Carnegie Observatories Astrophysics Series, Vol. 1:
Coevolution of Black Holes and Galaxies, 2004
ed. L. C. Ho (Cambridge: Cambridge Univ. Press)

Accretion onto Black Holes

ROGER D. BLANDFORD
California Institute of Technology

Abstract

Recent improvements in the observational study of quasar luminosity functions and the relationship between the masses of massive black holes and the properties of their host galaxies motivate the development of a more specific and quantitative model of accretion onto black holes. Some recent ideas in the theory of black hole accretion are summarized. Three cases are distinguished. A: The mass supply rate is well below the Eddington rate, the flow is adiabatic, and most mass is driven away by a combination of thermal and hydromagnetic stress. B: The mass supply rate is comparable with the Eddington rate, and the gas is mostly accreted with high radiative efficiency. C: The mass supply rate is high, the flow is again adiabatic, and most of the mass supplied is blown away by radiation pressure. A fluid model of adiabatic accretion is described, and it is conjectured that it captures the essential dynamical features of magnetic and radiative accretion. A procedure for incorporating the results from existing and future numerical simulations of accretion disks into global models is outlined. In particular, it is suggested that the boundary conditions at large and small radii determine the flow of mass, angular momentum, and energy at intermediate radii.

This (over)simple model of accretion implies a quantitative relationship between the time integral of the quasar luminosity function and the local mass function of dormant holes that ought to be testable.

1.1 The Cosmological Context

Although the notion that active galactic nuclei (AGNs, including quasars within this class) are powered by accretion onto black holes has been around since the discovery of quasars (Salpeter 1964; Zel'dovich & Novikov 1964), we have only recently developed a quantitative understanding of their relationship to their host galaxies and their role in galaxy evolution. In fact, we see quasars as far as the most distant galaxies, and four examples have been found with redshifts between $z = 6$ and 6.4 (Fan 2004). These high-redshift quasars are surprisingly luminous and indicate that holes more massive than a billion solar masses can accumulate in less than a Gyr (or the Eddington limit can be broken at will). Holes of this size cannot fail to exert a major influence on the dynamics and thermal properties of the nascent galaxy, and may even dictate its morphology (Blandford 1999).

AGNs, or more specifically quasars, are collectively and individually powerful. They can account for several percent of the luminosity density of the Universe (Fabian, this volume). A bright quasar radiates a bolometric power of $\sim 10^{13-14}\, L_\odot$ and can release enough energy

($\sim 10^{62}$ erg) to unbind $\sim 10^{12}\ M_\odot$ of interstellar gas a hundred times over. Equivalently, the $\sim 10^{73}$ ultraviolet photons that are released can reionize intergalactic gas out to a radius of ~ 30 Mpc, and so quasars contribute to the reionization of the intergalactic medium, particularly at later epochs, and therefore influence subsequent generations of galaxies. For these and other reasons, AGN are now seen as being in symbiosis with galaxies (Haehnelt, Somerville, this volume).

The study of accretion onto black holes has become much more quantitative in recent years, as it has become possible to relate the mass of the central black hole to the properties of the surrounding galaxy. The most recent version of this relationship has been the strong correlation reported between the hole mass and the bulge velocity dispersion (Richstone, this volume). A careful application of this relation to compare the local energy density of radiant energy from the totality of accreting black holes (corrected for redshift of the photon energies) to the mass density in local black holes leads to the conclusion that the overall radiative efficiency of accretion onto black holes is $\sim 0.2-0.3$ (Yu & Tremaine 2002; Yu, this volume). In other words, it looks very much as though $(2-3)\times 10^{20}$ erg of energy is produced for every gram by which a black hole increases its mass in a luminous AGN. This is a very powerful, though not unexpected, constraint on the nature of the accretion process.

1.2 Modes of Disk Accretion

I will follow Blandford & Begelman (1999, 2004) and take a simple-minded approach to describing the accretion of gas onto a black hole of mass M. We suppose that there is always sufficient angular momentum for the mass supplied to form a viscous disk. We distinguish three modes of accretion. These can be identified semi-quantitatively by the ratio of the mass supply rate to the Eddington rate. Accordingly, we define

$$\dot{m} = \frac{\dot{M}\sigma_T c}{4\pi G M m_p} = \frac{\dot{M} t_E}{M}, \tag{1.1}$$

where the Eddington time $t_E \approx 400$ Myr. Note that we distinguish the mass supply rate from the accretion rate onto the black hole, from which it may differ by many orders of magnitude.

1.2.1 Case A: Low Mass Supply Rate

When $\dot{m} < m_1$, with $m_1 \approx 0.3$, it appears that the energy dissipation associated with the viscous torque does not heat the electrons efficiently close to the black hole. In practice, it is supposed that the energy either is transported way from the disk by large-scale magnetic fields or is absorbed by the ions, which are coupled to the electrons by collisionless processes that are not more effective than Coulomb scattering. There is now quite a lot of phenomenological support for this conclusion from analyses of supernova remnant shock fronts, accreting systems, and laboratory plasmas, as well as theory (e.g., Quataert & Gruzinov 2000), although the details remain to be filled in. When $\dot{m} \ll \dot{m}_1$, this results in a very low radiative efficiency. For example, it appears that the rate of mass supply to the black hole at the center of our Galaxy is at least $\sim 10^{21}$ g s^{-1}, equivalent to $m_1 \sim 0.003$ for a three million solar mass hole (Ghez, this volume). Yet the bolometric luminosity is only $\sim 10^{36}$ erg s^{-1}, giving an apparent radiative efficiency of $\gtrsim 10^{15}$ erg g$^{-1} \approx 10^{-6}c^2$, hardly a good advertisement for "gravity power"! I will shortly argue that this efficiency is misleading, but the qualitative point remains true. Under these circumstances one can model the flow quite accurately by ignoring the cooling all together. We call this adiabatic accretion, in parallel with the usage in describing supernova remnants. The pressure is dominated by the ions, and the relevant specific heat ratio is $\gamma = 5/3$.

1.2.2 Case B: Intermediate Mass Supply Rate

When $\dot{m}_1 < \dot{m} < \dot{m}_2$, with m_2 estimated as ~ 30, a conventional, radiative accretion disk is supposed to form. Here the viscous torque G in the disk dissipates energy at a rate $-Gd\Omega/dr$, where Ω is the angular velocity, and transports energy outward through the disk at a rate $G\Omega$. The dissipated energy is converted into radiation locally within the disk, and this radiation escapes. The innermost parts of the disk, where most of the energy is liberated, are dominated by radiation pressure and electron scattering opacity (e.g., Frank, King, & Raine 1992). The overall radiative efficiency is $\sim 0.1-0.3$, as discussed above. The flow is radiative and most bright AGNs are probably in this state.

There has been quite a lot of progress in describing radiative accretion disk models in recent years. As most of the energy is released close to the hole, it is important to decide if the hole is spinning rapidly or not. This is because the disk can be stable closer to the horizon when the specific angular momentum of the hole approaches its maximal value $a = m$ in geometrical units. This increases the limiting binding energy of the disk and the overall radiative efficiency. The best direct evidence that we have that holes do spin rapidly is provided by *Chandra* and *XMM-Newton* observations of broad iron lines in the X-ray spectra of nearby Seyfert galaxies (e.g., Fabian 2002). It is generally envisaged that the disk itself creates a relatively soft, ultraviolet corona, which is Comptonized to form a power-law spectrum by hot electrons in an active disk corona.

The situation has become somewhat more complicated since the first *ASCA* observations, but there are still several *bona fide* observations of very broad lines, most notably by Wilms et al. (2001) in MCG –6-30-15. In the case of this last investigation, not only was it concluded that the hole had to be spinning rapidly but that energy had to be extracted from the spinning Kerr spacetime to power the line emission. However, these lines do not seem to be varying in response to the variation of the continuum that excites them through resonance fluorescence. This is unfortunate, as it had been hoped to perform reverberation analyses that would have improved the description of the disk geometry.

Another area where there has been good progress is in the creation of disk atmosphere models (e.g., Blaes et al. 2001; Hubeny et al. 2001). These are now able to reproduce the observed spectra of bright quasars like 3C273 quite well and provide good evidence radiative disks really do form.

1.2.3 Case C: High Mass Supply Rate

When $\dot{m}_2 < \dot{m}$, the mass supply rate is so high that, although there is no difficulty in emitting radiation, the photons are trapped in the gas as it flows inward onto the black hole and cannot escape. This trapping is effective out to a radius $\sim \dot{m}m$, where m is the gravitational radius and $\dot{m}$ is to be interpreted as the scaled local mass accretion rate. Broad absorption-line quasars, and perhaps most radio-quiet quasars, are examples of Case C accretion. (Galactic objects like SS433 and GRS1915+115 may also provide smaller scale examples of accretion flows in this regime.) For the purposes of understanding the gas dynamics, we can treat the flow as approximately adiabatic, with an important caveat, discussed below; as the gas is radiation dominated, $\gamma = 4/3$.

1.3 Transport Mechanisms in Disks

1.3.1 Fluid Disks

Before discussing these three accretion modes in more detail, I would like to describe the manner by which mass, angular momentum, and energy are transported within accretion disks. The fundamental principle, which has transformed our understanding of disk accretion, is that orbiting, conducting gas is generically unstable to the magnetorotational instability (MRI; Balbus & Hawley 1998). We are all magnetohydrodynamicists now! Although it is probably true that most disk torque derives from some variant of the MRI, I want to spend a lot of this review describing a model of disk accretion that is almost surely invalid, but which, I argue, allows one to discuss some general principles in a quite transparent fashion. I will start by ignoring the MRI altogether and just consider adiabatic accretion. Numerical simulations (e.g., Stone, Pringle, & Begelman 1999) show that adiabatic disks quickly evolve to become convectively unstable.

The most familiar example of convection is the Schwarzschild case, which is appropriate when the entropy decreases upward in a gravitational field. When two fluid elements are interchanged slowly on a dynamical time scale, but rapidly with respect to the time it takes heat to be exchanged, they will quickly come into pressure equilibrium, and the one moving up will find itself lighter than its surroundings and keep on going under the force of buoyancy. This is the physical content of a small-amplitude instability calculation. Stars are unstable when the entropy decreases outward, as happens in the outer envelopes of solar-type stars.

The next most familiar type of convection is the Rayleigh case. Here we imagine a thin, differentially rotating disk orbiting under gravity and ignore the entropy and the pressure. Imagine a thin ring displaced radially and axisymmetrically outward. If we suppose that this happens slowly on a dynamical time scale, though too fast for viscous torques to operate, the specific angular momentum L will be constant. If the centrifugal force L^2/R^3 is less than that of its surroundings, the ring will return to its initial radius. Disks are unstable when the specific angular momentum decreases outward. Consequently, Keplerian disks are quite stable with respect to the Rayleigh criterion.

Adiabatic, fluid disks are thick, as the pressure is large. They combine features of thin disks and stars, and their convective stability must be handled using considerations that combine the Schwarzschild and Rayleigh criteria. The entropy naturally increases inward, as this is where most of the energy is released, which is destabilizing However, the angular momentum increases outward, and this is stabilizing. To understand the stability, consider displacing a slender ring of gas in the poloidal plane. (Note that the displacement is now two dimensional.) The displacement $\delta\vec{r}$ must be made fast enough to be adiabatic and inviscid, though slow enough for the displaced gas to remain in pressure equilibrium with its surroundings and for inertial terms to be ignored. Next, evaluate the poloidal acceleration linear in the displacement and the virtual work per unit mass. This gives a quadratic form in the two-dimensional displacement, $\delta W = U_{ij}\delta r_i \delta r_j/2$, where

$$U = \nabla\left(\frac{P^{1-1/\gamma}}{1-1/\gamma}\right)\otimes\nabla S + \nabla\left(\frac{1}{2R^2}\right)\otimes\nabla L^2, \tag{1.2}$$

ρ is the gas density, and $S = P^{1/\gamma}/\rho$ is an entropy function that turns out to be more convenient to use than the true entropy and to which it is monotonically related. The flow will be stable if δW is negative definite (see Blandford & Begelman 2004 and references therein for details). There are two "Høiland" criteria for instability (e.g., Tassoul 1978).

The first of these is

$$\nabla\left(\frac{P^{1-1/\gamma}}{1-1/\gamma}\right)\cdot\nabla S+\nabla\left(\frac{1}{2R^2}\right)\cdot\nabla L^2>0. \tag{1.3}$$

This is really telling us if the flow is unstable to radial perturbations, and it turns out that most reasonable disk models are quite stable to this criterion. Rayleigh stabilization wins. The second instability criterion is

$$(\nabla P\times\nabla R)\cdot(\nabla S\times\nabla L)>0. \tag{1.4}$$

This describes vertical perturbations in the disk, and it seems that fluid disks can easily evolve to a condition of marginal stability where the isentropes, the surfaces of constant entropy, and the isogyres, the surfaces of constant specific angular momentum, almost coincide. It is easy to show that this implies that the surfaces of constant Bernoulli function,

$$B=L^2/2R^2+\Phi+H, \tag{1.5}$$

where Φ is the gravitational potential and H is the enthalpy, also coincide with the gyrentropes ($S=L=$ constant), as

$$\nabla B=H\nabla\ln S+\Omega\nabla L. \tag{1.6}$$

It is helpful to consider other surfaces inside the disk—isochores (ρ = constant), isobars (P = constant) and isorotational surfaces (Ω = constant)—which are nested within the gyrentropes (Fig. 1.1). We therefore argue that the principle governing the structure of adiabatic fluid disks is that they be gyrentropic, so that $S=S(L)$. This is analogous to modeling the solar convection zone as isentropic. An examination of the character of the unstable modes shows that the nature of the motion has slender ribbons of gas sliding past each other along the gyrentropes. Some axisymmetric, numerical simulations of adiabatic disks have shown this type of transport (Stone et al. 1999), with mass, angular momentum, and energy being carried toward the disk surface where they can drive an outflow. On this basis, it is possible to calculate two-dimensional, analytic, self-similar models of fluid accretion disks, in hydrostatic equilibrium (Fig. 1.2.). We develop this disk model further below.

1.3.2 Magnetorotational Instability

As emphasized above, real disks almost certainly evolve under the action of hydromagnetic stresses. Essentially a differentially rotating flow with a small vertical magnetic field will be generically unstable to radial motions, which cause the magnetic field to grow exponentially on a dynamical time scale. The nonlinear evolution of this instability and the asymptotic state of the gas are still not very well understood, although considerable progress has been made in recent years, thanks to numerical simulations of ever increasing sophistication (e.g., Hawley & Balbus 2002). It is very important to perform three-dimensional, as opposed to two-dimensional, simulations, as the latter may be quite misleading. In addition, it is necessary to include a serious energy equation when dealing with adiabatic disks, not just solve the momentum equations assuming a polytropic relation between the density and the pressure, as was done in early calculations.

We still lack a good prescription for characterizing the MRI torque G in terms of local

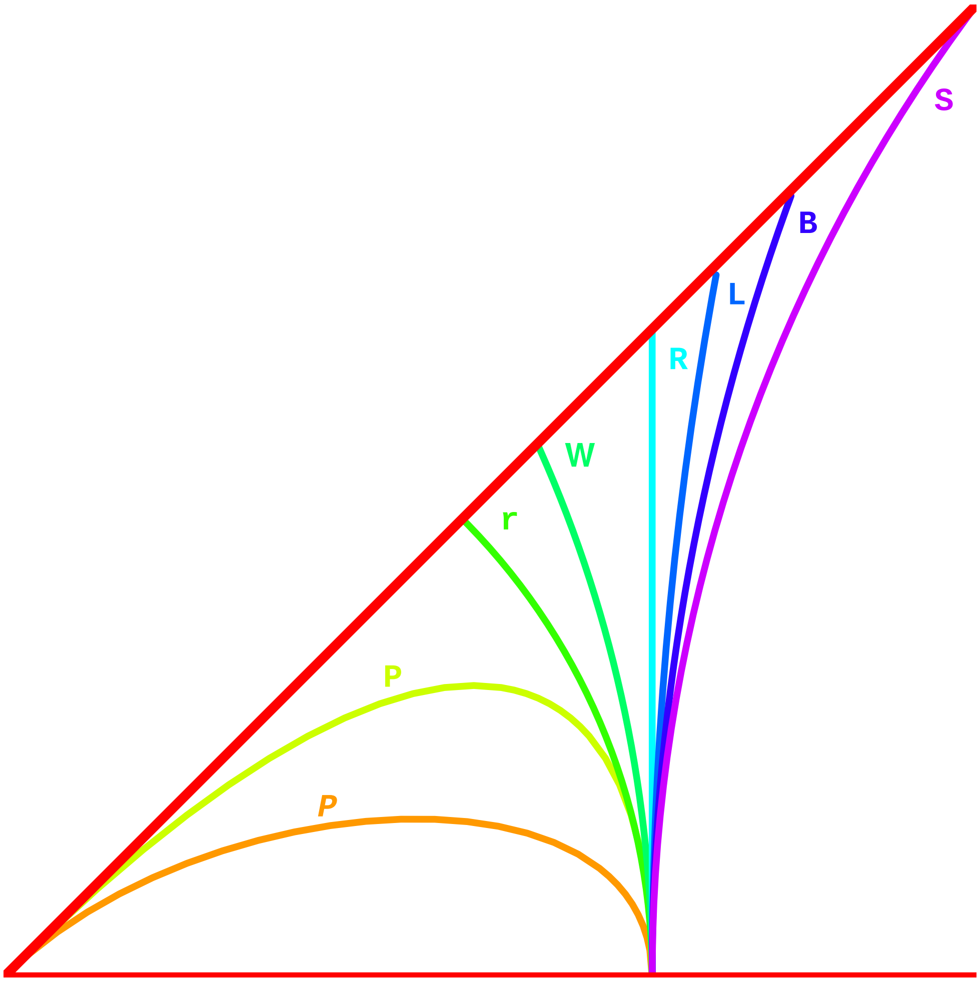

Fig. 1.1. Structure of a convectively stable fluid accretion disk. Level surfaces of density ($\mathcal{P}$), pressure (P), spherical radius (r), angular frequency (Ω), cylindrical radius (R), angular momentum (L), Bernoulli function (B), and entropy (S) passing through an equatorial ring of radius r. The configuration shown is dynamically stable according to both Høiland criteria. When the order of the isentropes and isogyres is reversed, the flow becomes convectively unstable.

physics. In general in these simulations, the magnetic stresses appear to saturate at a level corresponding to roughly several percent of the gas pressure. The overall momentum transport is probably not well described by a simple α prescription, in which the ratio of the shear stress to the total pressure P is supposed to be fixed number, α. Also, to repeat the point made above, we do not have a good understanding of how the associated dissipation proceeds. In one approach, it has been argued that an inertial range hydromagnetic turbulence spectrum is established with gas heating occurring at some inner scale (e.g., Quataert & Gruzinov 2000). Reconnection may also be important under some circumstances. In an

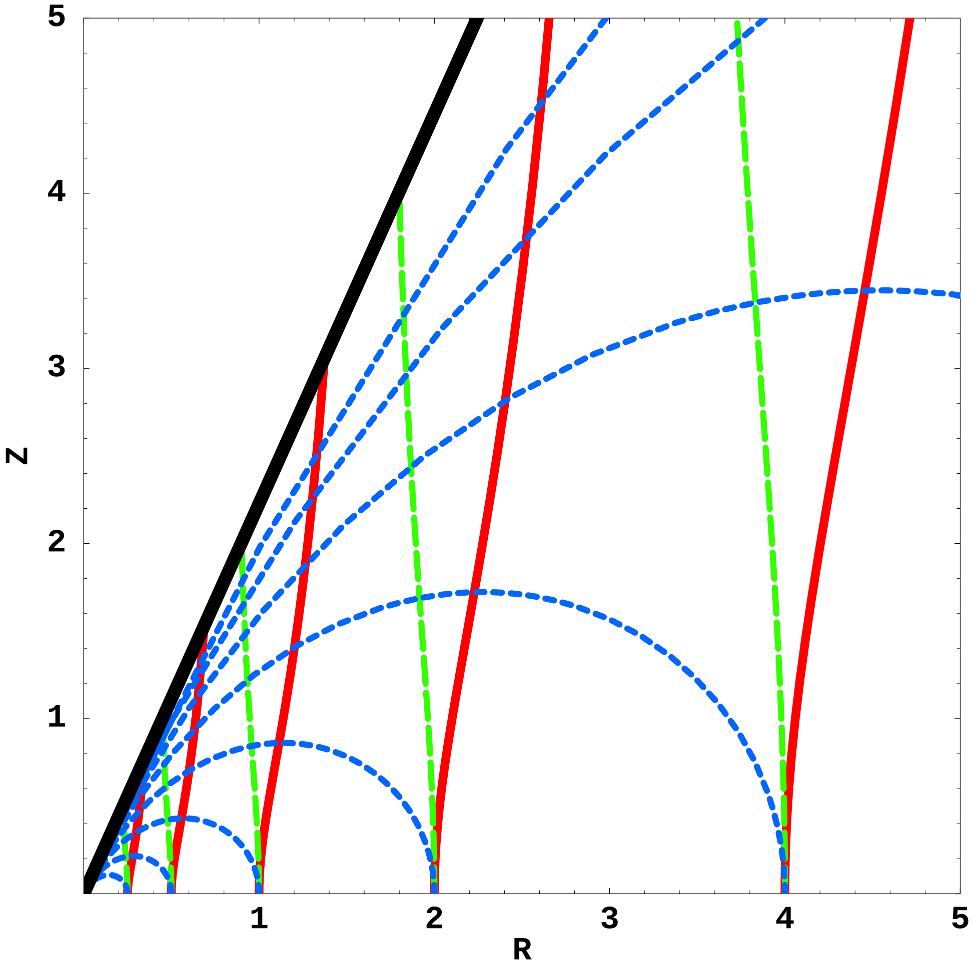

Fig. 1.2. Hydrostatic structure of an adiabatic, self-similar, ion-dominated, gyrentropic accretion disk. The gyrentropic surfaces, on which the entropy, specific angular momentum, and Bernoulli function are constant, are shown with solid lines. The isorotational and isobaric surfaces are denoted by long-dashed and short-dashed lines, respectively. Magnetic and radiation-dominated accretion disks are conjectured to have an analogous structure.

alternative proposal, there is an inverse cascade and little direct heating of the disk (e.g., Tout & Pringle 1996).

One expected feature of these disks, which is exhibited by some simulations, is the presence of outflow (e.g., Hawley & Balbus 2002). There are really two mechanisms for bringing this about. Firstly, there may be a large dissipation rate and entropy production in an active disk corona, and the outflow can be essentially thermally driven, although perhaps eventually guided by magnetic stress. Alternatively, the external magnetic field may be so strong that the MRI is suppressed on the disk surface, such that a cold outflow may be initiated as a consequence of centrifugal force flinging out plasma along open field lines.

1.3.3 Transport in Radiation-dominated Disks

There have been heroic attempts made in recent years to address the even harder problem of understanding the combined transport of momentum and radiation in radiation-dominated disks near the central black hole. This is important because this emission accounts for a major fraction of the the luminosity density of the Universe, rivaling that supplied by nuclear physics inside stars! The strength of the magnetic field within radiation-dominated disks has long been a matter of some controversy. For example, it has sometimes been argued that the magnetic field saturates at a value that is a fraction of the gas pressure, not the total pressure (Sakimoto & Coroniti 1981).

The way forward is to perform numerical simulations that combine gas dynamics with radiative transfer. To date these have had to be pretty limited, involving shearing boxes and ignoring vertical gravity (e.g., Agol et al. 2001), but still they have been quite instructive. The single most important feature of these simulations is that the disk naturally becomes quite inhomogeneous. This is not so surprising when one considers the large variety of linearly unstable wave modes that can be described analytically and demonstrated numerically (Blaes & Socrates 2001; Turner, Stone, & Sano 2001).

This strong inhomogeneity raises the possibility that radiation may be able to escape from the disk at a rate significantly in excess of the Eddington rate. One particular manifestation of this outcome is proposed by Begelman (2002), who finds solutions in which magnetosonic modes propagate obliquely through the disk, forming low-density chimneys along which radiation can escape. These are sustained by the magnetic field. The gas that defines the high-density chimney walls is held in place by gravity. It is argued that it is possible to beat the Eddington limit by a factor ~ 100 in this manner. However, it is also necessary that the radiation escape anisotropically so that the supply route for the accreting gas does not intersect the beam of escaping radiation, and that the power that is scattered isotropically out of this beam does not exceed the Eddington limit. If these conditions can be satisfied, then super-Eddington accretion provides a possible explanation for the ultraluminous X-ray sources (see, e.g., van der Marel, this volume), as well as perhaps the narrow-line Seyfert 1 galaxies.

1.4 Adiabatic Accretion Models

I would now like to return to adiabatic fluid accretion under Case A, and, more controversially, Case C. However, in order to make a central point, I should first consider regular, radiative accretion. In the simplest model of a stationary, Keplerian disk, the mass flow *inward* $\dot{M}$ and the net angular momentum flow $F_L = \dot{M}L - G$ *inward* are conserved. As $L \propto r^{1/2}$, conservation of angular momentum over several octaves of radius implies that $F_L \approx 0$, except very close to the hole. Combining these two equations, we can obtain an energy equation

$$\frac{dF_E}{dr} = G\frac{d\Omega}{dr}, \tag{1.7}$$

where $F_E = G\Omega - \dot{M}E$ is the net flow of energy *outward*, and E is the sum of the kinetic and gravitational potential energy. The term on the right-hand side corresponds to the rate of loss of energy from the flow through radiative loss and is easily shown to be 3 times the rate of release of binding energy in a thin, Keplerian disk. The origin of this extra energy release is the work done by the torque G, and it derives from the binding energy release at small radii, where this equation has to be modified.

Now, consider what happens when the accretion is stationary, adiabatic, and conservative. The mass conservation and angular momentum conservation are unchanged. However, the gas is no longer cold and E must be augmented with the thermal energy so that $F_E \to G\Omega - \dot{M}B$. In addition, as the gas cannot radiate, there is no sink of energy, and so F_E must be conserved and must also be very close to zero, as $G\Omega \propto r^{-3/2}$. This then implies that

$$B = \frac{G\Omega}{\dot{M}} = \Omega L. \tag{1.8}$$

The Bernoulli function must be positive and the gas has enough energy to escape to infinity with positive kinetic energy. Three resolutions of this paradox have been proposed.

In the ADAF prescription (Narayan & Yi 1994), it is proposed that the flow is essentially conservative and closes off along the rotation axis so as to prevent the escape of the unbound gas. Gas then crosses the horizon with positive energy. However, this requires the velocity field to become singular. In another variation on the basic model, it is also proposed that the flow is quasi-spherical so that little angular momentum transport is required. However, this seems very difficult to arrange in a galactic nucleus.

In the CDAF prescription (e.g., Quataert & Narayan 1999), it is recognized that the flow becomes convective, and it is proposed that convection transports angular momentum inward, while transporting energy outward. Independent of the likelihood of this happening in practice, this prescription leads to a mass accretion rate that increases with radius. If the flow is conservative, then it must be non-stationary. The mass supply can back up but eventually it must find some other mode of accretion.

Finally, in the ADIOS prescription (Blandford & Begelman 1999, 2004), it is proposed that the flow is non-conservative. The most important requirement is that the energy be removed in an outflow of some form. In principle this could be a magnetohydrodynamic wind with very little mass content. However, in practice, we strongly suspect that the outflow involves a large mass loss, and the rate at which gas accretes onto the black hole is orders of magnitude smaller than the rate at which it is supplied. This is not a small effect, and if this model is essentially correct, then many of the papers describing the radiative properties of conservative flows will need to be modified. As discussed above, observations of Sgr A* and other nearby (inactive) galactic nuclei do suggest that the flow is non-conservative.

1.5 ADiabatic Inflow-Outflow Solutions

It is quite instructive to consider how to make a model of disk accretion onto a black hole. As discussed above, I will restrict attention, initially, to stationary, adiabatic, hydrodynamic, convective models as these appear to be simpler (Blandford & Begelman 2004).

1.5.1 Hydrostatics

The first step is to make a model of a hydrostatic disk. Following the discussion above, we suppose that the disk is convective and consequently gyrentropic so that $L = L(S)$, or equivalently $L = L(B)$. We can make things simple by assuming that the structure of the bulk of the Newtonian disk is self-similar:

$$P = r^{n-5/2}p(\theta), \quad \mathcal{P} = r^{n-3/2}\rho(\theta), \quad L = r^{1/2}\ell(\theta), \tag{1.9}$$

where θ is the polar angle and the fixed exponent n has a value in $[0,1]$. If we use the self-similar scaling of the radial velocity $v_r \propto r^{-1/2}$, then we can identify n as the accretion exponent $\dot{M} \propto r^n$. It turns out that the parameter n can be replaced by a parameter b_0, which measures the Bernoulli function of the disk in the equatorial plane. b_0 is negative for a bound disk (although it can be positive in a funnel). As the disk becomes thicker and loses more mass, $b_0 \to 0$. The disk models are further parametrized by the equatorial angular momentum in units of the Keplerian value, ℓ_0. These parameters suffice to specify the disk structure (Fig. 1.3). The disk surface, where $P \to 0$ and $\mathcal{P} \to 0$, is a cone with an opening angle that decreases with ℓ_0.

1.5.2 Convection and Torque

Gyrentropic convection also provides a rule for transporting mass, angular momentum and energy from the body of the disk to the disk surface—the direction of transport is parallel to the gyrentropic surfaces. However, in order to make a more detailed model we must make additional assumptions. A convenient and natural choice in this context is that the effective Prandtl number is large, which means that the moving ribbons are in much stronger thermal than mechanical contact with their environment. In other words, the convection is only efficient at transporting heat directly to the surface of the disk.

Now, a steady disk flow also requires that angular momentum be transported across the gyrentropes, and so it is necessary to supplement the convection with a phenomenological torque. Self-similarity dictates a constant value for the α parameter, or, equivalently, a kinematic viscosity that scales with the product of the disk thickness and the sound speed. The torque per unit length in the meridional plane then satisfies

$$\vec{G} = -2\pi\alpha(r\sin\theta)^3 P \nabla \ln\Omega. \tag{1.10}$$

1.5.3 Circulation

Given these rules, or others like them, it is possible to use conservation of mass, angular momentum, and energy to determine the mass flow in the disk. This is accomplished by solving the equations

$$\nabla \cdot \vec{J} = 0 \tag{1.11}$$

$$\nabla \cdot (L\vec{J} + \vec{G}) = 0 \tag{1.12}$$

$$\nabla \cdot (B\vec{J} + \Omega\vec{G} + \vec{Q}) = 0, \tag{1.13}$$

where $\vec{J}$ is the mass flux per unit length in the poloidal direction and $\vec{Q}$ is the convective energy flux along the gyrentropes. Solving these equations leads to the general conclusion that there is a general inflow of gas through the disk and that superimposed upon this is a meridional circulation. Combining the circulation with the convection, we find that there is a steady flow of mass, angular momentum, and energy to the disk surface. The most common pattern is to have inflow near the equatorial plane and a quadrupolar circulation that can lead to outflow near the disk surface. However this sense of circulation is not required, and it is possible to reverse it by changing assumptions like the latitude dependence of α. The net inflow of gas in the disk matches the net loss of mass at the disk surface. Likewise, the difference between the angular momentum and energy carried inward by the flow and transported outward across gyrentropes by the torque is equal to the energy and angular momentum flux crossing the disk surface.

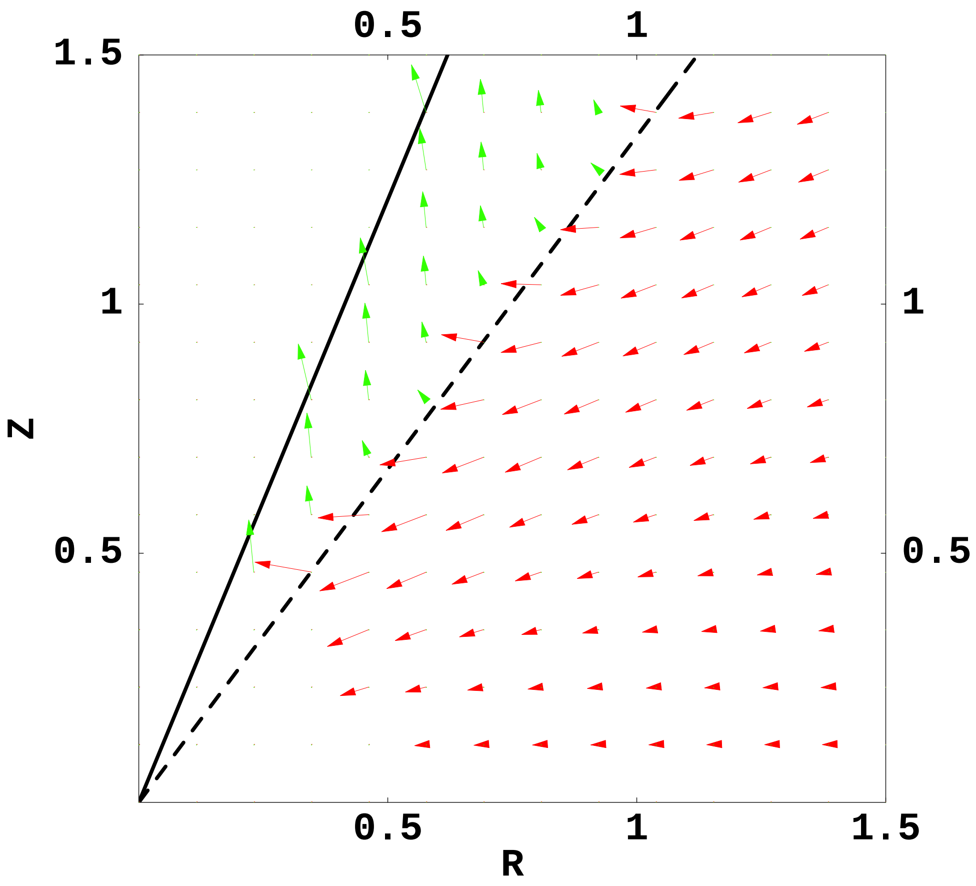

Fig. 1.3. Flow vectors for an axisymmetric, self-similar, fluid model of an ion-dominated, adiabatic accretion disk. The accretion disk is separated from the outflow by a thermal front, which is designated by a dashed line. The disk flow exhibits a poloidal circulation pattern with a net radial inflow. In addition, there is a net flow incident upon the disk surface, where entropy is generated, which feeds the outflow. Centrifugal forces acting in the outflow define an excluded cone which could be filled by a jet. As discussed in the text, radiation-dominated and hydromagnetic disk models should exhibit similar features.

1.5.4 Thermal Front

Now let us consider the consequences of having a flow of mass, angular momentum, and energy incident upon the disk surface. As the density vanishes at the surface, the flow velocity must diverge, signifying the ultimate failure of our model. What we actually believe would happen under these circumstances is that there is a rapid production of entropy where the flow speeds become supersonic. This allows gas that is bound in the disk to become unbound at the disk surface and supply an outflow. Something similar, but much more gentle, happens in the Sun, although the actual details of what is going on remain obscure. However, independent of the details, mass, momentum, and energy must be conserved and so we can introduce a "thermal front," analogous to a deflagration wave, across which we can compute the jump in these quantities, assuming that the convective energy flux Q vanishes downstream from the thermal front. Specifically, we solve the jump conditions

$$
\begin{aligned}
[J_\theta] &= 0 && (1.14)\\
[J_\theta V_r] &= 0 && (1.15)\\
[2\pi RP + J_\theta V_\theta] &= 0 && (1.16)\\
[G_\theta + J_\theta L] &= 0 && (1.17)\\
[G_\theta \Omega + J_\theta B + Q_\theta] &= 0. && (1.18)
\end{aligned}
$$

The approximation that is made here is that entropy is created quickly as gas flows across the front, analogous to what is commonly assumed in deriving the Rankine-Hugoniot conditions at a shock front.

It is important to emphasize that there can be no outflow from a fluid disk without dissipation near its surface. If there is no dissipation, then a circulation will be established with no net inflow and the flow resembles the CDAF solution, although we prefer to distinguish the organized poloidal circulation from the oscillatory motion associated with convection. A flow like this would be non-stationary—the mass supplied will back up all the way to the Bondi radius. Behavior like this can be seen in some recent numerical simulations. At this point the mass accretion rate would be shut down almost to zero. In practice, flows are likely to become radiative before they back up too far, except under conditions when the mass accretion rate is tiny in comparison with the Eddington rate. This radius, beyond which disks are radiative as opposed to adiabatic, is known as the "transition radius." It marks the outer boundary of an adiabatic flow. It is our contention that supersonic convection near the disk surface is replaced by strong dissipation and that, under the assumptions we have made in this model, there will be a strong outflow.

1.5.5 Outflow

The gas downstream from the thermal front is hot and flows away from the disk with a subsonic speed. Unlike the gas in the disk, it has positive Bernoulli function and can therefore escape to infinity with positive energy. We model the outflow as non-dissipative and ignore the effects of the torque as well as any convective motion. Clearly this is an approximation. Note that the quantities L, S, and B are conserved along the flow lines, and so the outflow is gyrentropic, just like the disk, though for a quite different reason. The disk speed quickly becomes supersonic, although there is no critical point in the solution due to the imposition of self-similarity, which only allows the velocity component resolved in the θ direction to vary freely, and this remains subsonic (Fig. 1.3).

This prescription for a simple, thermally driven wind does enable us to augment the disk model with an analytic, self-similar outflow that is asymptotic to the surface of a cone with a smaller opening angle than the disk. The excluded region owes its existence to the centrifugal barrier. If we allow a torque to develop in the outflow, the cone would be narrower, though should not disappear.

We have therefore described a large family of self-similar fluid disk models. At this point an interesting matter of principle arises. If we suppose that mass is supplied at a certain rate at an outer, transition radius and accretes adiabatically onto a massive black hole, and that microphysical parameters like γ and α are known, then our fluid disk models are described by two independent parameters, which we have chosen to be b_0 and ℓ_0. However, the disk structure ought to be unique. (Actually, in general there will be three independent parameters describing the central densities of mass, energy, and angular momentum. However, in a fluid disk the outflowing gas carries off only its own specific angular momentum and does not exert any significant reaction torque on the remaining gas. This implies that the transport of mass and angular momentum through the disk are not independent, and the number of independent parameters is reduced to two.)

So what fixes the remaining two degrees of freedom? The answer appears to be that the conditions in the innermost, relativistic portion of the accretion disk (especially the spin of the hole), where self-similarity breaks down, essentially act as a boundary condition and determine the variation of the Bernoulli parameter b_0 throughout the disk. To be more specific, a certain fraction of the energy released by the gas as it flows into the hole will be transported out into the disk by viscous torque and large-scale circulation. If the energy so transported is large, the next annular ring of gas will be relatively thick. It will then transport a large amount of energy to the next ring, and so on. Self-similarity propagates this information from small radius to large radius. Likewise, we suppose that the conditions at the outer radius of the adiabatic disk determine the angular momentum (and mass) flow inward.

Is this prescription an artificiality derived from self-similarity? I suspect that this is not the case. If the rule for deciding what fraction of the energy flow propagating outward through a disk actually leaves the disk over the next octave of radius really is independent of radius, then the structure of the disk will be dictated by the conditions at small radius. This is somewhat similar to what happens in a boundary layer on a solid surface in a flow, where the rule for determining the rate of thickening is local, but the actual thickness of the boundary layer will respond to what happens upstream where the surface first encounters the flow.

1.5.6 Radiation-dominated Disks

The modification of this model for the radiation-dominated outflow that exists under Case C is fairly straightforward. Firstly, it is necessary to change the specific heat ratio, γ, from $5/3$ to $4/3$. Secondly, the relevant rule for determining the torque may be quite different from that for an ion-dominated gas, and the influence of inhomogeneity may be large, as we described above. However, if we ignore this possibility, then the radiation will be trapped and the disk flow will be roughly adiabatic as long as $\dot{m} > r/m$. The trapping radius, $r_{\rm tr}$, where this inequality becomes an equality, is the effective transition radius. The power emitted from $r > r_{\rm tr}$ is roughly the Eddington limit. If the mass accretion exponent n is not much less than unity, the flow will be radiatively quite efficient with respect to the accreted mass, with most of the mass supplied leaving in a wind from within the trapping radius. Only a fraction $\sim 1/\epsilon\dot{m}$ of the mass supplied will actually accrete. The outflow will be driven by electron scattering, and the fluid model will certainly become invalid when the optical depth of the wind falls below c/V, where V is the outflow speed, so that the photons can escape on an outflow time scale. However, resonance lines are likely to be important in accelerating the outflow beyond this radius.

1.5.7 Magnetic Accretion

Now let us return to the more complex question of the accretion of magnetized gas. We have argued that the MRI implies that hydromagnetic stresses dominate the torque within the disk. However, despite considerable progress in performing numerical simulations, we still do not have a good, global model of the transport of mass, energy, and angular momentum. (There are some hints that the disk adjusts so as to become barytropic, rather than gyrentropic. If so, the implications for the disk structure are relatively minor.) However, the primary "convective" (i.e., oscillatory, as opposed to circulatory) transport is no longer necessarily vertical, as we have argued is the case in a fluid model. Another important change is that there can be a magnetic reaction torque connecting the disk to the outflow. In other words, magnetic outflows can be centrifugally as well as thermally driven. This means that there are now three eigenvalues that must be to be fixed through the imposition of an inner boundary condition on the energy and outer boundary conditions on the mass and the angular momentum.

Adopting the methodology advocated in this review, all of this means that we must first determine the rules for describing the *local*, averaged behavior of accretion disks. This will almost surely require numerical, hydromagnetic simulations with sufficient dynamic range to encompass many scale heights of the disk and the corona. In particular, it may be necessary to design the difference scheme so as to accommodate a structure like a thermal front in much the same way that schemes that have to handle supersonic flows are able to locate shock fronts. This scheme would have to create the necessary entropy at the disk surface. We note that the jump conditions at the thermal front naturally generalize to the hydromagnetic case.

The next task will be to use these prescriptions in global models of accretion disk evolution. In general, the flows will be time dependent. Let us denote the mass per unit spherical radius in the disk by $\mathcal{M}$. The angular momentum and energy per unit radius can likewise be defined by

$$\mathcal{S} = \langle \mathcal{M} L \rangle \quad \text{and} \qquad \mathcal{E} = \langle \mathcal{M}(B - P/\rho) \rangle, \tag{1.19}$$

respectively, where the averages are performed on spheres. If there are no interior radial scales in the adiabatic flow, then we expect there to be simple scaling relations such as $F_L \propto \dot{M} R^{1/2}$, etc.

We can also define the flows of mass inward, angular momentum inward, and energy outward across spheres in the disk by $\dot{M}$, $F_L = \langle \dot{M} L - G \rangle$, and $F_E = \langle G\Omega - \dot{M} B \rangle$, respectively. Naturally we want to know the functional forms of $\dot{M}(\mathcal{M}, L, B)$, etc. Finally, we need the mass, angular momentum, and energy carried off by the outflow per unit spherical radius, which we denote by K_M, K_L, K_E. An adiabatic disk will then evolve approximately according to the three coupled "kinematic" one-dimensional equations

$$\frac{\partial \mathcal{M}}{\partial t} = \frac{\partial \dot{M}}{\partial r} - K_M \tag{1.20}$$

$$\frac{\partial \mathcal{S}}{\partial t} = \frac{\partial F_L}{\partial r} - K_L \tag{1.21}$$

$$\frac{\partial \mathcal{E}}{\partial t} = -\frac{\partial F_E}{\partial r} - K_E. \tag{1.22}$$

These equations must be supplemented with boundary conditions on $\mathcal{M}$ and $\mathcal{S}$ at the outer radius and $\mathcal{E}$ at the inner radius, as discussed above. They make the implicit assumption that the local adjustment time scales are short compared with the flow times, which are not very accurate for thick disks.

It will be interesting to see if reasonable choices for these unknown functions lead to stationary or episodic inflows and outflows.

1.6 Implications for Black Hole Growth

Let us return to our simple scheme for describing accretion onto black holes. We suspect that, early in the life of an AGN, the hole mass will be low and the mass supply rate will be high, $\dot{m} \gg \dot{m}_2$, so that we are solidly in Case C. If our conjecture about these flows being adiabatic is correct, and it is not, in practice, possible to exceed the Eddington limit by a large factor, then almost all of the mass supplied must be blown off in an outflow and the hole will grow (exponentially) on a time scale $\sim \epsilon t_E \approx 40$ Myr. The accretion is "demand limited" instead of "supply driven," as in the "feast or famine" model of Small & Blandford (1992). There is time for the black hole to grow by a factor of at least a thousand before the highest redshift quasars are observed. (There is a possible exception to this rule, which may operate when the hole mass is small. If the trapping radius exceeds the accretion radius, $\sim GM/s^2$, with s the effective sound speed in the gas, and typically this requires that $\dot{M} > (m/100M_\odot)\ M_\odot\ \mathrm{yr}^{-1}$, then the outflow may be trapped by the infalling gas and the energy radiated away at a rate much greater than the Eddington limit for the hole. This may allow a stellar mass hole to grow quickly to an intermediate mass, very early in the life of the nascent galaxy.)

In general, however, we expect that the mass supply will decline while the hole mass grows. There will then come a time when the trapping radius shrinks so that it is not much larger than the event horizon, and Case B accretion will commence. Let us suppose that this only happens once in the lifetime of a galaxy. (If holes are frequently resurrected, then what follows must be modified.) Most of the black hole mass will have been built up around the time of this transition. Let us further assume that, at the time when holes of final mass M were mostly assembled, the luminosity function for the corresponding Eddington luminosity is dominated by holes transitioning from Case C to Case B. (Note that this does not have to be true now.) This implies that the local, hole mass function $N(M)$ should reflect the integral of the comoving quasar luminosity function $\Phi(L)$, evaluated at $L \approx L_{\mathrm{Edd}}(M)$ over time. Specifically,

$$N(M) \approx \mu \int dt \frac{\Phi(\lambda M c^2/t_E)}{t_E^2 c^2}, \tag{1.23}$$

where the coefficients $\lambda \approx 1$ and $\mu \approx \epsilon^{-1}$ can, in principle, be determined observationally and used to parameterize a more detailed model.

This somewhat impressionistic portrait of hole growth should eventually lead to a more specific prediction concerning the relationship between the contemporary hole mass function and the historical quasar luminosity function than the integral Sołtan (1982) relationship. This the next challenge in the study of black demography—to determine both when *and* how the dormant holes we see around us acquired their masses. Success in this endeavor will allow us to understand in more detail both the physics of the accretion process and its environmental impact on the host galaxy.

Acknowledgements. I am indebted to my collaborator Mitch Begelman for his contributions to the above research and for his patience. I thank Luis Ho for organizing a stimulating meeting and also for his patience. Support under NASA grant NAGW5-12032 is gratefully acknowledged.

References

Agol, E., Krolik, J. H., Turner, N., & Stone, J. 2001, ApJ, 558, 543
Balbus, S. A., & Hawley, J. F. 1998, Rev. Mod. Phys., 70, 1
Begelman, M. C. 2002, ApJ, 568, L97
Blaes, O., Hubeny, I., Agol, E., & Krolik, J. H. 2001, ApJ, 563, 560
Blaes, O., & Socrates, A. 2001, ApJ, 553, 987
Blandford, R. D. 1999, in Astrophysical Disks, ed. J. A. Sellwood & J. Goodman (San Francisco: ASP), 265
Blandford, R. D., & Begelman, M. C. 1999, MNRAS, 303, L1
——. 2004, MNRAS, in press
Fabian, A. C. 2002, Phil. Trans. R. Soc., 360, 21035
Fan, X. 2004, in Carnegie Observatories Astrophysics Series, Vol. 1: Coevolution of Black Holes and Galaxies, ed. L. C. Ho (Pasadena: Carnegie Observatories, http://www.ociw.edu/ociw/symposia/series/symposium1/proceedings.html)
Frank, J., King, A. R., & Raine, D. J. 1992, Accretion Power in Astrophysics (Cambridge: Cambridge Univ. Press)
Hawley, J. F., & Balbus, S. A. 2002, ApJ, 573, 738
Hubeny, I., Blaes, O., Krolik, J. H., & Agol, E. 2001, ApJ, 559, 680
Narayan, R., & Yi, I. 1994, 428, L13
Quataert, E., & Gruzinov, A. 2000, ApJ, 539, 809
Quataert, E., & Narayan, R. 1999, ApJ, 520, 298
Sakimoto, P. J., & Coroniti, F. V. 1981, ApJ, 247, 19
Salpeter, E. E. 1964, ApJ, 140, 796
Small, T. A., & Blandford, R. D. 1992, MNRAS, 259, 725
Sołtan, A. 1982, MNRAS, 200, 115
Stone, J. M., Pringle, J. E., & Begelman, M. C. 1999, MNRAS, 310, 1002
Tassoul, J.-L. 1978, Theory of Rotating Stars (Princeton: Princeton Univ. Press)
Tout, C. A., & Pringle, J. E. 1996, MNRAS, 281, 219
Turner, N. J., Stone, J. M., & Sano, T. 2002, ApJ, 566, 148
Wilms, J., Reynolds, C. S., Begelman, M. C., Reeves, J., Molendi, S., Staubert, R., & Kendziorra, E. 2001, MNRAS, 328, 27
Yu, Q., & Tremaine, S. 2002, MNRAS, 335, 965
Zel'dovich, Ya. B., & Novikov, I. D. 1964, Sov. Phys. Dokl., 158, 811