VARIABILITY OF ACTIVE GALACTIC NUCLEI

Marie-Helene Ulrich
European Southern Observatory, Karl-Schwarzschild Strasse 2, D-85748 Garching bei München, Germany

Laura Maraschi
Brera Astronomical Observatory, via Brera 28, 20121 Milan, Italy

C. Megan Urry
Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA

KEY WORDS: Seyfert galaxies, quasars, blazars, black holes, emission lines, accretion disks, jets

ABSTRACT

A large collective effort to study the variability of active galactic nuclei (AGN) over the past decade has led to a number of fundamental results on radio-quiet AGN and blazars. In radio-quiet AGN, the ultraviolet (UV) bump in low-luminosity objects is thermal emission from a dense medium, very probably an accretion disk, irradiated by the variable X-ray source. The validity of this model for high-luminosity radio-quiet AGN is unclear because the relevant UV and X-ray observations are lacking. The broad-line gas kinematics appears to be dominated by virialized motions in the gravity field of a black hole, whose mass can be derived from the observed motions. The “accretion disk plus wind” model explains most of the variability (and other) data and appears to be the most appropriate model at present. Future investigations are outlined.

In blazars, rapid variability at the highest energies (gamma-rays) implies that the whole continuum is relativistically boosted along the line of sight. The general correlation found between variations in TeV gamma rays and in X rays for Mrk 421, and between variations in GeV gamma rays and in the IR–optical–UV bands for 3C 279, two prototype objects, supports models in which the same population of relativistic electrons radiates the low-frequency continuum via synchrotron and the high frequency continuum via inverse Compton scattering of soft photons.
Identifying the dominant source of soft photons, which is at present unclear, will strongly constrain the jet physics.

1. INTRODUCTION

Active Galactic Nuclei (AGN) produce enormous luminosities in extremely compact volumes. Large luminosity variations on time scales from years to hours are common. The combination of high luminosity and short variability time scale implies that the power of AGN is produced by phenomena more efficient in terms of energy release per unit mass than ordinary stellar processes (Fabian 1979). This basic argument leads to the hypothesis that massive black holes are present in the cores of AGN. Accretion of matter onto a black hole or extraction of its rotational energy can in fact yield high radiative efficiencies (Rees et al 1982, Rees 1984).

The basic AGN paradigm developed thus far consists of a central supermassive black hole, surrounded by an accretion disk, or more generally optically thick plasma, glowing brightly at ultraviolet (UV) and perhaps soft X-ray wavelengths. In the innermost region, hot optically thin plasma surrounding and/or mixed with the optically thick plasma gives rise to the medium and hard X-ray emission. Clouds of line-emitting gas move at high velocity around this complex core and are in turn surrounded by an obscuring torus or warped disk of gas and dust, with a sea of electrons permeating the volume within and above the torus.

In some systems, highly relativistic outflows of energetic particles along the poles of the rotating black hole, accretion disk, or torus form collimated radio-emitting jets that lead to extended radio sources. These AGN are called radio loud because their radio emission is comparatively strong; AGN without collimated jets, which therefore have weaker (but detectable) radio emission, are called radio quiet.

Variability studies have been essential in understanding the physics of the central regions of AGN, which in general cannot be resolved even with existing or planned optical/infrared (IR) interferometers. The time scales, the spectral changes, and the correlations and delays between variations in different continuum or line components provide crucial information on the nature and location of these components and on their interdependencies.

In recent years progress has been made on two fronts. First, for a handful of objects, large international collaborations have led to improved sampling, duration, and wavelength coverage in AGN monitoring campaigns. Second, the availability of uniform data sets like the International Ultraviolet Explorer (IUE) archive and various X-ray archives has made statistical comparisons possible.
among different classes of AGN and different wavelength bands. This article describes these recent advances, with particular emphasis on multiwavelength variability studies. Several excellent reviews have covered or touched on the subject in previous years in this series: intraday variability (Wagner & Witzel 1995), X-ray spectra and time variability of AGN (Mushotzky et al. 1993), unified models for AGN (Antonucci 1993), and the earlier presentation of the black hole models for AGN (Rees 1984). We also note reviews elsewhere on AGN continuum and variability (Bregman 1990, 1994), on the properties of the gas in the inner regions of the AGN (Collin-Souffrin & Lasota 1988), on reverberation mapping of the emission line regions (Peterson 1993), and an overview of the AGN field (Blandford et al. 1990).

The AGN that are the subject of this review are those in which the central optical, UV, and X-ray source and the broad emission line region (if present) are viewed directly. The word AGN is used here regardless of redshift and luminosity and therefore encompasses the words Seyfert 1 and QSO or Quasar, which are often used to designate low- and high-luminosity AGN separately. In radio-loud AGN seen at small angles to the axis of the jet, the highly nonthermal radiation produced in the jet is strongly amplified by relativistic beaming and dominates the observed continuum. In these sources, called blazars, variability is the most violent and affects the whole electromagnetic range from the radio to the gamma-ray band.

The fundamentally different character of the radiation emitted by radio-quiet AGN and by blazars dictates different observational goals and techniques. For radio-quiet AGN the focus is on (a) the emission mechanisms of the optical–UV–X-ray continuum and (b) the kinematics of the gas, with the ultimate aim of investigating the mass accretion and mass loss, and of deriving the mass of the central black hole if, indeed, it can be shown that the kinematics is dominated by virialized motions. For blazars, the goal is to understand the structure and physical state of the plasma in the jet, i.e. the geometry, acceleration, and radiation processes.

Accordingly, this review is organized into two main parts, radio-quiet AGN (Sections 2–5) and blazars (Sections 6–8), with general conclusions in Section 9.

2. VARIABILITY OF THE CONTINUUM IN LOW-REDSHIFT RADIO-QUIET AGN

The continuum of radio-quiet AGN varies on all observable time scales, with amplitude up to a factor of 50 or so. The variability character depends on the wavelength, and there are correlations among the variations in different energy bands, as well as astrophysically important upper limits to the time delays between these variations, as detailed below.
The electromagnetic spectrum of radio-quiet AGN, after subtraction of the stellar continuum, extends from $\sim 1$ mm to $\sim 100$ keV with a prominent broad peak in the UV–extreme UV (EUV) range (for $\nu f_\nu$ versus $\nu$; Sanders et al 1989). The spectrum from 1200–5000 Å strongly suggests that the broad peak is primarily thermal emission from a very dense medium, probably an accretion disk (Lynden-Bell 1969, Shields 1978, Malkan & Sargent 1982; but see Ferland et al 1990). The temperature of the disk, however, is not set primarily by viscous effects, as initially thought, but by irradiation from the central X-ray source, as demonstrated by the recent variability studies (see below). The spectrum is not known in the range 1000 Å to $\sim 0.1$ keV.

2.1 Continuum Variability in the Optical, UV, EUV, and IR Ranges

THE DATA BASE: RICH AND INHOMOGENEOUS Most results come from three AGN—NGC 4151, NGC 5548, and 3C 273—that have been observed extensively in numerous coordinated campaigns in the IR, optical, UV, and X-ray ranges. NGC 4151 is an exceptional AGN in terms of the richness of its phenomenology and the quantity and variety of the data available. Of the brightest AGN, it varies on the shortest time scale. NGC 5548, another nearby bright AGN, has been observed in several multiwavelength campaigns, including 7 years of nearly daily spectrophotometric monitoring in the optical (Peterson et al 1994).

The quasar 3C 273 is the only extensively studied very high-luminosity AGN, with $L_{bol} \sim 2.5 \times 10^{47}$ erg s$^{-1}$. It displays some blazar characteristics (jet and VLBI source) but also has a strong blue bump and emission lines that indicate that the optical–UV is dominated by the non-blazar component. Two dozen AGN have been observed in smaller campaigns, some with multiwavelength coverage. The results are in agreement with and extend those obtained from the three intensively observed objects.

IR–EUV TIME SCALES, AMPLITUDES, AND SPECTRAL SHAPE On time scales of many decades, data are available only for NGC 4151. Multifractal analysis of B-band photometry from 1911–1991 clearly suggests a nonlinear intermittent behavior that, if confirmed, would rule out processes based on shot noise or on the superposition of a very large number of independent events (Longo et al 1996). The data were also searched for evidence of periodicity—none was found, confirming earlier results (Lyutyi & Oknyanski 1987, and references therein).

Figures 1 and 2 show examples of variability on time scales of years to a few hours. Remarkable “low states” or “minimum states” characterized by an exceptionally weak continuum flux and the quasi-absence of the broad components of the emission lines have been observed in some AGN (e.g. NGC 1566: Alloin et al 1986; Fairall 9: Clavel et al 1989, Recondo-González et al 1997;
Figure 1  Long- and short-term continuum variations in three low-luminosity AGN: (a) UV light curve of F9 over 14 years (Recondo-González et al 1997). (b) UV light curve of NGC 4151 over 17 years. The passage through the deep minimum was interrupted by short excursions to medium bright level. The vertical groups of points, unresolved on this scale, are IUE campaigns with an adequate sampling interval of typically three days. (c) Optical and UV light curves of NGC 5548 over 8 months (December 1988–July 1989; Clavel et al 1991, Peterson et al 1994). Ordinates in $10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ Å}^{-1}$.

NGC 4151) and, given time, could perhaps occur in all AGN. In NGC 4151 (Figure 1b) the prolonged minimum lasted from 1981–1987 with short spells at medium bright states (Perola et al 1986, Bochkarev et al 1991, Gill et al 1984, Ulrich et al 1985, 1991).

These extreme minima provide an opportunity to observe the central non-varying continuum (such as an underlying starburst or an extended scattered
Figure 2  Continuum light curves during the 10-day intensive period of multiwavelength observations of NGC 4151 in December 1993. The light curves are shifted vertically for convenience so a 10% change is indicated by the bar. The largest amplitude variations are in soft X rays. In the optical–UV, the amplitude decreases systematically with increasing wavelength. [From Edelson et al (1996).] Ordinates in erg s\(^{-1}\) cm\(^{-2}\) \AA\(^{-1}\).

An example of optical–UV variability on time scales of months and weeks is shown in the light curves of NGC 5548 from December 1988 to July 1989 (Figure 1c). The power spectrum of the 1400-Å continuum variations is “red,” with an exponent between \(-2\) and \(-3\) (Krolik et al 1991). A general feature of all radio-quiet AGN continuum variability is that the amplitude is inversely correlated with the time scale (see Table 1 for examples). In addition, the maxima and minima of the light curves appear to have statistically symmetrical shapes (see Figure 1c), but this property should be investigated more thoroughly.
Table 1  Examples of continuum and emission line variations

<table>
<thead>
<tr>
<th>Object</th>
<th>Variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fairall 9</td>
<td>Continuum variations at 1360 Å, July 1978–Oct 1984:</td>
</tr>
<tr>
<td></td>
<td>( r_{max} = 24 ) in ( \Delta t \sim 5.5 ) years (^a)</td>
</tr>
<tr>
<td></td>
<td>Line variations during the 1978–1987 IUE Campaign:</td>
</tr>
<tr>
<td></td>
<td>Ly(\alpha), C(\text{iv}), and Mg(\text{ii}) varied by factors 10, 7, and 3, respectively, with approximately the same (t_d) of (\sim 160) days (Clavel et al 1989, but see Recondo-Gonzalez et al 1997) (^c)</td>
</tr>
<tr>
<td>3C 273</td>
<td>Continuum variations at 1400 Å:</td>
</tr>
<tr>
<td></td>
<td>( r_{max} = 2 ) in ( \Delta t \sim 2 ) years</td>
</tr>
<tr>
<td></td>
<td>Variations of Ly(\alpha) &lt; 10% in 10 years (Ulrich et al 1993)</td>
</tr>
<tr>
<td>NGC 4151</td>
<td>Short- and long-term continuum variations at 1400 Å:</td>
</tr>
<tr>
<td></td>
<td>Line variations during the Nov.–Dec. 1991 IUE Campaign:</td>
</tr>
<tr>
<td></td>
<td>(t_d) of C(\text{iv}) blue wing, red wing, and whole line are 2.6, 1.9, and 2.4 days, respectively (Ulrich &amp; Horne 1996)</td>
</tr>
<tr>
<td></td>
<td>During the Dec. 1987–July 1988 optical Campaign:</td>
</tr>
<tr>
<td></td>
<td>(t_d) of H(\beta) was 9 days for variations by a factor 1.7 (Maoz et al 1991)</td>
</tr>
<tr>
<td>NGC 5548</td>
<td>Continuum variations at 1400 Å, Dec. 1988–Aug. 1989:</td>
</tr>
<tr>
<td></td>
<td>Factor 2.45 in 50 days</td>
</tr>
<tr>
<td></td>
<td>Line variations during the same IUE Campaign:</td>
</tr>
<tr>
<td></td>
<td>Ly(\alpha), C(\text{iv}), and He(\text{ii}) lambda 1640 varied by a factor 1.8, 1.8, and 4 respectively with (t_d) of about 12, 8, and 4 days (Krolik et al 1991). During the same period H(\beta) varied by a factor 1.7 with (t_d) of 19 days (Peterson et al 1994).</td>
</tr>
</tbody>
</table>

Other time responses of H\(\beta\):

<table>
<thead>
<tr>
<th>Object</th>
<th>(t_d) values</th>
</tr>
</thead>
<tbody>
<tr>
<td>NGC 3227</td>
<td>(t_d \sim 15) days (Salamanca et al 1994, Winge et al 1995)</td>
</tr>
<tr>
<td>NGC 3516</td>
<td>(t_d \sim 7) days (Wanders et al 1993)</td>
</tr>
<tr>
<td>NGC 3783</td>
<td>(t_d \sim 8) days while (t_d) of C(\text{iv}) was 5 days in the same period</td>
</tr>
<tr>
<td></td>
<td>(Stripe et al 1994, Reichert et al 1994)</td>
</tr>
<tr>
<td>PG 0804+762</td>
<td>(t_d \sim 93) days (Kaspi et al 1996b)</td>
</tr>
<tr>
<td>PG 0953+414</td>
<td>(t_d \sim 110) days (Kaspi et al 1996b)</td>
</tr>
</tbody>
</table>

\(^a\)\(\Delta t\): interval of time between successive maximum and minimum.  
\(^b\)\(t_d\): time delay between the UV lines and UV continuum or between optical lines and optical continuum variations.  
\(^c\)Measurement errors and other uncertainties are detailed in references.
On time scales of days and hours, by far the best variability data were collected during the 10-day intensive optical, UV, and X-ray monitoring of NGC 4151 in December 1993 (Figure 2; Crenshaw et al 1996, Kaspi et al 1996a, Warwick et al 1996 Edelson et al 1996). The fastest variations disappear and the overall variability amplitude decreases toward longer wavelengths, as is seen in all AGN (e.g. Kinney et al 1991; Figures 1c and 2). The variability observed with the Extreme Ultraviolet Explorer (EUVE) satellite at 65–120 Å is consistent with this trend (Marshall et al 1997).

The shape of the optical–UV continuum always hardens when the nucleus brightens (Kinney et al 1991, Paltani & Courvoisier 1994), in contrast with BL Lac objects where the spectral shape changes little when the flux varies (Section 6.3). The spectral change occurs because two components with different spectra and variability time scales make up the continuum: the “small blue bump” (2300–4000 Å), which varies with the smallest amplitude and is a sum of permitted Fe II UV and optical multiplets plus the Balmer continuum (Wills et al 1985), and the more variable primary continuum component, whose spectral shape is difficult to determine because of various contaminations. In NGC 5548, this continuum appears to harden when the nucleus brightens (Wamsteker et al 1990, Maoz et al 1993), whereas in Fairall 9 (F9), the optical spectral slope was unchanged while the intensity increased by a factor of 20 (Lub & de Ruiter 1992).

Infrared variations, when detected at all, are of smaller amplitude and longer time scale than in the optical (Neugebauer et al 1989, Hunt et al 1994) and are consistent with being the delayed response of dust around the central source to large amplitude long-term variations of the UV flux, at least for nearby AGN (e.g. Clavel et al 1989, Nelson 1996). Neugebauer and Matthews (G Neugebauer & K Matthews, private communication) have monitored the continua of 25 quasars (5 radio loud) between 1 and 10 µ over a period of up to 25 years. If the 10-µ continuum is caused by thermal emission from dust grains that symmetrically surround the quasar and are in equilibrium with its radiation, the size of the emitting region is such that significant variations would have time scales longer than 100 years. Variations with time scales on the order of years are seen (not surprisingly) in all the radio-loud quasars. In at least one radio-quiet quasar, significant variability on the order of 5–10 years is seen, with correlated variations in all bands. These variations suggest either that the mid-IR emission in some radio-quiet quasars is nonthermal or that quite complicated structures are involved.

2.2 Variability of the X-Ray Emission of Seyfert Galaxies

Here we mention only the key points and most recent results because an excellent review is given by Mushotzky et al (1993). The X-ray emission of Seyfert
galaxies consists of several components, including a power law in the medium
energy X-ray range (1–10 keV; $\alpha \simeq 0.9$, where $f_\nu \propto \nu^{-\alpha}$), a soft excess usually
below 1 keV, and a reflection hump in the 10- to 30-keV range. Superimposed
on this continuum is a prominent Fe line and often (50%) absorption edges from
highly ionized oxygen (Fabian 1996). In the hard X-ray range (>50 keV), the
few OSSE data available indicate that the power law steepens, possibly with a
cutoff around 100 keV (Johnson et al 1994).

Both the soft excess and the medium X-ray power-law component are vari-
able, albeit differently; the soft excess is more strongly variable and is often
but not always correlated with the medium energy X-rays. There are very few
cases where the soft and medium X-ray variations appear uncorrelated. The best
example is NGC 5548, observed 25 times at approximately daily intervals in
December 1992 to January 1993. The soft excess component varied (factor 10)
independently of the hard X-ray flux (factor 3), most noticeably in a soft X-ray
flare lasting 8 days which had no medium X-ray counterpart (Done et al 1995).

The medium energy power-law component in general varies in intensity with
very little change of the spectral index. In some well-observed AGN, the flux
variations are accompanied by a softening of the spectrum with increasing
intensity (e.g. Perola et al 1986, Mushotzky et al 1993 and references therein,

The most extreme soft X-ray variability occurs in Narrow-Line Seyfert 1
galaxies (NLS1), a subset of AGN with very steep soft X-ray spectra ($1 \lesssim \alpha \lesssim 4$
in the range 0.1–2.4 keV), narrow optical emission lines with full width at
half maximum (FWHM) $\lesssim 2000$ km s$^{-1}$, and prominent optical Fe II emission
(Osterbrock & Pogge 1985, Boller & al 1996). To explain the absence of broad
lines in NLS1, it has been proposed that the intense soft X-rays could blow away
the inner broad-line region (BLR) or ionize it to states currently undetectable
(Pounds et al 1995, Guilbert & Rees 1988). The steep soft X-ray spectrum of
NLS1 may indicate a high accretion rate or a small black hole mass. In
broad-line AGN, the medium–hard X-ray power law generally has luminosity
significantly higher than the soft excess (Pounds & Brandt 1996), but in the
NLS1 RE 1034+39, for example, the soft excess exceeds the luminosity of the
medium–hard component, and moreover, the source has an exceptionally
steep medium X-ray spectrum—characteristics shared with Galactic black hole
candidates. This suggests that in NLS1, or at least in some of them, the accretion
rate is close to the Eddington limit and the soft X-rays represent viscous heating
of the accretion disk (Pounds et al 1995). Alternatively, a small black hole with
an accretion rate of $\sim 0.1$ Eddington accretion rate could also emit a very hot
spectrum with such an intense soft X-ray component.

Although the NLS1 display the most extreme soft X-ray variations, their
range of variability merges with that of “classical” broad-line AGN. Here we
summarize a few of the most spectacular examples of variability. The largest variation observed in one year was by a factor of 70 in RE J 1237+264; this object has remained weak and has the very same steep slope as in the high state (Brandt et al 1996). An optical spectrum taken a few months after the soft X-ray outburst shows emission lines of [FeX] and Hα that are approximately 10 times brighter than those observed before or well after the burst (Pounds & Brandt 1996, Brandt et al 1996).

Strong variations are also seen in the Fe II strong AGN PHL 1092; if the radiation is isotropic, the rapid variability requires that mass be transformed into energy with an efficiency of at least 0.13, exceeding the theoretical maximum for a nonrotating black hole (Forster & Halpern 1996). Somewhat in contrast, a drastic spectral change from ultrasoft to typical soft X-ray Seyfert spectrum in RX J0134-42 occurred without change in count rate (Mannheim et al 1996; see also Pounds et al 1995).

An exceptional case is the persistent giant and rapid soft X-ray flux variability of the radio-quiet, ultrasoft, strong Fe II, narrow-line Seyfert 1 galaxy IRAS 13224-3809 (Boller, Brandt, Fabian & Fink 1997). In the first systematic monitoring of an ultrasoft NLS1, a 30-day Roentgen Satellite (ROSAT) High Resolution Imager observation revealed at least five giant amplitude variations, with the maximum observed amplitude of about a factor of 60. A variation by about a factor of 57 was detected in just two days. Variations by a factor of about 30 were also seen to occur during a 1994 observation with the Advanced Satellite for Cosmology and Astrophysics (ASCA; Otani et al 1996). Relativistic boosting effects provide the most plausible explanation of the X-ray data and may be relevant to understanding the strong X-ray variability of some steep spectrum Seyferts more generally. The variability is probably nonlinear in character, which suggests that flares and spots in the accretion disk interact nonlinearly or are affected by nonlinear flux amplification.

X-ray variability by a factor of about 50 has also been observed in one broad-line AGN, E1615+061 (Piro et al 1997). The high-state spectrum was very steep (α ∼ 3), whereas the low-state slope was near normal (α ∼ 1). There is only an upper limit to the variation time scale (16 years) and the soft X-ray activity of this broad-line AGN may not be related to the extremely fast soft X-ray variability of the NLS1 class. Finally, another extraordinary extragalactic X-ray transient is the NLS1 WPVS007, which decreased by a factor of more than 400 between the ROSAT All Sky Survey and the ROSAT pointed observations three years later (Grupe 1996).

2.3 Simultaneity of the Flux Variations at Various Energies

Simultaneity of the UV and Optical Flux Variations: Consequences for the Nature of the Optical–UV Continuum

The long viscous time
scale for a standard optically thick, geometrically thin accretion disk (Shakura & Sunyaev 1973, Pringle 1981) is incompatible with models in which rapid flux variability is caused by variable fueling (Clarke 1987). That variable fueling does not cause rapid optical–UV–EUV variability is confirmed by the simultaneity of the variations in those wavelength bands; the current tightest limit on the time delay between the UV and optical flux variations is 0.2 days in NGC 4151 (Crenshaw et al 1996) and 0.25 days between the UV and EUV in NGC 5548 (Marshall et al 1997).

Consider the example of a disk around a black hole of mass $4 \times 10^7 M_\odot$.

Because information cannot travel through the disk faster than the sound speed ($3 \times 10^6 \text{ cm s}^{-1}$ for $T = 10^5 \text{ K}$), the delay between the rings in the disk emitting at 5400 Å and 1350 Å should be $>3$ years, in clear contradiction to the simultaneous flux variations in the UV and optical ranges (Krolik et al 1991, Collin-Souffrin 1991, Courvoisier & Clavel 1991).

SIMULTANEITY IN THE UV, SOFT-, AND HARD-X-RAY RANGES: REPROCESSING IS CURRENTLY THE BEST MODEL

Given that viscous heating cannot be the emission process of the rapidly variable optical–UV continuum, the best alternative, suggested by the X-ray variations, is reprocessing. Simultaneity and proportionality on time scales of weeks to months between the medium energy X-ray flux and optical–UV flux have been observed in several AGN (NGC 4151: Perola et al 1986, Warwick et al 1996; NGC 5548: Clavel et al 1992, Tagliaferri et al 1996). This is readily explained by a model in which the optical–UV flux is emitted by an optically thick dense medium irradiated by a nearby variable central X-ray source (Collin-Souffrin 1991, Krolik et al 1991, Nandra et al 1991, Molendi et al 1992, Haardt & Maraschi 1993, Rokaki et al 1993, Petrucci & Henri 1997); this dense medium could be the accretion disk.

In one of the most actively investigated models, fueling occurs through an accretion disk, with angular momentum removed magnetically by field lines threading the disk surface. Explosive reconnections dissipate significant power via magnetic flares within the disk corona (Galeev et al 1979, Blandford & Payne 1982), and X-ray emission is produced via inverse Compton emission in the hot corona surrounding the cooler accretion disk (Haardt & Maraschi 1993).

These flares and variations in the optical depth of the corona (Haardt et al 1997) could be the primary cause of variability on time scales of days or less and could cause variations of the X-ray emission even with a constant accretion rate. This Comptonization model explains the average power-law slope at medium energy, the high-energy cutoff, the reflection hump, and the iron line emission. Reprocessing readily explains the correlation among power-law X rays, soft X rays, and UV emission on short time scales. In fact, it seems at present to be the only viable explanation.
If the corona is not uniform but is patchy (Haardt et al 1994, 1997), the reprocessed radiation is only a fraction of the UV emission and the rest is presumably accretion energy dissipated within the optically thick disk. The fraction of the UV emission due to reprocessed radiation can change “secularly.” This is quite clear in NGC 4151, as can be seen in Figures 1b and 3. A major test of the reprocessing scenario is to verify the energy budget among the variable components, i.e. to check that the energy of the medium–hard X-ray component exceeds that of the the UV (and perhaps also the soft X-ray) component. This has to be checked with care and is at present very uncertain (cf Perola et al 1986, Ulrich 1994, Edelson et al 1996).

The origin of the soft X-ray emission in the reprocessing model is not definite. An attractive possibility is that the soft excess is due to reprocessing of the Comptonized emission into thermal radiation. This clearly has strong implications on the expected correlation of variability in different bands.

On time scales of years, the proportionality of the UV and medium X-ray fluxes breaks down, as is evident in NGC 4151 (Figure 3), NGC 5548, and F9, with the UV varying more than the X-ray flux (Morini et al 1986b, Perola et al 1986, Clavel et al 1992, Warwick et al 1996). In NGC 4151, a large slow-varying UV component has disappeared and reappeared during the lifetime of IUE, between 1978 and 1996 (Figures 1b and 3).
Variability of AGN

On time scales of days, the UV and X-ray variations are correlated, though not in a detailed way. In NGC 5548 and NGC 4151, the short time-scale variability amplitude is smaller in the UV than in the X-rays (Türl et al. 1996 for NGC 5548; see Figure 2 for NGC 4151). The discrepancy from proportionality could be caused by variations of the soft X-ray emission due to viscous effects.

High-luminosity AGN tend to have larger UV/X-ray flux ratios when compared to low-luminosity AGN, which could pose an energy budget problem for the reprocessing scenario. This could be solved if the X-ray source is anisotropic and the disk receives more X-ray flux than can be inferred from the observed flux. Anisotropy of the Comptonized emission could account for a deficiency of only a factor of 2–3, but not more. In any case, at present, for high-luminosity AGN, the data are insufficient to establish whether or not there is simultaneity between the optical–UV and X-ray variations, and thus the relevance of the irradiation model for high-luminosity AGN is still uncertain.

Variability and the Starburst Model

A contrasting model views AGN as giant young stellar clusters (Terlevich et al. 1992). In this case, variability results from the random superposition of “events”—supernova explosions generating rapidly evolving compact supernova remnants (cSNRs) due to the interaction of their ejecta with the high density circumstellar environment. This model is supported by the striking similarity between the optical spectra of AGN and of cSNRs (e.g. Filippenko 1989). The characteristics of an event (i.e. its light curve, amplitude, and time scale) result from the combination of complicated processes (Terlevich et al. 1992). Still, the light curves of AGN of various absolute luminosities and redshifts can be predicted from this model and are found to be consistent with the observed dependence of the structure function (the curve of growth of variability with time) on luminosity and redshift (Cid Fernandes et al. 1996, Cristiani et al. 1996).

The light curves of cSNRs are still poorly known. At present, they appear to be consistent with the optical light curves of the low-luminosity AGN NGC 4151 and NGC 5548 (Aretxaga & Terlevich 1994). Much more detailed light curves of cSNRs are required to make a definitive check. The production of strong and rapidly variable X-rays is still difficult in this model.

3. Variability of the Optical Continuum in High Redshift AGN

3.1 Observed Characteristics of the Optical Variability

The high surface density of high-z AGN, \( \sim 100 \) per square degree at \( m_B \) brighter than 22, is such that a significant number of radio-quiet AGN can be recorded on a single image taken with a Schmidt telescope or a large telescope with a wide field. The technique of choice to study the optical variability of high-z
radio-quiet quasars is thus the definition of large optically selected samples and their repeated broadband imaging at regular intervals (a few months to a year) over many years. This results in a light curve for each quasar in the sample, with the advantage that all the light curves have the same number of data points.

The very large effort in monitoring high-z quasars has recently come to fruition. There are now enough data in various samples to separate the effects of luminosity and redshift on the variability, avoiding the inherent correlation that exists in magnitude-limited samples. The observation that optical–UV spectra of both low- and high-luminosity AGN vary more at short than at long wavelengths (as found for high-luminosity AGN by Cutri et al 1985) accounts completely for the observed increase of variability with redshift (Giallongo et al 1991, Trevese et al 1994, Di Clemente et al 1996, Cristiani et al 1996). It also explains the long-puzzling absence of a time-dilation effect \( \Delta t_{\text{rest-frame}} = \frac{\Delta t_{\text{observed}}}{1 + z} \), wherein quasars at higher redshift are sampled more frequently and for shorter time intervals in their rest frames, because this effect is compensated by the intrinsic increase of variability with decreasing rest-frame wavelength.

The recent developments come mostly from the four largest on-going programs (the first three based on Schmidt plates): (a) the South Galactic Pole sample of 300 radio-quiet quasars observed over 16 years (Hook et al 1994); (b) the monitoring program in field ESO/SERC 287 (Hawkins 1993, 1996), with more than 200 plates since 1975 (Hawkins & Véron 1993), which has produced the best quasar light curves to date; (c) the sample in SA 94 comprising 183 quasars observed in B over 10 years (Cristiani et al 1990); and (d) the sample in SA 57 based on prime focus plates at the Mayall 4-m telescope at Kitt Peak (Koo et al 1986, Trevese et al 1989).

The main result is that, in a given proper time interval and at a fixed rest-frame wavelength, more luminous AGN vary with a smaller fractional amplitude than less luminous AGN. Also, the maxima and minima of the light curves are symmetric, a result also suggested by the light curves of low-z AGN.

The analysis of the variability is normally done using the structure function, which is the curve of growth of variability with time. It is defined in slightly different ways in the literature but generally has the form \( S(\Delta t_j) = < |m_{ik} - m_{i\ell}| > \), where \( \Delta t_j = |t_k - t_\ell| \), \( m_{ik} \) is the magnitude of the quasar \( i \) at epoch \( k \), and brackets signify the median of the ensemble (Hook et al 1994) or the average of the ensemble (Di Clemente et al 1996, Cristiani et al 1996).

Parametrizations of the structure function usually assume that the dependences on \( \Delta t \) and luminosity are separable. In Hook et al (1994) for example (also in Trevese et al 1994 and references therein), the best-fit model has the form

\[
|\Delta m| = |a + b(M_B + 25.7)|\Delta t_{\text{rest}}^p.
\]
with \( b \) about 0.022 and \( p = 0.18 \pm 0.02 \). This value of \( p \) corresponds to a power spectrum of the light curve \( P(\omega) = \omega^{-(1+\alpha)} \), with \( \alpha = 0.36 \pm 0.04 \). For comparison, in this description, \( p = 0, 0.5, \) and 1 correspond to uncorrelated measurements, random walk variations, and linear variations, respectively.

3.2 **Microlensing as a Possible Cause of Quasar Variability**

Variability of distant quasars could result from microlensing by compact bodies in intervening galaxies. The cosmological implications would be very important: If there were such a lens along all sight lines to high-z AGN, then \( \Omega \) in compact objects would be close to one (Press & Gunn 1973, Blandford & Narayan 1992). (This potential cause of variability does not pertain to the nearby, rather faint AGN discussed in Section 2, as they have a low probability of having a lensing object in the line of sight.)

For the ESO/SERC 287 sample, the main predictions of microlensing for variability are borne out by the observations (Hawkins 1996). The smoothness of the light curves constrains the quasar emission region to be commensurate with the Einstein radius of the lenses. With a typical variability time scale of \( \sim 2 \) years and for a transverse velocity of 600 km s\(^{-1}\), the Einstein radius is \( \sim 8 \times 10^{-4} \) pc, corresponding to a Jupiter mass (Hawkins 1996).

Another view is that the quasar light curves give an upper limit to the lensing effect because some or all of the quasar variability can be intrinsic. Recently Schneider (1993), following an idea of Canizares (1982), made a theoretical analysis of the characteristics of the light curves in Hawkins’ sample and calculated the properties of the lenses that can reproduce them. He can set an upper limit to the mass of dark matter in the form of compact objects in the specific mass range \( \Delta M \) 0.001–0.03 M\(_{\odot}\). This limit corresponds to \( \Omega(\Delta M) < 0.1 \).

At present, the results of quasar variability studies appear to be consistent with microlensing. It is difficult to disprove that quasar variability is dominated by microlensing because the properties of the lenses—in particular, their space density and their Einstein radii—are free parameters.

4. **EMISSION LINE VARIABILITY: RATIONALE AND METHODS**

4.1 **Rationale for Emission Line Variability Study**

The exciting prospect of determining the mass of the central black hole is the main motivation for monitoring emission line variations in AGN. The mass is estimated in the following way. Variations in the emission line strengths of AGN are observed to echo the continuum variations with a time delay, which can be interpreted as the light travel time between the central source and the surrounding high velocity gas clouds (the BLR). Combining the radial distance
to the line-emitting gas with its velocity (assuming virialized motions) allows
determination of the mass of the central black hole.

It is clearly important to assess whether the gas is gravitationally bound, as
well as to search for kinematic evidence of accretion and/or ordered gas motions
such as infall or a rotating disk. Reverberation (or echo) mapping is a technique
for inferring the structure and velocity field of the BLR from the time delays
between continuum and line variations. The basic assumptions are that the BLR
is ionized by a central continuum point source, the light travel time between
continuum and gas clouds is much longer than the ionization or recombination
times, and the line intensity is linearly correlated with the incident continuum
flux (e.g. Peterson 1993).

4.2 Inversion Methods, Cross-Correlations, and Modeling

Methods of echo mapping are presented in numerous papers (e.g. Blandford
& McKee 1982; Horne, Welsh & Peterson 1991; Krolik et al 1991) and in
conference proceedings (Gondhalekar et al 1994). The light curve of a given
line, $L(t)$, can be considered as a convolution of the continuum light curve, $C(t)$,
and a transfer function (TF), $\Psi(\tau)$:

$$L(t) = \bar{L} + \int_{\tau_{\text{min}}}^{\tau_{\text{max}}} \Psi(\tau) \left[ C(t - \tau) - \bar{C} \right] d\tau,$$

(2)

where $\bar{L}$ and $\bar{C}$ are the mean of the line and continuum intensities, respectively.
The transfer function (the kernel) is a map showing where the line emission is
produced in each interval of time delay by the gas along the paraboloid surfaces
of constant delay. This can be generalized to give the time delay distribution at
each velocity in the line profile.

A powerful and flexible inversion method is the Maximum Entropy Method,
MEM (Horne 1994), especially in its velocity-resolved form, but less computa-
tionally expensive linear methods have also been proposed (Pijpers 1994, Krolik
& Done 1995). The results of numerical techniques for inverting Equation 2
have been limited by the intrinsic indeterminacy of any inversion problem in
the presence of noisy data collected at uneven time intervals (e.g. Vio et al 1994).
Several authors have calculated theoretical velocity-delay maps for plausible
BLR configurations and velocity fields (Perez et al 1992a,b, Goad et al 1993)
for comparison with the result of the inversion of Equation 2.

A simpler estimate of the linear scale of the BLR comes from cross-correlating
the line and continuum light curves. The peak of the cross-correlation function
represents the material closest to the ionizing source, while the centroid gives
an emissivity-weighted average time delay over the emission region (Edelson
& Krolik 1988, Perez et al 1992a,b). These estimates are not unique, as they
try to encapsulate in a single number the time-delay distribution of the emitting
clouds. Note that the centroid of the transfer function can be obtained directly from the line-continuum cross-correlation function (Koratkar & Gaskell 1991, Robinson 1995).

5. EMISSION LINE VARIABILITY: RESULTS

Before reviewing the progress realized by the study of line variability, we briefly recall two critically important results obtained from spectroscopy alone. First, the high- and the low-ionization lines (HIL and LIL) are emitted by two different gas phases (Kwan & Krolik 1981, Collin-Souffrin et al 1986, Netzer 1987). The LIL come from a very dense medium with $N_e \geq 10^{11}$ cm$^{-3}$, an ionization parameter much less than 0.1, and a column density exceeding $10^{24}$ cm$^{-2}$. This medium must have a flat geometry, and there is the attractive possibility that the LIL are emitted by the accretion disk (Collin-Souffrin 1987). The HIL come from a more dilute medium, possibly a wind (van Groningen 1987, Collin-Souffrin & Lasota 1988) or an ensemble of clouds in a broad cone above and below the disk, with the gas density in the clouds not larger than a few $10^{10}$ cm$^{-3}$ and an ionization parameter of the order of 0.3.

Second, the gas velocity and its degree of ionization are correlated. Among the HIL, the broadest FWHM and the most extensive wings are those of the most highly ionized species (NGC 5548: Krolik et al 1991; NGC 4151: Antonucci & Cohen 1983, Ulrich et al 1984a). Similarly, in a few well observed AGN the H$\beta$ line has more extended wings than H$\alpha$. (We recall the nomenclature in this context: the main LIL are the Balmer lines, optical and UV Fe II multiplets, Mg II $\lambda\lambda 2796,2803$, He I $\lambda 5876$. Among the HIL, the strongest are C IV$\lambda\lambda 1548,1551$, then C III$\lambda 1909$, Si IV$\lambda\lambda 1394,1403$, He II$\lambda 4686$, N V$\lambda\lambda 1239,1243$, and Fe[ X]$\lambda 6375$. The Ly$\alpha$ line is special and has an intensity about twice that of C IV).

The crucial element contributed by variability studies is the linear scale of the LIL and HIL emission regions. For well-sampled high S/N data, the three-dimensional architecture and velocity field of the gas clouds can be reconstructed from the velocity-resolved transfer function. The results are at present dominated by the two bright, well-observed, strongly variable and intrinsically weak AGN, NGC 4151 and NGC 5548, and by 3C 273. Other AGN give consistent results. Most of the observational data on line variability come from the sources quoted for the continuum data in Section 2.

5.1 Results of Variability: The Stratification of the Broad Line Region in Velocity and Degree of Ionization

OBSERVED STRATIFICATION Among the HIL, the more highly ionized lines have shorter delays and larger amplitudes. The He II lines at $\lambda\lambda 1640,4686$ and
Figure 4  Top: The optical spectrum of the nucleus of NGC 3783 on February 8 and March 23, 1992. Bottom: The difference spectrum showing the strong decrease and change of spectral shape of the continuum and the varying component of the main broad emission lines. Note the variations of the He II λ4686. [From Stirpe et al (1994).] Ordinates in 10^{-15} erg s^{-1} cm^{-2} A^{-1}.


The picture that emerges is that of a stratified highly ionized BLR with the most highly ionized and fastest moving gas closest to the center, and the degree of ionization and velocity of the gas decreasing outwards. This is a “soft” stratification with ample overlap of the various ionization states. The absence of extended wings in the C III] λ1909 line implies an electron density exceeding 10^{10} cm^{-3} in the innermost region of the HIL BLR.

The Hβ and Hα lines have, in general, a longer time response and smaller amplitude than the C IV line. In NGC 5548, we see some stratification in the
Balmer line emission region as well: The wings vary faster than the core, and the higher Balmer lines vary with larger amplitudes and shorter time delays than the lower ones (Dietrich et al 1993, Wamsteker et al 1990, Kollatschny & Dietrich 1996, van Groningen 1984). Representative values of the time delay and of $r_{\text{max}}$, the ratio of maximum to minimum emission during a given fluctuation, are given in Table 1. The correlation coefficient is usually larger than 0.7.

**PHYSICAL INTERPRETATION** The variations of the HIL intensity ratios are generally consistent with the photoionization of ionization-bounded clouds, but there is also evidence for matter-bounded clouds: (a) a progressively weaker response of the C IV line to increases of the continuum flux above a certain level and (b) a flattening or decrease of the ratio C IV/Ly$\alpha$ at high-ionizing flux levels (NGC 3516: Ulrich & Boisson 1983; F9: Wamsteker & Colina 1986; NGC 5548: Dietrich & Kollatschny 1995; see also Binette et al 1989, Sparke 1993, Shields et al 1995). The BLR appears to be a mix of optically thin and optically thick gas clouds.

Some large variations of the Balmer decrement are associated with variations of the spectral shape of the optical continuum on time scales of 5–10 years and are entirely consistent with transient, strong, and variable dust extinction, possibly caused by clouds torn from the molecular torus (Goodrich 1989,1995, Tran 1995, Villar-Martin 1996).

**A CAVEAT** The BLR can respond only to continuum variations that last long enough to penetrate its volume significantly, and the amplitude of the continuum variations must also be large enough to alter the gas clouds’ emissivity. That is, the BLR filters out continuum variations that are too fast or too small.

For example, in NGC 4151 the continuum variations occurring in $\sim$1 day (Figure 2, December 1993) did not result in any detectable variations of the C IV line intensity (Crenshaw et al 1996), although their amplitude, by a factor 1.3, was sufficient to produce line intensity variations in slower conditions.

Because the velocity and line emissivity vary with the radial distance (“stratification”), the line intensity and profile variations differ according to the duration and the amplitude of the continuum event (Netzer & Maoz 1990). Care should be exercised when comparing delays of line responses during different episodes or in different AGN. Only comparisons between events with similar continuum amplitudes are valid.

**THE TREND WITH ABSOLUTE LUMINOSITY** Although the data are few, spectroscopic observations of high-luminosity AGN show that time delays of lines with respect to the continuum appear to increase with intrinsic luminosity (Zheng et al 1987, Pérez et al 1989, Gondhalekar 1990, Hooimeyer et al 1992). The recent measurement of the H$\beta$ time delay in two high-luminosity PG quasars,
combined with data for low-luminosity AGN (Table 1), suggests the BLR radius scales as $L^{-0.5}$ (Kaspi et al 1996b). This derived scaling is subject to the caveat above, but as the Kaspi et al (1996b) data have amplitudes in the range 1.4–4, it is probably valid.

3C 273 is the highest luminosity AGN (Table 1) for which long-term spectroscopic monitoring has been organized (regularly observed with IUE since 1978, and every other week since 1985). While the continuum flux has varied several times by a factor of 2 on time scales of ~2 years, no intensity variations of Ly$\alpha + N\,V$ above 10% have been detected (Ulrich et al 1993 and references therein). Therefore, only a very small fraction of the Ly$\alpha + N\,V$ emitting gas can be within 2 light-years of the continuum source. [The small amplitude Ly$\alpha + N\,V$ variations were judged by Ulrich et al (1993) not to be enough above the measurements errors to produce a reliable value for the time delay. With the same data set, time delays of $74 \pm 33$ days and $118 \pm 57$ days were found by Koratkar & Gaskell (1991) and O’Brien & Harries (1991). Whatever the robustness of these values of the delay, they apply to only a minute fraction of the Ly$\alpha + N\,V$ region. The most important result on 3C 273 is that the variations of the Ly$\alpha + N\,V$ line are, at most, of very small amplitude.]

This contrasts with lower luminosity AGN where a continuum flux variation by a factor of 2 always produces a response of the lines of comparable amplitude (e.g. Ulrich et al 1993, Figures 1 and 2). This implies the quasi-absence in 3C 273 of Ly$\alpha$-emitting gas at a distance less than $c\Delta t$ from the continuum source, $\Delta t$ being the characteristic time scale of the continuum variations, defined here as the time separating two maxima.

5.2 Mapping the Velocity Field from the Emission Lines

That the fastest moving gas is the closest to the center implies that radiative acceleration is less important than gravity and rotation. This dynamical information supports the connection of AGN line variability to black hole mass, providing that radial motions are not dominating the velocity field.

The search for radial motions from the high-ionization lines To first order, the HIL profile variations do not show the systematic change of one wing before the other that would be expected for purely radial flows (spherical winds, spherical accretion). This requires the main motions of the HIL clouds to be circulatory with only minor components of net infall or outflow. The data are consistent with the HIL clouds being in circular orbits in a disk, or having “chaotic” motions along randomly oriented orbits in the gravitational field of the central mass.

The velocity-resolved MEM inversion of Equation 2 applied to the best-sampled data of NGC 4151 and NGC 5548 shows that the time delays vary as
function of velocity in a way that is roughly consistent with virialized motions. The central mass so derived is \( \sim 10^7 M_\odot \) for NGC 4151 (Ulrich & Horne 1996). In NGC 5548, it is between \( 2 \times 10^7 M_\odot \) (for two-dimensional random motions) and \( 8 \times 10^7 M_\odot \) (for three-dimensional random motions; Done & Krolik 1996).

In both NGC 4151 and NGC 5548 (and also in F9, Recondo-González et al 1997), however, the best-sampled data show small differences in the transfer functions of the blue and the red wing of C IV, with the response of the red wing being the stronger at small delays. This small asymmetry could, at first sight, be taken as a subtle indication of radial motions. On the other hand, optical depth effects in a rotating/outflowing wind produce a differential response across the C IV profile, with the red wing leading the blue wing (Bottorff et al 1997, Murray & Chiang 1997).

DIFFERENT EMITTING REGIONS FOR THE HIGH- AND LOW-IONIZATION LINES? As stated in the introduction, the LIL come from very dense gas with high column density, whereas the HIL come from a more diffuse photoionized medium. Can the variability data establish the existence of these two different emission regions? The answer is positive but provisional: In the few AGN with good variability data on C IV and H\( \beta \), the inversion of Equation 2 gives a transfer function for H\( \beta \) that is small near zero lag, indicating the near absence of matter along the line-of-sight, whereas in contrast, the transfer function of the C IV wings peaks near zero lag (Ulrich & Horne 1996, Done & Krolik 1996). This is what is expected if the Balmer lines come from a disk at small inclination (there is no matter close to the line of sight) and the HIL come from a broad cone or cylinder, in which case some matter lies along the line of sight producing a nonzero response near zero lag.

According to Wanders & Peterson (1996), however, the lack of response at zero lag for the transfer function of the H\( \beta \) line is spurious and due to the combined effects of noise in the data and lack of resolution. On the other hand, Keith Horne, responding to a friendly challenge, has run his MEMECHO program on sets of simulated noisy data prepared by Dan Maoz and successfully identified which data corresponded to a transfer function peaking at zero and which ones did not (K Horne, private communication; Maoz 1997). This exercise gives weight to the results obtained from inversion of Equation 2, but, clearly, a final answer to this question awaits data of higher quality than presently available.

THE NATURE OF THE MOTIONS OF THE HIL GAS—IS THE DISK OPAQUE? If indeed a disk is present in AGN (and emits the Balmer lines and the Fe II multiplets), then two important points about the disk need to be specified before interpreting the velocity information on the HIL. First, is the disk opaque or
transparent (can we see what is happening below the disk)? Second, could HIL clouds survive if they cross the disk in their chaotic orbits?

Calculations of accretion disk structure (Hure et al. 1994) suggest that the disk becomes self-gravitating at 100 $r_S$ ($r_S = 2G M_{BH}/c^2$ is the Schwarzschild radius), much smaller than the radius of the region emitting the Balmer lines, although plausible mechanisms could stabilize the disk at a much larger radius (magnetic fields or internal dissipation; e.g. Sincell & Krolik 1997). It is possible, and we believe likely, that the disk (whether in a continuous structure or broken up in clouds in its outer parts) joins to the molecular torus, and nowhere is it sufficiently transparent that we can see the far side. This has profound consequences for the interpretation of the observed velocity field. In a biconal flow, only the near half would be visible, and if there were truly chaotic motions, only one half of each orbit would be seen. In addition to being opaque, the disk could have a column density such that free-flying clouds would be destroyed when crossing the disk. In this case, only stars can have chaotic motions, and if the HIL BLR clouds are the (modified) atmospheres of stars (Penston 1988, Alexander & Netzer 1994), they could collectively partake in pure gravitational motions, but we would still see only those clouds on the near side of the disk.

A solution to this puzzle is offered by the fact that in radio-quiet AGN, the HIL are blueshifted (by 0 to $\sim 1500$ km s$^{-1}$) with respect to the low HIL, which themselves are at the host galaxy redshift (Gaskell 1982, Wilkes 1984, Corbin 1995, Sulentic et al. 1995a, Marziani et al 1996). This indicates that the gas emitting the highly ionized lines is moving towards us, probably emanating from the disk, and still retaining a large part of the angular momentum it had in the disk (thus allowing a derivation of the central mass). The origin of the observed range of the HIL blueshifts is unclear—orientation can explain only part of it. Additional evidence for the presence of outflowing material in AGN includes (a) the blueshifted absorption lines in the emission line profile of a significant number of AGN, (b) the blueshift of the coronal lines (Wagner 1997) and (c) the shift of the reflected broad lines in Seyfert 2.

Magnetically accelerated outflows from accretion disks and radiatively driven winds are promising models for the formation and evolution of the highly ionized gas clouds (Blandford & Payne 1982, Emmering et al 1992, Königl & Kartje 1994, Murray & Chiang 1995, Bottorff et al 1997). Clouds, probably filaments, are pulled from the originally dense low-ionized material of the disk and subjected to the intense ionizing field, forming a more diffuse highly ionized outflowing medium. The densest, coolest inhomogeneities form the BLR clouds emitting the prominent lines (and if one such cloud happens to be crossing our line of sight to the AGN center, it should produce a blueshifted absorption line as, in fact, often observed in AGN). The hottest phase is detected as the fully ionized component of the warm absorber, which could also produce the blueshifted
absorption lines in the wings of the HIL. Many features of these promising models remain unspecified and can be adjusted to accommodate the observations.

Two other models are compatible with an opaque accretion disk and the simultaneous response of the blue and red wings of the HIL. Both solve the confinement problem and include elements important for the formation and evolution of the BLR, but they do not specifically include magnetic forces. In the bouncing gas clouds model (Mathews 1993), the clouds congregate at a preferred radius where radiation forces and gravity are balanced. Fluctuations in the gas pressure move the clouds radially in and out, but most of the clouds tend to come back to the preferred radius set by the radiation level. In the bloated stars scenario, the BLR clouds are the modified atmospheres of some of the stars of the central stellar core and thus move along Keplerian orbits around the black hole (Penston 1988, Alexander & Netzer 1994).

5.3 Other Issues Concerning the BLR

PARTIAL REDISTRIBUTION OF THE GAS IN THE BLR

Comparison of line profile variability during month-long campaigns separated by one or more years reveals that the line response is not stationary. This phenomenon, observed in the HIL and the LIL, is probably caused by changes in the distribution of the BLR gas in a few years (NGC 4151: Ulrich et al 1991, Perry et al 1994; NGC 3516: Wanders & Horne 1994; NGC 5548: Wanders & Peterson 1996). Although it is not possible at this time to offer a definitive interpretation of these changes, they suggest the presence of time variable inhomogeneities on the surface of the disk (possibly due to the magnetic field) which can (a) enhance the Balmer lines emission in some locations (e.g. bumps on the disk surface would intercept more continuum flux), thus altering the Hβ profile and (b) strengthen the gas extraction from the disk surface (at local enhancements of the magnetic field), thus lifting additional HIL clouds above the disk and producing shoulders and other features in the C IV line (Figure 5).

That the line response evolves with time reduces the value of any single observation like the infall seen in NGC 4151 (November–December 1991) and in NGC 5548 (March–May 1993). Only if a certain behavior repeats itself over the years and for several AGN can it be considered to represent a standing feature of the velocity field or of the distribution of matter.

EMISSION LINES WITH DOUBLE PEAKS: ACCRETION DISKS, OUTFLOWS, OR BINARY BLACK HOLES?

A small number of AGN show double-peaked Balmer lines, suggestive of emission from a rotating disk. (Double peaks are present only in the LIL.) This profile appears preferentially among broad-line radio galaxies and galaxies with a very compact central radio source (Eracleous & Halpern 1994, Gaskell 1996b). A relativistic disk would have an extended red
Figure 5  Examples of long- and short-term variations of the C iv line in NGC 4151. Variations within days are shown on each panel. One can appreciate the variations on time scales of years by comparing the two panels and considering that between 1985 and 1990 the C iv line was observed to be perfectly symmetrical, e.g. November 1988–January 1989. [From Ulrich et al (1991).] Ordinates in $10^{-14}$ erg s$^{-1}$ cm$^{-2}$ Å$^{-1}$.

wing because of gravitational redshift, a stronger blue peak because of beaming, and a range of peak intensity ratios, B/R, that is well constrained. Several double-peak profiles have been successfully fitted with an axisymmetric accretion disk model at some epochs (Halpern & Filippenko 1988, Rokaki et al 1992, Eracleous & Halpern 1994), but later observations revealed changes of B/R inconsistent with the disk model (Miller & Peterson 1990). Elliptical disks, warps and spiral shocks have been proposed but with no definitive conclusions (Chakrabarti & Wiita 1994, Eracleous et al 1995b, Bao et al 1996).

The model where the BLR is a biconal inhomogeneous flow illuminated by a variable double beam appears unlikely because of the difficulties in accelerating the very dense matter that produces the LIL, and the model is little constrained (3C 390.3; Zheng et al 1991). Double peaks have also been interpreted as the signature of two orbiting black holes, each with its own BLR (Gaskell 1996a), resulting from the merger of two galaxies with central black holes. In 3C 390.3, the period is estimated to be $\sim$300 years, corresponding to a total mass of $7 \times 10^9 M_\odot$. A difficulty of this model is that no other drift has been well
observed (except possibly in OQ 208; Marziani et al 1993) and also that one would expect the double peaks to appear in the HIL as well as in the LIL, in contrast to the available observations (Arp 102B: Halpern et al 1996; 3C390.3: Wamsteker et al 1997).

Distinct double peak profiles form but one category of the complex profiles displayed by many AGN (Eracleous et al 1995a, Stirpe et al 1988). This suggests that the complex profiles may simply be caused by transient inhomogeneities and asymmetries in the emissivity and/or distribution of the emitting matter. This was also implied by the changes in the BLR gas distribution observed in NGC 4151 and NGC 5548 over several years (see Partial Redistribution of the Gas in the BLR). Symmetric inhomogeneities resulting in double peaks could be produced preferentially in the disk of radio sources through conditions related to radio jet formation.

**Emission Lines During Minimum States; The Effect of the Long-Term Variations**

The long minimum of NGC 4151 in 1981–1988 (interrupted by short episodes at medium bright states, Figure 1b), has caused the regular decrease of the central part of the C IV, C III, and Mg II lines between 1978 and 1991 (Ulrich et al 1991). A similar variation of a medium width component of Lyα has also been observed in 3C 390.3 (Clavel & Wamsteker 1987) and reveals, as in NGC 4151, the presence of an intermediate line region with a size of a few light-years and velocities of the order of 2000–3000 km s⁻¹.

During the deepest period of minimum NGC 4151 was the subject of two multi-month campaigns. The IUE spectra reveal the unexpected presence of two narrow emission lines, whose intensity varies by a factor of 2–3 in a few days apparently in phase with the small amplitude variations of the weak continuum (Ulrich et al 1985, Ulrich 1996). These lines, at $\lambda\lambda$(rest) 1518.5, 1594.4 Å and with FWHM less than 7 and 16 Å, respectively, are too narrow to be emitted by the entire BLR and must arise instead from two localized regions that have a special excitation mechanism.

These lines are best measured at minimum when the broad wings of C IV have faded but can also be seen, albeit with less contrast against the broad wings, at medium bright state (Clavel et al 1987, Kriss et al 1992). Their origin is unclear: That they are C IV components (at $-6100$ km and $+8500$ km s⁻¹) emitted by a two-sided flow is an attractive possibility, but it is based on the uncertain assumption that the accretion disk is transparent.

**The Appearance and Disappearance of Broad Emission Components**

Very large intensity changes in broad emission lines have been observed in different sets of circumstances, implying different origins for these large amplitude variations. First, some are in direct response to the large amplitude fluctuations of the ionizing continuum observed in a few AGN (see Sections 2.1 and 5.3).
The nonzero delay between continuum and line variations rules out, in these cases, obscuration by a dust cloud. Second, unexpected appearances of broad components have been reported in three AGN that, while displaying definite signs of activity, had before the event rather narrow emission lines or somewhat broad lines with triangular profiles. Interestingly, in these three cases (Pictor A, NGC 1097, and M81), the prominent new broad component is double-peaked. The circumstances that produce these lines and the cause of their double-peaked profiles remain unclear, although the accretion of a star followed by the formation of an elliptic disk has been proposed (Halpern & Eracleous 1994, Sulentic et al 1995b, Bower et al 1996, Storchi-Bergmann et al 1996).

The third case of unexpected appearance of a broad emission component occurred in the nucleus of NGC 4552, an apparently normal elliptical. Between two HST observations separated by two years, the continuum brightened by a luminosity of $10^6 L_\odot$ (Renzini et al 1995). Spectroscopic observations carried out after the brightening revealed the presence of a broad Mg II line. Was it accretion of a passing star or of an interstellar cloud by a dormant central black hole? A systematic search for such events in apparently quiescent galaxies could be a new way to discover black holes in the general galaxy population. Finally, as mentioned in Section 5.1, dust clouds crossing our line of sight to the BLR cause large reddening and flux variations of the continuum with simultaneous characteristic Balmer decrement variations.

5.4 Summary, Perspectives, and Emerging Fields in Emission Line Variability Studies

The most important results

1. The dimension of the line emitting region can be derived from the time delays between continuum and line variations. The delay depends on the line and on the velocity.

2. The velocity field of the broad line gas appears to be dominated by virialized motions in the gravity field of the black hole (the fastest moving gas is closest to the center; the blue and the red wings of the emission lines, to first order, vary simultaneously, as expected if radial motions are not important). The black hole mass can then be derived from the observed motions.

3. The variability observations can be best understood in the frame of the “accretion disk plus wind” model.1 The results of variability studies strengthen

---

1This model is the simplest architectural sketch explaining the two basic results of spectroscopic studies: 1. The HIL and the LIL come from two media with distinctly different physical conditions, and 2. these media have different velocity fields because the HIL are blueshifted (a small shift when compared with the full width of the lines). This model attributes the emission of the LIL to the accretion disk, whereas the HIL are emitted by a hotter more diffuse outflowing medium; the disk is implicitly opaque in this model.
or are consistent with this model. Specifically, the first-order similarity between the responses of the blue and red wings of the lines supports the model. A second-order effect is observed in three AGN in the form of some asymmetry between the blue- and red-wing responses, with the red wing leading the blue; such a difference has a natural explanation in wind models. Finally, hints for different values of the transfer function at zero lag in HIL and LIL have been found in several AGN. If this effect is confirmed in those and other AGN, it will add strong support to this model.

In conclusion, the gas in the radio-quiet AGN central region has a roughly ordered velocity field: rotation in a disk for the low-ionization medium and rotation plus outflow for the high-ionization medium, which is pulled out from the disk by magnetic or radiative forces. The details of the disk plus wind structure are complex. For example, changes in the HIL and LIL transfer functions over several years are best understood as changes in the distribution of the HIL and LIL gas on time scales roughly commensurate with the dynamical time of the inner BLR in low-luminosity AGN. Together with the complexity of the line profile variations (Figure 5), they suggest the presence of inhomogeneities on the disk surface and in the outflowing medium.

THE LIMITS AND THE REMEDIES Derived transfer functions should in general be regarded with caution in spite of the successful challenge of Horne (Section 5.2). The physical situation is bound to be complicated by the gas inhomogeneities, local overlap in velocities (e.g. Done & Krolik 1996), and imperfect correlation between the measured continuum and the ionizing continuum. The calculation of the transfer function itself is limited not by the techniques but by the data. “Improving the Signal-to-Noise ratio and sampling by modest factors of 3 would greatly sharpen the velocity-delay maps” (Horne 1994).

Such an improvement represents a rather tall order but appears to be necessary in order to reach firm conclusions on the distribution and velocity field of the gas and the anisotropy of the line emission in each cloud (Horne 1994). In the future, such excellent data will allow full utilization of the information in the profile variations (something that has not yet been done systematically) and will motivate the development of inversion methods incorporating more specific geometry and physics into the models, such as photoionization codes. Meanwhile, the universality of results obtained predominantly from a few low-luminosity radio-quiet AGN selected as targets because they were the most active AGN about 10 years ago should be regarded with caution.

THE EMERGING FIELDS A number of avenues are ripe for future variability investigations. First, the entire parameter space defined by the black hole mass and the accretion rate should be explored. Several subsets of AGN populate the faint end of the optical/UV luminosity function. Intrinsically low-luminosity
broad-line AGN (evidently with small accretion rates but with unknown black hole masses) have been discovered in 13% of nearby galaxies (Ho et al 1995). Their continuum and line variability is essentially unknown, and what little we know is puzzling, as evidenced by the fact that in M81, after about 15 years of no detectable change in the Balmer lines, a broad double-peaked component has recently appeared (Peimbert & Torres-Peimbert 1981, Ho et al 1996, Bower et al 1996).

From HST observations, LINERS (galaxies with low-ionization narrow emission-line regions whose spectra resemble neither an HII region nor the narrow-line spectrum of a broad-line AGN; Heckman 1980) have recently been found to harbor point-like nuclear UV sources in 25% of the cases (Maoz et al 1995). The absence of a nuclear UV source in 75% of the observed LINERS may reflect their duty cycle (Eracleous et al 1995a). The study of the continuum variability and the search for broad lines will greatly improve our understanding of LINERS.

Finally, for the NLS1—where the accretion rate may be exceptionally high (Section 2.2)—it is important to determine the central mass through systematic investigations of line variability.

All of the above AGN have in common that the broad lines are weak or only moderately strong. Their spectral variability studies require the use of HST or of large or very large optical telescopes. The question of the duty cycle of AGN, and the passage of a given AGN from one subset to another (from LINERS to “classical” broad-line Seyfert 1 to NLS1), requires good statistics on the number of AGN in each subset and long-term monitoring to witness the passage from one class to another, as occurred with Pictor A and NGC 1097 (Section 5.3, Appearance and Disappearance of Broad Emission Components). In fact, with the development of electronic archives, one could imagine light curves extending several centuries or more. (And we would not even be pioneers because records of supernovae can be found in Chinese and other archives that are a millennium or older).

At the bright end of the optical/UV luminosity function, the monitoring of line and continuum variability will allow us to follow the dependence of the line response on the AGN luminosity (cf the non-detection of significant Ly$\alpha$ variations in 3C 273, Section 5.1), and perhaps determine the central mass in the brightest objects in the universe. Powerful large field multi-object spectrographs (for example, installed on Schmidt telescopes) should be the instruments of choice.

Second, the recent observations of the profile and variability of the Fe K line (Tanaka et al 1995, Yaqoob et al 1996) and of the absorption lines originating in the hottest part of the warm absorber (Iwazawa et al 1997) are new ways to explore the innermost regions of AGN and will develop into a flourishing
field with the Advanced X-Ray Astrophysics Facility, the X-Ray Multi-mirror Mission, and other X-ray missions.

The third promising avenue is that of theoretical investigations. This is hardly an emerging field, but new data will elicit new models and calculations. The accretion disk plus wind model appears particularly promising as it offers explanations for the variability of the broad lines, the two phases of the BLR gas, the blueshift of the emission and the absorption lines, and the emission of the UV continuum. This model and others should be pursued to the point where they can predict line intensity and profile variability patterns that can be compared with observations.

The effect of the certain presence of the central star cluster should also be investigated. The stars can be accreted (cf NGC 4552, Renzini et al 1995), collide with one another (Courvoisier et al 1996), or be trapped by the disk (Artymowicz et al 1993), and their atmospheres contribute to the BLR (Alexander & Netzer 1994, Armitage et al 1996). All these processes cause line and continuum variations.

6. VARIABILITY OF BLAZARS

6.1 Overview and Relativistic Beaming

Blazars exhibit the most rapid and the largest amplitude variations of all AGN (Stein et al 1976, Angel & Stockman 1980). The combination of extreme variability and relatively weak spectral features suggests the continuum is emitted by a relativistic jet close to the line of sight and hence that the observed radiation is strongly amplified by relativistic beaming (Blandford & Rees 1978). Here we take the point of view that all blazars, whether weak-lined like BL Lac objects or strong-lined like flat spectrum radio-loud quasars (FSRQ), contain essentially similar relativistic jets. They may differ in other aspects of nuclear activity; in particular, BL Lac objects may have less luminous accretion disks and BLRs than FSRQ.

Early multiwavelength studies provided the first global support for the idea of bulk relativistic motion in blazars: The observed radio emission was sufficiently luminous and rapidly variable that, assuming it was due to synchrotron radiation, high X-ray fluxes would be expected from Compton upscattering of the synchrotron photons (the so-called synchrotron self-Compton process), unless the radio emission was relativistically beamed (Hoyle et al 1966, Jones et al 1974a,b).

In addition, many blazars exhibited large-amplitude extremely fast X-ray variations (e.g. Morini et al 1986a, Feigelson et al 1986) that for isotropic emission would violate the limits on \( \Delta L/\Delta t \) for Eddington-limited accretion (Fabian 1979), implying that relativistic effects are important (see also Bassani
et al 1983). Finally, direct evidence for relativistic bulk motion has been obtained with VLBI observations of many blazars, which show that apparent superluminal motion is the rule (Mutel 1990, Vermeulen & Cohen 1994).

Here we summarize the data in wavebands that are particularly relevant to understanding the overall continuum (Sections 6.2–6.6) and discuss multiwavelength studies, emphasizing the newest results (Section 7). Models are discussed in Section 8.

6.2 Spectral Shape and Variability of the Blazar Continuum

The continuum emission of blazars is remarkably smooth and steepens gradually towards shorter wavelengths from the radio to the UV range (Impey & Neugebauer 1988). The power emitted per decade exhibits a broad peak (where $\alpha \sim 1$) at IR through X-ray wavelengths. This long-wavelength component is almost certainly synchrotron emission, as evidenced by its high polarization from radio through optical wavelengths. The radio core is opaque, so that radio observations do not generally probe the region of the jet closest to the central engine; in general, the synchrotron continuum becomes optically thin at submillimeter or shorter wavelengths.

The wavelength of the peak of the synchrotron luminosity in blazars anticorrelates with the ratio of X-ray to radio flux. On this basis Padovani & Giommi (1995) divided BL Lac objects into high-frequency peaked (HBL) and low-frequency peaked (LBL) objects, according to whether $\alpha_{rx}$ (from 5 GHz to 1 keV) is $<0.75$ or $>0.75$, respectively (see schematic examples in Figure 6). The two subclasses, HBL and LBL, are sometimes referred to as XBL and RBL, respectively (see discussion in Urry & Padovani 1995).

Here we extend these definitions to all blazars. It is not yet known whether LBL and HBL represent two distinct populations or extremes of a continuous distribution of synchrotron peaks because X-ray selection favors HBL and radio selection, LBL. Arguments for continuity are given by Sambruna et al (1996a).

6.3 Far IR–Optical–UV: The Thin Synchrotron Emission

IRAS observations have shown that for LBL, a large fraction of the bolometric luminosity is radiated at mid-IR wavelengths (Impey & Neugebauer 1988) and that rapid (time scales of weeks) large-amplitude (factors of two) far-IR variability is common, whereas the nonblazar AGN do not vary significantly in the far-IR band (Edelson & Malkan 1986, 1987). Multiwavelength observations of 3C 345 from radio to UV wavelengths, close in time to the IRAS pointing, define clearly a peak in the power per decade between $10^{13}$ and $10^{14}$ Hz (Bregman et al 1986), one of the few LBL cases in which the peak is actually measured rather than inferred.
Figure 6  Schematic broadband spectra of blazars from radio through TeV gamma rays. The low-energy component is probably due to synchrotron radiation and the high-energy component to Compton scattering of lower-energy seed photons, possibly the synchrotron photons or ambient UV/X-ray disk or line photons. Two different curves represent the average spectral shapes (Sambruna et al 1996) of HBL (high-frequency peaked BL Lac objects; dotted line) and LBL (low-frequency peaked BL Lac objects; dashed line) as defined by their ratios of X-ray to radio flux (Section 6.2). Strong emission-line blazars (i.e. flat-spectrum radio quasars, or FSRQ) have continua like those of LBL (Sambruna et al 1996).

The spectra of LBL fall off rapidly at wavelengths shorter than the synchrotron peak, i.e. in the optical–UV bands, whereas the spectra of HBL in the same bands are much flatter. Because of their relative brightness, LBL have been monitored extensively in the optical band, whereas at UV and X-ray wavelengths, mostly HBL have been observed.

The “historic” optical light curves of blazars on time scales of years show a variety of behaviors, from slow long-term trends to rapid repeated flares (Webb et al 1988). In general, long-term trends have typical time scales of 5 years in the rest frame, both for BL Lacs and FSRQ (Smith et al 1993, Smith & Nair 1995). Structure functions of the best data show slopes somewhat flatter than for radio light curves, indicating a transition from shot noise in the radio to flicker noise in the optical (Hufnagel & Bregman 1992), with relatively more power on short time scales. There are no obvious differences between BL Lac objects and FSRQ within the limited blazar samples available.
Optical variability extends to very short time scales, and intra-night small-amplitude variability has been observed in a number of blazars (Jang & Miller 1995, Heidt & Wagner 1996). The short time scale variations of radio-selected (LBL) are systematically larger in amplitude and have shorter duty cycles than those of X-ray selected (HBL) BL Lac objects (Heidt 1996). Within the radio-selected sample, there is a tendency for greater optical activity among higher luminosity sources (Heidt & Wagner 1996).

HBL are in general less polarized than LBL (Jannuzi et al 1993); however, the constancy of the polarized fraction and wavelength dependence over large flux variations observed in PKS 2155–304 argues against dilution by either starlight or an accretion disk continuum as the cause of this effect (Smith et al 1992).

High-quality UV light curves for more than a dozen blazars, obtained with IUE, show significant variability (amplitudes of 8–80%), which correlates with degree of optical polarization and with luminosity (Edelson 1992). This trend, which is opposite to the case of Seyfert galaxies, has been interpreted to mean that the most variable objects are the most beamed. Alternatively it could result from an inverse correlation between luminosity and peak frequency.

The densest IUE monitoring has revealed UV variations on extremely short time scales. In PKS 2155–304, the observed UV flux doubled in 1 h on one occasion (Pian et al 1997), comparable to the fastest X-ray variability observed in this object. A doubling time scale of 10 days is common (Urry et al 1993). At extreme ultraviolet wavelengths, rapid variations of slightly larger amplitude have been seen, although only two blazars have been monitored extensively with EUVE (HL Marshall et al, in preparation).

The spectral variability in the UV band is generally small, with only a weak tendency for larger amplitude variability at shorter UV wavelengths (Edelson 1992, Pian & Treves 1993, Paltani & Courvoisier 1994). The two blazars with the most UV observations, Mrk 421 and PKS 2155–304 (both HBL), show spectral hardening with increasing intensity only in a statistical sense (Ulrich et al 1984b, Maraschi et al 1986, George et al 1988a, Urry et al 1988). We note that LBL are faint in the UV and their slopes are difficult to measure with IUE.

6.4 \textbf{X Rays: The Crossing of Different Emission Components}

For HBL, the medium energy X-ray emission (2–6 keV) is typically steep, extrapolating smoothly from the UV and comprising part of the downward curving synchrotron spectrum. In LBL, the synchrotron component curves down well below the X-ray band, which is then dominated by a much flatter component (see Figure 6). Emission line blazars tend to fall in the LBL category. Borderline objects may have both components contributing in the X-ray range with the steep one prevailing in the soft X-ray range.
In HBL, rapid large-amplitude X-ray variability is the rule (flux doubling on time scales of hours). The spectra harden systematically with increasing intensity (Urry et al 1986, Treves et al 1989, George et al 1988b, Giommi et al 1990, Sembay et al 1993, Sambruna et al 1994). Comparing the UV and X-ray variability of HBL suggests that both spectral changes and variability amplitude are greater beyond the synchrotron peak, which is in the soft X rays for these objects.

ROSAT observations of a complete sample of radio-selected BL Lac objects (mostly LBL) show, in general, flatter spectra than for HBL and different variability behavior as well. In three cases there were “inverse” spectral changes, that is, a softening of the spectrum with increasing intensity (Cappi et al 1994, Urry et al 1996). This can be understood in terms of the relative variation of two spectral components that intersect each other in the X-ray range: a soft, highly variable one that swamps (high soft intensity) or uncovers (low soft intensity) a less-variable flatter component (see Figure 6). Such objects should be intermediate between LBL and HBL.

Less is known about the X-ray variability of emission line blazars because they are relatively weak X-ray sources. Einstein observations revealed extremely hard spectra in the 0.2–4 keV band (Worrall & Wilkes 1990), much flatter than the extrapolation of the optical–UV spectrum. Short-term variations were not detected in 3C 279 and NRAO 140, the only two blazars with sufficient intensity in the Ginga data base (2–20 keV), while long-term variations were seen in three out of four sources observed repeatedly (Tashiro 1995).

In summary, when a flat component is present in the X-ray band (as for LBL), it appears to vary on longer time scales and with lower amplitude than the steep X-ray component that is thought to be an extension of the longer wavelength synchrotron emission. However, this statement may be partly biased by the lower X-ray fluxes of LBL compared to those of the better-observed HBL.

6.5 High Energy Gamma Rays: Where the Action Is

Perhaps the most important progress in the last decade was the discovery with the Compton Gamma-Ray Observatory (CGRO) EGRET instrument that many blazars (presently 40–60, Thompson et al 1995; RC Hartman, private communication) emit enormous power in rapidly variable GeV gamma rays. This gamma-ray emission indicates a second peak in the overall spectral power distribution (Maraschi et al 1994, von Montigny et al 1995; Figure 6).

Practically all EGRET blazars with sufficient statistics and observations are strongly variable on time scales of months (Hartman 1996) and in several cases significant large amplitude flares have been observed on time scales of days, (3C 279: Kniffen et al 1993; PKS 0528+134: Hunter et al 1993). The most extreme example is PKS 1622–297, which brightened by at least a factor of 10
in two days, reaching a peak gamma-ray intensity 5 times that of any previously observed blazar (Mattox 1995, Mattox & Wagner 1996). The gamma-ray spectra of individual blazars appear to harden with increasing intensity (3C 279: Kniffen et al 1993; PKS 0528+134: Mukherjee et al 1996).

In some cases the gamma-ray emission extends to the TeV range (Figure 7). Mrk 421 was the first extragalactic source detected at such energies with the Whipple Observatory (Punch et al 1992), although it is only weakly detected at GeV energies with EGRET (Lin et al 1992). The TeV flux of Mrk 421 is variable by up to a factor of 10 on time scales of a day (Kerrick et al 1995) and by a factor of 5 on a time scale of 30 min (Gaidos et al 1996). Shorter time scales are not presently accessible owing to low event rates. Large-amplitude TeV variability must be frequent as such variations are commonly detected, in contrast with the quiet behavior of the same source in the GeV range.

The TeV power emitted by Mrk 421 (see Figure 7) dominates its bolometric luminosity, at least in the high state. Another nearby BL Lac object, Mrk 501
(also a modest EGRET source), has also been detected at TeV energies (Quinn et al 1996). Absorption by IR-optical photons locally or along the intergalactic path may prevent the detection of more distant blazars (Stecker et al 1996). Alternatively HBL, of which Mrk 421 is the brightest in the northern sky, may be intrinsically stronger TeV sources than LBL owing to the higher peak frequency of both spectral components (see Figures 6 and 7).

Some important conclusions can be drawn from the EGRET and Whipple discoveries. First, the gamma-ray emitting region must be transparent (i.e. the optical depth to pair production must be low), yet for minimal assumptions about ambient X-ray photon densities, the size limit imposed by the rapid gamma-ray variability implies very high optical depths. Therefore the gamma rays must be relativistically beamed (Maraschi et al 1992, Becker & Kafatos 1995, Dondi & Ghisellini 1995, Gaidos et al 1996).

Second, blazars emit a large fraction of their luminosity at very high energies. If the long-term gamma-ray light curve of 3C 279 is typical, even if the EGRET detections are biased toward exceptional states of gamma-ray activity, the “average” power output in gamma rays is comparable to that at all other wavelengths.

Third, the second peak of the spectral power distribution remarkably seems to fall at the highest energies for those objects whose first peak is also at high frequency: it probably lies near 0.1–1 GeV for LBL and 10–100 GeV for HBL (see Figure 7), although this is based on very incomplete data. A likely origin of the gamma-ray emission is Compton scattering of lower energy photons by the same relativistic electrons producing the low frequency component (Section 8.4). The correlation of variability at high and low frequencies is a crucial test for this class of models.

6.6 Periodicity of OJ 287

In one exceptional blazar, OJ 287, there is evidence for periodic flaring on a time scale of $\sim 12$ years (Takalo 1994). Only three cycles have been seen during the epoch of dense monitoring, but the third of these was predicted to within six weeks from the light curve of the preceding 100 years (Sillanpaa et al 1988, Babadzhanyants et al 1992, Kidger et al 1992). Including the most recent data, the Fourier transform of the optical light curve shows six peaks, the two strongest of which correspond to 12.13 years, and its first harmonic, 6.07 years (Sillanpaa et al 1996). The period, if real, has a range of at least 1%; also, OJ 287 exhibits outbursts outside of the periodic oscillations, including some of comparable amplitude. A better assessment of the statistical significance of this period remains to be done.

If the light curve of OJ 287 does indeed contain a 12-year periodic component, its explanation in terms of precession of orbiting black holes is not
straightforward: It requires two black holes with very high masses, $10^8$ and $1.7 \times 10^{10} M_\odot$, and high eccentricity, $e = 0.7$, with a very short lifetime of $10^4$ years (Lehto & Valtonen 1996).

In no other blazar is there convincing evidence for either short- or long-term periodicities, so this is an extremely important precedent. Other periodicities—rotation of the position angle of the optical polarization with an apparent period of 27–35 days (Efimov & Shakhovskoy 1996), smaller fluctuations in intensity with periods of 10–20 min (Carrasco et al 1985, Komesaroff et al 1988, De Diego & Kidger 1990)—have also been reported for OJ 287 but have been (at best) of a transitory nature. We note that the observed rotation of the polarization position angle, even if not strictly periodic, is a remarkable phenomenon, which may be related to a lighthouse effect within a magnetized relativistic jet (Camenzind & Krockenberger 1992).

6.7 Variability of Emission Lines

Compared to quasars, blazars have generally weak emission lines (Miller et al 1978). This is especially true in BL Lac objects, where by definition the lines are practically absent ($W_\lambda < 5 \text{ Å}$). Because the intrinsic line luminosities of BL Lac objects are systematically low compared to FSRQ (Padovani 1992), BL Lacs cannot be quasars with exceptionally beamed continua. Rather, an astrophysical explanation must be found for the intrinsic weakness of their line emission.

The detected lines in BL Lacs are usually of the narrow/forbidden type, which may suggest some fundamental difference between BL Lacs and strong emission line blazars, motivating a separate classification. However, broad $H\alpha$ has been seen in quite a few BL Lac objects (Miller et al 1978, Ulrich 1981, Sitko & Junkkarinen 1985, Moles et al 1987), including BL Lac itself (in which the broad line was clearly absent at earlier epochs; Vermeulen et al 1995, Corbett et al 1996); also, significant variations in the intensity of broad $H\alpha$ or $Ly\alpha$ have been seen (Ulrich 1981, Kidger et al 1996).

These observations demonstrate that the presence or absence of broad lines can be a transitory phenomenon, underscoring the continuity of properties between BL Lac objects and other blazars and supporting a unified view of the blazar phenomenon (Maraschi & Roverini 1994, Bicknell 1994, Sambruna et al 1996a).

7. MULTIWAVELENGTH STUDIES OF BLAZARS

7.1 Broadband Continuum Snapshots

Spectral variability studies make it apparent that blazars are more variable at wavelengths shorter than the peak of the synchrotron emission, with amplitude
increasing with decreasing wavelength (Bregman et al. 1982, 1984, Maraschi et al. 1983, Makino et al. 1987, George et al. 1988b, Falomo et al. 1988, Treves et al. 1989, Kawai et al. 1991, Pian et al. 1994). This is particularly clear for HBL in the X-ray range (Sambruna et al. 1994, 1995), whereas for LBL this trend is observed at IR through UV wavelengths (Impey & Neugebauer 1988).

Recently, multiwavelength snapshots have been extended to gamma-ray energies. For example, the BL Lac object AO 0235+164 has a gamma-ray spectrum that matches well with an extrapolation from the flat X-ray spectrum (Madejski et al. 1996), whereas nonsimultaneous ROSAT observations show that the soft X-ray flux is strongly variable and has a steep spectrum. The simplest interpretation is that in this object the synchrotron emission extends to the soft X-ray band, arguing for a classification intermediate between LBL and HBL.

In a few cases there exist simultaneous multiwavelength spectra of the same source at two epochs, one with high and one with low gamma-ray flux, notably 3C 279 (Maraschi et al. 1994), Mrk 421 (Schubnell et al. 1996; Figure 7), and PKS 0528+134 (Sambruna et al. 1996b; Figure 7). The overall spectral variations of these three blazars show remarkable similarities. In all cases when the source was brighter in gamma rays, it was also brighter at longer wavelengths. The largest variations occurred at wavelengths shorter than the synchrotron and gamma-ray peaks. For Mrk 421 and PKS 0528+134, above gamma-ray peak frequency ($10^{25}$ Hz and $10^{22}$ Hz, respectively), the spectrum was harder in the brighter state. (For 3C 279, the gamma-ray spectrum in the faint state could not be determined owing to the low intensity.) Note that the spectral variation of Mrk 421 in the 0.5- to 10-keV range is analogous to that of 3C 279 in the IR to UV range. (PKS 0528+134 is highly reddened and therefore weak in the optical.)

In sum, the broadband variability behavior of different objects is similar when "normalized" to their respective peak frequencies. In each object, variations of the two spectral components are larger above the peak frequencies and appear to be correlated.

### 7.2 Multiwavelength Light Curves and Correlations

**RADIO–OPTICAL** On time scales of years, the optical emission from blazars is weakly correlated with the radio emission, with lead times of roughly one year (and with large uncertainties; Hufnagel & Bregman 1992). A stronger correlation appears between optical and high frequency (37 GHz) radio light curves, with lead times of months or less (Tornikoski et al. 1994). Whereas the low frequency radio light curves are well sampled with respect to the variability time scales, at high radio and optical frequencies the flares are faster and variations may be simultaneous within the sampling (≤1–2 months). Nevertheless, not every optical flare has a radio counterpart even at high frequency, indicating
that, although the particles radiating in the IR–optical and at submillimeter to centimeter wavelengths are physically related, most likely by a propagating shock, they occupy distinct spatial regions. Simultaneous multiwavelength spectra of blazars from radio to IR bands usually show a self-absorption turnover, in some cases with multiple structure, evolving with time (Robson et al 1983, Gear et al 1986, Valtaoja et al 1988, Brown et al 1989). Systematic long-term monitoring of a number of objects from millimeter to IR bands shows that flares typically propagate from short to long wavelengths (Roellig et al 1986, Stevens et al 1994). A particularly well-studied example is 3C 279, in which the IR spectral index flattened as the intensity increased by a factor of \( \sim 5 \) (Litchfield et al 1995). The delay of the submillimeter flare with respect to the IR appears to be \( \sim 1 \) month, although the sampling is poor on such short time scales. Similar data for 3C 345 can be reconciled with a shock model only if the jet is nonadiabatic or curves away from the line of sight during the decline (Stevens et al 1996).

At millimeter wavelengths, blazar spectra generally flatten with increasing intensity, with BL Lacs (here, LBL) having flatter spectra in the submillimeter range than FSRQ (Gear et al 1994). This may depend on the different line strengths or could be related to the different luminosities, peak wavelengths, and redshifts of the two groups.

On very short (intraday) time scales, radio and optical variability seems to be correlated with no lag, at least for some blazars (Wagner & Witzel 1995).

MILLIMETER–X RAY Some well-monitored blazars, notably 3C 279, have shown a correspondence between millimeter and X-ray emission. The first suggestion of such a correlation was for 3C 279 (eight epochs over 1988–1991, showing two flares; Makino et al 1993). Subsequent observations in 1992–1996 showed another prolonged flare at 22–37 GHz, while the X-ray intensity remained low; however, the 1994 flare had a much lower self-absorption frequency than in 1991 or 1988 (Makino et al 1996). One could speculate that the lack of a corresponding X-ray flare is due to the latest millimeter flare originating in the outer regions of the jet, where the Compton process may be less important.

The X-ray flux of BL Lac also appears to correlate with the submillimeter flux (Kawai et al 1991). Both 3C 279 and BL Lac have similar continuum spectra, with synchrotron peaks in the far-IR and flat X-ray spectra that lie above the extrapolation of the UV spectrum.

A study of the historic light curves of 3C 279 at radio, millimeter, optical, UV, X-ray, and gamma-ray wavelengths (Grandi et al 1994) confirms that on time scales of months to years, the X-rays correlate with the high-frequency radio flux (particularly when a linear trend is subtracted from the radio light curves), whereas the optical–UV fluxes are correlated with the gamma-ray flux.
Regardless of the obvious importance, it has proved very difficult to obtain simultaneous coverage of gamma-ray flares at optical wavelengths. In one case, PKS 1406–076, the optical flux rises by about 60% while the gamma rays increase by a factor of \(\sim 3\), with the optical flare apparently leading the gamma-ray flare by about a day (Wagner et al 1995b). In another case, PKS 0420–014, gamma-ray high states correspond to optical flares, while gamma-ray nondetections coincide with optically faint states (Wagner et al 1995a).

Over long time scales, the optical and gamma-ray fluxes in 3C 279 are well correlated (Grandi et al 1994). During the rapid flare of 3C 279 in June 1991 a definite optical flare was observed, although only three optical measurements are simultaneous to the gamma-ray light curve. These show a maximum coincident with the gamma-ray maximum and a steep decline of 0.36 mag in a day, corresponding to a simultaneous decline of a factor \(\sim 3–5\) in gamma rays (Hartman et al 1996).

**GAMMA RAY–RADIO** Comparison of gamma-ray data for blazars (detections and nondetections) with 37 GHz light curves shows that in a statistical sense, gamma-ray detections correspond to rising radio fluxes (Valtaoja & Teräsranta 1995). Large gamma-ray flares may be connected with the birth of new VLBI components, traditionally associated with the beginning of strong radio flares, as seems to occur in 3C 279 and PKS 0528+134 (Wehrle et al 1996, Zhang et al 1994).

### 7.3 **Intensive Multiwavelength Campaigns**

Most blazars vary substantially on very short time scales (hours to days), at least from the optical through the gamma-ray band. Therefore increasing efforts have been devoted to frequent, and when possible continuous, observations at many wavelengths. This has been possible only for very few objects.

The best candidates for UV–X-ray monitoring are the BL Lac objects PKS 2155–304 and Mrk 421 (HBL). Both can be sampled with \(\sim 1\)-h resolution due to their brightness at those wavelengths. Mrk 421 can also be monitored with daily sampling at TeV energies. In contrast, 3C 279, a highly luminous quasar with an LBL-like continuum, is too faint to monitor rapidly at UV wavelengths but is the brightest blazar at GeV energies and has a long history of multiwavelength observations. Finally, the LBL PKS 0716+71 is a well-monitored intraday variable source. These four blazars, the targets of the most extensive multiwavelength monitoring, are discussed in turn below.

**PKS 2155–304** The first intensive campaign on PKS 2155–304 (November 1991), with \(\sim 5\) days of quasicontinuous coverage with IUE and \(\sim 3.5\) days with ROSAT, yielded a number of unprecedented results (Smith et al 1992,
Variations at optical, UV, and X-ray wavelengths were extremely rapid. The fastest variations observed had $\sim 10-30\%$ changes over several hours, and the autocorrelation function of these fluctuations had a peak at $\sim 0.7$ days. Optical, UV, and X-ray light curves were closely correlated and the amplitude of variation was essentially independent of wavelength, ruling out an accretion disk origin of the UV continuum (as did also the polarization characteristics). The X-rays led the UV by a small but significant amount, roughly 2–3 h. Over the full month, the UV intensity changed by a factor of 2, as seen in the optical/IR light curves.

A second, longer campaign (May 1994)—with $\sim 10$ days of IUE, $\sim 9$ days of EUVE, 2 days of ASCA, and 3 short ROSAT observations—showed at least one prominent isolated flare rather than the “quasiperiodic” low amplitude variations seen in the first campaign. The light curves from this second campaign are shown in Figure 8 (Kii et al. 1997, Pesce et al. 1997, Pian et al. 1997, Urry et al. 1997; HL Marshall and collaborators, in preparation).

![Figure 8](image-url)
The overall variability amplitude was much higher than in the first epoch, particularly in the X-ray band, where the flux rose and fell by a factor of ~2.5 in about a day. The variability was strongly wavelength dependent: The central UV flare, which can most plausibly be identified with the strong X-ray flare, has an amplitude of ~35%, a factor of ~5 smaller than the X-ray flare. (Note however that the wavelength range covered by ROSAT is closer to that of EUVE than to ASCA, and the EUVE variations were only slightly larger than the IUE ones.) The X-ray flare appears to lead the EUV and UV fluxes by 1 and 2 days, respectively, an order of magnitude longer than the lag detected in the first epoch. Within the ASCA data, the 0.5- to 1-keV photons lagged the 2.2- to 10-keV photons by 1.5 h (Makino et al 1996).

Despite the differences, these two campaigns give the first direct evidence that the variations from 10 keV to 5 eV are correlated on short time scales and that high frequencies lead the lower ones.

**MRK 421** The extraordinary high-energy spectrum of Mrk 421 motivated repeated campaigns of simultaneous observations with ASCA, CGRO, and the Whipple Telescope in May 1994 and in May 1995. In both cases, large variations of TeV and X-ray fluxes were observed. The 1995 campaign had better sampling in the TeV range and quasicontinuous coverage with EUVE. The spectral snapshots shown in Figure 7 derive from the 1994 data, and the 1995 light curves obtained with Whipple, ASCA, EUVE, and a ground-based optical telescope are shown in Figure 9 (Buckley et al 1996).

The X-ray and TeV light curves are highly correlated with no apparent lag, while the EUVE and optical light curves lag the TeV (and likely the X-ray) maximum by about a day. This behavior is very reminiscent of the PKS 2155–304 flare, at least in the spectral bands that were covered in both objects. (Unfortunately the southern declination of PKS 2155–304 has prevented observations in the TeV range up to now.) Moreover, as for PKS 2155–304, the 0.5- to 1-keV photons are found to lag the 2- to 7.5-keV photons by about 1 h (Takahashi et al 1996).

**3C 279** 3C 279 was the first blazar discovered to emit strong and variable GeV gamma rays, by EGRET in June 1991. A two-week multiwavelength monitoring campaign followed in December 1992–January 1993. Because 3C 279 was then at a very low intensity level, the scheduled daily sampling could not be performed with IUE and the gamma-ray observations yielded only an average flux; however, several simultaneous gamma-ray, X-ray, UV, optical (BVRI), millimeter, and radio observations were obtained, yielding a well-measured spectral energy distribution of 3C 279 in the low state. Compared to the high state in June 1991, the low-state spectrum decreased dramatically.
Figure 9  TeV, X-ray, and optical light curves of Mrk 421 from April–May 1995. The largest amplitude variability is at TeV and X-ray energies; the optical amplitude is much smaller, but the optical light curve is well correlated with TeV light curve. [From Buckley et al (1996).]

and in a highly correlated fashion at frequencies above $10^{14}$ Hz (Maraschi et al 1994). The variability amplitude increased with frequency from the IR to the UV band, with a steeper spectral shape in the low state. Regarding the second spectral component, the variability amplitude was rather small at the lowest X-ray energies but increased with increasing frequency and was largest in gamma rays.
The optical monitoring during the 1992–1993 campaign evidenced large amplitude variability, almost a factor of 2 in 10 days (Grandi et al 1996), while at the same time the X-ray flux measured with ROSAT varied by less than 20%. This shows that the optical and X-ray emissions were not well correlated on short time scales.

In January 1996, intense gamma-ray flaring in 3C 279 was discovered and followed with multiwavelength coverage. These new data will yield information about the multiwavelength correlations on short time scales in this and, by extension, other LBL. Simultaneous observations with the Infrared Space Observatory (ISO) should determine the position of the synchrotron peak, which from ground-based data can only be constrained to lie between $10^{12}$–$10^{14}$ Hz.

S5 0716+71 After the discovery of intraday variability in S5 0716+71 (see Wagner & Witzel 1995), this BL Lac object was followed intensively at radio and optical wavelengths. IUE and X-ray variations were also seen but with sparser sampling (Wagner et al 1996). There is a definite trend for bluer colors during short flares, whereas the long-term variations occur at constant color (Ghisellini et al 1997). The spectrum in the radio band flattens when the optical flux brightens, indicating that the flaring component is self-absorbed. The UV and X-ray fluxes both seem to vary with the optical flux on short time scales (hours), but the optical to X-ray ratio is not constant from day to day.

The gamma-ray emission from S5 0716+71 is rather weak and shows a relatively flat spectrum, and it varies by factors of $\sim 2$ over year-long time scales (Lin et al 1995). The broadband spectral properties of this object suggest that it is intermediate between LBL and HBL.

8. INTERPRETATION OF BLAZAR VARIABILITY

8.1 Summary of Variability Results

Recent observations of blazars have yielded six key points critical for understanding the continuum. First, blazar spectra are characterized by two broad spectral components (Section 6.2, Figures 6 and 7), one with peak power at low frequencies (IR to soft X rays) and one with peak power at very high energies (GeV to TeV gamma rays).

Second, variability in the low frequency component is much more pronounced above the peak frequency, and the variability amplitude increases with frequency so that the spectra harden with increasing intensity. Below the peak, both the amplitude of variability and spectral changes are much smaller.

Third, the high-energy component also seems to vary more and to harden with increasing intensity above its peak frequency. This inference is, however, based on very few observations.
Fourth, simultaneous snapshots of the full broadband spectra suggest that the intensities in the two spectral components are correlated, in the sense that when the short-wavelength (gamma-ray) component is bright, the long-wavelength emission is also bright. These snapshots do not constrain possible lags significantly.

Fifth, finite lags have been measured among flares in the optical through X-ray light curves of two HBL. For both PKS 2155–304 and Mrk 421, the soft X-ray photons lag the hard X-ray photons by \( \sim 1 \) h (Makino et al 1996, Takahashi et al 1996). In addition, for PKS 2155–304 the EUV and UV light curves lag the X-ray curves by \( \sim 1 \) and \( \sim 2 \) days, respectively (Urry et al 1997; Figure 8). For Mrk 421, the EUV and optical light curves lag the TeV by \( \sim 1 \) day (Schubnell et al 1996; Figure 9); a big gap in the X-ray light curve close to the flare peak prevents quantitative statements, but it is plausible that the behavior of Mrk 421 closely resembles that of PKS 2155–304.

Sixth, HBL and LBL exhibit completely analogous variability with respect to the (different) peak frequencies in their spectral distributions. For example, both vary more above their respective power peaks; it is just that this makes LBL highly variable in the optical and GeV gamma rays, whereas HBL are highly variable in X rays and in TeV gamma rays (and relatively quiescent in the optical and GeV ranges).

Below we discuss current theories of blazar spectra and variability (Sections 8.2–8.5) in the light of these six points. We start with the radio-emitting outer jet, which is reasonably well understood, and work progressively inward to smaller scales, higher energies, and lesser knowledge, with the goal of relating the observed variability to the physics of the jet. Mechanisms for variability involve essentially two possibilities, shock waves moving along the jet and rotation of the beaming cone across the line of sight; both could be relevant at different times or in different objects. In the first case, spectral variations are expected, whereas in the second achromatic variability can be produced (see Marscher 1996 and Dreissigacker & Camenzind 1996 for physical models).

8.2 The Relativistic Jet: Synchrotron Radiation from the Outer Regions

Self absorption is important in blazars even at high radio frequencies. The flat radio spectra likely derive from inhomogeneity along the jet—that is, from contiguous jet regions with different self-absorption cutoffs (Blandford & Königl 1979, Marscher 1980, Königl 1981). Often, discrete spectral components can be recognized, sometimes related to individual VLBI components or to the unresolved core. The spectral evolution of these components—in particular, the shift of the self-absorption cutoff to longer wavelengths and concurrent changes in polarization angle—can be well modeled as enhancements in the local radio emission due to shocks propagating along a relativistic jet whose
particle density and magnetic field decrease outward (Marscher & Gear 1985, Hughes et al 1989, 1994, Litchfield et al 1995, Stevens et al 1996). The ability of such models to account at the same time for both total flux and polarization variability is encouraging, although the derived angle of view tends to be larger than for simple unification schemes (Kollgaard 1994, Urry & Padovani 1995).

The high polarization of the optical emission and the spectral continuity from the radio (albeit with increasing spectral index) imply that synchrotron radiation is the dominant emission mechanism in most blazars up to the UV range and in some cases up to the X rays. It is not yet clear, however, whether the relativistic electrons responsible for the high-frequency synchrotron emission are simply the high energy tail of those emitting at radio wavelengths or represent a distinct population in a different part of the jet, presumably closer to the central engine. Radio through optical light curves show some correlation, but the lags have not been quantified even for the best-observed sources (Stevens et al 1994, Tornikoski et al 1994). Uncorrelated flares are also present (Hufnagel & Bregman 1992). An understanding of the radio-optical connection is also hampered by the far-IR observational gap (now accessible with ISO). The self-absorption turnover of the optical synchrotron component cannot be determined unambiguously, and therefore its physical parameters are poorly constrained.

8.3 High-Frequency Synchrotron Emission: Energy Stratification?

At high frequencies (optical–UV for LBL, X-ray for HBL), the synchrotron spectrum steepens considerably and the largest spectral variability is observed. It is natural to associate the steepening with increasing radiative energy losses. Particle acceleration must also occur or the high-energy emission would vanish in a short time. Therefore the average spectrum probably represents an equilibrium among energy gains, losses, and escape. If the “injected” particle spectrum is a power law, its equilibrium spectrum (in a homogeneous region) is steeper by $\Delta \beta = 1$ above the energy $\gamma_b$ at which the radiative lifetime equals the escape time (Kardashev 1962). The peak in the synchrotron power would occur at $\nu_b \propto \gamma_b^2$, and the associated change in spectral slope would be $\Delta \alpha = 0.5$.

Spectral variability above $\nu_b$ (flattening with increasing intensity) could be due to fluctuations in the acceleration process on time scales shorter than those over which equilibrium can be established; the flatter injected spectrum would be visible briefly before energy losses set in. Another possibility is that increased power dissipation causes the injection of particles of higher energy. The second hypothesis is favored in model fits (assuming the simple homogeneous case) to the spectral variability of Mrk 421 (Mastichiadis & Kirk 1997). In both cases, the variations at the highest synchrotron and inverse Compton
frequencies are quasisimultaneous, whereas at lower frequencies the time scales are longer and the peaks are delayed.

Acceleration and loss processes can depend on position in the jet, causing different electron energy distributions as a function of location (inhomogeneous model). The steeper spectrum above the peak can be reproduced (even if higher energy electrons occupy progressively smaller volumes (Ghisellini et al. 1985)). This occurs naturally behind a shock where the acceleration process is localized at the shock front (Marscher 1996). Thus, a disturbance propagating along an inhomogeneous jet or across a shock front would cause spectral variability, usually with shorter time scales and larger amplitudes at higher energies (Celotti et al. 1991, Marscher & Travis 1991).

Indeed, the soft photons do lag the hard photons in both Mrk 421 and PKS 2155–304 (the two brightest HBL). In X rays, the frequency dependence of the measured lags is consistent with a $\nu^{-1/2}$ law, as expected from a radiative lifetime; the EUV and UV could also be consistent within the large uncertainties. At each frequency the decay times are longer than the lags, which is not easy to understand in a pure radiative case. A realistic model likely involves both radiative and travel time effects.

In any case, the data require a fast acceleration process injecting high-energy particles. This is not easy to reconcile with diffusive shock acceleration, whereby high energies build up through a cumulative process (see spectral evolution computed for time-dependent shock acceleration, Fritz & Webb 1990). A viable alternative could be particle acceleration through a large scale electric field (Kirk & Bednarek 1996). Note that the achromatic variability found in the first epoch of multiwavelength monitoring of PKS 2155–304 is not expected from a synchrotron flare in a jet and may be caused by a different physical process such as rotation or microlensing.

8.4 Inverse Compton Models: The Gamma-Ray Jet

Understanding the gamma-ray emission in blazars is particularly important because, at least during flare states, it dominates the bolometric luminosity. Here we discuss models that attribute the low-energy component to synchrotron emission and the high-energy component to inverse Compton emission from the same population of relativistic electrons.

For a homogeneous emission region, neglecting Klein-Nishina effects and assuming the seed photon spectrum is not too broad, the Compton emission closely mimics the synchrotron spectral shape—both depend primarily and in the same way on the electron spectrum. To first order, the peak of the Compton spectrum is due to the same electrons (of energy $\gamma_e$) as the synchrotron peak, independent of the seed photons. Pairs of wavelengths in the two spectral components in the same ratio as the peak wavelengths are also produced by the
same electrons. Thus, if variations are due to changes in the electron spectrum, this class of models predicts a close correlation between flux variations for each wavelength pair. The coherent variability of the synchrotron and Compton components of LBL and HBL summarized at the beginning of the present section supports this simple scheme in broad terms. More realistic models should take into account the effects of inhomogeneity in the jet.

The seed photons could come from various sources. Synchrotron photons are produced copiously within the jet (Maraschi et al. 1992) but we measure only their apparent density. The greater the beaming (bulk Lorentz factor \( \Gamma \)), the lower the intrinsic photon energy density in the jet frame (\( \propto \Gamma^{-4} \)). If an accretion disk is present, it is a strong source of photons (Dermer & Schlickeiser 1993), but at the distance implied by the gamma-ray transparency condition, their intensity as seen in the jet frame is greatly reduced. Alternatively, photons produced at the disk or nucleus can be reprocessed and/or scattered and therefore isotropized in a region of appropriate scale (Sikora et al. 1994); their energy density is then amplified in the jet frame by a factor \( \Gamma^2 \).

The relative importance of the possible sources of soft photons—the jet itself, the accretion disk, or the broad emission line region, illuminated either by the disk or by the jet—must depend on the particular characteristics of the source. In different blazars, the origin of the seed photons may well be different: Perhaps the external-Compton (EC) case is more important in strong emission-line objects like 3C 279 and PKS 0528+134 (Sambruna et al. 1996b), whereas the SSC model is applicable to weak-lined blazars like Mrk 421 and PKS 2155–304 (Zdziarski & Krolik 1993, Ghisellini & Madau 1996, Ghisellini & Maraschi 1996).

In a homogeneous SSC model, the ratio between the two peak wavelengths uniquely determines the energy of the radiating electrons: \( \lambda_S/\lambda_C \sim \gamma^2 \). For EC models, the relation is less straightforward and obviously leads to different physical parameters for the emitting region. For 3C 279, homogeneous SSC models yield lower magnetic fields than EC models and have difficulties accounting for the shortest variability time scales (Ghisellini & Maraschi 1996). Both models are consistent with the (presently) observed spectral variability; the larger amplitude in the Compton component compared to the synchrotron component is explained naturally by the SSC model (Maraschi et al. 1994) but is also consistent with an EC scenario where nonlinearity can be caused by a variation of the bulk flow speed or by an external mirror effect for the synchrotron photons (Ghisellini & Madau 1996).

8.5 The Invisible Jet Core

The EGRET detections of blazars demonstrate that the observed power is in many cases dominated by high-energy gamma rays. Although the degree of
dominance could result in part from selection effects—i.e. if there is a scatter, possibly due to variability, in the intrinsic gamma-ray to radio flux ratio of the population, objects with the highest ratio will be "selected" by gamma-ray observations (Impey 1996)—it is clear that gamma rays are a fundamental component of blazar power.

Paradoxically, rapid variability at gamma-ray energies implies that the observed radiation cannot be produced too near the center or the high-energy photons would never escape (Section 6.5). This raises the question of how power is transported from the central engine, and how and why it is radiated at a given distance. In FSRQ the gamma-ray variability time scale is $\sim 1$ day, similar to that in the optical for some well-studied objects, so that the minimum radiative zone occurs near $\sim 10^{16} (\delta/10)$ cm. At roughly this scale, a magnetohydrodynamic wind from a rotating disk around a black hole would be collimated and accelerated to relativistic speed (Appl & Camenzind 1993); the energy transport would be magnetohydrodynamic and the radiative region would start "naturally" where a high bulk velocity is achieved.

Alternatively, the energy transport could be either purely electromagnetic, via a Poynting flux or through mildly relativistic ions with low internal entropy (Blandford 1993, Blandford & Levinson 1995). Assuming the emitting region is initially opaque to gamma rays, there will be a "gamma-ray photosphere" (due to interactions with lower energy photons) that is larger for higher energy gamma rays. The concept of a photosphere makes the emergent spectra independent of the details of the energy transport and conversion, and also allows a very strong prediction that gamma-ray flux variations are slower and/or later at higher energies. This prediction is violated for Mrk 421, where TeV gamma rays vary more rapidly and more frequently than GeV gamma rays. That the gamma-ray spectra of FSRQ are harder in brighter states also appears difficult to reconcile with this model.

Another way of transporting energy without dissipation is highly relativistic protons ($\gamma_p \sim 10^7$) (Mannheim & Biermann 1992, Mannheim 1993, 1996, Protheroe & Biermann 1997). The associated primary electrons, whose break energy is much smaller than that of the protons, produce the low frequency spectral component via the synchrotron mechanism. The very high-energy protons interact with these photons, initiating pair cascades that yield very flat spectra ($\alpha \sim 1$) in the MeV–TeV range. This model predicts that high-energy variations should always follow low energy ones.

In an intermediate scenario, the jet starts out at high bulk Lorentz factor, $\Gamma_j \sim 10^4$, then decelerates through interactions with the local (disk) radiation field (Coppi et al 1993). This model has a difficulty in that the energy lost in the deceleration phase—i.e. the observed radiation—is much larger than that
Variability of AGN 493

carried by the decelerated jet, which may therefore be insufficient to power the radio lobes.

All these models have difficulties, and in particular none of them yields a good fit to the overall spectral energy distribution. Nonetheless, they are of interest because they attempt to address crucial points about the origin of the radiating particles, which remain unexplained in the more phenomenological models.

9. CONCLUSIONS AND FUTURE INVESTIGATIONS

Fundamental results have been obtained from variability studies of AGN, but many important questions are still unanswered. Substantial progress is expected in the near future.

Regarding the radio-quiet AGN, the simultaneity of the continuum flux variations at optical and UV wavelengths and the very short doubling time scales (one week to several months) observed in low-luminosity AGN are incompatible with models wherein the rapidly variable component of the optical–UV continuum is emitted through viscous effects in an accretion disk. Because the amplitude of the variations on time scales of weeks is more than a factor of 2 at 1400 Å, at least half of the blue bump luminosity is involved. The best current model for the rapidly variable continuum emission is reprocessed emission from a dense medium irradiated by the variable X rays.

For high-luminosity AGN, the importance of this irradiation process is uncertain because the relevant observations have not yet been carried out. The most interesting question—whether and how the continuum fraction emitted through reprocessing varies with absolute luminosity (accretion rate) and/or with the mass of the central black hole—remains unanswered. The organization of a series of quasisimultaneous optical–UV–X-ray observations is therefore a priority in order to advance in this field.

The rather good knowledge of the variability characteristics that we now have, at least for low-luminosity AGN, should stimulate new theoretical investigations on the origin of the continuum variability, both time scales and amplitudes. While turbulence and magnetic reconnection may be a sufficient explanation for the short time-scale X-ray variability, the mechanisms for longer term variations are essentially unknown.

Results crucial for our understanding of the inner region of radio-quiet AGN have been obtained from variability studies of the broad-line gas emission, including (a) a measurement of the dimension of the emitting region (dimension that depends on the line and on the velocity); (b) indications that the gas has roughly virialized motions in the gravity field of the central black hole, whose
mass can then be derived from the observed motions; and (c) support for the accretion disk plus wind working model wherein the LIL come from the accretion disk, and the HIL are emitted by a wind (which could be magnetically or radiatively driven from the disk).

Among future investigations, we see three priorities. The first is to confirm (or refute) the accretion disk plus wind model, which at present seems to be the most promising model, by collecting extensive data sets combining high S/N and large number of epochs. Second, and possibly even more important, is to verify if the models developed for low-luminosity AGN are valid for high-luminosity AGN, the brightest objects in the universe. This can be done only if we collect data on high-luminosity AGN that are of a quality similar to the low-luminosity AGN data. This requires more powerful instruments and a consistent effort over much longer periods than for low-luminosity AGN. The third priority is to explore the entire parameter space defined by mass and mass accretion rate. This requires monitoring the continuum and line variations in very low-luminosity AGN (low accretion but unknown mass), such as the LINERS and the galaxy centers displaying faint UV sources or very weak broad lines. Similar observations should be carried out on NLS1, which might be the AGN with the highest accretion rate.

Closely related to these questions are the recent observations of broad, variable Fe K lines (Tanaka et al 1995, Yaqoob et al 1996) and of absorption features originating in a warm absorber (Iwazawa et al 1997). All these observations will add impetus to the active field of investigations of the structure and stability of AGN accretion disks, the role of the magnetic fields, and the magnetically or radiatively accelerated outflows.

As for the strongly nonthermal activity of blazars, it is now reasonably well understood in phenomenological terms. The basic model consists of a relativistic jet closely aligned with the line of sight, filled with energetic particles that emit synchrotron and Compton-scattered radiation spanning radio through gamma-ray wavelengths. The radiating particles must have a bulk relativistic motion, beaming the radiation toward us and making the blazar appear more luminous and more rapidly variable. Beyond this, models differ, notably in the structure and physical parameters of the jet and in the nature of the seed photons to be upscattered to the gamma-ray range.

The detection of short lags in the UV to X-ray range in Mrk 421 and PKS 2155–304 indicates that, in HBL at least, energy injection in the jet occurs in a top down scenario. Because the spectral variability of LBL objects in the IR to UV range appears similar to that of Mrk 421 and PKS 2155–304 in X rays, it is important to look for similar lags in LBL at IR to UV wavelengths. In principle, this could be done with intensive ground-based monitoring but would be much
more likely to succeed with space instrumentation (HST and ISO) owing to the more extended wavelength range accessible and to the quasicontinuous coverage and even sampling. Comparison of flare evolution in wavelength and time in these very different sources, HBL and LBL, will give important clues to the differences in their jets.

Determining the nature of the seed photons (possibly different in different objects) is important because it will allow better determination of the physical conditions in the jet and eventually understanding of whether and how the environment influences the jet properties. Because model predictions for correlated variability between the synchrotron and Compton emission differ, multiwavelength campaigns offer the best means to discriminate. Substantial progress is expected over the next few years. Repeated multiwavelength campaigns on the brightest sources should enable discrimination among different sources of seed photons, especially with studies of multiwavelength variability on short time scales, during which the external photon field should not change substantially.

It is unfortunate that the available time for these programs is limited by the residual amount of gas for the EGRET experiment. One of the critical difficulties is that ground-based observations in the optical (as IUE is not available any longer) cannot be arranged quickly enough after the discovery of inherently unpredictable gamma-ray flares, given their rapid intraday time scales. In order to measure cross correlations and possible lags between gamma-ray and optical flares, optical-UV monitors on board the next generation gamma-ray experiments are needed.

Jet models imply that energy is transported in a dissipationless way from very small scales, within a few Schwarzschild radii of the central black hole, to relatively large distances along the jet, a light day or more. At present we have no clear picture of how this happens. We can infer from X-ray and gamma-ray observations that the innermost region is unlikely to be filled with photons, as the resulting pair cascades would redistribute energy from gamma-ray to X-ray energies. Further work in this area is critical, as it holds the key to understanding how energy is extracted from the central black hole, the fundamental issue in the study of AGN.

ACKNOWLEDGMENTS

It is a pleasure to thank our colleagues for comments and discussions at various stages of the preparation of this review, in particular R Blandford, S Collin-Souffrin, T Courvoisier, G Ghisellini, K Horne, B Hufnagel, L Lucy, G Madejski, S Phinney, E Pian, P Schneider, A Treves, R Terlevitch, and S Wagner.
Literature Cited


Heidt J. 1996. See Miller et al 1996, pp. 64–69


VARIABILITY OF AGN

Mattox JR. 1995. IAU Circ. No. 6181
Sambunam RM, Urry CM, Maraschi L,
VARIABILITY OF AGN

502  ULRICH ET AL