1
The Observational Picture of AGN

1.1
From *Welteninseln* to AGN

The dawn of extragalactic astronomy can be attributed to the year 1750, in which Thomas Wright speculated that some of the nebulae observed in the sky were not actually part of the Milky Way, but rather independent Milky Ways themselves (Wright, 1750). A few years later, Immanuel Kant introduced the term "Welteninseln" for these distant nebulae ("island universes"; Kant, 1755). It was François Arago in 1842 who first called the attention of astronomers to Kant,[1] whom he calls "the Astronomer of Königsberg," and declared that his name in that connection did not deserve the oblivion into which it had fallen (Arago and Barral, 1854). Thus the extragalactic hypothesis spread rapidly in the scientific community, although it was still not completely accepted as true. One main difficulty was the fact that some of the nebulae were actually of galactic origin, such as planetary nebulae and globular clusters. A significant step forward was the compilation of a large catalog of some 5000 nebulae assembled by William Herschel in the late eighteenth and early nineteenth century. Another advance was made by Lord Rosse, who constructed in 1845 a new $72''$ telescope in Ireland, managing to distinguish individual point sources in some of the nebulae, and therefore giving further support to Kant's and Wright's hypothesis. Spectroscopic observations by Vesto Slipher of nebulae in the early twentieth century revealed that some of these show redshifted lines indicating they are moving relative to the Milky Way at velocities exceeding the escape velocity of our Galaxy (Slipher, 1913).

The issue, whether some of the observed nebulae are actually extragalactic, was finally settled in the 1920s. In 1920 Heber Curtis summarized a number of arguments why the Andromeda "nebula" M31 is a galaxy of its own, similar to the Milky Way (Curtis, 1920). For example, he noticed a Doppler shift in M31 due to its rotation and absorption by dust similar to what was observed in our Galaxy. Finally, a distance estimation of M31 was given with $d = 450\,\mathrm{kpc}$ (Öpik, 1922), about a factor of 2 lower than its actual value, but placing the Andromeda nebula clearly as an ex-

---

[1] Arago wrote: "Kant condensait ses idés dans le moindre nombre de mots possible, quand il appelait la Voie lactée le Monde des Mondes."

Active Galactic Nuclei, First Edition. Volker Beckmann and Chris Shrader.

tragalactic object. Using the 100″ Mt. Wilson telescope, Edwin Hubble was able to observe Cepheids in M31 and M33. Cepheids are variable stars with characteristic light curves which allow to determine their absolute brightness. Using the distance modulus of several Cepheids in these nearby galaxies, Hubble confirmed the large distance of these objects, although again underestimating their distance by a factor of ~ 3 to be about 285 kpc. Based on his observations, he also established a system of classifying galaxies, the so-called Hubble sequence (Hubble, 1926), and laid the starting point for cosmology assuming an expanding Universe (Hubble, 1929).

The first evidence that some galaxies were hosting some additional strongly emitting component in their center was found by Carl Seyfert in the 1940s. He obtained spectra of six galaxies, showing high-excitation nuclear emission lines superposed on a normal star-like spectrum (Seyfert, 1943). He also noticed that some galaxies showed broad emission lines, while others exhibited only narrow ones. The nature of the strong emission from the center of some galaxies remained a mystery. A common hypothesis was the assumption that a large number of stars would produce the observed features. Woltjer (1959) pointed out though that the observed concentration of the emission within the central 100 pc of the galaxies would require a mass of a few $10^8\ M_\odot$. A step closer to current understanding was the idea that in the center of these galaxies resides a stellar type object of very large mass, which then would emit mainly by accretion processes of a surrounding disk of gas (Hoyle and Fowler, 1963). It was not until a year later that the idea was put forward to assume that in the center of an AGN there could lie a black hole as opposed to a hypermassive star (Salpeter, 1964; Zel'Dovich and Novikov, 1964).

The hypothesis that there might exist objects in the Universe whose gravity would be sufficient to trap even light was discussed first by John Mitchell[2] in the late eighteenth century (Mitchell, 1784). Independently Pierre-Simon Laplace developed the concept of "dark stars," speculating that the most massive stars would be invisible due to their strong gravity (Laplace, 1796). The concept of the black hole was ignored though in later years, as light was considered to be made of massless particles with no interaction with a gravitational field. When Albert Einstein (1916) formulated the general relativity theory the possible existence of black holes was shown to be a solution for the gravitational field of a point mass and of a spherical mass by Karl Schwarzschild (1916). Nevertheless, this solution to Einstein's theory was thought to be merely hypothetical. Only when solutions had to be found to explain phenomena like AGN, and the fact that massive stars had to collapse into a black hole (Oppenheimer and Volkoff, 1939), was the existence of black holes accepted by a continuously growing fraction of the scientific community.

The idea of a supermassive black hole in the center of active galactic nuclei (Salpeter, 1964; Zel'Dovich and Novikov, 1964; Lynden-Bell, 1969) and also in the center of our own galaxy (Lynden-Bell and Rees, 1971) was a powerful model. It

---

2) Mitchell wrote: "If the semidiameter of a sphere of the same density as the Sun were to exceed that of the Sun in the proportion of 500 to 1, a body falling from an infinite height towards it would have acquired at its surface greater velocity than that of light, and consequently supposing light to be attracted by the same force in proportion to its *vis inertiae*, with other bodies, all light emitted from such a body would be made to return towards it by its own proper gravity."

explained not only the large energy output based on the release of gravitational energy through accretion phenomena, but also the small size of the emitting regions and connected to it the short variability time scales of AGN. The field was now open to study the physics involved in the accretion phenomenon, to observe and explain AGN emission throughout the electromagnetic spectrum, and to study the distribution in space, the origin, the evolution and fate of these elusive objects.

1.2
Broad Lines, Narrow Lines, and the Big Blue Bump

The first notably distinct observational characteristic of AGN was the presence of emission lines with widths upwards of $1000\,\mathrm{km\,s^{-1}}$ and far in excess of any known class of objects. Furthermore, the centers of these broad emission lines did not correspond to the laboratory wavelengths of any known atomic species and certainly not to the well known hydrogen Balmer series or other common lines known to be of astrophysical origin. This dilemma was resolved in the 1960s leading to the basic AGN paradigm described in the previous section of a distant and highly luminous object powered by a massive, accreting black hole. The deep gravitational potential of the black hole was responsible for the dynamical broadening of the observed lines and for radiatively efficient accretion leading to the extreme luminosities. The line identification dilemma was solved with the realization that the distances involved were of such magnitude that the cosmological expansion of the Universe redshifted atomic emission lines to the observed values including some high-ionization UV lines were.

Fast forwarding ahead several decades, it became evident that these broad emission line spectra could be exploited as a diagnostic of the physical conditions in the environment ambient to the central black hole. As we discuss in some detail later in the text, correlated variability of line and continuum emission components have been applied to “reverberation mapping” analyses leading to constraints on the broad-line emission region size and on the mass of the central black hole, for example Peterson and Horne (2004); Bentz *et al.* (2009a). This led to a dramatic revision of our basic understanding of AGN. Additionally, this knowledge has been used to cross-calibrate alternative black hole mass estimation methods and to better constrain physical models of the broad-line emission media as virialized gas clouds in photoionization equilibrium with the central engine radiation field.

Another distinguishing observational feature of some AGN is the presence of narrow, nonvariable forbidden emission lines. The similarity to nebular line emission in our galaxy was noted and some of the atomic physics and computational formalism developed and to study those objects was employed (Osterbrock, 1989). There were some significant differences between the galactic nebulae and the AGN as well. In particular, the AGN narrow lines required a broad-band ionizing continuum extending far bluewards of the stellar radiation fields responsible for photoionizing the galactic nebulae.

At present, with an improved understanding of the narrow-line region morphologies and thermodynamics, insight into the gas and dust distributions of the central regions of AGN can be gained. The dynamics of the inner AGN region can also be probed in cases using the narrow lines. For example, approximately 1% of low-redshift ($z \simeq 0.3$) optically selected type 2 AGNs exhibit a double-peaked [O III] narrow-line profile in spatially resolved spectra Shen *et al.* (2011). These types features have been interpreted in the context of kinematics, such as biconical outflows (Kraemer *et al.*, 2008) or rotation of the narrow-line region about the central black hole, or to the relative motion of two distinct NLRs in an ongoing AGN merger event.

Another observational aspect of AGN that was evident in early observations was that the continuum spectral distribution was very distinct from an integrated stellar continuum characteristic of normal galaxies (Oke and Sargent, 1968). Observationally, AGN were comparatively very blue. In fact, radio-quiet AGN would later be identified and cataloged primarily by performing multicolor imaging of sky regions, sorting the results in color-color plots, and performing spectroscopic follow-ups on the blue excess subpopulation thus identified (Green *et al.*, 1986, it is now known that this approach omits redder objects that are picked up in X-ray surveys).

The blue colors were due to both the fact that the continuum emission extended into the UV and beyond and that structure was often seen in the blue continua – the so-called "big blue bump" (Richstone and Schmidt, 1980). The big blue bump spectral component was a positive flux excess relative to an underlying power-law continuum. It exhibited curvature that suggested a thermal origin. This was interpreted as the first observational evidence for the presence of an accretion disk (Malkan and Sargent, 1982a), thus lending support to the basic paradigm of Lynden-Bell (1969). The quest to further corroborate this basic paradigm and to gain a deeper fundamental understanding of putative AGN accretion disks was the driver behind many subsequent observational campaigns and theoretical efforts during the decades which followed.

1.3
Jets and Other Outflows

The sixteenth century French astronomer Charles Messier published a catalog comprising 103 spatially extended or *nebular* objects. Among the most prominent, roughly spherical, examples in his catalog was object number 87, thus its designation as M87. In 1918 the American astronomer Heber Curtis noted that the presence of a "curious straight ray" that protruded from the nebula and apparently traced back to its nucleus (Curtis, 1918). As noted, there was at that time still disagreement as to whether or not the nebulae were external to or contained within our galaxy, but as that issue was soon resolved and it became clear that M87 was a giant elliptical galaxy, we consider Heber's observation to be the first documented example of an AGN *jet*.

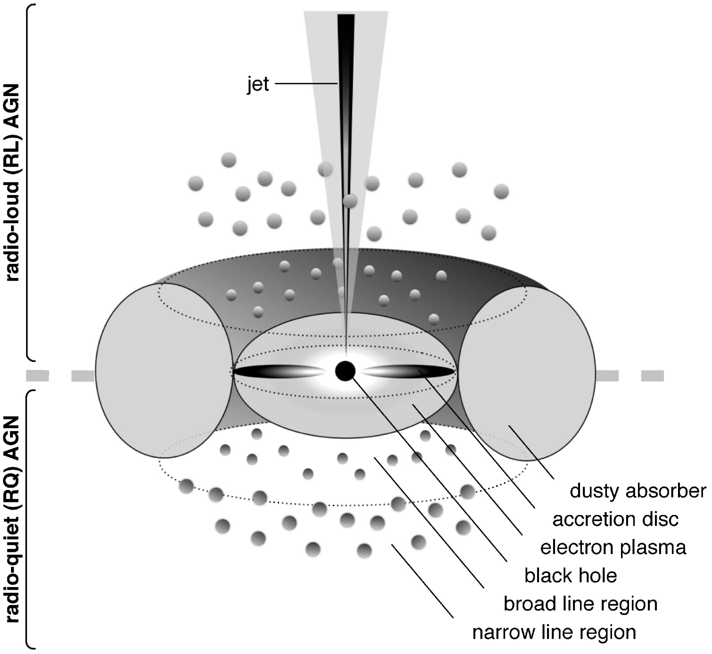

Figure 1.1 Schematic representation of our understanding of the AGN phenomenon and its main components. Note that this is a simplified view and not to scale. Graphic courtesy of Marie-Luise Menzel.

There were relatively few of examples of these highly collimated, bipolar outflows or jets from the centers of AGN until the observational techniques of radio interferometry matured in the decades subsequent to the Second World War. There are now AGN subclasses, which are believed to be related to each other within the context of the unification scenarios described in subsequent chapters. The physics underlying the launching and propagation of these AGN jets – their remarkable energetics, enormous size and plethora of associated phenomenology (hot spots, knots, bends) – remains enigmatic in many regards and its pursuit may be considered a “holy grail” of modern astrophysics.

Our general understanding of the main components of an AGN is shown in Figure 1.1. The schematic representation distinguishes between sources which display a jet and are therefore bright in the radio band, and those which do not show strong radio dominance and in which case one assumes no or weak jet emission.

1.4
X-ray Observations: Probing the Innermost Regions

During the same decade of the 1960s when the basic AGN paradigm was being developed, the first cosmic X-ray source, known as Scorpios X-1, was discovered using a rocket-borne detector (Giacconi *et al.*, 1962). It quickly became evident that X-ray emission was characteristic of compact, accretion-powered sources associated with galactic binaries. With the realization that the deep gravitational wells of massive black holes were likely the source for the extreme energetics exhibited by quasars, the generation of X-rays seemed natural. However, the rocket-borne experiments had limited capabilities and the first significant breakthrough came with the launch of the Uhuru satellite (also known as SAS-A) in 1970. A catalog of sources detected with Uhuru ultimately included nearly 340 objects, mostly galactic binaries, but also including about a dozen AGN (Forman *et al.*, 1978).

The field progressed rapidly during the 1970s, with detectors having larger collecting area, increased spectral coverage and improved spectral resolution, for example the Ariel-5 satellite launched in 1975 and the OSO-7 (1974), OSO-8 (1975) and HEAO-1 (1977) (see, e.g., Tucker and Giacconi, 1985). This led to the characterization of AGN as a class of X-ray sources and to the first detection of the iron K_α line emission from an extragalactic source. The biggest breakthrough however came later in that decade with the launch of the Einstein Observatory (originally called HEAO-2). This was the first true orbiting X-ray *telescope*, in that it utilized a concentric array of grazing incidence mirrors to focus $\sim$ keV photons onto its focal plane detectors. The resulting images provided, in addition to vastly improved spatial localization of sources on the sky, a large leap in sensitivity since the source and celestial background could be effectively separated.

It had become clear that X-ray emission was a common property of different subclasses of active galaxies with the X-ray flux comprising a significant fraction (about 5–40%) of the bolometric emission from such objects (Ward *et al.*, 1987). Rapid variability was also found to be a prominent feature of the X-ray emission with kilosecond timescale X-ray flux variations seen in local Seyfert galaxies. This imposed new and increasingly stringent constraints on the size of the X-ray emission region and strongly supported the idea that it occurs very close to the active nucleus (e.g., Pounds *et al.*, 1986). The origin of X-rays from close to the central black hole means that X-ray data offer a chance to study the immediate environs of supermassive black holes and the poorly understood accretion process that fuels them. Although the angular scale of the X-ray emission region is too small to image with current instrumentation, timing analysis and spectroscopy offered methods to probe these regions indirectly.

Specific spectral signatures were attributed to characteristics of the gas inflow and outflow near the central most regions in AGN. The X-ray observations also provided signatures of reprocessing of radiation in material within approximative distance of hundreds of gravitational radii and thus the potential for discerning signatures of the accretion disk at even smaller radii. Features such as the weak, broad emission lines due to low-ionization states of iron as well as other structured deviations from simple power laws had been identified in the spectra of AGN. In 1991, George and Fabian (1991) offered an interpretation of these features in terms of X-irradiation of relatively cold, dense gas in the vicinity of the central black hole. The emergent spectrum then consists of direct radiation from the central source plus a scattered or "reflected" spectrum that includes imprinted photoabsorption, fluorescent emission and Compton scattering from matter within the surrounding accretion flow. This basic idea has withstood the scrutiny of improved observational data and has become a tenant of the AGN paradigm.

X-ray observations of AGN are also being applied to address issues of fundamental black hole physics. The shapes of line profiles have also been applied to models which in principle allow one to infer an intrinsic property of the central black hole, namely its intrinsic angular momentum or *spin* (e.g., Brenneman and Reynolds, 2009). The basic idea is that the asymmetry of a line profile produced in the inner AGN accretion disk depends in a predictable manner on the shape of the gravitational potential which in turn depends on the black hole *spin*.

In the chapters that follow, we discuss these issues in further detail highlighting a number of results from the current astrophysics literature. We also speculate on the possibilities offered by future orbiting X-ray observatories, which are currently under discussion.

1.5
Up, Up and Away: from Gamma-Rays toward the TeV Range

The impact of gamma-ray astronomy on AGN research did not emerge as rapidly as did X-ray astronomy, although the fields were initiated more or less concurrently with 1960s rocket flights followed by satellite-borne experiments in the 1970s. The reasons for this are several-fold. There are fewer gamma-ray photons than lower energy photons emitted even though the overall energy budget for some AGN may be dominated by the gamma rays. There are substantial instrumental and celestial backgrounds at gamma-ray energies that need to be understood and modeled or subtracted. Gamma-ray detectors tend to be more massive for a given effective collection area than X-ray detectors and gamma rays cannot be focused. Additionally, it became apparent that only the radio-loud AGN, which comprise $\sim$ 5% of the overall population are prolific emitters of gamma radiation. We should note here that the term “gamma rays” encompasses a huge swath of the electromagnetic spectrum. Here we will designate photons with energies above $\sim$ 100 keV as gamma rays. AGN have been detected at $\sim$ TeV, thus we are considering over 7 decades in our discussion of gamma-ray studies. The energy range above about 100 MeV has, somewhat surprisingly, provided the richest bounty of results as we will further discuss in later chapters.

In the 1970s the ESA mission COS-B, along with NASA’s SAS-2, provided the first detailed views of the Universe in gamma rays. COS-B, launched in August 1975, was originally projected to last two years, but it operated successfully for nearly seven. It made the first gamma-ray measurement of an AGN, that being 3C 273 (Swanenburg *et al.*, 1978). However, it was not until nearly 20 yr later with the launch of the Compton Gamma-Ray Observatory (CGRO) that additional gamma-ray detections were made, starting with the discovery in 1991 of bright gamma-ray emission from 3C 279 (Hartman *et al.*, 1992). New results came quickly after that leading ultimately to the identification of some 70 high-latitude CGRO gamma-ray sources with radio-loud AGN. Specifically, BL Lac objects and flat-spectrum radio quasars (FSRQs), known collectively as blazars, comprised the entire gamma-ray sample. It was also clear that the radiative output of the blazars was typically dominated by the gamma rays. The gamma-ray emission was also found to be variable on time scales less than a day.

These observations had several immediate implications for physical models. The emission had to emanate from a compact region. For example, a factor of 2 flux variation limits, approximately, the size r of a stationary, isotropic emitter to $r \lesssim c\delta t_{\text{var}}/(1+z)$ where δt_{var} is the variation time scale. The implications from the early CGRO results, which by this line of reasoning necessitated a very compact emission region, were problematic in any scenario in which the gamma-ray production involves such a stationary isotropic source. The problem involved the transparency of a compact region such as inferred here. If X-rays are produced cospatially with the gamma rays, attenuation of the gamma rays due to the process $\gamma\gamma \rightarrow e^+e^-$ for which the cross-section for attenuation of $\sim$ 100 MeV gamma rays is in the X-ray range $\sim$ keV X-ray range. The inferred gamma-ray opacity from the CGRO observations would exceed unity in many instances. Either the radiating particles were strongly beamed or the emitting plasma was undergoing bulk relativistic motion. Thus beaming was very strongly implied.

Models that had been previously favored to explain the radio-to-optical continua in these objects, for example Blandford and Konigl (1979), implied that we are viewing nearly along an axis of a relativistic plasma jet ejected from near the central black hole, involving nonthermal synchrotron emission as we will discuss in later chapters. An extension of this scenario invoking a distinct second spectral component was now clearly required. The basic idea was that gamma rays emitted by blazars are produced by the same population of electrons that produced the synchrotron emission via Compton scattering of ambient low-energy photons. The ambient photon field could be the synchrotron photons themselves (e.g., Maraschi *et al.*, 1992) or from an external source such as the accretion disk or broad-line clouds (e.g., Dermer *et al.*, 1992).

Shortly after the CGRO results began to emerge, another major discovery followed from ground-based Cherenkov gamma-ray telescopes, which measured gamma rays in the $\sim$ TeV range. Blazars such as Markarian 421 (Punch *et al.*, 1992) and Markarian 501 (Quinn *et al.*, 1996) were detected during a high-amplitude variability episode. These discoveries established this subclass of AGN as emitters over $\sim$ 20 decades of the electromagnetic spectrum. As such, they were a striking example of the value, indeed the necessity, inherent in the multiwavelength approach to studying AGN. The high-energy gamma-ray observations also fit in naturally with the synchrotron plus Comptonization model scenarios. They also had other potentially significant implications, not only on the blazar AGN themselves, but on the gamma-ray transparency of the universe and thus in turn the background radiation fields to the cosmic star-formation history.

In the two decades since these discoveries, gamma-ray studies of AGN have expanded enormously. The Fermi Gamma-Ray Space Telescope, launched in 2008, has cataloged approximately 900 gamma-ray AGN. Advances in ground-based

Cherenkov telescope facilities, as well as in detection and analysis methodologies, has produced a similar order-of-magnitude increase in the TeV gamma-ray sample. Multiwavelength campaigns have begun to reveal how jet formation and propagation may be correlated with the gamma-ray flux variations. Clearly, gamma-ray astronomy will continue to be a vital component of our quest to better understand the AGN phenomenology for the foreseeable future.