COSMOLOGICAL MESTEL DISKS AND THE ROSSBY VORTEX INSTABILITY: THE ORIGIN OF SUPERMASSIVE BLACK HOLES

Stirling A. Colgate,1 Renyue Cen,2 Hui Li,1 Nathan Currier,1 and Michael S. Warren1

Received 2003 March 31; accepted 2003 October 6; published 2003 October 31

ABSTRACT

A scenario is put forth for the formation of supermassive black holes at the centers of galaxies. It depends upon the formation of a Mestel disk with a flat rotation curve, \( M_\odot \propto r \) and \( \Sigma \propto 1/r \). Such disks could form from the collapse of uniformly rotating, isolated, gaseous clouds, either protogalactic, galaxy-mass damped Ly\( \alpha \) clouds or the gas that survives galaxy mergers. We propose that in either case the disk will be unstable to the Rossby vortex instability (RVI). This instability grows from any large, steep pressure gradient in an optically thick disk. Such pressure gradients either occur adjacent to compact objects or could be triggered by heating from individual supernovae in and around the disk. Upon excitation, the RVI transports angular momentum outward, accreting nearly all mass within the initiation radius. We have calculated that in very thin disks, the nonlinear vortices initiated by the RVI can transport angular momentum far more efficiently than turbulence. Compared to a viscosity-based Shakura-Sunyaev disk, the RVI transports angular momentum out to a much larger radius, so more mass is accreted into the central black hole. A typical galaxy rotational velocity is \( v_{\text{rot}} = 200 \text{ km s}^{-1} \), and the critical column density, necessary to initiate the RVI, is \( \Sigma_{\text{CCD}} \approx 100 \text{ g cm}^{-2} \). For \( M_\odot = 2\pi r^2 \Sigma \), we have \( \Sigma_{\text{CCD}} = v_{\text{rot}}^2/(2\pi \Sigma_{\text{CCD}} G) \), and the mass accreted becomes \( M_{\text{BH}} = v_{\text{rot}}^2/(2\pi \Sigma_{\text{CCD}} G^2) = 3 \times 10^7 M_\odot \). Both the black hole mass \( M_{\text{BH}} \) and its \( v_{\text{rot}} \) dependence are in good agreement with recent observations, because \( v_{\text{rot}} = \gamma^3 \sigma_\perp \), where \( \sigma_\perp \) is the velocity dispersion of the bulge at the radius of mutual contact.

Subject headings: accretion, accretion disks — black hole physics — galaxies: formation — galaxies: kinematics and dynamics — hydrodynamics — instabilities

1. INTRODUCTION

There is mounting evidence that nonactive galaxies, including our own, harbor central supermassive black holes (SMBHs) with masses from \( 10^6 \) to \( 10^9 \) \( M_\odot \) (Richstone et al. 1998; Tremaine et al. 2002). Furthermore, observations indicate a close correlation between the SMBH mass and the velocity dispersion, \( \sigma_\perp \). How and when the SMBHs form has remained unknown, although there are many possibilities (Rees 1984; Ostriker 2000; Adams, Graff, & Richstone 2001; Loeb 1993; Eisenstein & Loeb 1995; Umemura, Loeb, & Turner 1993).

Mestel (1963) developed a model of cloud collapse that preserves specific angular momentum. His spherical cloud collapsed to a disk with a nearly flat rotation curve, but a uniformly rotating inviscid cylindrical cloud should collapse to a disk with a perfect flat rotation curve. This mass distribution, \( M_\odot \propto r \), should extend inward to where pressure forces become important, e.g., where the mass enclosed within a solid radius \( r_\odot \) becomes \( M \approx 0.2 M_\odot \). The Helmholz cooling time at every radius is \( \sim 10^4 \text{ yr} \), which is short compared to the black hole (BH) formation time of \( 10^8 \text{ yr} \), so the disk is thin. Turbulence has been invoked to form a Shakura-Sunyaev (SS) disk at larger radii. However, for a \( 10^8 M_\odot \) BH, an SS accretion disk would become subject to the self-gravitating instability (SGI) at \( \sim 0.01 \text{ pc} \) (Begelman, Frank, & Shlosman 1989; Shlosman et al. 1990; Goodman 2003; Shakura & Sunyaev 1973, Gammie 2001). At 0.08 pc, the disk would have more mass than the BH, because of the excessive column density of the SS disk, required to transport angular momentum at a rate corresponding to near Eddington luminosity. Hence, even if the SGI turbulence is excited and maintained, the disk mass is still greatly excessive. The large-scale vortices of the Rossby vortex instability (RVI) disk are one solution to this problem.

An Eddington limit (1 \( M_\odot \) yr\(^{-1} \)) Thompson opacity SS disk surrounding a central mass of \( M_\odot \approx 10^8 M_\odot \) results in \( H_{\odot}/r \approx 10^{-4} \) (Shakura & Sunyaev 1973; Frank, King, & Raine 1985; Goodman 2003; Papaloizou & Pringle 1984). Therefore, with the critical column density (CCD) necessary to initiate the RVI at \( \Sigma_{\text{CCD}} \approx 10^8 M_\odot \), all turbulent viscosity mechanisms, no matter how robust, will lead to a disk mass too large for the disk to be consistent with the observed mass, growth time, and luminosity of active galactic nuclei. The RVI disk does not rely on turbulence but on vortices with transport lengths of \( l \sim r/10 \), which is much greater than \( l \sim H \sim r/10^4 \) for the SS disk.

2. THE COLLAPSE OF A GASEOUS CLOUD AND FORMATION OF A MESTEL DISK

A self-gravitating cloud is parameterized by its mass, \( M_\star = (4\pi/3)\rho_r^3 \), initial (virial) radius, \( r_v \), initial (virial) velocity dispersion, \( \sigma_v \), and the dimensionless spin parameter,

\[
\lambda \equiv \frac{J}{|E|^{1/2} G^{-1} M_\star}^{5/2}.
\]  

Here \( J \) is the total angular momentum, and \( |E| = \int M_\odot \times (G/r) dM \) describes the potential energy (Peebles 1969). For a lone test particle with tangential velocity \( v_t \) as it orbits a fixed mass, equation (1) gives \( \lambda = v_t v_r / v_\perp = 1 \), where \( v_\perp \) is the Keplerian velocity. We are interested in the final collapsed state of an initially rotating cloud. The mass at its outer boundary will form the outer boundary of the equilibrium disk. However, simulations calculate the mass average of \( \lambda \), \( \lambda_{\text{av}} \), not its value...
at the outer surface. For our purposes we renormalize $\lambda$ to $\lambda_{eq}$ so that a particle in a Keplerian orbit tracing an object's equator has $\lambda_{eq} = 1$.

We also greatly simplify Mestel's analysis by treating a sphere as a right circular cylinder with the sphere's mass, uniform density, and equal total angular momentum. The cylindrical approximation introduces only a 3% change in radius, yet the expected variation in precollapse cloud shapes is very much greater. The advantage of the cylindrical approximation is that in Mestel's model it forms a flat rotation disk, because each cylindrical shell at $r$ can be treated as a separate test particle that collapses to its own Keplerian radius, $r_{K}$. For a cylinder rotating as a solid body, the specific angular momentum starts as $j(r) \propto r^2$ and $M(<r) \propto r^2$, so $j(r) \propto M(<r)$. After collapse, $j(r_c) \propto (M_{eq})^{1/2}$, so $M \propto (M_{eq})^{1/2}$, and one obtains $M \propto r_{K}$; i.e., a flat rotation curve as seen in the calculations of Bullock et al. (2001) and R. Cen et al. (2003, in preparation).

Using our equatorial definition of $\lambda_{eq}$, a typical gaseous cloud, collapsing with no transport of angular momentum, becomes rotationally supported at $r_c \sim \lambda_{eq} \rho_{cloud}^{-1/2} = 1 \text{ kpc}$ starting from a typical cloud with $\rho_{cloud} = 300 \text{ pc}, v_c \sim 100 \text{ km s}^{-1}$, and $M_{eq} = 10^{12} M_{\odot}$ of dark and (15%) baryonic matter. A typical value of $\lambda_{eq}$ for a cloud formed in cosmological structure formation is 0.05 (Warren et al. 1992; Steinmetz & Bartelmann 1995; R. Cen et al. 2003, in preparation). This radius of support is about 1/10 of a typical $L^*$ galaxy radius of 10 kpc, a well-recognized dilemma, leaving in doubt this explanation of BH formation based on the transition from inviscid to vortex flow at the critical column density. However, we see two factors that should reduce this discrepancy: (1) the conversion from $\lambda_{eq}$ to $\lambda_{eq}$ and (2) the reduction of interior mass as the dark matter and baryonic matter separate during the collapse.

For a sphere, the mass averaged $\lambda_{eq} = 0.31$ if the equator rotates at the Keplerian velocity. This is because an integration of $\lambda$ combines the moment of inertia of a sphere, $I = (2/5)M R^2 \omega$, the self-energy $|E| = (3/5)M^2 G/R \omega = |E|^{1/2}R$, and $v_{eq} = \omega R$. Thus the disk radius predicted by $r_{eq}$ above should be larger by $(1/0.31)^2 = 10.4$, giving agreement with observed disk size because of this correction alone. However, a further correction may be made for the decrease of total angular momentum as the baryonic matter falls deep inside the dark matter Navarro-Frenk-White distribution, $\rho(r) \rho_{CDM} = b \rho [(r/r_c)(1 + r/r_c)]^3$. For $r \ll r_c$, the interior mass of dark matter decreases as $M_{eq} \propto r^2$, but the baryonic Mestel disk follows $M_{eq} \propto r$, and so we expect $\rho_{dark} \propto \rho_{CDM}$. We note that the collapsed radius depends sensitively on the initial spin parameter $\lambda_{eq}$, which implies that the total amount of gas and hence the SMBH masses could have a large dispersion, corresponding to the dispersion in $\lambda_{eq}$.

### 2.1. Making a Gaseous Disk by Galaxy Mergers

In addition to the collapse of a single gas cloud as described above, in galaxy-galaxy mergers the gas fraction of the progenitors should collide and make a uniformly rotating gas cloud. When the two clouds collide, they produce a gas pressure corresponding to the kinetic energy of collision. This occurs before their subsequent collapse by cooling. The $\lambda_{eq}$ of the now combined cloud will depend sensitively on whether the collision was prograde or retrograde (van den Bosch et al. 2002), but in either case a merged cloud will rotate as uniform body, collapse, and form a disk with nearly flat rotation.

A typical bulge might have $10^{10} M_{\odot}$ or 10% of the mass of a massive galaxy. However, because we expect mergers to take place after initial galaxy formation, the gas fraction should be $10^3 M_{\odot}$ or $\sim 10\%$ of the bulge stars. The outer boundary of a typical bulge merges with the flat disk at $\sim 500 \text{ pc}$ where the velocity is Keplerian. The relaxation of a uniformly rotating bulge of stars and gas should leave the merged stars in inclined orbits, thus forming the visible bulge. The gas will collapse, shock, and cool just as in the collapse of the original cloud, forming a subdisk of $10^3 M_{\odot}$ with a flat rotation curve within $\sim 500 \text{ pc}$. Barnes & Hernquist (1991) and Mihos & Hernquist (1996) have simulated such a merger producing a more compact mass inside $\sim 200 \text{ pc}$.

### 3. The RVI in a Mestel Disk

The RVI is a global hydrodynamic instability in thin disks, excited by a radial extremism in an entropy-modified version of potential vorticity (Lovelace et al. 1999; Li et al. 2000). Nonlinear evolution has been studied using global two-dimensional hydrodynamic simulations in which the flow of matter through the vortices is well illustrated in Figure 6 of Li et al. (2001). One may rightfully ask why the SGI is not sufficient to transport the angular momentum as Gammie (2001) has discussed, but if the RVI works, then the strong shocks may disperse slowly growing self-gravity perturbations. In fact, the SGI might even be useful, because it could help initiate the RVI. The RVI, once initiated, has been shown to efficiently transport angular momentum outward, especially by the large-scale vortices that it produces. In order to initially excite the instability, there must be a large enough local pressure gradient associated with the potential vorticity extremum. The RVI can be initiated provided the pressure gradient condition is met. Also, the column density of the disk has to be high enough so that locally generated heat can be trapped for at least several orbits before radiative cooling damps the fluid velocity stresses. This thickness condition is actually quite generic for many hydrodynamic instabilities in disks since they rely on pressure forces to drive fluid motions. The only hurdle is the need for positively correlated Reynolds stresses for outward transport of angular momentum.

The growth rate of the RVI depends on the sound speed, but if the collapse were to reach equilibrium at a temperature of $\sim 10^4$, the sound speed would be $c_s \sim 10^7 \text{ cm s}^{-1}$. This low sound speed would combine with an orbit time of $t_{orb} \approx 3 \times 10^5 \text{ yr}$ at 10 pc to yield an RVI growth time of $\sim 10^9$ yr, which is unrealistically long. However, during the initial collapse there is both a time and a radius at which the column density is great enough to trap a major fraction of the heat of collapse, allowing for rapid growth of the RVI. This would ideally lead to a sound speed $c_s = (\gamma_{adiab} - 1)^{1/2} = v_f/3$, where $\gamma_{adiab} = 4/3$. A disk with $\Sigma = 100 \text{ g cm}^{-2}$ collapsing at 1 pc radius is likely to be initially supported by radiation pressure when collapsing through a disk height $H \approx 0.1 \text{ r},$ with a corresponding $T \approx (\rho c_s^2)^{1/4} \approx (v_f/10)^3 \approx 740 \text{ K}$. At this temperature the opacity is of the order of $\kappa \approx 1 \text{ cm}^2 \text{ g}^{-1}$, and the radiation pressure is $\approx 10 \times n_k T$. Consequently, the cooling time becomes $t_{cool} \approx H(\Sigma k)(c/3) = 10^5 \text{ yr}$, far too fast for the RVI to be initiated. However, as the disk cools to 100 K, the height shrinks to $H \approx 0.01 \text{ r}$, the disk becomes supported by matter pressure below $T = 340 \text{ K}$, and with an opacity of $\kappa \approx 1 \text{ cm}^2 \text{ g}^{-1}$, the cooling time increases to $\tau_{cool} \approx H(\Sigma k)(c/3)(TC_p\rho a T^4) = 0.4 \text{ yr}^{-1}$. If we consider somewhat greater thickness at correspondingly smaller $r$, then the cooling time scales as $\tau_{cool} \approx H(\Sigma k)(c/3)(TC_p\rho a T^4) = (rl_{100})^{1/3} \Omega^{-1} = (\Sigma/\Sigma_{sph})^{1/3} \Omega^{-1}$. Thus there is a thickness $100 \text{ k} < \Sigma_{RVI} < 1000 \text{ k} \text{ cm}^2$ where the RVI should maintain itself with an instability.
growth time proportional to \((r/r_{\text{iso}})^3\), which is a fraction of the SMBH formation time. Three-dimensional simulations are planned, a laboratory experiment has been proposed (Colgate & Buchler 2000) based on evidence that Rossby vortices dominate the transport of angular momentum in Earth’s atmosphere.

3.1. Angular Momentum Transport and Triggering of the RVI

Li et al. (2001) have studied various initial disk pressure profiles that are unstable to the RVI in a disk where \(\Sigma\) is uniform. (Since the initiating vortices have radii \(r \ll r\), we expect the Mestel disk, with \(\Sigma \propto r\), to behave similarly.) The RVI can have a growth rate of \((0.1-0.3)\Omega\) when the initiating scale length associated with the local pressure gradient \(P/(dP/dr)\) is approximately \(3H\) (Fig. 6, Li et al. 2001). Large pressure gradients produce strong nonlinear vortices with strong shocks. In these calculations the shocks reach pressure ratios of \(P_\text{shock}/P \approx 3p_c\) with \(c_s \approx 0.1c_s\). The pressure contrast of such shocks is large compared to that expected in strong turbulence where \(c_{\text{shock}} \leq 0.3c_s\), so \(P_\text{shock}/P \approx 3p_c/c_s\sim 30\) . Thus we postulate that the RVI generates the necessary pressure contrast in shocks such that the large-scale \(l = r/10\) transport of angular momentum persists regardless of how thin the disk is.

The torque produced by the RVI becomes \(H_{\text{RVI}} = (r^2H)(P/l3\). The ratio of torque produced by the RVI to that produced by turbulent viscosity is \(H_{\text{RVI}}/H_{\text{visc}} \approx 0.03P(r/\rho c_s^2)H/r\), where \(H_{\text{visc}} = (r^2H)(\rho c_s^2)/9H/r\). Both azimuthal pressure gradients \(P/l3\) and \(\rho(r/c_s)^2H/r\) act with a torque arm \(r\) over an area \(Hr\), but the turbulent stress is smaller because the turbulent eddy scale is limited to \(H\) (Pringle 1981). Thus in the special case of a massive BH’s accretion disk where \(H \approx 10r\), the RVI torque may be greater than the SS turbulent torque by a factor of \(\sim 10^4\).

During the cloud collapse, we expect the SMBH to be seeded by a BH from a giant central star. Each new stage of the star’s evolution causes a steep pressure gradient at the inner boundary of the disk. If the RVI starts on the inside, it must progress to the outside where all the mass resides. There should also be a large dispersion in SMBH masses since the variance-to-mean ratio of \(\lambda\) from simulations is \(\sigma_\lambda = 0.4\) (Warren et al. 1992), and \(M_{\text{BH}} \propto M^2\lambda^3\). We note that for a cloud with uniform density and uniform rotation, any subregion has the same \(\lambda\) as the whole cloud, and so the collapse of subregions will follow equation (2).

The rotation velocity of the disk \(v_{\text{rot}}\) is related to the disk column density by \(v_{\text{rot}}^2 = 2\pi G\Sigma_0r_0\), where \(G\) is the gravitational constant. From equation (2), the mass within a fixed column density \(\Sigma_{\text{ccd}}\) of such a disk is

\[
M(\Sigma_{\text{ccd}}) = \frac{1}{2G} \frac{M_0^2}{\pi \Sigma_{\text{ccd}} (\lambda_{\text{ccd}} r_0)^2}.
\]

5. DISCUSSION AND CONCLUSIONS

We have proposed a simple model for the formation of a SMBH, which is synchronous with the formation of its host galaxy from the gravitational collapse of a massive cloud. We rely on two assumptions: (1) the initial local conservation of angular momentum leading to a flat rotation disk and (2) the subsequent initiation of the transport of angular momentum at a given column density of the disk. Together, these lead to a simple explanation of the previously puzzling mass of the SMBH and its mass-\(\sigma\) relation. We have related the critical column density for the onset of angular momentum transport with our work on a large-scale instability in accretion disks, the RVI. Star formation may follow exactly the same scenario in molecular clouds as SMBHs in Ly\(\alpha\) clouds. There is growing evidence that gravitational tidal torquing of “cores” in star formation regions of molecular clouds leads to the same values of \(\lambda_{\text{ccd}} \approx 0.05\) as in Ly\(\alpha\) clouds for galaxies. Then the mass of the cores within the RVI limit becomes \(\Sigma_{\text{ccd}} \pi r_{\text{ccd}}^2 \approx 1 M_\odot\) for \(r_{\text{ccd}} = 0.04\) pc and \(\lambda_{\text{ccd}} = 0.05\).
Gaseous disks formed at the centers of galaxies are not exactly flat rotation disks even after gas collisions following mergers. Furthermore, mergers of binary SMBHs are expected to evolve somewhat differently because of differing mass ratios and impact parameters. These effects as well as the dispersion in $\lambda_*$ will introduce a finite dispersion in the bulge velocity, as the $M_{\text{BH}}$-$\sigma$ relation indicates. This cosmological explanation of the $M_{\text{BH}}$-$\sigma$ relation will be explored in a subsequent paper deriving the SMBH mass distribution from the $\lambda_*$ and Press-Schechter distributions.

The authors are deeply indebted to Ethan Vishniac as reviewer. One of us, R. C., would like to thank Jeremy Goodman. This research was performed under the auspices of the Department of Energy. It was supported by the Laboratory Directed Research and Development Program at Los Alamos and in part by grant AST 02-06299. S. A. C. and H. L. acknowledge the hospitality of the Aspen Center for Physics while part of the research was carried out during the summer of 2002.

REFERENCES

Adams, F. C., Graff, D. S., & Richstone, D. O. 2001, ApJ, 551, L31
Barnes, J., & Hernquist, L. 1991, ApJ, 370, L65
Begelman, M. C., Frank, J., & Shlosman, I. 1989, in Theory of Accretion Disks, ed. F. Meyer et al. (NATO ASI Ser. C, 290; Dordrecht: Kluwer), 373
Bragg, A. E., Greenhill, L. J., & Moran, J. M. 2000, ApJ, 535, 73
Bullock, J. S., Dekel, A., Kolatt, T. S., Kravtsov, A. V., Klypin, A. A., Porciani, C., & Primack, J. R. 2001, ApJ, 555, 240
Claussen, M. J., Heiligman, G. M., & Lo, K.-Y. 1984, Nature, 310, 298
Colgate, S. A., & Buchler, J. R. 2000, in Ann. NY Acad. Sci., 898, Coherent Transport of Angular Momentum: The Ranque-Hilsch Tube as a Paradigm, ed. R. Buchler & H. Kantrup, 105
Eisenstein, D. J., & Loeb, A. 1995, ApJ, 443, 11
Ferrarese, L. 2002, ApJ, 578, 90
Frank, J., King, A. R., & Raine, D. J. 1985, Accretion Power in Astrophysics (Cambridge: Cambridge Univ. Press)
Gammie, C. F. 2001, ApJ, 553, 174
Goodman, J. 2003, MNRAS, 339, 937
Harkel, C., Güsten, R., Wilson, T. L., Bierman, P., Downes, D., & Thum, C. 1984, A&A, 141, L1
Li, H., Colgate, S. A., Wendroff, B., & Liska, R. 2001, ApJ, 551, 874
Li, H., Finn, J. M., Lovelace, R. V. E., & Colgate, S. A. 2000, ApJ, 533, 1023
Loeb, A. 1993, ApJ, 403, 542
Lovelace, R. V. E., Li, H., Colgate, S. A., & Nelson, A. F. 1999, ApJ, 513, 805
Mestel, L. 1963, MNRAS, 126, 553
Mihos, J. C., & Hernquist, L. 1996, ApJ, 464, 641
Ostriker, J. P. 2000, Phys. Rev. Lett., 84, 5258
Papaloizou, J. C. B., & Pringle, J. E. 1984, MNRAS, 208, 721
Peebles, P. J. E. 1969, ApJ, 155, 393
Pringle, J. E. 1981, ARA&A, 19, 137
Rees, M. J. 1984, ARA&A, 22, 471
Richstone, D., et al. 1998, Nature, 395, A14
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
Shlosman, I., Begelman, M. C., & Frank, J. 1990, Nature, 345, 679
Steinmetz, M., & Bartelmann, M. 1995, MNRAS, 272, 570
Tremaine, S., et al. 2002, ApJ, 574, 740
Umemura, M., Loeb, A., & Turner, E. L. 1993, ApJ, 419, 459
van den Bosch, F. C., Abel, T., Croft, R. A. C., Hernquist, L., & White, S. D. M. 2002, ApJ, 576, 21
Warren, M. S., Quinn, P. J., Salmon, J. K., & Zurek, W. H. 1992, ApJ, 399, 405