V. CONFERENCE SUMMARY

https://doi.org/10.1017/S0074180900162564 Published online by Cambridge University Press

Another Meal

Mary Kaiser and Byurakan employee

https://doi.org/10.1017/S0074180900162564 Published online by Cambridge University Press

Active Galactic Nuclei and Related Phenomena
IAU Symposium, Vol. 194, 1999
Yervant Terzian, Daniel Weedman, Edward Khachikian, eds.

Future Directions in AGN Research

Julian H. Krolik
Physics and Astronomy Department, Johns Hopkins University, Baltimore MD 21218

Abstract.
A review is given of the principal successes in AGN research (black holes as the central engines, radiation mechanisms, population studies) and the most important open questions (creation of AGN, dynamics of the accretion flow, acceleration and collimation of relativistic jets). Some recent work that gives hope for progress towards answering these questions is also discussed.

1. Successes

Before considering where the field may be headed in the future, it would be worthwhile to review how far we have already come. Although many important questions remain unanswered, there have also been a number of notable successes.

1.1. Black Holes

Perhaps the single most important of these is the identification of accretion onto massive black holes as the fundamental energy source for AGN. The most powerful argument for this idea is that nothing else surpasses the possible energy production efficiency, $\eta \sim 0.1$ in rest-mass units, when matter accretes into a relativistically deep potential.

Moreover, even when such a high efficiency is realized, the accumulated mass must still be very large. By counting how many AGN we see as a function of the flux they deliver at Earth, we can "scoop up" a sample of the total photon density created by AGN over the lifetime of the Universe (Sołtan 1982). If we ascribe a mean fuel efficiency to the production of these photons, we may then estimate the mean density today of AGN "ash". Carrying out this exercise with modern data yields a mean remnant mass per bright galaxy

$$\langle M_{\rm rem}\rangle \simeq 1.6\times 10^7 \left(\frac{F_{\rm bol}}{10F_{\rm B}}\right)\left(\frac{h}{0.75}\right)^{-3}\left(\frac{\langle 1+z\rangle}{3}\right)\left(\frac{\eta}{0.1}\right)^{-1}\frac{M_\odot}{L_*\text{–galaxy}}, \quad (1)$$

where the ratio of bolometric flux to B-band flux is scaled to a factor of 10, the Hubble constant is $100h$ km s^{-1} Mpc^{-1}, and the mean redshift of AGN light is scaled to 2.

Several important conclusions follow from this estimate. First, the mean remnant mass is minimized if *every* galaxy once housed an AGN; if only select

galaxies ever contained an AGN, the remnant mass per once-active galaxy must be even greater. Even with an efficiency close to the maximum conceivable, and with the assumption that the remnant mass is spread as evenly as possible over contemporary galaxies, the mass budget for AGN light production is interesting on a galactic scale. If the efficiency were substantially smaller, or if the remnants were found only in some galaxies, the mass per remnant would be even greater.

This theoretical argument for black holes as the prime movers of AGN has been strongly bolstered by recent observations. Fe Kα profiles with velocity widths several tenths of c have now been measured in many nearby AGN (Nandra et al. 1997). Clearly, at least mildly relativistic dynamics are required; orbits just outside a black hole are a very plausible location.

A second line of approach toward the identification of black hole central engines and AGN remnants has been to study the dynamics of the inner-most regions of galaxies, searching for the telltales of a Keplerian potential. Several different methods have all proved fruitful in different contexts: observing the kinematics of H_2O mega-maser spots (as reviewed, for example, by Willem Baan in this volume); using Hα emission to measure the rotation curves of orbiting gaseous disks (e.g. Harms et al. 1994); and studying how the stellar surface brightness and velocity dispersion behave as functions of radius from the galactic center (Magorrian et al. 1998). The specific results in terms of mass within a given radius vary from case to case, but discovering $\sim 10^8 M_\odot$ within ~ 10 pc of a galactic nucleus appears to be quite common.

1.2. Radiation Mechanisms

Major advances have also been made in the quest to work out the specific mechanisms by which different portions of the photon spectrum are made. Ordering by increasing frequency, we now have good evidence that

- Relativistic electrons generate the radio spectrum by synchrotron radiation. In blazars, the importance of synchrotron radiation extends to much higher frequencies, sometimes even into the X-ray band (see the reviews by Rita Sambruna and Marina Romanova in these proceedings).

- In non-blazar AGN, the infrared continuum can be safely attributed to thermal dust emission, although there is some uncertainty about how much of the energy can be attributed to the AGN proper, and how much to a nearby starburst region.

- The optical/ultraviolet continuum is very likely due to quasi-thermal emission from the surface of an accretion disk, although there are a great many detailed discrepancies between model predictions and observed spectra (Koratkar & Blaes 1999).

- Photoionization is almost certainly the culprit for both powering, and controlling the selection of, the rich spectrum of optical and ultraviolet emission lines generally seen in AGN. Although this conclusion had commanded general agreement long before, the monitoring experiments of the past few years have strongly confirmed this belief.

- Thermal Comptonization appears to provide an excellent explanation for the character of the hard X-ray spectrum (again excepting blazars).

- Non-thermal inverse Compton scattering is by far the most likely way to produce the high-energy photon continua of blazars, which often extend all the way up to GeV, and sometimes even to TeV energies (this subject is also discussed in the reviews by Sambruna and Romanova).

1.3. Population Studies

We also have a reasonable understanding of the character of AGN populations, and how they have evolved over the lifetime of the Universe. Both Pat Osmer and Vahé Petrosian reviewed for us the state of the art in determining how the luminosity function of AGN has changed from $z > 3$ to the present epoch. We have learned that extremely strong quasar activity was a distinctive mark of the period from $z \simeq 3$ to $z \simeq 1$, the time when, as we are now beginning to understand, the Universal star formation rate was at its peak (e.g., Madau, Pozzetti & Dickinson 1998).

The last decade has also seen a great improvement in our ability to recognize when two AGN are truly intrinsically different, and when they merely appear different due to anisotropic radiation and the accident of different viewing angles. Several different mechanisms are now known to create anisotropic appearance, and we can now, in many cases, trace their impact quite reliably.

Compact, flat-spectrum radio sources are almost always found in those AGN exhibiting dramatic variability, strong polarization, and powerful high-energy γ-ray emission. Objects such as these might appear to be quite distinct from extended, steep-spectrum radio sources, which are generally far less variable, much more weakly polarized, and undetectable above a few tens of keV. However, we now understand that these AGNs are in fact one and the same. They simply look very different because they contain jets of plasma moving at relativistic speed. The former phenomenology appears when the jet is moving almost straight at us; the latter is what we see when our line-of-sight is directed otherwise.

Similarly, there is now strong evidence for extremely optically thick belts of dusty gas occluding the majority of solid angle around the nuclei of both low-luminosity radio-quiet AGN (Seyfert galaxies) and high-luminosity radio-loud AGN (radio-loud quasars and radio galaxies). When our line of sight passes along the system axis, we are privileged to see the nucleus in all its glory, and we call it a type 1 Seyfert or a radio-loud quasar; when our line of sight is blocked by obscuring gas, we cannot easily detect the true nucleus, and rename the object a type 2 Seyfert or a radio galaxy.

2. Open Questions

That we have learned much does not negate the fact that much remains to be learned. We have made but little progress toward answering quite a number of the most fundamental questions about AGN that we might wish to ask.

https://doi.org/10.1017/S0074180900162564 Published online by Cambridge University Press

2.1. Why do they exist?

Perhaps the most basic of these unanswered questions is why AGN should exist at all. Although it somehow seems "natural" that stars should form from the collapse of interstellar gas clouds, we could easily imagine a Universe without any AGN at all. If massive black holes ready to accrete are indeed the *sine qua non* of active galactic nuclei, this question may be rephrased as "What makes massive black holes?"

We still cannot say whether the origin of massive black holes may be found in the direct collapse of very large interstellar gas clouds, or in stellar collapse and subsequent growth. Whichever mechanism creates the original event horizon, we do know that their growth must be, in a well-defined sense, very "rapid". Because the maximum mass accumulation rate due to an efficiently radiating accretion flow (i.e., the Eddington accretion rate) scales in proportion to the central mass, black holes accreting in this fashion grow exponentially with a characteristic timescale called the "Salpeter time"

$$t_S \simeq 4 \times 10^7 \left(\frac{\eta}{0.1}\right) \left(\frac{L}{L_E}\right)^{-1} \text{ yr.} \tag{2}$$

Note that to grow from, say, 10 $M_\odot$ to $\sim 10^8$ $M_\odot$ requires 16 e-foldings. On the other hand, the age of the Universe at the beginning of the quasar epoch ($z = 5$) was only

$$t_U \simeq (5\text{—}10) \times 10^8 \left(\frac{h}{0.75}\right)^{-1} \text{ yr,} \tag{3}$$

where the range corresponds to a range of reasonable values for q_o. Sixteen e-foldings barely fit within the time available (perhaps that's why quasars may have first appeared around then).

A closely related question is why certain galaxies are active (at least at any one time), and why certain kinds of galaxies prefer certain kinds of nuclear activity. In the contemporary Universe, essentially all radio-quiet AGN are in disk galaxies (Adams 1977, Huchra & Burg 1992) while essentially all radio-loud AGN are in ellipticals (Martel et al. 1998). On the other hand, at slightly earlier times ($z \simeq 0.3$), there are indications that this clear division may break down (McLure et al. 1998).

We might likewise ask, "How long does activity last?" and "Once it ceases in a certain galaxy, does it recur?" The answers to these questions are bound up with the inquiry into what controls the fueling of active nuclei. The gravitational influence of even a very large central black hole cannot extend far enough into the host galaxy for it to be able to control its own fuel supply; the radius outside of which the deepening of the gravitational well due to the central black hole becomes negligible is

$$r_* \simeq 40 \left(\frac{M}{10^8}\right) \left(\frac{\sigma_*}{100 \text{ km s}^{-1}}\right)^{-2} \text{ pc,} \tag{4}$$

where σ_* is the *rms* stellar orbital speed. There simply isn't enough fuel within that small a portion of a galaxy to supply what is needed, so it must be events farther out, connected with the ordinary life of the galaxy, that ultimately control how much fuel the nucleus is permitted to consume.

https://doi.org/10.1017/S0074180900162564 Published online by Cambridge University Press

As mentioned earlier, the whole phenomenon of galactic nuclear activity reached its peak somewhere between $z \simeq 3$ and $z \simeq 1$. This fact immediately raises the suspicion that there is something about the youth of galaxies that encourages nuclear activity. Just what that might be, however, remains unknown. Many have speculated that this epoch was specially favorable to AGN ignition and fueling because it was a time when the ratio of gas mass to stellar mass in galaxies was relatively high. Because fluids are much more dissipative than collisionless systems like stars, it might therefore have been much easier to move matter inward then. On the other hand, others have favored the hypothesis that AGN activity was so much stronger then because the rate of major galaxy-galaxy encounters was very high. Even if the co-moving number density of galaxies hasn't changed since that time, the physical density was, of course, larger by a factor $(1+z)^3$, so the rate of encounters should have been that much greater. Moreover, many of the currently most fashionable theories about galaxy formation also posit that galaxies are assembled from small pieces, so that the *co-moving* number density of galaxies was also much larger then than now. We do not know which of these suggestions is more nearly correct, or whether some other effect was more important.

2.2. Accretion dynamics

A second area where we are truly far short of our goals is our understanding of accretion dynamics. That we do not genuinely understand accretion is immediately demonstrated by the gap between what we might predict for the output spectrum of AGN and reality.

According to the standard picture of accretion dynamics, the gas settles down into a geometrically and optically thick system. It should therefore radiate in a quasi-thermal fashion, with the temperature declining outward roughly as $r^{-3/4}$:

$$T_d \simeq 6.8 \times 10^5 \eta^{-1/4} L_{46}^{-1/4} \left(\frac{L}{L_E}\right)^{1/2} x^{-3/4} R_R^{1/4}(x) \text{ K}, \tag{5}$$

where L_{46} is the total luminosity scaled to 10^{46} erg s^{-1}; $x = rc^2/GM$, the radius in gravitational units; and R_R is a correction factor that includes both the outward flow of energy carried with the outward flow of angular momentum and general relativistic effects. R_R approaches unity at large x, and falls to zero at the innermost ring of the disk. This model would predict that AGN would radiate predominantly in the UV, with little power emerging at wavelengths either much longward of the optical band, or much shorter than a few tens of eV.

The reality is, of course, dramatically different. AGN radiate their power almost even-handedly in luminosity per logarithm of wavelength from the mid-infrared to at least hard X-rays, and sometimes beyond. There is often a modest local maximum in the UV, where accretion disk models predict that it should be found, but this maximum does not dominate the total luminosity anywhere near as thoroughly as the simple disk model would lead us to expect. Although, as mentioned in §1, the infrared continuum can be ascribed to reradiation of a primary continuum carried initially in higher energy photons, distant thermal reprocessing of ultraviolet photons cannot explain the strength of even the EUV continuum, much less hard X-rays or γ-rays.

https://doi.org/10.1017/S0074180900162564 Published online by Cambridge University Press

There are also smaller scale problems. Virtually every detailed calculation of AGN accretion disk atmospheres has predicted some sort of Lyman edge feature, although effects such as oblique view of an accretion disk around a rapidly-spinning black hole can reduce the apparent strength of such features.

The failure to explain the very broad-band character of AGN emission may be interpreted in another way as a fundamental failure to understand why, in AGN accretion disks, a substantial fraction of the dissipated heat is concentrated onto a small fraction of the mass. Such a segregation of the dissipated energy is the only way to channel enough energy into "coronal" plasma capable of creating the observed EUV and X-ray continua.

In fact, there is an even deeper dynamical problem. Canonical disk models (in which the local stress is proportional to the total pressure) predict that the inner regions of AGN disks should be dominated by radiation pressure, and are therefore both thermally and viscously unstable. No satisfactory model has yet been found for these regions that is free of instability, although a number of speculative alternatives have been proposed.

2.3. Formation of relativistic jets

Yet another mystery about AGN on which we have made little headway is the nature of relativistic jets. That we see such jets is incontrovertible: from VLBI observations of superluminally-separating radio spots to the strength of high-energy γ-ray emission in blazars (a fact that requires relativistic outflow in order to reduce the apparent γ-γ opacity of the source plasma), we see unmistakable signatures of relativistic motion. The problem is that we have no answers to the most basic questions about them.

Although we can estimate global quantities such as total mechanical power and momentum flux, we do not know the nature of the plasma inside them: Is it an ordinary electron-ion plasma, or are most of the positive charges provided by positrons? Likewise, is most of the energy carried by matter or by Poynting flux?

We are also entirely ignorant about how they are accelerated and collimated. Most workers in the field believe that somehow they are expelled from the immediate vicinity of the central black hole by some combination of MHD effects and rotation, but there is no consensus beyond that vague statement.

Nor have we been able to identify clearly the energy source for jets. Some argue that their energy comes predominantly from the same accretion flow that powers everything else; others insist that the energy is drawn from the rotational energy of the black hole.

Finally, there is the basic mystery of why strong jets occur in only about 10% of all AGN, the radio-loud minority. This question is, of course, coupled to the one raised earlier about why (at least at the present epoch) radio-loud AGN (i.e., strong jet AGN) can be found only in elliptical galaxies, and never in disks.

3. Hopes for Progress

Having posed these questions, it is time to point out the directions in which we may hope for some progress toward answering them.

3.1. Existence

Several different lines of approach may soon yield clues to the creation of quasars. Major programs are being pursued to image the inner regions of AGN hosts, both near and far. Of course, the central technical problem to overcome is eliminating the bright point source that is the AGN itself. Starlight in the host can only be seen clearly after the contribution of the AGN is isolated and removed. Consequently, the best data come from those instruments with the highest angular resolution; in this case, that means the HST.

One way to further ensure complete removal of the AGN light is to look at AGN in which the nucleus is thoroughly obscured by opaque matter very close to the galactic center. Type 2 Seyfert galaxies are, of course, ideal for this purpose. HST observations (González Delgado et al. 1998, Storchi-Bergmann in these proceedings) have now shown that bright starbursts often occupy the inner few hundred parsecs of the hosts of type 2 Seyfert galaxies. One naturally wonders whether the existence of these starbursts has any causal relation to the creation of an active nucleus.

When looking at higher luminosity AGN, i.e. quasars, the number of known obscured nuclei is so small as to preclude making use of such a natural "coronagraph". Instead, one has no choice but to try to subtract out the point source light. Combining the results of Bahcall, Kirhakos & Schneider (1996) with those of McLure et al. (1998), we have learned that the hosts of quasars are an unusual lot, with many showing signs of disturbance.

Further pursuit of this program should bring further rewards, but genuine progress will require moving from the collection of examples to systematic statistical studies. Fortunately, the construction of such studies is about to become much easier, for several enormous surveys are about to burst onto the scene. Over the next five years, the Sloan Digital Sky Survey will image 1/4 of the entire sky, collecting spectra for 10^6 galaxies and 10^5 quasars. Meanwhile, the 2MASS (Two Micron All Sky Survey) is roughly halfway through covering the entire sky in the near-infrared. It will ultimately produce a catalog of $\sim 10^6$ galaxies brighter than $K \simeq 13.5$ mag. In a few years, the GALEX spacecraft will obtain UV images of 10^7 galaxies and 10^6 quasars, and UV spectra for 10^5 galaxies and 10^4 quasars. The FIRST (Faint Images of the Radio Sky at Twenty-one centimeters) survey has produced VLA images of roughly 10% of the sky, with an ultimate goal of covering about 25%. Its ultimate catalog should contain $\sim 10^6$ entries.

These enormous databases should be useful in several ways. Clear sample definition will become much easier because they have well-defined, homogeneous selection rules. Because of their tremendous size, it should be possible to find statistically significant samples even after taking "cuts" in a variety of parameters. Such overwhelmingly large catalogs should also turn up examples of unusual and rare varieties, some of which may be especially instructive.

Yet another positive trend, remarked upon by numerous speakers at this Symposium (Malcolm Longair, Dave Sanders, and Gene Smith), is the convergence between galaxy formation and AGN studies. With our newfound ability to observe galaxies in the same redshift range where the quasar activity peak is located, we can now seriously begin to develop the connection between galaxy formation and nuclear activity that has so long been merely a matter of spec-

ulation. Real data about the character of inactive galaxies at $z \simeq 2$ may be extremely helpful in understanding what made so many other galaxies' nuclei active at that time.

3.2. Accretion dynamics

There is also a great deal of activity in the area of accretion dynamics. Not so long ago, state-of-the-art accretion disk atmosphere codes incorporated only Thomson and free-free opacity. Now they include: bound-free opacity for both H and He, with a non-LTE treatment of their ionization balance; opacity due to heavy element resonance lines and ionization edges; non-diffusive Comptonization; and explicit treatment of vertical force balance, including radiation pressure (Koratkar & Blaes 1999). These dramatic improvements in calculational quality should permit the confrontation between theoretical predictions and observations to become ever more pointed. Soon we should be able to say clearly whether the almost-universal absence of Lyman edge features truly poses a fundamental problem for conventional ideas about disk structure. Similarly, theorists should soon be able to make clear statements about when they would predict HeII edges, and about the character of UV polarization to be expected.

The last few years have also seen the emergence of a number of important new ideas about fundamental disk dynamics. The invention of the Balbus-Hawley mechanism (reviewed in Balbus & Hawley 1998), has finally given us a plausible physical picture of how angular momentum is transported in accretion disks. With this advance, we are now poised to learn how and where the accretion energy is dissipated into heat, so that it may eventually be radiated as photons. A description from first principles of where the dissipation occurs will remove a major systematic uncertainty from disk atmosphere models.

Another important new development in fundamental disk physics is a deeper understanding of disk equilibria and stability. Narayan & Yi (1995) emphasized the importance of the fact that low accretion rates may permit "advection-dominated accretion", i.e., a condition in which the gravitational potential energy of the accreting matter is dissipated into heat, but is advected into the black hole (or used to drive a wind) rather than being lost in photons. Others have stressed that high accretion rate disks are subject to such strong radiation pressure-driven instabilities that the conventional Shakura-Sunyaev equilibrium likely does not exist in Nature. One suggestion is that the result is a limit-cycle (Szuszkiewicz & Miller 1998); another is that the disk adopts a qualitatively different equilibrium, in which most of the matter is found in relatively high-density clumps, and the rest is more smoothly distributed. In this latter picture, the low density gas rises to a temperature so high that it cools by Compton scattering and produces the observed hard X-ray emission (Krolik 1998).

In fact, the interplay between X-rays produced by Comptonization and their reprocessing in the (presumably) adjacent accretion disk has opened up a whole new field of observational studies, coordinated monitoring of X-ray and ultraviolet monitoring (as reported, for example, in the talks by Toshihiro Kawaguchi and Thierry Courvoisier). By following the variations in both bands, we can reasonably hope to understand better the double feedback loop (ultraviolet photons from the disk upscattered to X-rays, X-rays partially absorbed by the disk and reprocessed into the ultraviolet) that connects the two systems.

Still more information on disk dynamics should be forthcoming as the quality of Fe Kα line profiles improves. Hitherto, these have been badly limited by the inability of ASCA data to clearly define the continuum shape (ASCA's sensitivity falls drastically above 6 keV). However, data should be available very soon from coordinated ASCA and RXTE observations, in which RXTE spectra define the continuum shape that can then be subtracted from the ASCA spectra in order to isolate the Kα profile. Moreover, the order of magnitude increase in sensitivity promised by XMM should permit both much higher S/N Fe Kα profiles in bright AGN, and the extension to high redshift of the entire field of study. Indeed, there may be some quasars whose redshifts bring the line within the range spanned by the high-resolution spectrometer on XMM; from those quasars (if they are bright enough), we may be able to obtain a much finer measure of the shape of the Kα line.

3.3. Jets

The complexities of jet acceleration and collimation physics demand numerical simulation (reviewed in this volume by Dick Lovelace). Fortunately, the rapid growth in computational power makes possible today calculations that even a few years ago were only pipedreams. Reasonable resolution 2-d MHD simulations in a general relativistic background are quite feasible today; 3-d should not be far in the future.

Moreover, advances in disk physics should contribute to new steps in understanding the origins of jets. The Balbus-Hawley mechanism doesn't only explain inter-ring torques in disks; it also explains (and predicts) the strength and structure of disk magnetic fields. A physical understanding of dissipation in disks should also be of great aid when we try to discover how disk heating can help launch jets.

On the observational side (as discussed by Thierry Courvoisier), monitoring campaigns have much to offer this subject, just as they do for disk-corona relations. Multiwavelength monitoring campaigns can trace the propagation of shocks down jets, constraining our ideas about how and where relativistic electrons are accelerated and cooled, as well as ideas about the source and transmission of the seed photons later inverse Compton scattered up to high energies.

References

Adams, T.F. 1977, ApJS 33, 19

Bahcall, J.N., Kirhakos, S. & Schneider, D.P. 1996, ApJ 457, 557

Balbus, S.A. & Hawley, J.F. 1998, Revs. Mod. Phys. 70, 1

González Delgado, R.M., Heckman, T., Leitherer, C., Meurer, G., Krolik, J., Wilson, A.S., Kinney, A. & Koratkar, A. 1998, ApJ505, 174

Huchra, J.P., and Burg, R. 1992, ApJ 393, 90

Koratkar, A. & Blaes, O.M. 1999, PASP in press

Krolik, J.H. 1998, ApJ 498, L13

Madau, P., Dickinson, M. & Pozzetti, L. 1998, ApJ 498, 106

Magorrian, J., Tremaine, S.D., Richstone, D., Bender, R., Bower, G., Dressler, A., Faber, S.M., Gebhardt, K., Green, R.F., Grillmair, C., Kormendy, J. & Lauer, T. 1998, AJ 115, 2285

Martel, A.R., Baum, S.A., Sparks, W.B., Wyckoff, E., Biretta, J.A., Golombek, D., Macchetto, F.D., de Koff, S., McCarthy, P.J. & Miley, G.K. 1998, ApJS in press

McLure, R.J., Dunlop, J.S., Kukula, M.J., Baum, S.A., O'Dea, C.P. & Hughes, D.H. 1998, preprint astro-ph/9809030

Nandra, K., George, I.M., Mushotzky, R.F., Turner, T.J. & Yaqoob, T. 1997, ApJ 477, 602

Narayan, R. & Yi, I. 1995, ApJ 452, 710

Sołtan, A. 1982, M.N.R.A.S. 200, 115

Szuszkiewicz, A. & Miller, J.C. 1998, MNRAS 298, 888