 10^{15-16}$ cm where the effective temperature is indeed of a few thousand degrees, the disk should be unstable. For the parameters of AGNs, the implied timescales are 10^4–10^7 yr, making the direct observation of the instability impossible, but predicting that in many systems the disk should not be in viscous equilibrium and that many AGNs should be in a quiescent state (see Siemiginowska et al. 1996; Siemiginowska & Elvis 1997). It was also immediately realized that, as in dwarf-novae, the character of putative AGN outbursts strongly depends on the assumptions one makes about the disk viscosity (Mineshige & Shields 1990). However, while one is guided in the case of dwarf-novae by the *observed* outburst properties when fixing the viscosity prescription, it is not even clear in the case of AGN that outbursts are present, as the variability of these objects could be just due to mass-supply variations. This state of affairs gave rise to various, more or less arbitrary, prescriptions for how viscosity varies (or not) with the state of the accretion flow (Mineshige & Shields 1990; Menou & Quataert 2001; Janiuk et al. 2004). In addition, results of numerical calculations of AGN outbursts were marred by the insufficient resolution of the grids used. As shown by Hameury et al. (1998), low grid resolution often leads to unreliable results[1].

The aim of the present article is to systematically study how the disk-instability model (DIM) can be applied to AGNs. Instabilities other than the thermal-viscous instability may

[1] Mayer & Pringle (2006) make the mischievous remark in this context that "mathematical convergence does not necessarily imply more accurate modeling of physical reality". While this might be true, it is clear that the lack of convergence of a mathematical model makes it useless for physical applications.

exist in AGN disks (beyond the MRI instability thought to be the source of viscosity, Balbus & Hawley 1991) and in particular the gravitational instability that arises when self-gravity exceeds the combined action of pressure and Coriolis forces (Toomre 1964; Safronov 1960); conditions for the onset of this instability are met at large distances from the black hole (see e.g. Shlosman 1990). The outcome of this instability in the AGN case is most probably the fragmentation of the accretion disk (see e.g. Gammie 2001; Goodman 2003; Rafikov 2007) since the cooling time is likely to be short in the AGN case. Duschl & Britsch (2006) suggest that the gravitational instability might instead be a source of turbulence, which could be the case if the non linear development of the instability does not lead to fragmentation, not a likely outcome in the AGN case as mentioned above. Other local or global instabilities may arise, such as the Lightman-Eardley instability (Lightman & Eardley 1974), but it is far beyond the scope of this paper to discuss them all, and we consider parameters such as these instabilities do not occur.

2. Vertical disk structure

We recall here the vertical-structure equations adapted to AGN parameters. We consider only the case where the viscosity ν is proportional to the gas pressure (not the total pressure, in order to avoid the Lightman & Eardley 1974, instability). The vertical structure of an α disk in which the viscosity ν is assumed to be proportional to the gas pressure is given by the standard disk equations (see e.g. Frank et al. 2002, and references therein):

$$\frac{\mathrm{d}P}{\mathrm{d}z} = -\rho g_z = -\rho\Omega_\mathrm{K}^2 z, \tag{1}$$

$$\frac{\mathrm{d}\varsigma}{\mathrm{d}z} = 2\rho, \tag{2}$$

$$\frac{\mathrm{d}\ln T}{\mathrm{d}\ln P} = \nabla, \tag{3}$$

$$\frac{\mathrm{d}F_z}{\mathrm{d}z} = \frac{3}{2}\alpha_\mathrm{eff}\Omega_\mathrm{K}P_\mathrm{g} \tag{4}$$

where $P = P_\mathrm{g} + P_\mathrm{rad}$, ρ and T are the total (gas plus radiation) pressure, density, and temperature, respectively, ς is the surface column density between vertical coordinates $-z$ and $+z$, $g_z = \Omega_\mathrm{K}^2 z$ the vertical component of gravity, Ω_K the Keplerian angular frequency, F_z the vertical energy flux, and ∇ the temperature gradient of the structure. This is generally radiative, with $\nabla = \nabla_\mathrm{rad}$, given by

$$\nabla_\mathrm{rad} = \frac{\kappa P F_z}{4P_\mathrm{rad}cg_z}\cdot \tag{5}$$

When the radiative gradient is superadiabatic, ∇ is convective ($\nabla = \nabla_\mathrm{conv}$). The convective gradient is calculated in the mixing length approximation, in the same way as in Hameury et al. (1998), with a mixing length taken as $H_\mathrm{ml} = \alpha_\mathrm{ml}H_\mathrm{P}$, where H_P is the pressure scale height:

$$H_\mathrm{P} = \frac{P}{\rho g_z + (P\rho)^{1/2}\Omega_\mathrm{K}}, \tag{6}$$

which ensures that H_P is smaller than the vertical scale height of the disk. Here, we use $\alpha_\mathrm{ml} = 1.5$.

We have neglected the disk self gravity. This approximation is valid as long as the ratio of self gravity to that of the central object is low:

$$\frac{g_\mathrm{s}}{g_\mathrm{c}} = \frac{\Omega_\mathrm{K}^2 H_\mathrm{P}}{2\pi G\Sigma} < 1. \tag{7}$$

If this not the case, the disk is gravitationally unstable, which, as mentioned in the introduction, is likely to lead to fragmentation if the cooling time is short enough, or may significantly change the angular momentum transport by introducing non local terms (see e.g. Lin & Pringle 1987; Balbus & Papaloizou 1999). In both cases, the thermal-viscous instability can no longer apply (in the first case for obvious reasons, and in the second one because non local effects cannot be approximated by viscosity, which is local); in our calculations, we always make sure that the condition (7) is fulfilled.

The parameter α_eff is an effective viscosity, equal to the standard viscosity coefficient α when the disk is in thermal equilibrium, but it also accounts for the time-dependent terms that are assumed to also be proportional to the pressure (see Hameury et al. 1998, for a detailed discussion).

The equation of state of matter is interpolated from the tables of Fontaine et al. (1977); in the low temperature regime (below 2000 K), which is not covered by these tables, Saha equations are solved iteratively, as described by Paczyński (1969). The Rosseland mean opacities are taken from Cox & Tabor (1976) above 10 000 K, and from Alexander (1975) below (more modern opacities introduce changes that are not important in the present context, see Lasota et al. 2008).

The boundary conditions are $\varsigma = 0$ and $F_z = 0$ at the disk midplane, and $\varsigma = \Sigma$ at the surface. The standard photospheric condition $\kappa P_\mathrm{g} = 2/3g_z$ has to be slightly modified, as (1) radiation pressure can de dominant; and (2) g_z can vary in the photosphere. Integrating the vertical hydrostatic equilibrium equation and using the Eddington approximation leading to $T^4(\tau) = 3/4T_\mathrm{eff}^4(2/3 + \tau)$ where τ is the optical depth, one obtains

$$\kappa\left(P_\mathrm{g} + \frac{1}{2}P_\mathrm{rad}\right) = \frac{2}{3}g_z\left(1 + \frac{1}{\kappa\rho z}\right)\cdot \tag{8}$$

The term $1/\kappa\rho z$ is close to the relative thickness of the photosphere relative to the total disk thickness. It is usually not very important except when the disk luminosity is close to its local Eddington limit, in which case the photosphere can be quite extended.

The thermal equilibrium corresponds to $Q^+ = Q^-$, where Q^+ and Q^- are the surface heating and cooling rates, respectively (see Eq. (14) below). Figure 1 presents two examples of thermal equilibrium curves in the $\Sigma - T_\mathrm{eff}$ plane, showing the characteristic S shape. Also plotted are the conditions $g_\mathrm{s}/g_\mathrm{c} = 1$ and $h/r = 0.1$. As can be seen, self gravity becomes important at radii larger than $1 - 2\times10^{16}$ cm, in agreement with the findings of Cannizzo & Reiff (1992) and Cannizzo (1992). The condition that self gravity be small can be quite severe; for example, we note that in several of the simulations by Janiuk et al. (2004) this condition is not fulfilled and the corresponding results are therefore invalidated. The thin disk approximation condition ($h/r \ll 1$) is usually less stringent; it may, however, break for high accretion rates, in which case radiation pressure gradient almost balances vertical gravity in a significant fraction of the disk vertical extent.

The values Σ_min and Σ_max that are the minimimun (resp. maximum) values of Σ on the upper (resp. lower) branches of the S curve can be fitted by

$$\Sigma_\mathrm{min} = 2.90\times10^3\alpha^{-0.74}\left(\frac{r}{10^{15}\ \mathrm{cm}}\right)^{1.04} M_8^{-0.35}\ \mathrm{g\ cm^{-2}} \tag{9}$$

and by

$$\Sigma_\mathrm{max} = 3.85\times10^3\alpha^{-0.82}\left(\frac{r}{10^{15}\ \mathrm{cm}}\right)^{0.99} M_8^{-0.33}\ \mathrm{g\ cm^{-2}} \tag{10}$$

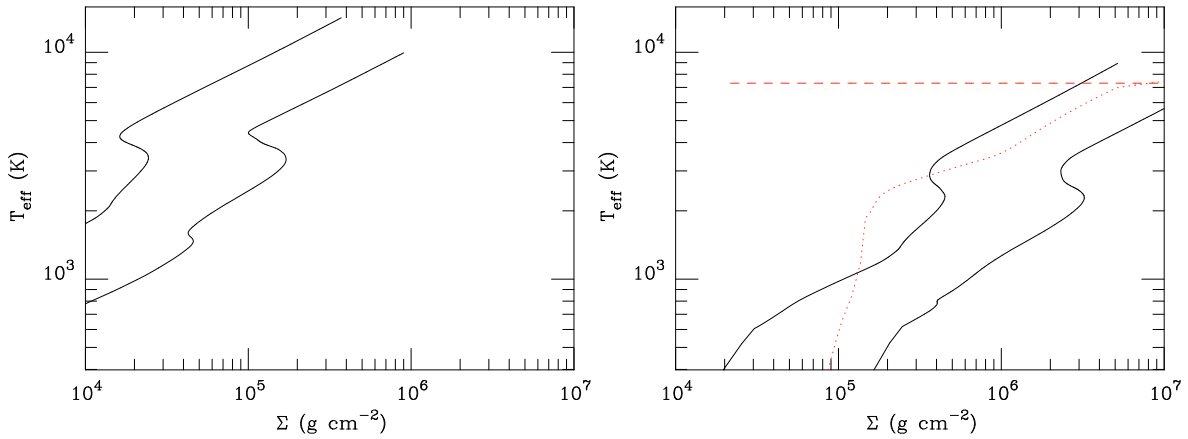

Fig. 1. Examples of S curves in the $\Sigma - T_{\rm eff}$ plane, for $M = 10^8\ M_\odot$ and $r = 10^{15}$ cm (*left*) and for $r = 2 \times 10^{16}$ cm (*right*). In both cases, S curves obtained for $\alpha = 0.1$ and 0.01 are shown. The dotted curve corresponds to ratio of self to central gravity equals 1, and the dashed curve to $h/r = 0.1$. Only regions below the dashed curved and above the dotted one are allowed. For $r = 10^{15}$ cm, these limits lie outside the portion of the $\Sigma - T_{\rm eff}$ plane shown here.

where $M_8 = M/(10^8\ M_\odot)$, M the black hole mass. The corresponding effective temperatures are $T_{\rm eff}(\Sigma_{\rm min}) = 4300(r/10^{15}\ {\rm cm})^{-0.12}$ K and $T_{\rm eff}(\Sigma_{\rm max}) = 3300(r/10^{15}\ {\rm cm})^{-0.12}$ K, respectively. These are independent of α, as expected, and their radial dependence is quite weak. As compared to disks around stellar mass objects, the surface densities are much higher, hence somewhat smaller effective temperature at the turning points of the S-curve (the upper stable solution ends at 3000–4000 K instead of 7000–8000 K), even though the corresponding mid-plane temperatures are quite similar.

It should also be noted that, because we are restricted to a region where self gravitation is small, the disk extension, as measured by the ratio $r_{\rm in}/r_{\rm out}$ is not very large. For the case of a $10^8\ M_\odot$ black hole, this is about 100, i.e. comparable to disks in CVs, but much smaller than for LMXBs. It is also worth noting that the disk thickness H,

$$\frac{H}{r} \simeq \frac{c_{\rm s}}{v_{\rm k}}, \tag{11}$$

where $c_{\rm s}$ and $v_{\rm k}$ are the sound and Kepler velocities, is small in regions where the thermal-viscous instability can propagate. As compared to the CV or LMXB case, $c_{\rm s}$ is unchanged because the central temperature at the turning points of the S curve are similar, and on the order of 10^6 cm s^{-1}, but $v_{\rm k}$ is much larger, since in the AGN case, we are restricted to regions close to the black hole as mentioned above; here, we have $H/r \sim 10^{-3}$ or smaller. As the width of the heating and cooling fronts are a few times H (Menou et al. 1999), this can be a source of numerical problems. In particular, these fronts have been completely unresolved in all previous studies, casting some doubt on their results.

2.1. Critical points and the viscosity prescription

It should be noted that, for some choices of parameters, the equilibrium curves show two "wiggles" (see e.g. the case $\alpha = 0.1$ at $r = 10^{15}$ cm, Fig. 1) on the lower branch. This also happens in accretion disks around stellar mass black holes, but, in contrast to Janiuk et al. (2004), we do not find that this is always the case for AGNs. This discrepancy can be due to a difference in the treatment of convection or to different opacities. The small wiggle at low temperature is not related to a strong change in the opacities, but instead to a strong change in the adiabatic

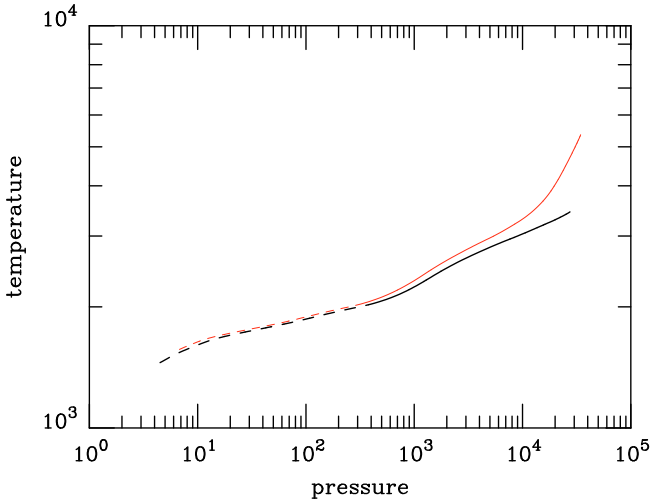

Fig. 2. Vertical structure of the accretion disks, for $M = 10^8\ M_\odot$, $r = 10^{15}$ cm, $\Sigma = 4.3{\times}10^4$ g cm^{-2}, $\alpha = 0.01$, and $T_{\rm eff} = 1400$ K (thick curve) and 1570 K (thin curve). The line is dashed when energy transport is radiative. These two cases correspond to the upper and lower stable parts of the small wiggle shown in Fig. 1.

gradient when molecular hydrogen becomes partially dissociated, as is shown by Fig. 2. Two vertical structures, corresponding to the same r, Σ, and α, but two different effective temperatures on the upper and lower stable branches of the small wiggle of Fig. 1, differ essentially by a much stronger temperature gradient in the disk midplane. In one case, $\nabla_{\rm ad} \sim 0.10$, while $\nabla_{\rm ad} \sim 0.40$ in the other. This effect occurs only if molecular hydrogen becomes partially dissociated in the convective zone. Since the opacities are relatively low for the corresponding temperatures and densities, there are cases where the transition between molecular and atomic hydrogen occurs in a radiative zone, in which case no wiggle is found.

In the standard dwarf-nova model, it is assumed that the α-parameter changes rapidly when the disk temperature reaches the ionization instability; this is required for the amplitude of the modeled outburst to be comparable to the observed one. It is often stated that the physical reason for such a change is the change in the ionization parameter of the gas, hence α is assumed to remain constant when transiting this secondary wiggle. This also

seems to be a reasonable hypothesis in the AGN case, and does not require the physics of accretion disk to be different in different environments, even though the temperatures and densities are similar. In the following, contrary to Janiuk et al. (2004), we assume therefore that the critical $\Sigma_{\rm max}$ of the cold stable branch corresponds to the ionization instability and that the lower wiggle is not associated with a change in α. This point is of importance, since as shown by Hameury (2002), the shape of the resulting S curve and hence the outcome of the model is by far dominated by the change in the viscosity parameter α.

Finally, one should point out that whatever arguments are used to justify the change in α, i.e. the use of an $\alpha_{\rm cold}$ and an $\alpha_{\rm hot} \approx (4-10)\alpha_{\rm cold}$, the real reason is the necessity to produce the required outburst amplitude. It has been argued (Gammie & Menou 1998) that the difference between viscosities in the high and low (quiescent) states of dwarf-nova disks is due to the "decay" of the MRI mechanism that is supposed to be the source of turbulence in accretion disks (Balbus & Hawley 1991). In the environment of AGN disks, the MRI is supposed to be operating also in cold disks, which was used to argue that $\alpha_{\rm hot} \approx \alpha_{\rm cold}$ in this case (Menou & Quataert 2001).

However, as noted by Steven Balbus (private communication), because of the fact that numerical simulations treat the turbulent dynamics of disks at a level far beyond anything that can be approached with strictly analytic techniques, there has been a tendency to grant simulations a level of certainty that they do not merit yet. A careful treatment of realistic energetics still remains beyond the capabilities of current codes, and even simple polytropic shearing box calculations need to be run at much higher resolutions and for much longer times than were once thought necessary.

Therefore the values of critical Reynolds numbers deduced only from numerical simulations (Gammie & Menou 1998; Menou & Quataert 2001) are highly uncertain and we opted for using the standard dwarf-nova DIM also in AGNs.

3. Disk evolution

3.1. Basic equations

The standard equations for mass and angular momentum conservation in a geometrically thin accretion disk can be written as

$$\frac{\partial \Sigma}{\partial t} = -\frac{1}{r}\frac{\partial}{\partial r}(r\Sigma v_{\rm r}) \tag{12}$$

and

$$j\frac{\partial \Sigma}{\partial t} = -\frac{1}{r}\frac{\partial}{\partial r}(r\Sigma j v_{\rm r}) + \frac{1}{r}\frac{\partial}{\partial r}\left(-\frac{3}{2}r^2\Sigma\nu\Omega_{\rm K}\right) \tag{13}$$

where $v_{\rm r}$ is the radial velocity in the disk, $j = (GM_1 r)^{1/2}$ is the specific angular momentum of material at radius r in the disk, $\Omega_K = (GM_1/r^3)^{1/2}$ is the Keplerian angular velocity

The energy conservation equation is taken as (see Cannizzo 1993; Hameury et al. 1998, for details):

$$\frac{\partial T_{\rm c}}{\partial t} = \frac{2(Q^+ - Q^- + J)}{C_P\Sigma} - \frac{P_{\rm c}}{\rho_{\rm c}C_P}\frac{1}{r}\frac{\partial(rv_{\rm r})}{\partial r} - v_{\rm r}\frac{\partial T_{\rm c}}{\partial r}, \tag{14}$$

where $P_{\rm c}$ and $\rho_{\rm c}$ are the midplane pressure and density, and Q^+ and Q^- are the surface heating and cooling rates, respectively. They are usually taken as $Q^+ = (9/8)\nu\Sigma\Omega_{\rm K}^2$, and $Q^- = \sigma T_{\rm eff}^4$, $T_{\rm eff}$ is the effective temperature. The term J accounts for the radial energy flux carried by viscous processes,

$$J = 1/r\partial/\partial r(rF_{\rm e}), \tag{15}$$

where $F_{\rm e}$ is the flux carried in eddies with characteristic velocity $v_{\rm e}$ and size $l_{\rm e}$:

$$F_{\rm e} = C_P\Sigma v_{\rm e}\frac{\partial T_{\rm c}}{\partial r}l_{\rm e} = \frac{3}{2}\nu C_P\Sigma\frac{\partial T_{\rm c}}{\partial r}\,. \tag{16}$$

These are identical to the equations of a disk in a binary system, except that there are no tidal torques and no tidal dissipation.

The inner boundary condition is also unchanged from the binary case:

$$\nu\Sigma = 0 \quad \text{at } r = r_{\rm in} \tag{17}$$

where $r_{\rm in}$ is the radius of the inner edge of the disk and can be larger than the radius of the innermost stable orbit if the disk is truncated by the formation of an ADAF, in which case $r_{\rm in}$ is a given function of the mass accretion rate (see e.g. Hameury et al. 1997). As we use $\ln(\Sigma)$ as a variable, the $\Sigma = 0$ boundary condition is not applicable. Instead, we take

$$\nu\Sigma = 1.1\Sigma_{\rm min} \quad \text{at } r = r_{\rm in} \tag{18}$$

so that this allows for the disk to be both in the hot or cold states (see below for a more detailed discussion on the effect of using this boundary condition).

The outer boundary condition is more problematic, as the disk extends to large distances where all the usual approximations are invalid (thin disk, neglect of self gravity, etc.). We instead assume that at some distance $r_{\rm out} \sim 10^{16}$ cm for $M = 10^8\ M_\odot$, the mass-transfer rate is given and constant. This approximation is valid provided that the heating front does not reach this outer radius.

The heat Eq. (14) requires two additional boundary conditions. As discussed in Hameury et al. (1998), these are of little importance, and we take $J = 0$ at $r = r_{\rm in}$ and $r = r_{\rm out}$.

3.2. Results

Figure 3 shows an example of the evolution of the accretion disk. We have considered here a $10^8\ M_\odot$ black hole accreting at 10^{24} g s^{-1}, about one hundredth of the Eddington limit:

$$\dot{M}_{\rm Edd} = \frac{L_{\rm Edd}}{c^2\eta} = \frac{4\pi GMm_{\rm p}}{\sigma_{\rm th}c\eta} \simeq 1.4\times10^{26}M_8 \text{g s}^{-1} \tag{19}$$

where $\eta \sim 0.1$ is the efficiency of accretion, and $\sigma_{\rm Th}$ the Thompson cross section. The mass is the same as in Janiuk et al. (2004), but the mass-transfer rate is lower by a factor ~10 in order to avoid (or try to avoid) the instability occuring in the self-gravitating part of the accretion disk. We have assumed a varying α between $\alpha_{\rm h} = 0.2$ and $\alpha_{\rm c} = 0.04$. We have also taken the outer disk radius to be 10^{16} cm, as the disk becomes self-gravitating at larger distances. As can be seen, the disk can never be brought completely to the cold regime; as a consequence, relatively low amplitude oscillations are seen in the visual magnitude and in the accretion rate onto the black hole. This situation is reminiscent of what happens in the case of soft X-ray transients when no disk truncation or irradiation is assumed (Menou et al. 2000; Dubus et al. 2001), or in the case of symbiotic stars (Duschl 1986). A cooling front is reflected at some radius much larger than the disk inner radius and as a result a heating front starts propagating outwards, but it cannot quite reach the outer disk edge and a cooling front forms again. Such reflections occur when the surface density Σ behind the cooling front reaches $\Sigma_{\rm max}$, which triggers a new instability. The resulting heating front propagates outwards until the post-front density reaches $\Sigma_{\rm min}$.

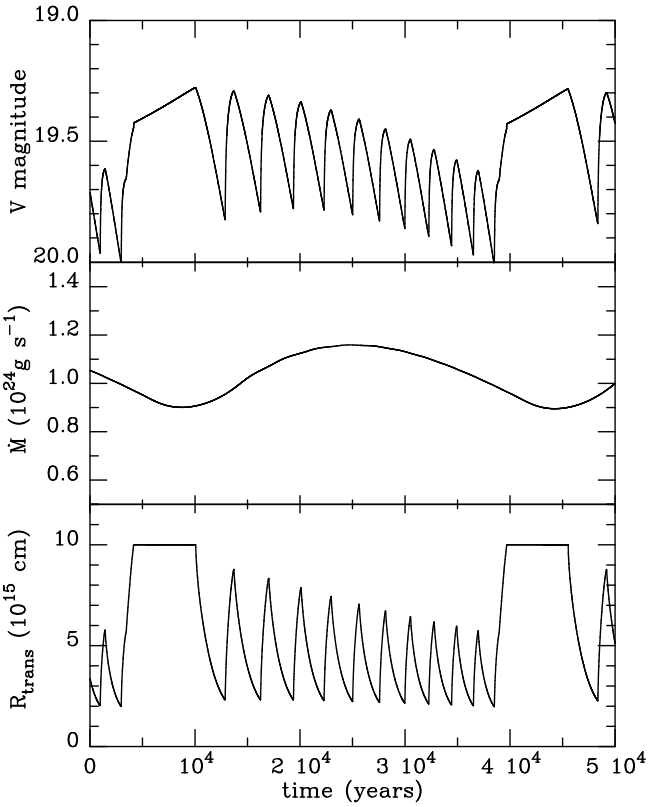

Fig. 3. Time evolution of an accretion disk with the following parameters: black hole mass: $10^8\ M_\odot$, inner and outer radius: 10^{14} and 10^{16} cm, respectively, and mean mass transfer rate: 10^{24} g s^{-1}. *Top panel*: visual magnitude; intermediate panel: accretion rate onto the black hole; *lower panel*: radius at which the transition between the hot and cold regimes takes place.

Then a new cooling front starts going down the disk. There is, however, a significant difference in that the short time oscillations do not result in oscillations of the mass accretion rate on the same time scale. $\dot{M}$ fluctuates only on the longer time scale of the front oscillation pattern. The basic reason for this is that the front propagates at approximately α times the sound speed, i.e. on a time scale

$$t_{\rm front} = \frac{r}{\alpha c_{\rm s}} = \frac{r}{h} t_{\rm th}, \tag{20}$$

where $t_{\rm th}$ is the thermal time scale. $t_{\rm front}$ is shorter than the viscous time $t_{\rm visc} = (r/h)^2 t_{\rm th}$ by a factor r/h, i.e. by several orders of magnitude. The cooling front therefore propagates so rapidly that the surface density at smaller radii does not change; to a first approximation, it cannot propagate in regions where $\Sigma > \Sigma_{\rm max}(\alpha_{\rm c})$. In the CV case, $t_{\rm front}$ is shorter than $t_{\rm visc}$, but not by such a large amount, and strong gradients in the disk make the effective viscous time comparable to the front propagation time.

It must also be noted that the front occasionally reaches the outer disk edge; then the outer boundary condition that dictates in particular that there is no outward mass flow is not valid, so that the correct sequence is probably different. The back and forth propagation of heating fronts on a short time scale is, however, a firm prediction of the model.

For lower mass-transfer rates, the outer part of the disk can remain on the cold, stable branch, in which case the front propagation is restricted to the innermost parts of the disk. Figures 4 and 5 show the evolution of a disk with the same parameters

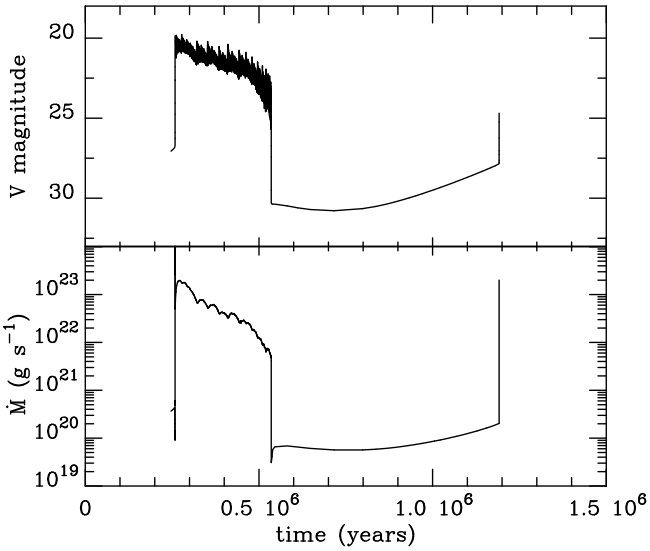

Fig. 4. Time evolution of an accretion disk with the following parameters: black hole mass: $10^8\ M_\odot$, inner and outer radius: 10^{14} and 10^{16} cm, respectively, and mean mass transfer rate: 2×10^{22} g s^{-1}. *Top panel*: visual magnitude, *lower panel*: accretion rate onto the black hole.

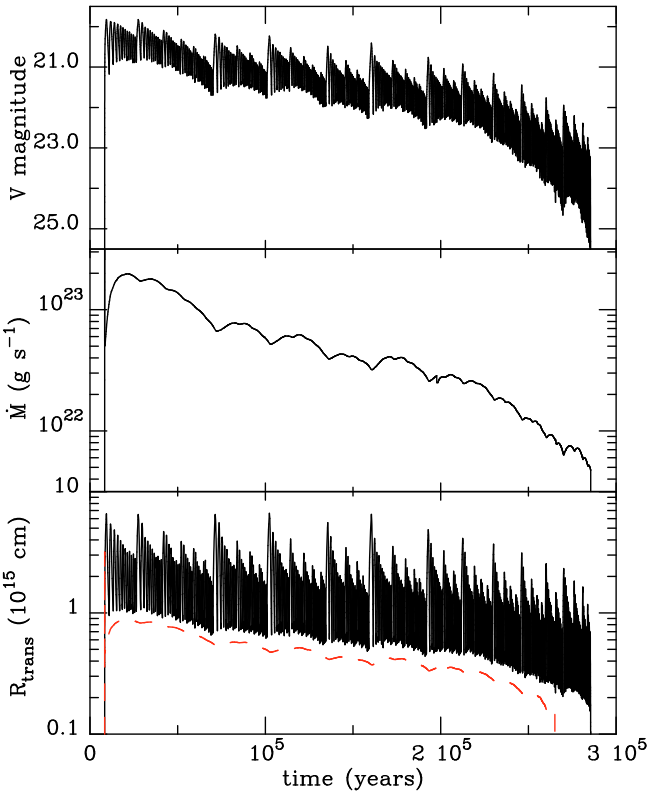

Fig. 5. Details of the outburst shown in Fig. 4. *Top panel*: visual magnitude; intermediate panel: accretion rate onto the black hole; *lower panel*: radius at which the transition between the hot and cold regimes takes place. The red-dashed line is the semi-analytic value of the minimum transition radius given by Eq. (23).

as in Fig. 3, but with a mass-transfer rate of 2×10^{22} g s^{-1}, 5×10^{-3} times the Eddington limit. As can be seen in Fig. 5, heating and cooling fronts propagate in a restricted fraction of the accretion disk. They do not reach radii larger than that at which the disk can sit on the stable cool branch, given the externally imposed mass-transfer rate. They also do not reach the innermost

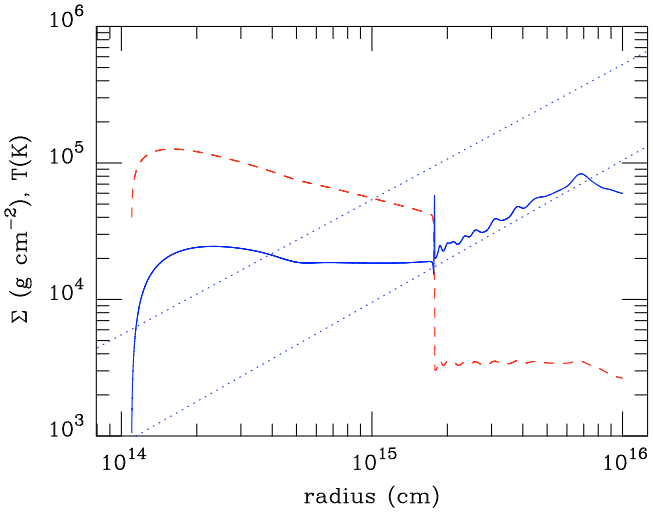

Fig. 6. Radial structure of the disk. The blue solid line represents the surface density, the red dashed one the central temperature. The dotted lines are the critical $\Sigma_{\rm min}$ and $\Sigma_{\rm max}$. See text for details.

regions where the surface density always remains high enough for the disk to be stable on the hot branch, except when entering a quiescence period, which happens when the disk finally empties on a much longer viscous time. The active phase lasts for 3×10^5 yr in the case presented here, with more than 400 consecutive oscillations. These are not random, but show relatively regular sequences of decreasing oscillations that are interrupted by an oscillation with a larger amplitude, clearly visible in Fig. 5. Note also some sort of hierarchical structure for these oscillations. Figure 6 shows the radial structure of the disk during the oscillating phase (compare wih Fig. 2 of Dubus et al. 2001). The semi-stable inner and outer regions are clearly visible. The central unstable zone is divided into two parts: an inner unstable one, and an outer marginally stable one, where $\Sigma \approx \Sigma_{\rm min}$, resulting from the successive passage of heating fronts that die at radii decreasing with time. A leftover of the death of these fronts is the little wiggle in Σ that gets smoothed with time as a result of diffusion, or when a heat front is able to reach this region. Note also the spike in the unstable region, which carries a small amount of mass that will cause the small wiggles in the marginally stable region.

3.3. Minimum radius reached by cooling fronts

The minimum radius reached by the cooling front can be determined by noting that the front propagates down to a point where $\Sigma = \Sigma_{\rm max}(\alpha_{\rm c})$ and that the innermost parts of the disk are in quasi viscous equilibrium. This means that the surface density is determined by the accretion rate which is almost constant in this hot inner region. This is equivalent to stating that, at the reflection point, the dissipation rate Q^+ is

$$Q^+ = \frac{3GM\dot{M}}{8\pi r^3} f = \sigma T_{\rm eff}^4(\Sigma_{\rm max}, \text{hot}) \qquad (21)$$

where $f = 1 - (r/r_{\rm in})^{-1/2}$. Note that $T_{\rm eff}$ is calculated on the hot branch and is *not* given by the analytic fits obtained in Sect. 2. An examination of Fig. 1 shows that $T_{\rm eff}(\Sigma_{\rm max}, \text{hot})$ is about 3.2 times that of the turning point on the cool branch, $T_{\rm eff}(\Sigma_{\rm max}, \text{cold})$ for $\alpha_{\rm c}/\alpha_{\rm h} = 0.1$. As Q^+ is proportional to $\nu\Sigma$, hence proportional to α, one can guess that

$$T_{\rm eff}^4(\Sigma_{\rm max}, \text{hot}) = 3.2^4 \frac{\alpha_{\rm h}}{10\alpha_{\rm c}} T_{\rm eff}^4(\Sigma_{\rm max}, \text{cold}), \qquad (22)$$

which is also a very good approximation even for $\alpha_{\rm c}/\alpha_{\rm h} = 1$, as can be seen from Fig. 1. Now, from the fits of $T_{\rm eff}^4(\Sigma_{\rm max}, \text{hot})$ we have

$$r = 1.7 \times 10^{15} \left(\frac{\dot{M}}{10^{23} \text{g s}^{-1}} M_8 \frac{\alpha_{\rm c}}{\alpha_{\rm h}} f \right)^{0.4} . \qquad (23)$$

Figure 6 shows that there are two points where Σ crosses the $\Sigma_{\rm max}$ line, and Eq. (23) indeed has two solutions, one for which f is small, and another one in which $f \simeq 1$. The first one corresponds to the transition between the very inner disk, where Σ is vanishingly small because of the boundary condition and therefore the cool branch solution applies, and nearby regions where Σ is large enough for the hot solution to apply. This is discussed in the next section. The second one corresponds to the radius at which the cooling front is reflected and becomes a heating front.

From Eq. (23), it appears that, for low enough $\dot{M}$ or large enough $r_{\rm in}$, the cooling front can reach the inner radius, in which case the system will enter a quiescence phase. More precisely, this happens when Eq. (23) has no solution. Simple algebra shows that the critical $\dot{M}$ is

$$\dot{M} = 5 \times 10^{20} \left(\frac{r_{\rm in}}{10^{14} \text{ cm}} \right)^{2.5} \frac{\alpha_{\rm h}}{\alpha_{\rm c}} M_8^{-1} \text{ g s}^{-1}, \qquad (24)$$

and the corresponding critical radius is $r = 1.44\, R_{\rm in}$, at which $f = 1/6$. It is interesting to rescale this relation as

$$\frac{\dot{M}}{\dot{M}_{\rm Edd}} = 2.7 \times 10^{-6} \frac{\alpha_{\rm h}}{\alpha_{\rm c}} M_8^{0.5} \, (r_{\rm in}/3r_{\rm s})^{2.5} , \qquad (25)$$

which shows that, if the disk is not truncated, low states will be found only for low mass-transfer rates, whatever the black hole mass.

The critical rate given by Eq. (25) refers to the accretion rate onto the black hole and not to the mass-transfer rate. For high mass-transfer rates, both are almost equal, as seen above (see e.g. Fig. 3). For lower values of $\dot{M}_{\rm transf}$, they may differ by up to one order of magnitude, as shown for example in Fig. 4 where the accretion rate at maximum is 10 times higher than the mass supply rate to the disk. In this case, $\dot{M}$ ultimately falls below the critical value given by Eq. (25), and the disk enters a quiescent state. The duration of this state is short, however, close to the duration of the outburst state, as the average mass accretion rate during the active state is $\sim 5.7 \times 10^{22}$ g s^{-1}, i.e. not very different from the steady mass-transfer rate. (The duty cycle expected for an outburst with an average accretion rate 2.85 times higher than the transfer rate is 0.35, very close to the value given by the simulation (0.30), showing that the disk is almost relaxed.) We therefore expect that outbursts exist only for low mass-transfer rates, that these outbursts are weak – never reaching anything close to the Eddington limit – and that the duty cycle cannot be large.

3.4. Innermost disk instability

The very inner parts of the disk, where the density is very low because of the inner boundary condition should therefore be on the cold branch. The transition between this cold region and more distant, hotter regions should also be unstable; indeed, when one

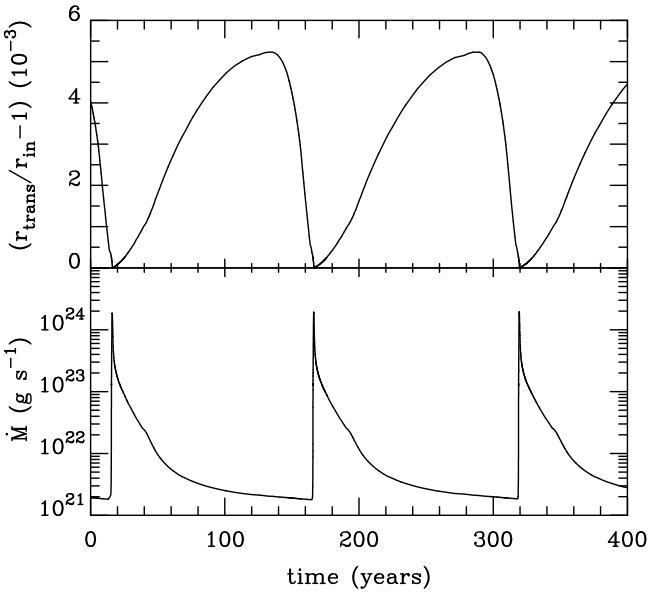

Fig. 7. Instability of the inner disk edge when the boundary condition $\Sigma = 0$ is used. The *top panel* shows the position of the transition radius as a function of time, the *lower panel* the mass accretion rate onto the black hole. The parameters are those of Fig. 4. These large amplitude are unphysical, since mass flows into the black hole from a region much larger than the width of the zone in which these fluctuations occur.

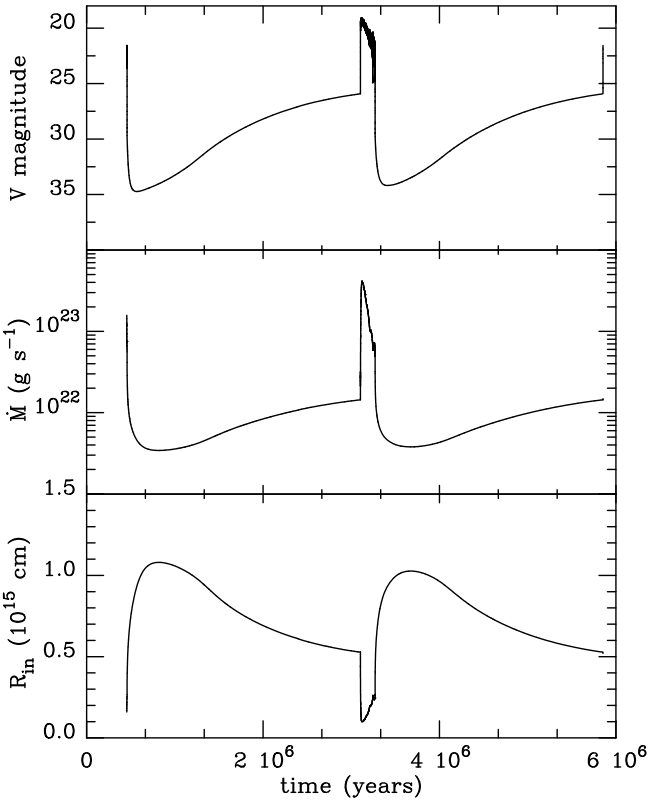

Fig. 8. Long-term evolution on an AGN accretion disk when r_{in} can vary as a result of e.g. evaporation or the formation of an ADAF. The system alternates between active phases in which heating and cooling fronts propagate back and forth in the disk and quiescent phases lasting about one million years. *Top panel*: visual magnitude; *intermediate panel*: accretion rate onto the black hole; *lower panel*: inner disk radius.

assumes that the inner boundary condition is not $\Sigma = 1.1\Sigma_{min}$ at $r = r_{in}$, but is smaller than Σ_{min}, oscillations are found. For the sake of completeness, we show in Fig. 7 the effect of these oscillations in such a case. Cooling/heating fronts propagate in a very restricted region, whose radial extent is greater than the vertical scale height so that the thin disk approximation is still valid, but presumably much smaller than the zone from which matter flows into the black hole. Also, this region is so small that the total disk luminosity remains constant. These oscillations are possible only when the width of heating/cooling fronts is less than the width of the region over which Σ catches the boundary condition $\Sigma = 0$, which is one to a few percent of r_{in} (see Fig. 6), otherwise fronts would simply not exist. This is quite possible in the AGN case, because the fronts are so narrow and contrasts with the CV or LMXB case where the reverse is true and the condition $\Sigma = 0$ does not have such an effect. There oscillations are most probably not physical, because one assumes that (i) there is absolutely no torque at the inner disk edge; and (ii) that matter is lost from the disk only at $r = r_{in}$. It is very likely that the mechanisms leading to accretion at the inner disk edge (e.g. evaporation, etc.) will smooth oscillations there. To avoid these and to ease the numerical computations, we have assumed that Σ is not vanishingly small at the inner disk edge, but that instead it is very slightly larger than Σ_{min}.

3.5. Disk truncation

Disk truncation could be a solution to the absence of large outbursts. This was found to be an essential ingredient of the soft X-ray transient model (see e.g. Menou et al. 2000; Dubus et al. 2001). Truncation can be the result of the formation of an advection-dominated accretion flow (ADAF) or of one of its variants (see e.g. Narayan & McClintock 2008; Kato et al. 1998, for reviews of the ADAF), the important feature being that the flows becomes hot, geometrically thick, and optically thin close to the black hole. For the outburst cycle to be modified, one needs the disk not to extend down to the innermost stable orbit, but instead be truncated at a radius comparable to the minimum radius reached by the cooling front. The inner disk radius will then depend on the mass accretion rate onto the black hole. Many prescriptions can be derived; what really matters is whether Eq. (23) can be satisfied or not, since the details of the variations of r_{in} as a function of $\dot{M}$ are not important.

Figures 8 and 9 show an example in which $r_{in} = 2 \times 10^{14} (\dot{M}/10^{23}$ g s$^{-1})^{1/2}$. As can be seen, quiescent states are found, as well as active states that are not very bright though – only brighter than the active states for non truncated disks by a factor ~2, resulting in duty cycles that also differ by factors ~2.

3.6. Disk irradiation

Disk irradiation plays an essential role in soft X-ray transients (see e.g. van Paradijs 1996; Dubus et al. 2001, and references therein) and sometimes in CVs, see e.g. Hameury et al. (1999), and could also play an important role in the AGN context. We follow here the same procedure as in the case of irradiated disks in SXTs (Dubus et al. 2001). We assume that the irradiation flux F_{irr} onto the disk is given by

$$F_{irr} = \sigma T_{irr}^4 = C \frac{\dot{M}c^2}{4\pi R^2} \quad (26)$$

where C is a constant. In the case of SXTs, $C = 5 \times 10^{-4}$ has been adopted by Dubus et al. (2001). (We include here in C the

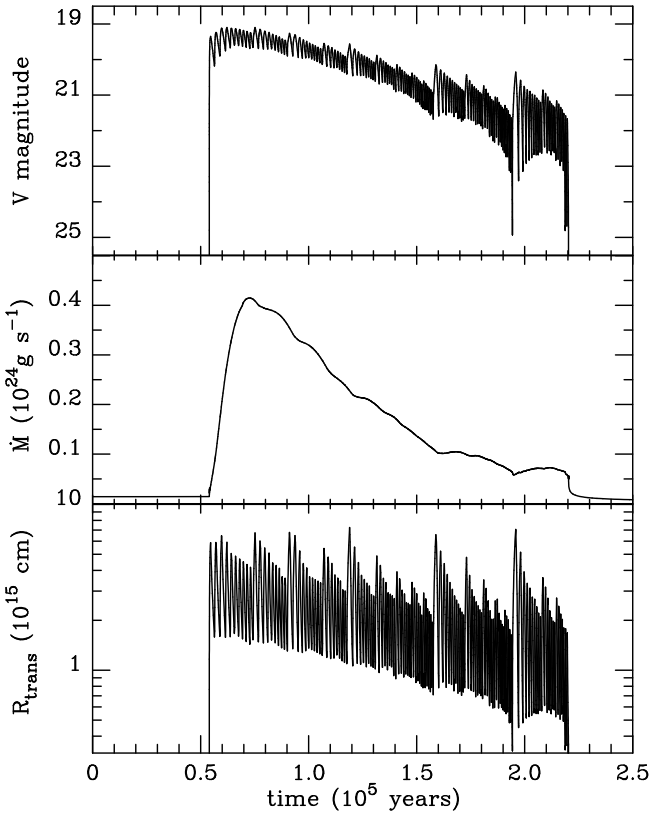

Fig. 9. The outburst shown in Fig. 8. The *two upper panels* are the same as in Fig. 8, and the *lower panel* shows the transition radius between the inner hot disk and the cooler outer parts.

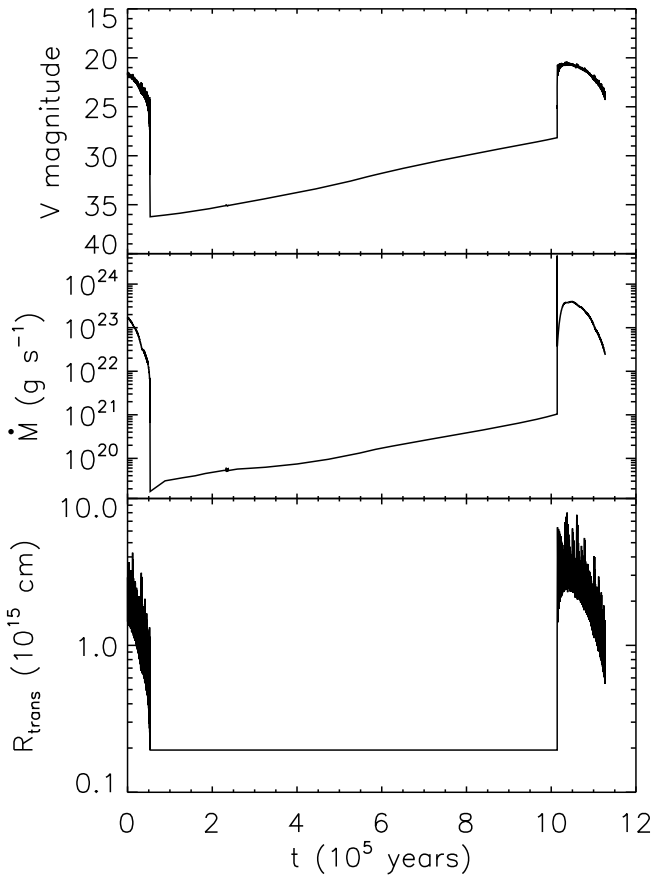

Fig. 10. Time evolution of an irradiated disk, with parameters identical to those of Fig. 4; here $C = 0.01$. *Top panel*: V magnitude; *intermediate panel*: accretion rate onto the black hole; *lower panel*: transition radius between the hot, inner disk and the cool outer disk.

efficiency of mass to energy conversion.) The ratio of the irradiating flux to the viscous flux is then given by

$$\frac{F_{\rm irr}}{F_{\rm visc}} = \frac{4}{3} C \frac{\dot{M}_{\rm acc}}{\dot{M}} \frac{r}{r_{\rm s}}, \tag{27}$$

which clearly shows, because we are interested in regions much closer to the black hole than in the case of SXTs, that C must be large if irradiation is to have any effect at all. To affect the central temperature of the disk, $F_{\rm irr}/F_{\rm visc}$ must exceed the optical thickness of the disk (in the radiative case). This is possible in principle because the X-ray emitting region could have a complex geometry, such as a corona above a cooler disk, in which case the irradiation flux can be large. It is however very unlikely that it could exceed 0.01, since the fraction of X-rays absorbed below the photosphere is at most 10%. In the following, we consider the case $C = 0.01$, which corresponds to a maximally irradiated disk. One should also note that, in this case, the X-ray luminosity would be linked in a complex way to the local properties of the accretion flow and would not be proportional to the accretion rate onto the black hole $\dot{M}_{\rm acc}$ alone. Equation (27) assumes steady state, and neglects the $r_{\rm in}/r$ terms in the energy dissipation equation, which can be significant close to the disk inner edge.

We then calculated a grid of vertical disk structures to determine the effective temperature as a function of Σ, $T_{\rm c}$, $T_{\rm irr}$, as described in Sect. 2, with a modified boundary condition at the disk surface:

$$F_z = \sigma(T^4 - T_{\rm irr}^4). \tag{28}$$

As in the case of SXTs, the effect of irradiation is a stabilization of the disk when the irradiation temperature is high enough, typically higher than $\sim 10^4$ K.

A reasonably good fit to the effective temperature at $\Sigma_{\rm max}$ is given by

$$T_{\rm eff}(\Sigma_{\rm max}) = 730 \left(\frac{r}{10^{15}\ {\rm cm}}\right)^{0.33} M_8^{-0.11} \left(\frac{\dot{M}_{\rm acc}}{10^{25} {\rm g\ s^{-1}}}\right)^{-0.29} \times \left\{1 - \left(\frac{7\times 10^{19}\ {\rm g\ s^{-1}}}{\dot{M}_{\rm acc}}\right)^{1/4}\right\}^{1/4} {\rm K} \tag{29}$$

in a situation where the disk is in viscous equilibrium, hence $T_{\rm irr}$ is directly given by $\dot{M}$. This fit is valid for irradiation temperatures higher than about 4000 K, but less than 10^4 K for which the disk becomes stable.

Figure 10 shows the time evolution of an irradiated disk with the same parameters as in Fig. 4, apart from the irradiation factor C set to 0.01. As can be seen, even in this maximally irradiated disk, the time evolution is not very different from that of an unirradiated disk. There is still a succession of rapid oscillations of the luminosity, with a heating/cooling front propagating back and forth. The main difference is here that the disk enters into a quiescent phase more rapidly than in the unirradiated case, but one should note that the initial structure was not exactly the same in both cases, and that because of the huge computing time required to follow the disk oscillations, a relaxed state cannot be attained in practice. However, the radial disk structure obtained at the end of active phases in the irradiated and unirradiated cases

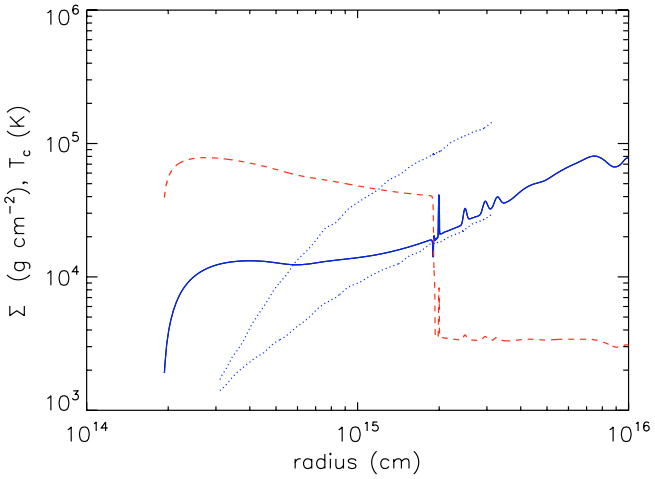

Fig. 11. Example of the disk radial structure in the irradiated case. The blue solid line is the surface density, the red dashed line is the central temperature. The blue dotted lines represent $\Sigma_{\rm min}$ and $\Sigma_{\rm max}$, and the effects of irradiation are clearly visible ar radii $\lesssim 10^{15}$ cm; at radii smaller than 3×10^{14} cm, the S shape of the cooling curve vanishes and the disk is stable.

do not differ much, with an outer disk on the cold stable branch, and most of the disk having $\Sigma = \Sigma_{\rm min}$. This similarity, and the fact that the unirradiated disk was almost relaxed makes us confident that, here also, the disk is close to relaxation.

Figure 11 shows the radial structure of the disk at time $t = 3.5\times10^4$ yr, when 3/4 of the first outburst have elapsed. It clearly shows the impact of irradiation on the innermost part of the disk, which is due both to a high irradiation temperature and to the decreasing viscous dissipation close to the inner disk edge (the f factor).

The conclusion that quiescent states are only possible for low mass-transfer rates is very general. An analysis similar to that described above in the non irradiated case leads to the conclusion that the disk can enter into quiescence only if the mass accretion rate is less than

$$\dot{M}_{\rm acc,crit} = 4\times10^{20}\left(\frac{r_{\rm in}}{10^{14}\ {\rm cm}}\right)^2 M_8^{-2/3}\left(\frac{\alpha_h}{\alpha_c}\right)^{0.46}\ {\rm g\ s^{-1}}, \qquad (30)$$

which is accurate to within a factor 2 when compared to the results of the numerical simulations. Although the dependence on M_8, $r_{\rm in}$, and α are different from the unirradiated case, the numerical value of this critical rate is not changed to the point where one could obtain large amplitude outbursts, during which the Eddington limit would be attained or approached.

4. Conclusion

We have shown that the accretion disks in AGNs can indeed be subject to the same thermal-viscous instability as in dwarf novae and soft X-ray transients, but the outcome of this instability is very different. This contrasts with previous findings that large amplitude outbursts reaching the Eddington limit were possible, and the reason for this discrepancy is the poor spatial resolution of the numerical codes that have been used to model the disk. In AGNs, the disk opening angle is much less than in DNs or SXTs, because the Keplerian velocity is not low compared to the speed of light, whereas the sound speed is, by construction, the same in both cases. This results in very thin transition fronts, which are quite difficult to follow numerically.

We do, however, predict time variations of the AGN luminosity by a few magnitudes on time scales ranging from a few thousand (the propagation time of a thermal front in the disk) to a few million years (the typical quiescent/outburst time), and these oscillations are enhanced by a possible truncation of the innermost parts of the disk. However, because of their small amplitudes and duty cycles, these variations cannot explain the statistical properties of quasar and AGN luminosity distribution.

One should also stress that, for high mass-transfer rates, the transition front can reach regions in the disk where self-gravitation becomes important, and there the assumption of a homogeneous disk becomes quite questionable, as a result of the development of a gravitational instability that is likely to result in the fragmentation of the disk.

References

Alexander, D. R. 1975, ApJS, 29, 363
Balbus, S. A., & Hawley, J. F. 1991, ApJ, 521, 650
Balbus, S. A., & Papaloizou, J. C. B. 1999, ApJ, 376, 214
Cannizzo, J. K. 1992, ApJ, 385, 94
Cannizzo, J. K. 1993, ApJ, 419, 318
Cannizzo, J. K., & Reiff, C. M. 1992, ApJ, 385, 87
Cox, A. N., & Tabor, J. E. 1976, ApJS, 31, 271
Dubus, G., Hameury, J.-M., & Lasota, J.-P. 2001, A&A, 373, 251
Duschl, W. J. 1986, A&A, 163, 61
Duschl, W. J., & Britsch, M. 2006, A&A, 653, L89
Fontaine, G., Graboske, H. C. Jr., & Van Horn, H. M. 1977, ApJS, 35, 293
Frank, J., King, A. R., & Raine, D. 2002, Accretion Power in Astrophysics, 3rd edn. (Cambridge University Press)
Gammie, C. F. 2001, ApJ, 553, 174
Gammie, C. F., & Menou, K. 1998, ApJ, 492, L75
Goodman, J. 2003, MNRAS, 339, 937
Hameury, J. M. 2002, in The physics of cataclysmic variables and related objects, ed. B. T. Gänsicke, K. Beuermann, & K. Reinsch, ASP Conf. Ser., 261, 377
Hameury, J.-M., Lasota, J.-P., MCClintock, J. E., & Narayan, R. 1997, ApJ, 489, 234
Hameury, J.-M., Menou, K., Dubus, G., Lasota, J.-P., & Huré, J. M. 1998, MNRAS, 298, 1048
Hameury, J.-M., Lasota, J.-P., & Dubus, G. 1999, MNRAS, 303, 39
Janiuk, A., Czerny, B., Siemiginowska, A., & Szczerba, R. 2004, ApJ, 602, 595
Kato, S., Fukue, J., & Mineshige, S. 1998, Black-hole accretion disks (Kyoto University Press)
Lasota, J.-P. 2001, New Astron. Rev., 45, 449
Lasota, J.-P., Dubus, G., & Kruk, K. 2008, A&A, 486, 523
Lightman, A. P., & Eardley, D. M. 1974, ApJ, 187, L1
Lin, D. N. C., & Pringle, J. E. 1987, MNRAS, 225, 607
Lin, D. N. C., & Shields, G. 1986, ApJ, 305, 28
Mayer, M., & Pringle, J. E. 2006, MNRAS, 368, 379
Menou, K., & Quataert, E. 2001, ApJ, 552, 204
Menou, K., Hameury, J. M., & Stehle, R. 1999, MNRAS, 305, 79
Menou, K., Hameury, J. M., Lasota, J. P., & Narayan, R. 2000, MNRAS, 314, 498
Meyer, F., & Meyer-Hofmeister, E. 1981, A&A, 104, L10
Mineshige, S., & Shields, G. A. 1990, ApJ, 351, 47
Narayan, R., & McClintock, J. E. 2008, New Astron. Rev., 51, 733
Paczyński, B. 1969, Acta Astron., 19, 1
Rafikov, R. R. 2007, ApJ, 682, 542
Safronov, V. S. 1960, Ann. Astrophys., 23, 979
Shlosman, I., Begelman, M., & Frank, J. 1990, Nature, 345, 679
Siemiginowska, A., & Elvis, M. 1997, ApJ, 482, L9
Siemiginowska, A., Czerny, B., & Kostyunin, V. 1996, ApJ, 458, 491
Toomre, A. 1964, ApJ, 139, 1217
van Paradijs, J. 1996, ApJ, 464, L139
Warner, B. 1995, Cataclysmic Variable Stars (Cambridge: Cambridge University Press)