AGN SURVEYS
ASP Conference Series, Vol. 284, 2002
R.F. Green, E.Ye. Khachikian, D.B. Sanders

The Formation and Feeding of Massive Black Holes in the Early Universe

Wolfgang J. Duschl

Institute for Theoretical Astrophysics, University of Heidelberg, Germany; wjd@ita.uni-heidelberg.de

Peter A. Strittmatter

Steward Observatory, The University of Arizona, Tucson, AZ, USA; pstrittmatter@as.arizona.edu

Abstract. It is still an open question whether the super-massive black holes thought to be present in quasars are of primordial nature, or whether there is a viable way of forming them in the very short time scale (less than a billion years) permitted by the observational data. In this contribution, we present a way in which a galaxy-galaxy merger can provide not only the "fuel" for quasar activity, but can also build a super-massive black hole, i.e., "the engine", in the first place.

1. Introduction

During the last two decades dynamical studies of the nuclei of galaxies in the local Universe have lead to the conclusion that super-massive black holes of some $10^{6\cdots9}\,\mathrm{M}_\odot$ are commonly present (Kormendy 2001, and references therein). Actually, no firm counterexample has been found yet. The growth of these black holes to their current mass may be understood as a consequence of (intermittent) accretion lasting over the entire lifetime of the galaxies (Duschl 1988a, 1988b) at average rates of $10^{-4\cdots-1}\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$.

Soon after its discovery, the quasar phenomenon was attributed to accretion onto super-massive black holes. Accretion rates of $10^{-1\cdots+1}\,\mathrm{M}_\odot\,\mathrm{a}^{-1}$ and black hole masses of typically $10^{8\cdots11}\,\mathrm{M}_\odot$ (Lynden-Bell 1969) were inferred. However, the predominant presence of quasars at red-shifts $z > 1$ and their paucity in the local, contemporary Universe, demand a much more rapid and closed ended growth process of quasar black holes.

In this contribution, we will argue that the quasar phenomenon is a direct consequence of a major merger. As a "major merger" we define the coalescence of two galaxies of about equal mass resulting in the deposition of large amounts of gas in a disk close to the ccenter of the merged galaxy (Barnes & Hernquist 1996). Other galaxies which did not undergo such major mergers may also exhibit phases of nuclear activity, for instance as Seyfert galaxies, but in general harbor central black holes of considerably smaller masses than galaxies which hosted quasars.

While the occurrence of mergers as an important process for feeding quasars has been under discussion now for quite some time (e.g., Stockton 1999), we will show that the merger is not only instrumental in "providing the fuel", but even more for "building the engine" to produce the quasar phenomenon and accounting for the absence of quasars at the current epoch.

Before proceeding, however, we will briefly describe recent developments in the theory of self-gravitating accretion disks, which suggest accretion rates which exceed by more than an order of magnitude previous estimates. Then, we will give estimates which show how—as a consequence of a major merger—a super-massive black hole may develop in a galactic center within approximately $5 \cdot 10^8$ years, and how the quasar activity may set in, reach its maximum, and cease after about 10^9 years. Finally, we will note some open questions and work in progress.

2. Massive gas/dust disks and accretion

Whenever in an astrophysical flow angular momentum is not negligible, accretion disks come into play. The qualitative concept of accretion disks dates back to Kant (1755) and Laplace (1796) and their models for the formation of our planetary system. The modern quantitative description was first developed by Lüst (1952). The major obstacle to progress has been the lack of detailed understanding of the physical processes giving rise to the radial transport of mass (towards the disk's center) and angular momentum (in the opposite direction). A very successful, albeit originally purely heuristic, parameterization of the viscosity coefficient ν is due to Shakura (1972) and Shakura & Sunyaev (1973). They assumed that the viscosity in the flow is due to turbulence, and that the turbulent velocity and length scales are limited by the sound velocity $c_{\rm s}$ and the disk's thickness h. Together with a constant $\alpha \leq 1$ this led to the now famous "α-viscosity":

$$\nu = \alpha h c_{\rm s}. \tag{1}$$

Based on this viscosity prescription, the outbursts of dwarf novae and related systems could be explained to a surprisingly detailed level (Warner 1995). It turned out that for most astrophysical applications where such models were applicable at all, α was generally of order or slightly less than unity.

Later, a physical basis for this parameterization was re-discovered for the case of magnetic disks ("Balbus-Hawley instability", Balbus & Hawley 1991; but see also Velikhov 1959, Chandrasekhar 1960), but not for purely hydrodynamic disks. The fact that a physical process was known only for the magnetic case, should not, however, have led to the conclusion that the interaction of a magnetic field and the disk material is a necessary condition for angular momentum and mass transfer to occur.

Again, based on α-viscosity it has been shown that accretion disk models lead to exceedingly long evolution time scales for disks in the centers of AGN. To remedy this, various—mostly non-axisymmetric—processes (bars, spiral waves, etc.) were investigated in order to obtain sufficiently short time scales (e.g., Shlosman, Frank, & Begelman 1989; Chakrabarti & Wiita 1993) to account for the observed phenomenon.

In the meantime, however, it has been noted that, independent of the origin of the viscosity, α-disk models lead to physically altogether inconsistent results as soon as the mass of the disk is no longer small compared to the central, accreting body's mass, i.e., when the disks are self-gravitating (Duschl, Strittmatter, & Biermann 2000, hereafter DSB). DSB have also pointed out that experimental data on rotating fluids suggest an alternative, hydrodynamic origin of turbulence and hence of the viscosity.

2.1. Massive vs. mass-less accretion disks

For an accretion disk, self-gravity becomes important globally as soon as, at some radius s from the disk's center, the disk mass $M_{\mathrm{disk}}(s)$ enclosed within this radius becomes comparable to or larger than the accreting body's mass M_{a} at the disk center. Then in the radial range where this is fulfilled, the disk is *fully self-gravitating* (FSG). But already at lower disk masses, self-gravity becomes important locally (in vertical direction). This happens when $M_{\mathrm{disk}}(s)$ becomes larger than $(h/s)\cdot M_{\mathrm{a}}$. For disk masses where local, but not yet global self-gravity is important, the azimuthal velocity v_φ is still given by a Keplerian rotation law, while the vertical structure is already dominated by self-gravity. Such disks are of the *Keplerian self-gravitating* type (KSG). Only for even lower masses the disks do not experience any important influence due to the gravitational forces of their own mass distributions, i.e., they are non-self-gravitating (NSG).

2.2. Viscosity parameterization in self-gravating disks

In both, KSG and FSG disks, a naive extrapolation of the α-parameterization leads to physically inconsistent results in that the disks would be isothermal in the radial direction. Recently a generalization of this parameterization was proposed (DSB) which solves the problem in the KSG and FSG cases, but recovers the α-parameterization for NSG disks (where this ansatz is so successful). Based on laboratory experiments and on theoretical considerations (Wendt 1933, Taylor 1936a, 1936b; see also Richard & Zahn 1999), the unconstrained viscosity ν is written as

$$\nu = \frac{1}{\Re_{\mathrm{crit}}} s v_\varphi = \beta s v_\varphi \tag{2}$$

with the critical Reynolds number $\Re_{\mathrm{crit}} \sim 10^{2\ldots3}$ and $\beta = 1/\Re_{\mathrm{crit}}$. As an additional constraint, the corresponding turbulent velocity $v_{\mathrm{turb}} \sim \sqrt{\beta} v_\varphi$ is always required to be less than or equivalent to the sound speed. In the limiting case of a NSG accretion disk this latter condition leads to the α-viscosity prescription but gives a different result in all other cases (i.e., for subsonic turbulence or shock limited selfgravitating disks).

The timescale τ_{visc} can then be estimated to be of the order of

$$\tau_{\mathrm{visc}} = \frac{s^2}{\nu}. \tag{3}$$

3. How to build and run a quasar engine

In the following, we assume that due to a major merger tidal forces have driven a large amount ($10^{9...10}\,M_{\odot}$) of accretable matter into the central regions (within a few 10^2 pc from the center) of the newly formed merged galaxy (Barnes & Hearnquist 1996) , where—due to its angular momentum—the material takes up a disk equilibrium configuration. This is a fairly robust assumption, unless the rare case happens that the net angular momenta of the two merging galaxies point (almost) exactly in the same direction. We also assume that there is no pre-existing super-massive black hole at the center of the merged galaxies. This scenario provides the starting point for our model.

3.1. A brief description of the physical model

We envisage that this self-gravitating gas (and dust) mass will start to accrete towards the center, independent of whether a seed black hole (of comparatively small mass) is present or not. Initially, the accretion disk is capable of radiating all energy liberated through dissipation. The huge amounts of mass moving towards the disk's center (see next paragraph for details) will lead to *(a)* the formation of a seed black hole (if none was present before), and *(b)* an initial phase of accretion into it which is Eddington-limited. While the details of the formation process of a seed black hole remain to be investigated, the assumption that it will form and that it does so very quickly seems to us to be unavoidable.

The black hole accretes at its Eddington rate as long as the disk delivers enough mass to maintain this rate. It is still an open question what happens to the mass which the accretion process supplies to the black hole, but which cannot be swallowed by it due to the effect of radiation pressure. One may speculate that this is an ideal source of material and energy for a jet and a broad line region to form. This process, however, is beyond the scope of this paper.

Ongoing accretion will deplete the mass of the accretion disk and thus decrease the mass delivery rate towards the black hole, while—at the same time—the black hole is growing in mass due to the same accretion process. This will continue in the way described above until the mass flow rate from the disk to the black hole has become smaller than the Eddington accretion rate. From this point on, free accretion sets in, and all the incoming mass may be accreted by the black hole. At the beginning of this phase, the accretion disk is still able to radiate all its liberated energy.

In the course of this evolution, however, the accretion rate drops, both in absolute terms as well as in units of the corresponding Eddington accretion rate. When the actual accretion rate falls below roughly 0.3 % of the Eddington rate, the flow becomes advection dominated (Beckert & Duschl 2002), and the radiation efficiency of the accretion process falls very quickly by several orders of magnitude. The luminosity decreases correspondingly.

Altogether, this leads to a three stage evolution of a quasar:

- **Eddington-limited phase:** During the phase in which accretion is limited by the Eddington rate, a sizable fraction of the central black hole is built up. This phase lasts as long as disk accretion is efficient enough that it exceeds the Eddington rate.

- **Free accretion phase:** As soon as the Eddington rate is no longer a relevant limit, all mass supplied by the disk can be accreted into the black hole. All the dissipated accretion energy can now be radiated. During this phase, the black hole accretion process runs at its maximum efficiency, although the rate itself and hence the luminosity declines with time (see Sect. 3.2.).

- **Advection-dominated phase:** When advection takes over, the actual quasar phase comes to an end (unless the galaxy encounters another major merger), and does so fairly abruptly. Due to the lack of large masses of accretable material, the growth of the black hole also effectively ceases. Minor episodes of accretion events may lead occasionally to comparatively short phases of enhanced activity. Such phases can also be experienced by non-quasar host galaxies and would be descibed as normal AGN of, for instance, Seyfert type.

A quasar model of this kind requires a very efficient underlying accretion mechanism. The process must be efficient enough to deliver sufficient mass from the disk towards its center so that the central black hole can grow quickly enough. Most quasars must reach their peak activity at a red-shift $z \sim 2$ (Hasinger 1998) in order to be compatible with the observed distribution of luminous quasars. Furthermore, the process involved must happen much more often in the $z > 1$ Universe and must almost shutdown at later epochs. This is indeed the case for major mergers.

3.2. Order-of-magnitude estimates

In the following, we will discuss a somewhat extreme example in order to show that the process discussed above is capable of giving rise to very massive quasar black holes. We assume an accretion disk of outer radius $s_{\rm disk} = 100\,{\rm pc}$ and initial mass of the disk of $10^{10}\,{\rm M}_\odot$. The choice of the seed black hole mass is not crucial as long as it is considerably smaller than the disk mass. We estimate the accretion rate $\dot{M}$ at time t as

$$\dot{M}(t) = \frac{M_{\rm disk}(t)}{\tau_{\rm visc}(t)}. \tag{4}$$

In the same spirit, we estimate v_φ at the disk's outer radius $s_{\rm disk}$ as

$$v_\varphi(t) = \sqrt{\frac{GM_{\rm disk}(t)}{s_{\rm disk}}}. \tag{5}$$

Equations 2 – 5 then lead to $\dot{M}(t) \propto M_{\rm disk}^{3/2}$. For the present estimates, we assume that the turbulent velocity is always subsonic. This, of course, has to be justified in later numerical model calculations.

With the above initial conditions, the mass flow rate deduced from Eq. 4 surpasses the Eddington rate by far at the beginning of the evolution. Subsequently, two counteracting effects come into play: The accretion process increases the black hole's mass and decreases the disk's mass. As a consequence, the Eddington rate increases, while the mass flow rate from the disk decreases.

In Eddington terms, the mass flow rate according to Eq. 4 becomes smaller and approaches unity. For the assumed initial disk parameters this transition point between Eddington-limited and free accretion is reached after $\sim 7 \cdot 10^8$ years, at a time when the central mass has reached $\sim 7 \cdot 10^9\,\mathrm{M}_\odot$. This is also the instant of peak accretion rate, which in the present case amounts to $\sim 100\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$.

From this point onwards, the mass flow rate is smaller than the Eddington rate, and the black hole accretion process is no longer Eddington-limited. What follows is the free accretion phase during which practically all mass supplied by the disk is accreted into the black hole. With a growing central mass and a shrinking disk mass, the mass flow rate decreases both in absolute as well as in Eddington terms. Again with the chosen parameters, the free accretion phase comes to an end after another $\sim 3 \cdot 10^8$ years when the mass flow rate drops to $\sim 0.5\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$, i.e., below approximately 0.3 % of its Eddington value and the flow becomes advection dominated. During this phase the central mass grows to $\sim 9 \cdot 10^9\,\mathrm{M}_\odot$, and the disk radiates all the dissipated energy.

The onset of advection dominated flow leads to a sharp drop in the radiation efficiency of the disk. Already within the first $\sim 1.5 \cdot 10^8$ years of the advection dominated phase, the radiation efficiency drops by about an order of magnitude while at the same time, the accretion rates drops by an additional factor of the same order. The disk radiates almost two orders of magnitude less than at the end of the free accretion phase, bringing the high luminous quasar phase to an abrupt end. In the course of time, the effect of advection becomes even stronger.

If we define—somewhat arbitrarily—the observable quasar phase to be the period during which the mass accretion rate is above $0.1\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$, then for this quasar its lifetime would have been somewhat less than 1 billion years.

4. Conclusions and Outlook

The essence of this scenario is that, within approximately one billion years the outcome of the merger has not only succeeded in "providing the fuel" for quasar activity, but also in "building the engine". As a consequence one then has to conclude that quasars occur in those galaxies which encountered major mergers during which large amounts of gas and dust were driven into the inner galaxies on short tidal time scales. Normal galaxies, in contrast, are those which never experienced a major merger. As a consequence, in today's Universe, the galaxies which once harbored a quasar have considerably more massive black holes in their centers ($10^9\,\mathrm{M}_\odot$ and more) then other galaxies.

For quasars an efficient accretion process is required which delivers enough material to the galaxy's center so that a black hole can grow quickly. This accretion process also has to be efficient enough not only to grow the black hole, but also to accrete away most of the available gas and dust and thus lead to a rapid end of the quasar phenomenon due to a combination of a drop in the accretion rate as well as in the disk's radiation efficiency. For this all to happen, the disk's viscosity is *the* crucial quantity. With the newly proposed generalization (DSB), self-gravitating disks are capable of doing this.

Currently we are carrying out numerical model calculations for the evolution of such disks in order to investigate in more detail the scenario presented here.

The results must also be compared with both quasar and merger statistics in order to test the assumed underlying physical processes.

While the scenario described above does not exclude the existence of primordial super-massive black holes in the young Universe, it makes their existence unnecessary as far as the quasar phenomenon is concerned.

References

Balbus, S.A., Hawley, J.F., 1991, ApJ 376, 214

Barnes, J.E., Hernquist, L., 1996, ApJ 471, 115

Beckert, T., Duschl, W.J., 2002, A&A (in press)

Chakrabarti, S.K., Wiita, P.J., 1993, ApJ 411, 602

Chandrasekhar, S., 1960, Proc. Natl. Acad. Sci. USA 46, 253

Duschl, W.J., 1988a, A&A 194, 33

Duschl, W.J., 1988b, A&A 194, 43

Duschl, W.J., Strittmatter, P.A., Biermann, P.L., 2000, A&A 357, 1123 (= DSB)

Hasinger, G., 1998, Astron. Nachr. 319, 37

Kant, I., 1755, Allgemeine Naturgeschichte und Theorie des Himmels oder Versuch von der Verfassung und dem mechanischen Ursprung des ganzen Weltgebäudes, Nach Newtonschen Grundsätzen abgehandelt

Kormendy, J., 2001, ASP Conf. Ser. 230, 247

Laplace, P.-S. de , 1796, Exposition du système du Monde

Lüst, R., 1952, Naturforsch. 7a, 87

Lynden-Bell, D., 1969, Nature 223, 690

Richard, D., Zahn, J.-P., 1999, A&A 347, 734

Shakura, N.I., 1972, AZh 49, 921 (translated in 1973, Soviet Ast. 16, 756)

Shakura, N.I., Sunyaev, R.A., 1973, A&A 24, 337

Shlosman, I., Frank, J., Begelman, M.C., 1989, Nature 338, 45

Stockton, A., 1999, IAU-Symp. 186, 311

Taylor, G.I., 1936a, Proc. Roy. Soc. London 157, 546

Taylor, G.I., 1936b, Proc. Roy. Soc. London, 157 565

Velikhov, E.P., 1959, Sov. Phys. JETP 36, 995

Warner, B., 1995, Cataclysmic Variable Stars, Cambridge University Press, Cambridge, UK

Wendt, F., 1933, Ingenieur-Archiv IV, 577