Bars within bars: a mechanism for fuelling active galactic nuclei

Isaac Shlosman*, Juhan Frank† & Mitchell C. Begelman‡

Joint Institute for Laboratory Astrophysics, University of Colorado and National Bureau of Standards, Boulder, Colorado 80309, USA

ACTIVE galactic nuclei (AGNs) are usually thought to be powered by accretion onto a supermassive black hole (SBH)[1,2]. Their luminosities, which may exceed 10^{46} erg s^{-1}, require mass accretion rates of $\gtrsim 1\, M_\odot\, \mathrm{yr}^{-1}$ for reasonable mass-to-energy conversion efficiencies. Although this quantity of fuel could be supplied by the interstellar medium of the host galaxy, it is not obvious how it could be transported from typical galactic radii, ~10 kpc, down to the scale of the SBH, $\lesssim$10 pc. We propose here a mechanism, applicable to AGNs and nuclear starburst galaxies, which brings in gas from large to small scales by successive dynamical instabilities. On the large scale, a stellar bar sweeps the interstellar medium into a gaseous disk of a few hundred parsecs in radius. Under certain conditions, this disk can become unstable again, allowing material to flow inwards until turbulent viscous processes control angular-momentum transport. This flow pattern may feed viscosity-driven accretion flows around a SBH, or lead to the formation of a SBH if none was present initially.

If the accretion flow is disk-like and is driven by local viscosity, then the radial inflow time can be expressed in terms of the standard α parameter[3], which is defined as a suitably averaged ratio of viscous stress to pressure inside the disk. For a disk with shear induced by a point-mass potential, the viscosity-controlled radial inflow time, t_{visc}, is

$$\begin{aligned} t_{\mathrm{visc}} &= \alpha^{-1}(v_\phi/c_s)^2 t_{\mathrm{dyn}} \\ &\approx 1.2\times 10^9 \alpha^{-1} v_{\phi 2} T_2^{-1} R_{10}\ \mathrm{yr} \end{aligned} \tag{1}$$

where $100 v_{\phi 2}$ km s^{-1} is the orbital speed, $10 R_{10}$ pc is the radius, t_{dyn} is the dynamical time R/v_ϕ, c_s is the sound speed in the disk, and $100 T_2$ K is the gas temperature[4,5]. An alternative viscosity formulation, obtained by setting the coefficient of kinematic viscosity to $\nu = \alpha c_s H$, where H is the disk thickness, gives identical results for a non-self-gravitating disk, but leads to a slower inflow rate when vertical self-gravity is important. Because of the effects of self-gravity and fragmentation[4-7], the disk feeding an AGN may resemble a system of gas clouds rather than a quasi-continuous fluid. In this case, the efficiency of angular-momentum transfer depends on the cloud–cloud collision timescale, t_{coll}, and the value of the effective α is of order $t_{\mathrm{dyn}}/t_{\mathrm{coll}}$. We expect α to be $\lesssim$1 at the onset of fragmentation, and to become smaller if the clouds contract further. If fragmentation is somehow suppressed, higher effective values of α may be possible[8], but a consideration of realistic heating and cooling processes suggests that this does not lead to a significant enhancement of inflow[4,5]. In regions where the stellar potential dominates over the potential of the SBH, the shear is reduced and the viscous inflow time is prolonged. We conclude, therefore that standard viscous processes are too slow to be effective in collecting gas liberated in the main body of the host galaxy. This can be overcome either if the gas is generated close to the SBH or if gas liberated at larger distances is brought in towards the SBH by a mechanism other than subsonic turbulent viscosity. For example, a large-scale non-axisymmetric disturbance of the galaxy, such as a stellar bar, is capable of inducing global shocks, thereby driving the inflow of gas towards the centre. The amount of H I in spirals is observed to lie in the range 10^8–$10^{10}\, M_\odot$, and molecular gas may be more abundant by a factor of ten. Both the total H I, CO and H_2 content and the ratio of molecular gas to H I are maximal for spiral galaxies of type Sb to Sbc[9,10]. There is certainly enough gas present in spirals to power the whole range of AGN activity, provided that this gas can be swept inwards from several kpc down to tens of pc in a short time.

It has been known for some time that barred spiral galaxies are associated with Seyfert galaxies. Adams[11] noticed that a large number of Seyferts are barred. The 12 objects in his highest-resolution class are all barred spirals at some level, or show inner rings thought to be generated by bar instabilities[12]. Apart from ten obvious SB and SAB galaxies, this class contains NGC1068 and Mk10, which are observed to be barred[13-15]. In the next resolution class, Adams found that two thirds of the galaxies are SB or have rings. The remaining third either have an "amorphous body" or are seen edge-on, or their observation was subject to instrumental errors.

In a survey of a sample of Seyfert galaxies, Simkin *et al*[16] concluded that there was an excess of bars, rings and oval distortions among Seyferts with respect to their control sample. A recent survey based on CCD images of a volume- and luminosity-limited sample by MacKenty[17] shows that Seyfert galaxies "nearly always possess mechanisms for transporting material into their nuclei (for example, peculiar, tidally interacting or barred galaxies)". In other respects the host galaxies appear normal, having colours, magnitudes and disk parameters typical of their morphological class. Clearly, the simplest interpretation of these observations is that all Seyferts are barred at some level.

The large-scale stellar bar cannot, however, be solely responsible for the AGN phenomenon. First, about half of the spiral galaxies in the sky are barred or show oval distortions whereas only a few per cent of them host AGNs (but see ref. 18). Second, the gas inflow generated by a barred rotating potential does not extend down to scales at which turbulent viscosity could take over. Numerical studies of the response of a compressible fluid to an imposed oval distortion of the potential indicate that, in these systems, the gaseous component loses angular momentum in a pair of shocks, and flows towards the centre of the galaxy in about ten rotation periods[19,20]. Given the fractional deviation of the potential from axial symmetry, $\Delta\Phi$, we can estimate the radius at which the shocks weaken and the inflow ceases to be dynamically driven. We assume that $\Delta\Phi \approx \varepsilon v_\phi^2$, where ε is of the order of 0.1. The condition for $\Delta\Phi$ to cause a shock in the gas can be expressed as[21] $c_g < 0.5(\gamma+1)\varepsilon v_\phi$, where c_g is the sound speed in the gas and γ is its adiabatic index. If the inner parts of the galaxy rotate as a solid body within a radius R_t, shocks will not be present within a radius R_d given by $R_d = 2R_t/(\gamma+1)\varepsilon\mathcal{M}$, where $\mathcal{M}$ is the Mach number (v/c_g) of the flow at R_t and we have assumed $c_g \ll v_\phi$. Numerical simulations show that the gas can be swept up into a ring if inner Lindblad resonances (ILRs) are present, yielding a ring-like configuration of H II regions. If the bar distortion is strong, however, the positions of the ILRs cannot be calculated using the circular velocity curve[22].

In summary, the gaseous inflow in a barred potential slows down at a radius R_d. We shall assume, for simplicity, that the net effect of a stellar bar on the gaseous component of the host galaxy is to sweep it into a disk with radial scale $R_d \approx 0.1 R_b$, where R_b is the radius of the bar. At R_d, viscous processes are still too slow to drive inflow. We propose instead that the gas which accumulates in the central kpc or so, as a consequence of the inflow in the stellar bar, forms a disk which, under certain conditions, becomes again dynamically unstable to form a gaseous bar. We derive a condition for this secondary instability in the context of a simple model, and argue that the resulting inflow can extend all the way into the centre of the galaxy, that is, to the inner ~10 pc or so.

No single criterion is known for global stability of self-gravi-

* Present address: Theoretical Astrophysics 130–33, California Institute of Technology, Pasadena, California 91125, USA.

† Permanent address: Max-Planck-Institut für Astrophysik, 8046 Garching bei München, FRG.

‡ Also at: Department of Astrophysical, Planetary, and Atmospheric Sciences, University of Colorado, Boulder, Colorado 80309, USA.

tating disks (gaseous or stellar) encompassing all possible disk models[22,23]. However, the stability of a large variety of disk models can be described by applying the semi-empirical criterion proposed by Ostriker and Peebles[24], which is known to have a theoretical analogue for Maclaurin ellipsoids. A sufficient condition for instability is given by $t = T_{rot}/|W| \geq t_{crit}$, where T_{rot} is the kinetic energy of rotation of the disk, W is the gravitational energy of the system and t_{crit} is the maximum value of t required for stability. This criterion seems to apply to gaseous thin disks, which have $t_{crit} \approx 0.14$ for secular instability and $t_{crit} \approx 0.26$ for dynamical instability[25,26]. We model the system after the bar has brought the gas within the radius R_d in terms of three components: a cool gaseous disk having a Kuzmin density distribution with radial scale a; a stellar disk having a similar density distribution but with a larger radial scale b; and a Plummer-model halo, also with radial scale b (see, for example, ref. 23). Let the total mass of the system be M, a fraction g of which is in the gaseous disk and a fraction s in the stellar disk. The halo therefore contains a mass $(1-g-s)M$. This particular choice of model is motivated by the fact that t can be evaluated analytically, but the conclusions drawn from it are expected to be at least qualitatively correct in a more general context.

We treat the potential of the stellar component as axisymmetric and fixed, and therefore include in the calculation of t only the kinetic energy of the gaseous disk:

$$T_{rot} = \frac{gf^2GM^2}{8a}\left[g + \frac{4(1-g)}{(1+\beta)^2}\right] \qquad (2)$$

where we have assumed that the disk rotates at every radius with a fraction f of the local circular speed for equilibrium ($f = 1$ indicates full rotational support), and have introduced $\beta \equiv b/a$, the ratio of length scales of the fixed stellar component to the gaseous component. From equation (2) and the expression for the potential energy,

$$|W| = \frac{GM^2}{4a}\left[g^2 + \frac{4g(1-g)}{1+\beta} + \frac{s(2-2g-s)}{\beta} + \frac{3\pi}{8\beta}(1-s-g)^2\right] \qquad (3)$$

one can obtain an expression for t in terms of the mass fractions in the various components and the radial-scale ratio β. The physically interesting question is: given a certain gas fraction, by how much must the gas disk shrink radially in order to become unstable? A rough answer to this question can be obtained by setting t equal to t_{crit} and solving for β. In the limit $g \ll 1$, we find that the gaseous disk is unstable ($t > t_{crit}$) if

$$\beta > \frac{2t_{crit}(1-h^2+3\pi h^2/8)}{g^2(f^2-2t_{crit})} \qquad (4)$$

where $h \equiv 1-g-s$ is the halo mass fraction. The r.h.s. of eq. (4) is not very sensitive to the value of h, and equation (4) holds to better than 10% for $g \leq 0.3$, when compared with the exact evaluation of the Ostriker–Peebles criterion. Any effects of non-axisymmetry will probably enhance the instability.

Two aspects of equation (4) are worth mentioning: (1) for a given f and h, β varies inversely with the gas fraction squared, and (2) the disk must be cold, $f > \sqrt{2t_{crit}}$ for it ever to become unstable. This places stringent limits on the proposed mechanism. For example, in a cold disk with $f = 1$, $h \approx 0.5$ and $t_{crit} = 0.14$, β must be $> 0.4/g^2$ for instability. If $\beta = 1$ initially and $g \approx 0.2$, then a reduction in size of the gas disk by a factor of ten would make it susceptible to global instabilities. It is important to note that the value of $t_{crit} = 0.14$ used above corresponds to dynamical instability of stellar (collisionless) disks, but to only secular instability of gaseous disks. This is because gaseous disks have fewer degrees of freedom than stellar disks. Secular instability of a gaseous disk would occur on the viscous inflow time, and would therefore be too slow to be of interest. Dynamical instability of a smooth gaseous disk would require $t_{crit} = 0.26$, corresponding either to contraction by an additional factor of 2–3, or to increase in g by a factor of about two. However, it is likely that the self-gravitating gaseous disk will be highly inhomogeneous, consisting of clouds which fill a rather small fraction of the volume[5]. If the collision frequency of clouds is small relative to the orbital frquency, the system will behave more like a stellar disk than a gaseous disk, and a lower value of t_{crit} will apply. Gravitational coupling of the gaseous disk to the stellar disk may also contribute to the instability[5,27].

Therefore, we have established that a large-scale stellar bar sweeping the gas inward yields a disk, which can become unstable, if it is not too hot, and can feed a viscosity-driven flow. The unstable configuration is likely to resemble a bar-driven spiral[25], which will transfer angular momentum outwards, leading to further contraction. At the same time, the velocity dispersion of inhomogeneities will tend to increase, but this energy will be dissipated through cloud–cloud collisions. To the extent that the system does not form stars, the inevitable consequence of sinking of the gaseous system towards the centre of the stellar system is that t will grow, and the gaseous system will become increasingly unstable. The degree to which the system forms stars before the gas reaches the centre must, in this model, determine whether the principal form of activity is a starburst or an AGN. The above argument implies that galaxies with a gas content of less than 10% will not display a bar within a bar and are therefore candidates for the generation of a starburst but not of a powerful AGN.

The process is not necessarily restricted to spirals. Elliptical galaxies are to some extent also supported by rotation. Material that is released at radii of a few kpc, and which is only 1% rotationally supported, would go into keplerian orbit at about one tenth of its original distance from the centre. Turbulent viscosity would still be insufficient to drive inflow, and the formation of a gaseous bar might occur even in this case.

Thus, a large-scale bar is a necessary but not sufficient condition for strong nuclear activity if the host is a spiral galaxy. Another stage of dynamical instability, which occurs in the gaseous disk that accumulates in the centre under the influence of the stellar bar, is required to feed an accretion disk at smaller scales, that is, such a process requires gaseous bars within stellar bars. A similar but tidally induced process was found in numerical simulations (L. Hernquist, preprint). Any process that suppresses this second stage of instability will prevent formation of the SBH and/or its subsequent fuelling, and is likely to lead to nuclear star formation or to low-level activity. This view is supported by recent evidence that most or perhaps all starbursts are barred[28]. The majority of spirals avoids becoming a conspicuous AGN or starburst. The most probable causes for this are relatively low gas content of the host, inefficiency of the stellar bar in sweeping all of the gas in, and the presence of ILRs.

The fact that activity seems to occur preferentially in intermediate morphological classes can be simply explained in terms of the relative abundance of gas-rich galaxies among those morphological types[9,10]. Additional factors, such as the length of the bar relative to the photometric galactic radius, may influence the correlation between activity and morphological type. It is plausible that nuclear activity in some galaxies may be suppressed at present because the SBH (or the nuclear starburst) has exhausted the available mass supply and must wait for new material to be brought in from large scales. The duration of the active phase of an AGN would be of the order of a few t_{dyn} at the bar radius, whereas the gas would be replenished only on the stellar-evolution timescale or on the timescale for capture of extragalactic clouds.

An important prediction of this model is that dynamical instabilities on scales of ~100 pc must be present in the central regions of active galaxies, including ellipticals (such as powerful radio galaxies). Perhaps some evidence of the processes discussed here can be found in the Sérsic–Pastoriza galaxies, whose nuclei have been long recognized as unusually bright and disturbed[29]. All of these galaxies are barred at some level and some of them are well-known today to be Seyferts and starbursts. More high-resolution optical and infrared observations of these objects and H I and CO observations of the innermost regions of moderately distant AGN are desirable.

We have not discussed the effects of environment on Seyferts and starbursts. The passage of a close companion may induce a stellar bar, but this may not be necessary for their production, as isolated galaxies may host bars which are relics from galaxy formation, or which have been generated spontaneously as a result of infall. □

Received 29 July 1988; accepted 9 January 1989.

1. Lynden-Bell, D. *Nature* **223**, 690–694 (1969).
2. Begelman, M. C., Blandford, R. D. & Rees, M. J. *Rev. mod. Phys.* **56**, 255–351 (1984).
3. Shakura, N. I. & Sunyaev, R. A. *Astr. Astrophys.* **24**, 337–355 (1973).
4. Shlosman, I. & Begelman, M. C. *Nature* **329**, 810–812 (1987).
5. Shlosman, I. & Begelman, M. C. *Astrophys. J.* (in the press).
6. Paczyński, B. *Acta Astr.* **28**, 91–109 (1978).
7. Shore, S. N. & White, R. L. *Astrophys. J.* **256**, 390–396 (1982).
8. Lin, D. N. C. & Pringle, J. E. *Mon. Not. R. astr. Soc.* **225**, 607–613 (1987).
9. Haynes, M. P. & Giovanelli, R. *Astr. J.* **89**, 758–800 (1984).
10. Verter, F. *Astrophys. J.* **65**, 555–580 (1987).
11. Adams, T. F. *Astrophys. J. Suppl.* **33**, 19–34 (1977).
12. Buta, R. *Astrophys. J. Suppl.* **61**, 609–630 (1986).
13. Telesco, C. M. & Decher, R. *Astrophys. J.* (submitted).
14. Scoville, N., Matthews, K., Carico, D. P. & Sanders, D. B. *Astrophys. J.* **327**, L61–L64 (1988).
15. Mazzarella, J. M. & Balzano, V. A. *Astrophys. J. Suppl.* **62**, 751–819 (1986).
16. Simkin, S., Su, H. & Schwarz, M. P. *Astrophys. J.* **237**, 404–413 (1980).
17. MacKenty, J. *Astrophys. J.* (in the press).
18. Keel, W. C. *Astrophys. J. Suppl.* **52**, 229–239 (1983).
19. Matsuda, T., Inoue, M., Sawada, K., Shima, E. & Wakamatsu, K. *Mon. Not. R. astr. Soc.* **229**, 295–314 (1987).
20. Schwartz, M. P. *Mon. Not. R. astr. Soc.* **212**, 677–686 (1985).
21. Landau, L. D. & Lifshitz, E. M. *Fluid Dynamics* 94 (Pergamon, New York, 1959).
22. Athanassoula, E. *Phys. Rep.* **114**, 329–403 (1984).
23. Binney, J. & Tremaine, S. *Galactic Dynamics* (Princeton University Press, 1987).
24. Ostriker, J. P. & Peebles, P. J. E. *Astrophys. J.* **186**, 467–480 (1973).
25. Bardeen, J. M. in *Dynamics of Stellar Systems* (ed. Hayli, A.) 297–320 (Reidel, Dordrecht, 1975).
26. Aoki, S., Noguchi, M. and Iye, M. *Publs astr. Soc. Japan* **31**, 737–774 (1987).
27. Jog, C. J. & Solomon, P. M. *Astrophys. J.* **276**, 114–126 (1984).
28. Jackson, J. M., Barrett, A. H., Armstrong, J. T. & Ho, P. T. P. *Astr. J.* **93**, 531–545 (1987).
29. Sérsic, J. L. & Pastoriza, M. *Publs astr. Soc. Pacif.* **77**, 287–289 (1965).

ACKNOWLEDGEMENTS. We are grateful to Oved Dahari, Don Osterbrock, John Papaloizou, John Stocke and Andrew S. Wilson for illuminating discussions and to Lars Hernquist for providing results before publication. This research was supported in part by NASA and the NSF and grants from Ball Aerospace Systems Division, Rockwell International Corporation and the Alfred P. Sloan Foundation. M.C.B. is a Presidential Young Investigator and Alfred P. Sloan Foundation Research Fellow.

A global assessment of natural sources of atmospheric trace metals

Jerome O. Nriagu

National Water Research Institute, Box 5050, Burlington, Ontario L7R 4A6, Canada

A PROPER inventory of atmospheric emissions from natural sources is basic to our understanding of the atmospheric cycle of the trace metals (and metalloids), and is also needed for assessing the extent of regional and global pollution by toxic metals[1]. It is generally presumed that the principal natural sources of trace metals in the atmosphere are wind-borne soil particles, volcanoes, seasalt spray and wild forest fires[2–6]. Recent studies have shown, however, that particulate organic matter is the dominant component of atmospheric aerosols in non-urban areas[7–10] and that over 60% of the airborne trace metals in forested regions can be attributed to aerosols of biogenic origin[11,12]. Here I estimate that biogenic sources can account for 30–50% of the global baseline emissions of trace metals. For most of the toxic metals, the natural fluxes are small compared with emissions from industrial activities, implying that mankind has become the key agent in the global atmospheric cycle of trace metals and metalloids.

In compiling the emission factors used in the calculations (Table 1), the most recent data have been used; for older data, the lowest reported concentrations have been preferred. For enrichment factors, a range of values[1], rather than a single value, has been used. A range in the flux of material from each source has also been employed so as to assess the errors inherent in the reported emission intensities of trace metals.

The episodic nature of volcanic emissions makes it difficult to derive trace-metal emission rates using the enrichment-factor strategy[12,13]. There have been numerous studies, however, of the release of sulphur and trace metals from volcanoes and fumaroles, and a number of authors have estimated the outputs of trace metals by normalizing the metal release to the sulphur flux from this source[15]. I have adopted such an approach here, using a global volcanic sulphur flux of $(15\text{-}50)\times 10^{12}\ \mathrm{g\ yr^{-1}}$ (refs 15, 16) and metal-to-sulphur ratios derived from published data[13–17].

The release of large quantities of volatile non-methane hydrocarbons (NMHC) is well documented, particularly in forested areas[18–20]. For example, the average emission rates for NMHCs in the forested areas of the United States range from 450 to $1{,}712\ \mu\mathrm{g\ m^{-2}\ h^{-1}}$ (ref. 18). Biogenic emissions of isoprene and terpenes in the Amazon rain-forests have been estimated as 233 and $1{,}040\ \mu\mathrm{g\ m^{-2}\ h^{-1}}$ respectively[19]. Rasmussen and Khalil[20] have calculated the global atmospheric flux of isoprene to be $450\times 10^{12}\ \mathrm{g\ yr^{-1}}$, and the combined releases of isoprene and terpenes have been suggested[19] to be equivalent to ~0.7% of the net global primary production. Thus the NMHC flux of $100\text{-}500\ \mu\mathrm{g\ m^{-2}\ h^{-1}}$ used here is clearly a conservative figure. Terpenes and isoprene, the two dominant components of NMHCs, are known to form strong complexes with many trace metals[21], and thus should play a part in the transfer of metals to the atmosphere. As far as I know, however, the NMHCs have not been analysed for trace metals, so that the emission factors listed in Table 1 are subject to a large uncertainty. They are based on reported flux measurements and the concentrations of methylated compounds of Hg, As, Pb and Se in the atmosphere[6,14,24]. For the other elements, the emission factors are based on metal concentrations in the surface organic microlayers of aquatic ecosystems[22,23].

The average concentration of particulate organic carbon (POC; derived from organic matter which is the dominant component of atmospheric aerosols) in the Amazon rain-forests has been reported[9] to be $8.9\ \mu\mathrm{g\ C\ m^{-3}}$ in the mixed layer and $2.6\ \mu\mathrm{g\ C\ m^{-3}}$ in the free troposphere. These reported POC concentrations in forest ecosystems are considerably higher than the mean value of $0.06\ \mu\mathrm{g\ C\ m^{-3}}$ in the marine atmosphere of the Southern Hemisphere[44]. The carbon isotope composition suggests, however, that most of the marine POC is also derived from natural land-based sources[44]. The annual flux of particulate organic matter (POM) to the atmosphere, Q, can be estimated from

$$Q = V\times B\times 2.2\times(3\text{-}6)\times 10^{13}$$

where V, the total deposition velocity of the aerosols, is estimated to be $1.0\ \mathrm{cm\ s^{-1}}$ (ref. 44), and b, the average POC concentration, is $3\text{-}9\ \mu\mathrm{g\ m^{-3}}$. The figure of 2.2 is used to relate organic carbon to organic matter[47], and the forested area is believed to be $(3\text{-}6)\times 10^{13}\ \mathrm{m^2}$ (ref. 51). These values yield a global atmospheric POM flux of about $(1\text{-}5)\times 10^{14}\ \mathrm{g\ yr^{-1}}$. The POM flux used here is thus higher than the value of $27\times 10^{12}\ \mathrm{g\ yr^{-1}}$ reported recently by Cachier *et al.*[44]. The global flux of material from the other non-biogenic sources are summarized in Table 1.

There is a wide range in the estimated total emission of each metal from any given source (Table 1), reflecting the large dispersion in the emission factors used and in the published estimates of the mass flux from individual sources. There are significant gaps in the data, demonstrating the need for further study of the role of biological processes in the emission of trace metals. The proposed limits on the fluxes of metals from each source serve to indicate that current data are sufficient for only order-of-magnitude estimates of the global emission intensities. In view of this, the central (or median) values of the ranges