Unification and Evolution of AGN

Moshe Elitzur

University of Kentucky, Department of Physics and Astronomy, Lexington, KY, USA
E-mail: moshe@pa.uky.edu

Observations give strong support for the unification scheme of active galactic nuclei (AGN), based on toroidal dusty obscuration of the central engine. Clumpiness of the obscuring dust is crucial for explaining the IR observations and has significant consequences for AGN classification: type 1 and type 2 viewing is an angle-dependent probability, not an absolute property. The broad line region (BLR) and the dusty torus are, respectively, the inner and outer segments, across the dust sublimation radius, of a continuous cloud distribution. All clouds are embedded in a disk wind, whose mass outflow rate is diminishing as the accretion rate, i.e., AGN luminosity, is decreasing. Both the torus and BLR disappear at sufficiently low luminosities, leaving radio jets as the sole release channel for the accreted mass that does not reach the central black hole. The enhanced prominence of the jets at low Eddington ratios may provide important studies for the CTA.
1. AGN Unification

The classification of AGN into types 1 and 2 is based on the extent to which the nuclear region is visible. The basic premise of the unification scheme is that every AGN is intrinsically the same object: an accreting supermassive black hole. This central engine is surrounded by a dusty toroidal structure so that the observed diversity simply reflects different viewing angles of an axisymmetric geometry (figure 1). Directions with clear sight of the central engine and the broad-line region (BLR) yield type 1 sources. Those blocked by the torus from direct view of the BLR result in type 2 objects, where the existence of the hidden BLR is revealed only in polarized light [2]. From basic considerations, Krolik & Begelman [18] concluded that the torus likely consists of a large number of individually very optically thick dusty clouds. Indeed, VLTI interferometric observations of the Circinus AGN provide strong evidence for a clumpy or filamentary dust structure [33].

In its standard formulation, the unification approach posits the viewing angle as the sole factor in determining the AGN type, and this is indeed the case for any smooth-density torus whose column density declines with angle \( \beta \) away from the equatorial plane. The AGN is obscured from directions that have \( e^{-\tau_V(\beta)} \gg 1 \), where \( \tau_V \) is optical depth at visual, and visible from those with \( e^{-\tau_V(\beta)} \ll 1 \). Because of the steep variation of \( e^{-\tau} \) with \( \tau \), the transition between these two regions is sharp, occurring around the angle \( \sigma \) where \( \tau_V = 1 \). Then, so long as \( \tau_V(0) \gg 1 \) and \( \tau_V(\frac{1}{2}\pi) \ll 1 \), all AGN viewed at \( 0 \leq i < \frac{1}{2}\pi - \sigma \), where \( i \) is angle from the axis, appear as type 1 sources, those at \( \frac{1}{2}\pi - \sigma \leq i \leq \frac{1}{2}\pi \) as type 2. If \( f_2 \) denotes the fraction of type 2 sources in the total population then \( f_2 = \sin \sigma \) for all smooth-density tori, irrespective of their specific angular profiles. This relation has been employed in all studies of source statistics performed to date. From statistics of Seyfert galaxies Schmitt et al [29] find that \( f_2 \simeq 70\% \), hence their estimate \( \sigma \simeq 45^\circ \). The issue is somewhat unsettled; Hao et al [15] later found that \( f_2 \) in Seyfert galaxies is only about 50\%, or \( \sigma \simeq 30^\circ \).

Two basic tenets of the standard unification analysis require modification: (1) The obscuration is clumpy, not smooth, and (2) even at the same luminosity, the torus is highly unlikely to have the exact same covering factor in all AGN. Each issue is now addressed separately.
Figure 2: AGN statistics—the fraction $f_2$ of obscured sources as a function of the torus angular width $\sigma$ (see [23]). In a uniform density sharp-edge torus (cf fig. 1) this fraction is determined uniquely by $\sigma$, and is shown with the dashed line. In contrast, in a clumpy torus with the Gaussian angular distribution $N(\beta) = N_0 e^{- (\beta/\sigma)^2}$, where $N$ is the number of clouds along radial ray at angle $\beta$ from the equatorial plane, $f_2$ depends on both $\sigma$ and $N_0$, which is marked on the various solid lines.

1.1 Clumpy Unification

Because of clumpiness, the difference between types 1 and 2 is not truly an issue of orientation but of probability for direct view of the AGN (figure 1, panel c): AGN type is a viewing-dependent probability. Since that probability is always finite, type 1 sources can be detected from what are typically considered type 2 orientations, even through the torus equatorial plane: if the number of clouds along radial rays in that plane is, for example, $N_0 = 5$ on average, the probability for that is $e^{-5} = 1/148$. This might offer an explanation for the few Seyfert galaxies reported in [1] to show type 1 optical line spectra together with 0.4–16 $\mu$m SED that resemble type 2. Conversely, if a cloud happened to obscure the nucleus from an observer, that AGN would be classified as type 2 irrespective of the viewing angle. In cases of such single cloud obscuration, on occasion the cloud may move out of the line-of-sight, creating a clear path to the nucleus and a transition to type 1 spectrum. Such transitions between type 1 and type 2 line spectra have been observed in a few sources (see [4] and references therein). It is worth while to conduct monitoring observations in an attempt to detect additional such transitions. The most promising candidates would be obscured systems with relatively small X-ray obscuring columns, small torus sizes and large black-hole masses [23].

A sharp-edge uniform-density clumpy torus has $f_2 = (1 - e^{-N_0}) \sin \sigma$, where $N_0$ is the average number of clouds along radial rays within angle $\sigma$ from the equatorial plane, and is practically indistinguishable from its smooth-density counterpart when $N_0$ exceeds $\sim 3–4$. However, the situation changes fundamentally for soft-edge distributions because at every viewing angle, the probability of obscuration increases with the number of clouds. As is evident from figure 2, the Gaussian
distribution produces a strong dependence on \( N_0 \) and significant differences from the sharp-edge case. Since the sharp-edge angular distribution is ruled out by observations [23], the fraction of obscured sources depends not only on the torus angular width but also on the average number of clouds along radial rays. While the fraction \( f_2 = 70\% \) requires \( \sigma = 45^\circ \) in the smooth-density case, it implies \( \sigma = 33^\circ \) in a Gaussian clumpy torus with \( N_0 = 5 \) clouds.

There are indications that the fraction \( f_2 \) of obscured sources decreases with bolometric luminosity \( L \) (see [19] and references therein). The possible decrease of \( f_2 \) with \( L \) has been interpreted as support for the “receding torus” model, in which \( \sigma \) decreases with \( L \) (see [31] and references therein). However, all the quantitative analyses performed thus far for the \( L \)-dependence of \( f_2 \) were based on sharp-edge angular obscuration. Removing this assumption affects profoundly the foundation of the receding torus model because the dependence on the number of clouds necessitates analysis with two free parameters, therefore \( \sigma \) cannot be determined without \( N_0 \). A decrease of \( N_0 \) with \( L \) at constant \( \sigma \) will also produce a decrease in \( f_2 \), the same effect as a decrease of \( \sigma \) (figure 2). An observed trend of \( f_2 \) with \( L \) may arise from a dependence on either \( \sigma \) or \( N_0 \) or both. There is no obvious a-priori means for deciding between the various possibilities. Shape variations of the 10\( \mu \)m dust silicate feature in IR spectra of AGN indicate that the decrease in torus covering factor with luminosity may reflect a decrease in \( N_0 \) rather than \( \sigma \) [23], but more data is needed to establish this effect with any certainty.

1.2 Covering-Factor Variations

In unification’s most extreme form, the viewing angle is the sole parameter responsible for the type 1/type 2 dichotomy. That is, “extreme unification” posits an identical torus for all AGN. The observational evidence for decrease of the torus covering factor with AGN luminosity shows that extreme unification cannot be universally correct. In fact, even AGN with the same luminosity cannot be expected to all have the exact same torus; there must be a spread in torus properties among AGN with the same \( L \). This has immediate consequences for unification statistics. In panel (a) of figure 1, observer 1 will see the AGN as type 1, observer 2 as type 2. In panel (b) the AGN is oriented exactly the same as in panel (a), only its torus has a smaller covering factor. Now the two observers will both see a type 1 AGN even though their viewing angles have not changed at all. Obviously, the torus covering factor \( C_T \) (same as the fraction \( f_2 \) of type-2 sources mentioned above) is central to the AGN classification:

**A:** An AGN whose torus has a larger covering factor has a higher probability to be viewed as type 2 by a random observer

From this statement it follows immediately that in a sample of AGN with any distribution of covering factors, those with larger \( C_T \) will have a higher probability to be viewed as type-2 by a random observer; namely,

**B:** AGN’s are drawn preferentially from the distribution of covering factors; type-1 are more likely drawn from the distribution lower end, type-2 from its higher end

Although this statement appears to contradict the basic idea of AGN unification, it does not; it is the exact same as the previous statement, which is obvious. Statement B, a complete equivalent
Figure 3: Clumpy torus covering factors. Contour plots are for a Gaussian cloud distribution (see fig. 2 caption). Each contour is the locus of \( N_0 - \sigma \) combinations that produce the labeled covering factor. The data points are from modeling of IR observations reported in [21], [25] and [27].

of statement A, implies that the only random variable in AGN unification statistics is the viewing angle.

Statement A, and thus its consequences, is even more obvious when considering a clumpy torus as shown in panel (c) of figure 1. If \( N(\beta) \) is the overall number of clouds encountered, on average, along radial ray at angle \( \beta \) from the equatorial plane then the probability for direct viewing of the AGN in that direction is \( e^{-N(\beta)} \), same as the probability for a photon emitted by the AGN to escape all clouds on its way out. So the probability for absorption in direction \( \beta \) is \( 1 - e^{-N(\beta)} \) and integrating over all directions gives the clumpy torus covering factor:

\[
C_T = 1 - \int_0^{\pi/2} e^{-N(\beta)} d\sin\beta.
\] (1.1)

For the Gaussian cloud distribution introduced in [22, 23] (see also figure 2), obviously the covering factor increases with both \( N_0 \) and \( \sigma \). From statement B above it follows that the covering factors of a random sample of AGN will not be distributed randomly in the \( N_0-\sigma \) plane; instead, type 2 sources will preferentially occupy the regions at the high end of \( N_0 \) and \( \sigma \), type 1 the lower end. Clumpy torus modeling of IR observations has already been performed for a number of sources, and figure 3 shows the results of these models in the \( N_0-\sigma \) plane together with the contour plots of
the covering factor. As expected, type 1 and type 2 AGN preferentially occupy different regions of the $N_0-\sigma$ plane, although more data, especially for type 2 sources, is needed to establish the effect with certainty. This has a number of immediate implications:

- Covering factors deduced from analysis of type 1 sources do not necessarily apply to type 2, and vice versa.
- The finding that at least 50% of Seyfert 2 galaxies have Compton thick X-ray obscuration need not imply a similarly large fraction among Seyfert 1.
- The torus IR luminosity obeys $L_T = C_T L$, where $L$ is the AGN UV/optical luminosity. With a larger covering factor, type 2 AGN have, on average, a larger fraction of their luminosity converted to IR.

2. AGN Evolution

Mass accretion into a black hole at the rate $\dot{M}$ yields a luminosity $L = \eta \dot{M} c^2$, where $\eta$ is the radiation conversion efficiency. A standard $\eta \sim 0.1$ yields $\dot{M} \sim 0.1L_{45} M_\odot \text{yr}^{-1}$, where $L_{45} = L/10^{45} \text{erg s}^{-1}$. The AGN phase lasts probably $\sim 10^7-10^8$ years, thus the overall mass accreted is $\sim 10^6-10^7 L_{45} M_\odot$. Since Seyfert luminosities do not exceed $\sim 10^{45} \text{erg s}^{-1}$ at most, the entire mass they typically accrete from the host galaxy during an AGN episode is no more than that contained in a single giant molecular cloud. Thus a Seyfert episode may represent the occasional capture of a giant molecular cloud by the supermassive black-hole that is now known to exist at the centers of most galaxies. In contrast, quasar luminosities reach upward of $10^{48} \text{erg s}^{-1}$. If QSO lifetimes are similar, the mass fed into their central black hole is comparable to that of an entire galaxy. This suggests that the quasar and Seyfert nuclear activities could be triggered differently. Starting a quasar might require interaction with another galaxy to trigger accretion of the prodigious amounts of fuel that can only be supplied by a full galactic mass. Indeed, Sanders et al noted a number of fundamental differences between the Ultra Luminous IR Galaxies (ULIRG) at the high luminosity end ($\geq 10^{12} L_\odot$) and their lower-luminosity counterparts, and proposed the following ULIRG–QSO evolutionary scheme: The most powerful ULIRG represent major mergers of gas-rich spiral galaxies. Collisions between giant molecular clouds of the approaching galaxies trigger strong starburst activity, followed by merger of the nuclei and the birth of a dust-shrouded quasar. Dust clear-up by the nuclear activity, probably because of powerful outflows, exposes the young QSO, which keeps growing by strong accretion. Subsequent quasar evolution reflects the decline in accretion rate and turn off of the AGN phase. Figure 4, from Haas et al, sketches a more recent variant of the “Sanders scheme”.

Powerful outflows, expected to clear up the dust shroud of the young quasar, also play a prominent role in what has become known as the “AGN-Galaxy Connection”. The main impetus for this concept is the finding that most nearby early-type galaxies contain massive black holes and that the black-hole mass $M_*$ scales as $M_* \propto \sigma^4$, where $\sigma$ is the stellar velocity dispersion in the host galaxy. It would be remarkable if this correlation between the black-hole mass and the

$^1$Earlier versions of this figure were presented in [7] and in [27]. The contour plots in both of these earlier figures are afflicted by the computer bug reported in [24].
Figure 4: Scheme of quasar evolution for the surrounding dust distribution and corresponding SEDs (from [14]). During the evolution (from top to bottom) the AGN strength grows, then stays high and finally declines, as is marked by the size of the central ⋆ and shown by the SEDs optical slope.

gravitational potential of the galaxy at large were a mere coincidence. More likely, the correlation reflects a causal connection, and it has been suggested that the same outflows quench both the black-hole growth and star formation in the surrounding galaxy. Winds sufficiently powerful were recently discovered by Arav et al [3] in quasar ionic spectra. However, to inhibit star formation in the host galaxy, outflows have to affect the molecular gas out of which stars form and there remains the question whether such winds would originate in the central QSO or the surrounding starburst. Recent Herschel observations by Sturm et al [32] shed some light on these issues. With the 79µm rotational transition of OH as a tracer they detected massive molecular outflows in far-infrared spectra of ULIRGs. In some of these objects the terminal outflow velocities exceed 1,000 km s\(^{-1}\), and their outflow rates (up to \(\sim 1,200 M_\odot \text{yr}^{-1}\)) are several times larger than their star formation rates. Comparing the outflow signatures in different types of objects they report preliminary evidence that ULIRGs with a higher AGN luminosity have higher terminal velocities and shorter gas depletion timescales. In contrast, the data do not support the correlations that would be expected if the outflows were powered by starbursts. The outflows in the observed ULIRGs are able to expel the cold gas reservoirs from the centers of these objects within \(\sim 10^6-10^8\) years. Identifying the physical mechanism behind these outflows is a challenge since the mass outflow rates exceed the mass accretion rates into the AGN black-holes by three orders of magnitude.
2.1 Low-Luminosity Evolution

In spite of the considerable success of the unification scheme there is now clear evidence [17] that the BLR is actually missing, and not just hidden, in many low-luminosity AGNs (LLAGNs); these sources have been named “pure” or “true” type 2 AGNs. The BLR disappearance finds a natural explanation in the disk-wind scenario, first proposed in [10]. The AGN accretion disk appears to be fed by a midplane influx of cold, clumpy material from the main body of the galaxy. Approaching the center, conditions for developing hydromagnetically- or radiatively-driven clumpy winds above this equatorial inflow become more favorable. The composition along each streamline reflects the origin of the outflow material at the disk surface. The disk outer regions are dusty and molecular, as observed in water masers in some edge-on cases. Moving inward, at some smaller radius the dust is destroyed and the disk composition switches to atomic and ionized. The outflow from the inner atomic/ionized region feeds the BLR while the dusty clouds in the wind outer regions obscure the inner zones. As clouds rise away from the disk they expand and lose their column density, limiting the vertical scope of both broad-line emission and dust obscuration and emission. The result is a toroidal geometry for both the BLR and the obscuring region, i.e., the torus, which may be more appropriately named in this scenario the Toroidal Obscuration Region (TOR). Thus the toroidal obscuration arises in this picture from a dynamic rather than hydrostatic structure. An immediate consequence of this scenario is the prediction that the TOR and BLR disappear at low bolometric luminosities (i.e., low accretion rates; [9, 6]). The reason is that, as the mass accretion rate decreases, the mass outflow rate of a disk wind with fixed radial column decreases more slowly and thus cannot be sustained below a certain accretion limit. This unavoidable conclusion follows from simple considerations of mass conservation. In accordance with this model predictions, data from a nearly complete sample of nearby AGNs show that the BLR disappears at luminosities lower than $5 \times 10^{39} \left( \frac{M_*}{10^7 M_\odot} \right)^{2/3} \text{erg s}^{-1}$ [8]; every source below this limit is a “true” type 2 AGN. The TOR disappearance has been verified, too, in a number of independent studies which show the lack of obscuration [5, 20] and of thermal dust emission [35, 26, 34] in LLAGNs.

Since the accreted mass cannot be channeled in full into the central black hole, with the disk-wind turning off the system must find another release for the excess mass, and the only remaining channel is the radio jets. Indeed, Ho [16] finds that the AGN radio loudness $R = \frac{L_{\text{radio}}}{L_{\text{opt}}}$ is inversely correlated with the mass accretion rate $L/L_{\text{Edd}}$. This finding is supported by Sikora et al [30], who have greatly expanded this correlation and found an intriguing result: $R$ indeed increases inversely with $L/L_{\text{Edd}}$, but only so long as $L/L_{\text{Edd}}$ remains $\gtrsim 10^{-3}$. At smaller accretion rates, which include all FR I radio galaxies, the radio loudness saturates and remains constant at $R \sim 10^4$. This is precisely the expected behavior if as the outflow diminishes, the jets are fed an increasingly larger fraction of the accreted mass and finally, once the outflow is extinguished, all the inflowing material not funneled into the black hole is channeled into the jets, whose feeding thus saturates at a high conversion efficiency of accreted mass. It is important to note that radio loudness reflects the relative contribution of radio to the overall radiative emission; a source can be radio loud even at a low level of radio emission if its overall luminosity is small, and vice versa. The increased jets prominence at low Eddington ratios may offer important studies for CTA.
References


Unification and Evolution of AGN

Moshe Elitzur


