Lecture Notes in Physics

The Lecture Notes in Physics

The series Lecture Notes in Physics (LNP), founded in 1969, reports new developments in physics research and teaching – quickly and informally, but with a high quality and the explicit aim to summarize and communicate current knowledge in an accessible way. Books published in this series are conceived as bridging material between advanced graduate textbooks and the forefront of research to serve the following purposes:

- to be a compact and modern up-to-date source of reference on a well-defined topic;
- to serve as an accessible introduction to the field to postgraduate students and nonspecialist researchers from related areas;
- to be a source of advanced teaching material for specialized seminars, courses and schools.

Both monographs and multi-author volumes will be considered for publication. Edited volumes should, however, consist of a very limited number of contributions only. Proceedings will not be considered for LNP.

Volumes published in LNP are disseminated both in print and in electronic formats, the electronic archive is available at springerlink.com. The series content is indexed, abstracted and referenced by many abstracting and information services, bibliographic networks, subscription agencies, library networks, and consortia.

Proposals should be sent to a member of the Editorial Board, or directly to the managing editor at Springer:

Dr. Christian Caron
Springer Heidelberg
Physics Editorial Department I
Tiergartenstrasse 17
69121 Heidelberg/Germany
christian.caron@springer.com

Danielle Alloin Rachel Johnson
Paulina Lira (Eds.)

Physics of Active Galactic Nuclei at all Scales

Springer

Editors

Danielle Alloin
CEA
Saclay-Orme des Merisiers
91191 Gif-sur-Yvette Cedex
France
E-mail: danielle.alloin@cea.fr

Paulina Lira
University of Chile
Department of Astronomy
Casilla 36D
Santiago, Chile
E-mail: plira@das.uchile.cl

Rachel Johnson
Oxford University
Department of Physics
Keble Road
Oxford, OX1 3RH, U.K.
E-mail: raj@astro.ox.ac.uk

D. Alloin et al., *Physics of Active Galactic Nuclei at all Scales*, Lect. Notes Phys. 693 (Springer, Berlin Heidelberg 2006), DOI 10.1007/b11550280

Library of Congress Control Number: 2006925838

ISSN 0075-8450
ISBN-10 3-540-31207-2 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-31207-9 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springer.com

Printed in The Netherlands

Typesetting: by the authors and techbooks using a Springer LaTeX macro package
Cover design: *design & production* GmbH, Heidelberg

Printed on acid-free paper SPIN: 11550280 55/techbooks 5 4 3 2 1 0

Preface

The study of the physical processes at work in Active Galactic Nuclei (AGN) has kept a large number of astronomers busy since the discovery, in the early sixties, of the first radio galaxies and quasars. After a few decades of debate there is now a clear consensus about the source of energy in AGN, namely gravitational energy released through matter accretion onto a massive black hole.

The most commonly accepted ideas regarding AGN are pictured in the so-called AGN unified model, although one should remember that this is still a "working" model. Tremendous progress has been made in unveiling, analyzing, and modeling the different components in AGN: the accretion disk, the jets of relativistic particles, the x-ray absorber close to the central engine, the "torus" that funnels the ionizing radiation, the surrounding clouds of dense material in the broadline region and narrowline region, the cool molecular and dusty material around the torus, as well as jet-induced effects on larger scales. Yet, it remains to incorporate all these components and physical processes into a self-consistent AGN model that would take into account, and predict, all the facets of AGN activity. The mere process of matter fueling is far from being totally understood and neither is the link of the AGN with its host galaxy. The birth and growth of massive black holes in the universe is a related question to be investigated and is of prime interest for cosmology. A wealth of new observational constraints about the physical processes taking place in AGN has recently been made available, thanks to ongoing key space missions (probing in particular their x-ray emission), large ground-based telescopes, and high spatial resolution techniques such as adaptive optics and interferometry. Therefore, it is both exciting and timely to see how these additional constraints fit in with model predictions and to explore the new perspectives they bring to the field.

In order to contribute to this goal, we organized the International Workshop, "Physics of Active Galactic Nuclei at all Scales", which was held at the European Southern Observatory in Santiago, Chile, from December 3 to 6, 2003. The meeting was sponsored jointly by the Conicyt/FONDAP Center for Astrophysics, the European Southern Observatory, and the Calan Observatory of Universidad de Chile. The Workshop was intended for training young researchers and students: it included a series of invited reviews, each followed by oral contributions of topical relevance, and was completed by a rich display of posters. Although the emphasis was on providing a solid theoretical base, appropriate results from recent observations across the electromagnetic spectrum were discussed, which made this Workshop extremely enlightening. The reviews have been prepared for publication and constitute the bulk of this volume. We are grateful to the reviewers for their generous and enthusiastic contributions.

The Workshop was opened with a tutorial by Hagai Netzer, providing a comprehensive picture of AGN physics. He pointed out that in spite of the remarkable insights into the AGN phenomenon at present, some key questions are still open. One of them is the "energy budget problem", which is the large discrepancy between the energy output required from the observed emission line spectrum and that extrapolated from the observed continuum energy distribution. Another question refers to the detailed understanding of the physical processes hidden within the so-called alpha parameter used to parameterize the viscosity in the accretion disk.

Following the tutorial, and in the framework of the AGN unified model, subsequent reviews and discussions about AGN subsystems were given by specialists in the field. Diana Worrall and Bradley Peterson focused on the physics of material in the close vicinity of the central source and on related radiative processes, discussing relativistic jets and the broadline region. Moving to larger scales, Bob Fosbury, Jack Gallimore, and Moshe Elitzur reviewed the physics of gas further away but still exposed to the intense ultraviolet radiation field from the AGN, as well as of molecular and dusty material surviving such a harsh environment.

Questions about the AGN fueling and the AGN relationship with its host galaxy were discussed by Sharda Jogee. Very recent results from space x-ray missions are shedding new light on the evolution of the AGN luminosity function with redshift and on the AGN-galaxy formation scenario: these were presented by Niel Brandt and Omar Almaini.

Thorough reviews on the forefront instrumentation available today in Chile-based observatories were given by Chris Lidman and Malcolm Smith, which, however, are not provided in this volume.

The workshop was attended by around seventy participants, with a remarkable delegation from South America. It provided the ideal format and size to boost exchanges and discussions and in this, it met the original goals of the organizers, Danielle Alloin, Poshak Gandhi, and Rachel Johnson at ESO; and Paulina Lira, Sebastian Lopez, and Jose Maza at Calan Observatory.

We gratefully acknowledge the valuable support from Ms. Pamela Bristow at ESO/Garching in the editing work of this volume.

Garching and Santiago
July 2005

Danielle Alloin
Rachel Johnson
Paulina Lira

Contents

Active Galactic Nuclei: Basic Physics and Main Components
H. Netzer 1
1 Terminology, Classification and Abbreviations 1
2 Black Holes 1
3 Accretion Disks 4
4 Physical Processes in AGN Gas 12
5 Main Components of AGN 25
6 Summary 37
References 37

Multiwavelength Evidence of the Physical Processes in Radio Jets
D.M. Worrall and M. Birkinshaw 39
1 Introduction 39
2 Radiative Processes 40
3 Effects of Bulk Relativistic Motion 46
4 The External Medium 48
5 Simple Radio-Jet Models 52
6 The Interpretation of Multiwavelength Data in Terms of Physical Parameters 56
7 Conclusion 71
References 71

The Broad-Line Region in Active Galactic Nuclei
B.M. Peterson 77
1 Introduction 77
2 Basic Inferences about the Line-Emitting Gas 78
3 Broad Emission-Line Variability and Theory of Reverberation Mapping 81
4 Reverberation Results 86
5 What is the BLR? 91
6 What will it Take to Map the Broad-Line Region? 95
7 Some Obvious Questions to Keep in Mind 97
8 Summary 97
References 99

Cold Gas Near Active Galactic Nuclei
J.F. Gallimore, M. Elitzur, S.A. Baum 101
1 Background 101
2 Radio Observations 104
3 Infrared SEDs and Modeling 113
4 Future Observations 117
References 118

AGN Beyond the 100pc Scale
R.A.E. Fosbury 121
1 Introduction 121
2 AGN–host Interactions 122
3 Observational Techniques and Examples 126
4 What Next? 139
5 Further Reading 140
References 141

The Fueling and Evolution of AGN: Internal and External Triggers
S. Jogee 143
1 Introduction 143
2 BH Demographics and BH-Bulge-Halo Correlations 144
3 Central Issues in Fueling AGN and Starbursts 147
4 Hubble Type and Colors of AGN Hosts 154
5 Interactions and AGN/Starburst Activity 156
6 Large-Scale Bars in Starbursts/AGN Hosts 160
7 Nuclear Bars 169
8 Nuclear Spirals and AGN/Starburst Activity 171
9 From Hundred pc to Sub-pc Scales 174
10 Summary and Future Perspectives 175
References 179

X-Ray Survey Results on Active Galaxy Physics and Evolution
W.N. Brandt, D.M. Alexander, F.E. Bauer, C. Vignali 185
1 Introduction 185
2 Chandra and XMM-Newton Extragalactic Surveys 186
3 Some Implications for AGN Physics and Evolution 191
4 Some Future Prospects 203
Some Recent Key References 207

The Links Between AGN and Galaxy Formation
O. Almaini 211
1 Supermassive Black Holes in Local Galaxies 211
2 Models for Galaxy Formation 214
3 Observations of the High-Redshift Universe 216
4 The Impact of AGN on their Environments 220
5 Models for the Black-Hole/Bulge Relation 223
6 Summary and Future Prospects 225
References 227

Appendix: Acronyms Used in this Book 229

Index 231

List of Contributors

David M. Alexander
Institute of Astronomy
Madingley Road
Cambridge CB3 0HA, UK
`dma@ast.cam.ac.uk`

Omar Almaini
School of Physics and Astronomy
University of Nottingham
University Park
Nottingham NG7 2RD, UK
`Omar.Almaini@nottingham.ac.uk`

Franz E. Bauer
Institute of Astronomy
Madingley Road
Cambridge CB3 0HA, UK
`feb@ast.cam.ac.uk`

Stefi A. Baum
Center for Imaging Science
Rochester Institute of Technology
54 Lomb Memorial Drive
Rochester, NY 14623, USA
`baum@cis.rit.edu`

Mark Birkinshaw
Department of Physics
Royal Fort, Bristol University
Tyndall Avenue
Bristol, BS8 1TL, UK
`mark.birkinshaw@bristol.ac.uk`

W. Niel Brandt
Department of Astronomy & Astrophysics
The Pennsylvania State University
525 Davey Lab
University Park, PA 16802, USA
`niel@astro.psu.edu`

Moshe Elitzur
Department of Physics and Astronomy
University of Kentucky
600 Rose Street
Lexington, KY 40506-0055
USA
`moshe@pa.uky.edu`

Robert Fosbury
Space Telescope –
European Coordinating Facility
Karl Schwarzschild Strasse 2
85748 Garching, Germany
`rfosbury@stecf.org`

Jack Gallimore
Department of Physics and Astronomy
Bucknell University
Moore Avenue
Lewisburg, PA 17837, USA
`gallim@stsci.edu`

Sharda Jogee
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218, USA
`jogee@stsci.edu`

Hagai Netzer
Wise Observatory
Tel Aviv University
69978 Tel Aviv, Israel
`netzer@wise.tau.ac.il`

Bradley Peterson
Department of Astronomy
The Ohio State University
140W 18th Avenue
Columbus, OH 43210-1173, USA
`peterson@astronomy.ohio-state.edu`

Cristian Vignali
INAF – Osservatorio Astronomico di Bologna
Via Ranzani 1
40127 Bologna, Italy
`cristian@anastasia.bo.astro.it`

Diana Worrall
Department of Physics
Royal Fort, Bristol University
Tyndall Avenue
Bristol, BS8 1TL, UK
`d.worrall@bristol.ac.uk`

Active Galactic Nuclei: Basic Physics and Main Components

H. Netzer

School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel
`netzer@wise.tau.ac.il`

Abstract. This review gives a brief introduction to the physics and the structure of active galactic nuclei (AGN). The main topics that are being discussed are general accretion, the structure and the spectrum of various types of accretion disks, photoionization and recombination, energy balance and gas temperature, the spectrum of ionized gas including the effect of dust, and the dynamics of ionized gas in a strong radiation field. These principles are then applied to various AGN components including the broad and the narrow line regions, the ionized X-ray gas, the torus and the central disk.

1 Terminology, Classification and Abbreviations

The study of active galactic nuclei (AGN) has produced many new terms and abbreviations. Only the most common ones will be used in this review, e.g., the narrow line region (NLR), the broad line region (BLR), the highly ionized gas (HIG) and the black hole (BH). I shall avoid, as much as possible, using names such as Seyfert 1s, Seyfert 2s and LINERs. Instead I shall refer to type-I AGN (those objects showing broad, strong optical-UV emission lines in their spectrum), and type-II AGN (those showing prominent narrow emission lines, very faint, if any, broad lines, and a large X-ray obscuring column).

2 Black Holes

Black holes (BHs) of all sizes are very common in the universe. This has been known for many years. However, the understanding that most galaxies, especially early type galaxies, contain massive BHs is more recent. New studies compare BH mass and evolution in high redshift quasars to the evolution of early galaxies. It seems that massive BHs, as large as $10^9\ M_{\odot}$, have already been present at $z \simeq 6$ [1]. Some of these sources seem to accrete matter at a very high rate. This raises important questions about the growth rate and the formation mechanism of those BHs. Any BH and accretion disk model, like the one presented below, must be compatible with such observation.

The basic properties of BHs can be expressed using their gravitational (r_g) and Schwarzschild (r_s) radii,

H. Netzer: *Active Galactic Nuclei: Basic Physics and Main Components*, Lect. Notes Phys. **693**, 1–38 (2006)
DOI 10.1007/3-540-34621-X_1

$$r_g = \frac{GM}{c^2} \ , \ r_s = 2r_g \ . \tag{1}$$

We can also define the BH angular momentum s

$$s \sim I\Omega \simeq Mr_g^2 \frac{v}{r} \simeq Mr_g c \ , \tag{2}$$

where Ω is the angular velocity, and the BH specific angular momentum (angular momentum per unit mass),

$$s/M \equiv ac \sim r_g c \sim 5 \times 10^{23} M_8 \ \mathrm{cm^2 s^{-1}} \ , \tag{3}$$

where M_8 is the mass in units of $10^8 M_\odot$. Considering massive BHs in galactic nuclei, we note that the typical specific angular momentum on galactic scales (e.g. r=1 kpc, v = 300 km s^{-1}) is many orders of magnitude larger than ac in (3). Thus, for BHs to grow substantially over time scales typical of galactic evolution, there must be an efficient mechanism to get rid of the excess angular momentum of the accreted gas (in order to bring it to the center). On large scales, BH growth depends on mechanisms such as galaxy collision and merging and bar instability that are capable of bringing far away gas into the vicinity of the BH.

Gas particles very close to the center, at $r_g \sim 1000$, are subjected to different types of forces and are not necessarily accreted onto the BH. The understanding of their motion, temperature and emission requires a discussion of those mechanisms including the balance between gravity and radiation pressure force, the general process of spherical accretion and the properties of thin and thick accretion disks.

2.1 The Eddington Luminosity

Assume a central source of mass M emitting monochromatic luminosity L_ν. The radiation pressure force acting on gas particles at a distance r is given by

$$f_{rad} = \frac{N_e \sigma_T}{4\pi r^2 c} \int_0^\infty L_\nu d\nu = \frac{N_e L \sigma_T}{4\pi r^2 c} L \ , \tag{4}$$

where L is the total luminosity, N_e is the electron density and σ_T the Thomson cross section. The gravitational force is

$$f_g = \frac{GM\mu m_p N_e}{r^2} \tag{5}$$

where μ is the mean molecular weight (mean number of protons and neutrons per electron; about 1.17 for a fully ionized solar composition gas). Spherical accretion of fully ionized gas can proceed as long as $f_g > f_{rad}$. The requirement $f_{rad} = f_g$ leads to the so-called "Eddington luminosity",

$$L_{Edd} = \frac{4\pi c G M \mu m_p}{\sigma_T} \simeq 1.5 \times 10^{38} (M/M_\odot) \ \mathrm{erg\,s^{-1}} \ , \tag{6}$$

which is the maximum luminosity allowed for objects that are powered by steady state accretion. This definition of L_{Edd} takes into account only one source of opacity, Compton scattering. This is the case for fully ionized plasma. More realistic situations may involve partly neutral gas and hence much larger opacity. Thus, the effective L_{Edd} can be significantly smaller than the value defined in (6).

Given the definition of L_{Edd}, and the accretion rate $\dot{M} = L/\eta c^2$, where η is the efficiency of converting gravitational potential energy to electromagnetic radiation, we can define several other useful quantities. Thus, the "Eddington accretion rate", $\dot{M}_{Edd}$, is the accretion rate required to produce L_{Edd},

$$\dot{M}_{Edd} = \frac{L_{Edd}}{\eta c^2} \simeq 3M_8 \left[\frac{\eta}{0.1}\right]^{-1} M_{\odot}\, y^{-1} \,, \tag{7}$$

The "Eddington time", t_{Edd}, is the typical time associated with this accretion rate and is given by

$$t_{Edd} = \frac{M}{\dot{M}_{Edd}} \simeq 4 \times 10^8 \eta\, \mathrm{yr} \,. \tag{8}$$

This terminology allows us to express the relative accretion rate (accretion rate per unit mass of the BH) in the following way:

$$\frac{L}{L_{Edd}} \propto \frac{\dot{M}}{\dot{M}_{Edd}} \propto \frac{\dot{M}}{M} \,. \tag{9}$$

The normalized Eddington accretion rate, $\dot{m} = \dot{M}/\dot{M}_{Edd}$, is used in various articles. Unfortunately, this is done in an inconsistent way since in some articles it includes the efficiency factor η (like in here) while in others η is not included (i.e. $\dot{m} = c^2 \dot{M}/L_{Edd}$).

2.2 Spherical Accretion

The simplest possible case involves spherical accretion (also known as "Bondi accretion"). In this case we assume that accretion starts from a large radius where the gas is at rest. We follow the gas radial motion under the combined force of gravity and radiation pressure. Detailed calculations [2], show that the optical depth of such gas is very large. For example, if the gas is completely ionized, its electron scattering optical depth is

$$\tau_e \simeq \frac{\dot{M}/\dot{M}_{Edd}}{[r/r_g]^{1/2}} \,, \tag{10}$$

which, for a critical accretion rate ($\dot{M} = \dot{M}_{Edd}$) with inflow at the free fall velocity, gives a Compton depth of order unity. The radiation conversion efficiency (η) in spherical accretion is very small. The reason is that, in such accretion, most of the radiation is emitted by two body collisional processes in the ionized plasma. Such processes are relatively inefficient compared with the release of gravitational energy. The result is that most of the gravitational energy is advected into the central object and the radiated luminosity is only a small fraction of the released gravitational energy. A good estimate is

$$\frac{L}{L_{Edd}} \simeq 10^{-4} \left[\frac{\dot{M}}{\dot{M}_{Edd}} \right]^2 . \tag{11}$$

3 Accretion Disks

The most efficient accretion processes are associated with accretion disks. Such systems are naturally formed by infalling gas that sinks into the central plane while retaining most of its angular momentum. Such disks, if dense and thick enough, can provide the necessary mechanism to transfer angular momentum out and to allow gas infall into the vicinity of the BH [3]. There are several types of accretion disks with different shapes and properties. The most important ones are discussed here in general terms.

3.1 Basic Disk Parameters

First we consider optically thick geometrically thin accretion disks. Such systems are formed and maintained for a large range of accretion rates, $10^{-4} \lesssim L/L_{Edd} \lesssim 0.3$. The important physical parameters of such systems are the angular momentum of the material in the disk at radius r, $s(r)$, the net torque associated with this angular momentum, $N(r)$, and the radial drift velocity of the gas, v_r [4, 5]. In general all these quantities can change as a function of time. Here we limit the discussion to time independent, stationary disks.

Consider first the angular momentum of a ring of radius r with a total (time independent) mass m,

$$s = m v_\phi r \, , \tag{12}$$

where v_ϕ is the azimuthal velocity. For most cases of interest $v_r \ll v_\phi$ and $v_\phi \simeq v_K$, where v_K is the Keplerian velocity. The associated Keplerian angular velocity is

$$\Omega_K = \frac{v_K}{r} = \left[\frac{GM}{r^3} \right]^{1/2} . \tag{13}$$

The radial mass flow rate of a stationary disk, $\dot{M}$, is constant and is given by

$$\dot{M} = -2\pi r \Sigma v_r \, . \tag{14}$$

Here Σ is the surface mass density of the disk (cm^2), which is also time independent and is given by

$$\Sigma = 2\rho H(r)\,, \tag{15}$$

where ρ is the mean density and $H(r)$ the scale height (see below) at a radius r.

Given these assumptions we can find the net torque (i.e. the torque between two adjacent rings) by considering conservation of angular momentum. This means that the angular momentum carried inward across radius r minus the angular momentum transported outward by the torque exerted on the disk outside of that radius, is constant. This requires that

$$N(r) = \dot{M}[GMr]^{1/2} + const. \tag{16}$$

Thin accretion disks are defined by those properties as well as their inner (r_{in}) and outer (r_{out}) radii. The innermost radius is where the assumption of bound Keplerian motion breaks down. This is normally referred to as the "last stable orbit" and is a function of the angular momentum of the BH. For non-rotating BHs $r_{in} = 6r_g$ while for a maximally rotating Kerr BH $r_{in} = 1 - 1.2r_g$. Regarding the outer radius, r_{out}, this is not well defined and depends on the pressure and the gravity at large distances (see Subsect. 3.5). The various disk parameters are shown in Fig. 1.

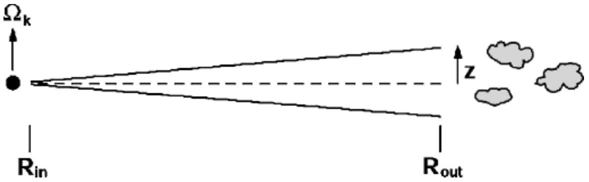

Fig. 1. Definition of the basic disk parameters (see text). The disk breaks into self gravitating blobs beyond the self-gravity radius, r_{out} (courtesy of K. Sharon)

3.2 Energy and Luminosity

The energy release in thin accretion disks is determined by the loss of gravitational energy of the inward going material and the work done by the torque that results in outward loss of angular momentum. The first of these can be written as

$$dL_g = -d\left[\dot{M}\frac{GM}{2r}\right] \tag{17}$$

and the second as

$$dL_N = -d(N\Omega_K) = -d\left[\dot{M}\frac{GM}{r}\right] \ , \tag{18}$$

where L_g and L_N are the luminosities associated with those terms. Combining the two and differentiating with respect to r we obtain an expression for the energy release at various radii,

$$\frac{dL}{dr} = \frac{3GM\dot{M}}{2r^2}\left[1 - \left(\frac{r_{in}}{r}\right)^{1/2}\right] \ , \tag{19}$$

where the second term on the right takes into account the innermost radius of the disk, r_{in}. Simple integration over parts of the disk far from its center shows that $L_r \propto GM\dot{M}/r$. The total radiated power, obtained by integrating over the entire disk, in the non-relativistic limit, is

$$L = \frac{1}{2}\frac{GM\dot{M}}{r_{in}} \ . \tag{20}$$

This is half the total available power meaning that, neglecting relativistic corrections, the matter at r_{in} still retains a kinetic energy which is half the potential energy it has lost.

Relativistic corrections are important for thin accretion disks since much of the radiation is released close to r_{in}. They change the geometric correction factor, $1 - (\frac{r_{in}}{r})^{1/2}$, in a significant way, especially for rapidly rotating BHs. Including those terms one finds that for a Schwarzschild BH, 50% of the energy is emitted between 6 and about 30 r_g while for extreme Kerr BHs, the limits are even closer. It is thought that in thin AGN accretion disks, much of the ionizing radiation is released at radii of order $\sim 10 r_g$ [6].

3.3 Emissivity and Temperature

The expression derived above for dL/dr enables us to obtain the emissivity per unit area of the disk,

$$D(r) = \frac{3GM\dot{M}}{8\pi r^3}\left[1 - \left(\frac{r_{in}}{r}\right)^{1/2}\right] \ . \tag{21}$$

Assuming local emission as a perfect blackbody with a temperature T ($D = \sigma T^4$), we get an expression for the disk temperature at all radii,

$$T(r) = \left(\frac{3GM\dot{M}}{8\sigma\pi r^3}\left[1 - \left(\frac{r_{in}}{r}\right)^{1/2}\right]\right)^{1/4} \ . \tag{22}$$

Expressing the radius in units of r_g and the accretion rate in normalized units of $\dot{M}/\dot{M}_{Edd}$ (i.e. for a given L/L_{Edd}) enable us to compare accretion disks around various mass BHs. This gives

$$T(r) \propto M_8^{-1/4} \left[\frac{\dot{M}}{\dot{M}_{Edd}} \right]^{1/4} \left[\frac{r}{r_g} \right]^{-3/4} . \tag{23}$$

For scaling purposes we note that in a thin accretion disk around a BH with $M_8 = 1$, the temperature is $T(6r_g) \sim 10^5$ K. Such disks emit most of their energy in the UV part of the spectrum. In a similar way, the maximum blackbody temperature associated with thin accretion disks around stellar-size BHs is about 10^7 K, with most of the energy emitted in the X-ray part of the spectrum. Thus, accretion disks around massive BHs are much cooler than stellar size accretion disks. Also, for standard geometrically thin optically thick disks, $T(r) \propto r^{-3/4}$.

3.4 Viscosity

The only variable which is not yet fully defined is the radial velocity v_r. This is determined by the viscosity of the disk which in turn depends on the micro-physics (atomic or molecular processes) of the disk material, the turbulence in the disk, the magnetic fields strength and structure, etc [4, 5, 7, 8, 9]. It is customary to define the viscosity stress, ν, such that

$$\nu = \alpha v_s H \tag{24}$$

where H is the scale height of the disk (to be defined later) and the parameter α includes all unknowns regarding the disk microphysics. This results in a family of disk models referred to as "α disk models". The simplest form assumes the same α for all parts of the disk. The radial velocity in α disks is found by combining the accretion rate (14) and the torque (16) equations with the definition of ν. This gives

$$v_r = \frac{3\nu}{2r} \ll v_s . \tag{25}$$

Thus, the radial inflow is subsonic.

We can combine the various parameters in a different way to obtain

$$\nu \rho \Omega_K = \alpha P , \tag{26}$$

where P is the total pressure. In this prescription the viscosity depends on the pressure. There are various variants of this prescription where α is proportional to the total pressure, the gas pressure, or some combination of the two.

The above definitions allow us to write an expression for the important "viscous time scale", t_{vis}, which is the radial drift time from radius r to r_{in},

$$t_{vis} \simeq \frac{r}{v_r} \simeq \frac{2r^2}{3\nu} . \tag{27}$$

For steady state AGN disks this time is very long (many years).

3.5 Disk Geometry

Accretion disks can be thin, slim or thick, depending on the mass accretion rate. This can be seen by considering the vertical component of the BH gravity, g_z, at some distance r,

$$g_z = \frac{GM}{r^2} \sin\theta \simeq \frac{GM}{r^3} z = \Omega_K^2 z \,, \tag{28}$$

and the hydrostatic equation

$$\frac{dP}{dz} = -\rho g_z = -\rho \Omega_K^2 z \,, \tag{29}$$

where

$$P = P_g + P_{rad} = \frac{\rho k T}{\mu m_p} + \frac{1}{3} a T^4 \,. \tag{30}$$

Assume first that the total pressure is determined by the gas pressure, $P = P_g$. In this case, $P = \rho v_s^2$, where v_s is the sound speed. The hydrostatic equation is

$$\frac{dP}{dz} = v_s^2 \frac{d\rho}{dz} = -\rho \Omega_K^2 z \tag{31}$$

with the simple solution

$$\rho = \rho_0 \exp\left[-\frac{z^2}{H^2}\right] \;;\; H \simeq \frac{v_s}{\Omega_K} \,, \tag{32}$$

where ρ_0 is the mid-plane density of the disk. Thus

$$\frac{H}{r} \simeq \frac{v_s}{v_K} \tag{33}$$

and the disk is geometrically thin since $v_s \ll v_K$.

Next, consider the self gravity of the disk in the z direction,

$$g_z^{self} = 2\pi G \Sigma = 2\pi G (2\rho z) \,, \tag{34}$$

and compare it with the vertical component of the central gravitational field, g_z. We expect the disk to break into small fragments at those radii where $g_z^{self} > g_z$, which gives a natural outer boundary for the disk, r_{out}. This boundary, together with the inner boundary r_{in}, and the disk scale height, H, specify the thin disk geometry (see Fig. 1).

3.6 Slim and Thick Accretion Disks

The situation is very different when the accretion rate, that determines the disk temperature and hence P_{rad}, exceeds a critical value where $P_{rad} > P_g$. The simple hydrostatic solution of (32) is no longer valid and a more accurate treatment gives much larger H/r. Such systems are sometimes called “slim” or “thick” accretion disks, depending on their exact geometry. Calculations show that AGN disks are dominated by radiation pressure [6]. However, for $L/L_{Edd} \lesssim 0.3$ the hydrostatic solution gives H/r which is small enough for the disk to be considered “thin”. This is no longer the case for larger accretion rate where more accurate solutions must be considered.

Large accretion rate disks can be inefficient emitters since the accretion (inflow) time scale can be shorter than the time it takes for radiation to diffuse to the surface of the disk. The photons in such disks can become trapped in the accretion flow leading to low radiation efficiency [10].

It is speculated that some AGN (e.g. narrow line Seyfert 1 galaxies, NLS1s, and very high luminosity quasars) are accreting at rates that are close or even exceed the Eddington accretion rate. These are the most likely candidates to host slim or thick accretion disks. A derivation of the mass of such systems based on the emitted radiation can lead to erroneous results, due to the processes discussed here.

3.7 The Spectrum of Accretion Disks

The expressions obtained above for the emissivity and local temperature of thin accretion disks can be combined to derive the local emitted monochromatic luminosity,

$$dL_\nu = 2\pi r(\pi B_\nu)dr \ , \tag{35}$$

and the total emitted monochromatic luminosity,

$$L_\nu = \int_{r_{in}}^{r_{out}} dL_\nu = \frac{4\pi^2 h\nu^3}{c^2} \int_{r_{in}}^{r_{out}} \frac{rdr}{exp(h\nu/kT) - 1} \propto \nu^{1/3} \, , \tag{36}$$

where the last step involves the substitutions $x = h\nu/kT$, $T \propto r^{-3/4}$ and the assumption that $x_{out} \gg x_{in}$. This simple $\nu^{1/3}$ dependence holds over a limited energy band because of the physical limits of the disk. The outer disk boundary imposes a minimal disk temperature. Thus, at low frequencies the spectrum resembles a blackbody with $T = T(r_{out})$. This gives, at low frequencies, $L_\nu \propto \nu^2$. At the high frequency end, L_ν drops exponentially with a functional dependence corresponding to the maximum disk temperature (emitted at a radius slightly larger than r_{in}). A schematic representation of the spectrum of such ideal disks is shown in Fig. 2.

3.8 Complex Disks and ADAFs

The application of the thin-disk theory to AGN is problematic for several reasons. Most important, such disks cannot explain the entire emitted spectral energy distribution (SED) since much of AGN energy is emitted in the X-ray band yet the maximum disk temperature cannot exceed $few \times 10^5$ K.

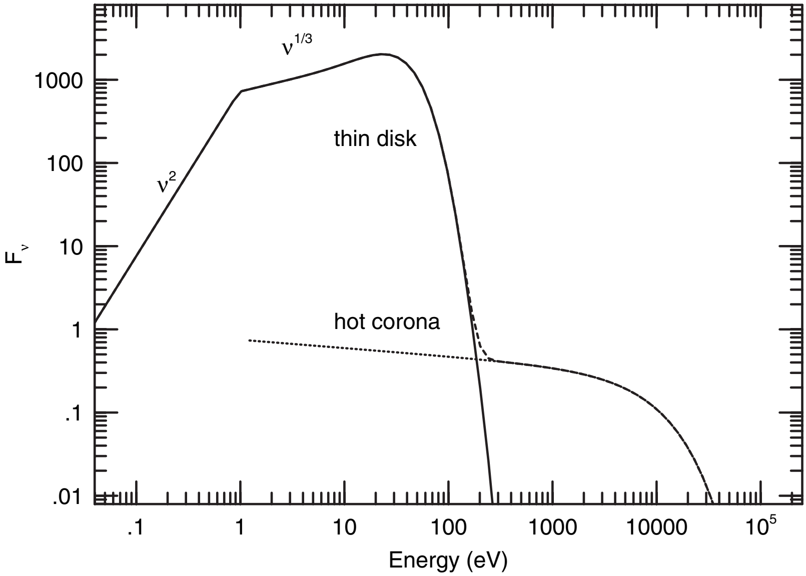

Fig. 2. Schematic representation of a disk-corona spectrum assuming a geometrically thin optically thick accretion disk with $T_{max} = 10^5\,\mathrm{K}$ and an optically thin corona with $T_{cor} = 10^8\,\mathrm{K}$

It is therefore plausible to assume an additional hot medium, in the vicinity of the disk, where more energetic photons are emitted. An appealing idea is that of a hot dilute gas where the soft disk-emitted photons are upscattered, via Compton scattering, to their observed X-ray energy. This process (Comptonization) can also explain the observed shape of the X-ray continuum which resembles a simple power-law over a large range of frequencies. The condition for the hot medium is to have a non-negligible Compton depth.

Accretion through the outer layers of the disk can result in a large energy dissipation and a sharp temperature rise. The outer skin expands and the disk develops a hot corona [11]. Such a structure can considerably modify the emitted spectrum. In particular, hard photons that are created (or scattered) in the corona can either escape the disk-corona structure or else be absorbed by the cool disk and hence change its energy balance. Likewise, soft disk photons can be absorbed by the corona in some parts or freely escape in other parts, depending on the geometry. All these effects are difficult to calculate.

A combined solution of the hydrostatic and radiative transfer equations for the disk-corona configuration is complicated and involves some ad-hoc assumptions [10]. A simplified treatment could include the assumption of an optically thin corona with a small scale height (i.e. geometrically thin disk-corona system). The corona temperature in this case is determined by the fraction of the total accretion power released in the corona. Even a very small fraction, such as 1% of the total accretion rate, can result in very high

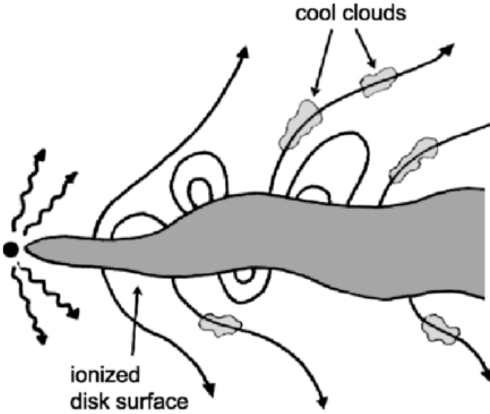

Fig. 3. Complex magnetic disks can include a non-uniform geometry like the one shown, schematically, in this diagram. Cool clouds can be flown from the surface of the disk along the magnetic field lines. Such gas is exposed to the central radiation field at some height above the disk surface (courtesy of K. Sharon)

temperatures, $\sim 10^8$ K. Such coronae radiate through free-free processes and their emission can dominate the spectrum at X-ray energies. As explained, depending on their Compton depth, they can also up-scatter the soft X-ray photons which results in additional X-ray flux. A schematic spectrum a corona dominated by free-free X-ray emission is shown in Fig. 2.

Real disks are likely to be very complex. In particular, strong magnetic fields can strongly influence the disk viscosity and may drive massive winds from the surface. A possible (schematic and highly simplified) configuration is shown in Fig. 3. Numerical calculations are needed to solve for the structure and the spectrum of such systems.

Very low accretion rate disks are interesting and important for other reasons, not necessarily related to AGN. Low accretion rate results in weak coupling between the ions (mostly protons) and the electrons which will lead to two different temperatures for the two types of particles. While protons carry most of the gravitational energy, cooling, and hence electromagnetic energy release is mostly due to electrons. In such systems, the heavier protons are advected into the BH taking with them much of the gravitational energy. Such advection dominated accretion flows (ADAFs) are very hot yet very inefficient emitters. It is speculated that some AGN disks can fluctuate between the states of high accretion rate and ADAF, depending on the fuel supply to the center.

3.9 Spectral Energy Distribution

The spectral energy distribution (SED) of AGN is complex, with significant variations between different energy bands and different objects. Figure 4

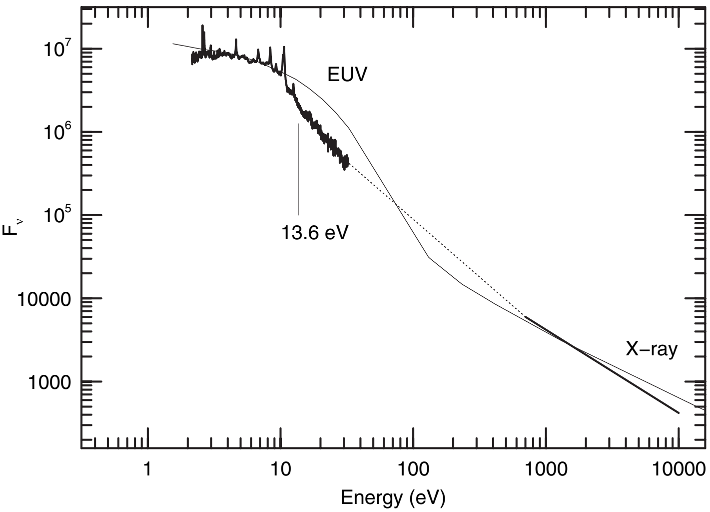

Fig. 4. Observed SED (thick solid lines connected with a thin dotted line) compared with a theoretical SED composed of a spectrum of a thin accretion disk around a 3×10^9 $M_\odot$ BH with $L/L_{Edd} = 0.1$. The theoretical spectrum has an additional component of an X-ray power-law. Note the strong "blue bump" (energy excess close to and beyond the Lyman limit) predicted by this model

shows a composite SED made of the combined spectrum of several hundred AGN in the 1–20 eV range and a "typical" X-ray powerlaw continuum.

The observed SED in Fig. 4 looks different than the schematic disk spectrum in Fig. 2. However, there are good reasons to believe that the optical-UV continuum of AGN is dominated by disk emission. First, real accretion disks do not emit as simple blackbodies. The run of density and temperature with distance from the mid-plane of real disks result in a thin atmosphere where conditions are quite different from thermal equilibrium. Integrating over the various parts of such disks results in a spectrum which is considerably different from the schematic $\nu^{1/3}$ dependence and is closer in shape to $L_\nu \propto \nu^{-s}$, with $s \sim 0.5$ in the optical-UV band. Second, the infrared part of the SED ($\lambda \geq 1\,\mu$m) is dominated by non-disk processes (thermal emission by dust and nonthermal synchrotron emission). This obscures the part of the spectrum where pure disks are predicted to show $L_\nu \propto \nu^2$. A theoretical spectrum that includes several of those effects is shown in Fig. 4 where it is compared with the composite observed spectrum.

4 Physical Processes in AGN Gas

Dilute ionized gas in AGN is found from very close to the central BH, at $r \simeq 1000 r_g$, to distances of several kpc [12, 13]. This gas is exposed to the strong central radiation field and its properties are well described by considering photoionization and recombination and neglecting most other ionization and excitation processes. Notable exceptions are regions of violent star formation where mechanical heating and ionization, due to super-novae explosions and fast stellar winds, are major sources of energy input to the gas. The spectrum of such regions is typical of thermal plasma although photoionization can be important too.

The following is a brief account of the principles and the more important processes necessary to analyze and model the spectrum of active galaxies. The main assumption is that the gas density is low, $N_H < 10^{12}$ cm^{-3}, such that collisional transitions among excited states are of secondary importance for most ions (but not for hydrogen and helium) [12].

4.1 Photoionization and Radiative Recombination

Consider gas at a distance r from a point source of total luminosity L and monochromatic luminosity L_ν. The fractional ionization of ion X is determined by the ionization rate per particle due to flux reaching the location of the ion, I_X, and by the recombination rate per particle, R_X. These rates are given by

$$I_X = \int_{\nu_X}^{\infty} \frac{(L_\nu/h\nu)\sigma_\nu e^{-\tau_\nu}}{4\pi r^2} d\nu \ , \tag{37}$$

and

$$R_X = \alpha_X(T) N_e \ , \tag{38}$$

where ν_X is the (ground state) threshold ionization frequency for ion X, $\alpha_X(T)$ is the summed recombination coefficient for the ion, and τ_ν the optical depth. The time dependent fractional ionization is obtained from the solution of

$$\frac{dN_X}{dt} = -N_X[I_X + R_{X-1}] + N_{X-1}I_{X-1} + N_{X+1}R_X \ , \tag{39}$$

where N_X is the number density of ion X. All other ionization and recombination processes have been neglected at this stage. The steady state solution is

$$\frac{N_{X+1}}{N_X} = \frac{I_X}{R_X} \tag{40}$$

which represents n equations for $n+1$ ions. The additional equation is

$$\Sigma N_X = N_{\mathrm{element}} \ , \tag{41}$$

where N_{element} is the total abundance of the element in question.

The above set of equations leads to the definition of two important time scales: the recombination time, $1/R_X$, and the ionization time, $1/I_X$. The first gives the typical time required for ion $X+1$ to recombine to ion X, when the radiation source is turned off. The second time scale is the time required for

X to get ionized to $X+1$ after a large increase in the ionizing flux. We note that the recombination time is density dependent but the ionization time is not. Thus, following a time dependent recombination provides a way to probe the local conditions while following a time-dependent ionization enables to set a limit on the distance r.

The recombination and ionization time scales can be quite different, depending on the radiation field and the gas density. Moreover, the coupling between the various stages of ionization suggests that, under various conditions, the fractional abundance of ion X may change very slightly, or remains almost constant, despite large changes in the radiation field. This leads to somewhat different definitions of the recombination time involving positive contributions (recombination of X^{+1}) as well as negative contributions (recombination to X^{-1}). Such expressions can be found in various other reviews [14]. Perhaps more important are the global recombination and ionization time scales which are the times required for the entire atmosphere to undergo a significant change in the level of ionization. Both are given roughly by $1/R_X$ where X is the *most abundant* ion of the element in question.

4.2 Additional Atomic Processes

While photoionization from the ground level and radiative recombination following cascade are the dominant processes in AGN ionized gas, other processes must be considered too. The most important ones are:

Dielectronic recombination: These are normally divided into "low temperature dielectronic recombination", due to $\Delta n = 0$ transitions, where n is the energy level under question, and "high temperature dielectronic recombination", $\Delta n = 1$ transitions. The first are important for many elements at $T_e = 1 - 3 \times 10^4\,\mathrm{K}$ and the second for hot ($T_e \sim 10^{6-7}\,\mathrm{K}$) plasma.

Auger ionization: Ionization from inner shells due to absorption of high energy X-ray photons is important for ions with three or more electrons. The process couples non-adjacent stages of ionization and results in X-ray fluorescence lines that are commonly observed in the spectrum of AGN.

Heating and ionization by secondary electrons: This process is important for gas with a very low level of ionization which is exposed to a large flux of high energy photons. The energetic electrons ejected due to the absorption of such photons can collide with neutral atoms (mostly hydrogen) and cause additional ionization before they are thermalized. They can also collisionally excite some lines, such as Lyα.

Charge exchange: Charge exchange collisions between various ions can be very fast, leading to significant changes in the level of ionization under some favorable conditions.

Collisional ionization and three body recombination: These processes are negligible for most ions for temperatures below $few \times 10^5\,\mathrm{K}$. Collisional ionization dominates the ionization of the gas in starburst regions. Both processes are important in high density large optical depth clouds where they can dominate the population of the high-n levels of hydrogen and helium.

4.3 Thermal Balance

The thermal structure of the gas is governed by a set of equations that involves energy terms rather than ionization and recombination terms. These will be noted by H for heating and C for cooling. In AGN gas, the most important heating is photoionization (bound-free) heating which, for ion X, is given by

$$H_{bf,X} = N_X \int_{\nu_X}^{\infty} \frac{L_\nu \sigma_\nu e^{-\tau_\nu}}{4\pi r^2} d\nu \ . \tag{42}$$

In this equation we sum over the absorbed energy while in the photoionization equation (37) we summed over the number of ionizing photons. Similarly, an important cooling term is “recombination cooling” given by

$$C_{bf,X} = N_{X+1} N_e \alpha_X(T)(h\nu_{1,\infty} + \langle h\nu_{free} \rangle) \tag{43}$$

where $h\nu_{1,\infty}$ is the ionization energy from the ground level of ion X and $\langle h\nu_{free} \rangle$ is the mean energy of the free electron (of order kT_e).

The above heating and cooling terms are different from those used in most text books where only *the energy of the free electrons* is considered [15]. In that case the term L_ν in the heating integral should be replaced by $(L_\nu/h\nu)(h\nu - h\nu_X)$ and the recombination equation does not include the $h\nu_{1,\infty}$ term. All other heating and cooling terms can be computed with either of the two equivalent treatments provided they are used consistently when calculating the heating-cooling balance.

In many cases, the most important cooling term is due to line cooling. This involves solving for the time dependent statistical equilibrium equations which result in the time dependent level population and the net cooling (net line emission). This is illustrated here for a two-level system and only three processes: recombination with a rate α_{eff}, collisional excitation and de-excitation with rates q_{12} and q_{21} respectively, and line (but no continuum) photon emission and absorption. The time dependent equations in this case are

$$\frac{dn_2}{dt} = n_1 q_{12} - n_2(A_{21}\beta_{21} + q_{21}) + N_{X+1} N_e \alpha_{\text{eff}} \tag{44}$$

and

$$n_1 + n_2 = N_X \ , \tag{45}$$

where the abundance of ion X, N_X, is also time-dependent (see (39)). Potentially important processes that are neglected in this simplified treatment are dielectronic recombination, collisional ionization, three body recombination, photoionization from both levels and continuum pumping (see below).

The term β_{21} in (44) is the "local escape probability", i.e. the probability of line photons to leave the point where they are created without being absorbed by 1–2 transitions in a different location. This simplified escape probability treatment replaces the more rigorous, full radiative transfer solution. It enables a full local solution but neglects radiative coupling of different locations in the gas. Detailed radiative calculations show that a reasonable approximation for β is $\beta = 1/(1 + b\tau_{12})$, where τ_{12} is the optical depth between the two levels and b is a constant of order unity [12].

Having solved for n_1 and n_2, we can now express the line emissivity, ϵ_{21}, which is also the bound-bound (bb) cooling term, C_{bb}, as follows:

$$C_{bb} = \epsilon_{21} = n_2 \beta_{21} A_{21} h\nu_{21} \,. \tag{46}$$

The general energy equation is $H = C$, where $H = \Sigma_i H_i$ and $C = \Sigma_i C_i$ are sums over all heating and cooling terms. The solution of this equation is used to obtain the kinetic temperature of the gas. This is relatively simple in steady state, where the level population and ionization fractions are not time dependent. The main complications are due to radiative transfer effects that couple level populations to the gas temperature through collisional excitation and de-excitation processes.

Mechanical Heating of AGN Gas

The general energetics and the physical conditions in AGN gas are controlled primarily by the central radiation field. This is in contrast to different environments, such as star-forming regions, where mechanical heating is the dominant heating source. There have been several suggestions that parts of the gas in the nucleus is affected by collisional processes, e.g. small-scale jets can transfer a large amount of mechanical energy into a very small region, producing strong shocks and bright emission knots typical of collisionally excited plasma. However, general considerations suggest that such processes cannot be globally important in the type of environment considered here.

The mechanical energy produced by a shock of velocity v_{shock} is roughly $m_{shock} v_{shock}^2$ where m_{shock} is the mass of the flowing material. Equivalently, if m_{acc} is the mass of the gas accreted by the BH then the energy associated with this process is about $\eta m_{acc} c^2$, where η is the accretion efficiency factor. The observed velocity of much of the AGN line emitting gas is of order 500 km s^{-1} or larger and, for accretion disks around massive BHs, $\eta \sim 0.1$. Thus the ratio of the two energy producing processes is approximately $10^5 (m_{acc}/m_{shock})$. For shock excitation to be energetically significant, the amount of flowing shocking gas must exceed, by many orders of magnitude, the amount of accreted material. This seems to be in contradiction with known properties of AGN.

4.4 The Spectrum of Ionized AGN Gas

Ionization Parameters

Given photoionization-recombination equilibrium, one can define an "ionization parameter" which is proportional to the level of ionization of the gas. One possible definition is

$$U = \int_{E_1}^{E_2} \frac{(L_E/E)dE}{4\pi r^2 c N_{\rm H}} \ , \tag{47}$$

where c is the speed of light (introduced to make U dimensionless) and E_1 and E_2 are energy boundaries to define the photon flux. U defined in this way gives a good indication for the level of ionization provided E_1 is close to the minimum energy required to ionize the most "important" (cooling-wise) ions and elements. For gas that is ionized by soft-UV radiation, E_1 should be chosen such that it is related to the UV ionizing field, e.g. the energy of the hydrogen Lyman limit. An appropriate name in this case is U(hydrogen). Similarly, a suitable choice for gas whose level of ionization is dominated by a strong soft X-ray source is E_1 inside that energy range. Such gas can be transparent to UV photons and U(hydrogen) is not a good indicator of its level of ionization. A suitable choice in this case is $E_1 = 0.54$ keV, corresponding to the K-shell threshold ionization of oxygen, the most important emitting and absorbing element under such conditions. A suitable name is U(oxygen).

There are other, somewhat different definitions of the ionization parameter that are used in various publications. A common one, ξ, is defined by the energy flux rather than the photon flux,

$$\xi = \int_{13.6 \text{ eV}}^{13.6 \text{ keV}} \frac{L_E}{r^2 N_H} dE \ . \tag{48}$$

This ionization parameter has been used in many X-ray related papers.

Using the "right" photon flux ionization parameter, we normally find that $U = 10^{-1} - 10^{-2}$ corresponds to gas that produces strong emission lines in the energy range under question. For example, $U(\text{oxygen}) = 0.1$ results in strong 0.5–3 keV emission lines. A much smaller U results in more neutral gas, with very little X-ray emission, and a much larger U in gas that is so highly ionized that line emission is insignificant.

Table 1 gives a list of several ionization parameters that are useful in various situations and are commonly used in AGN research. The right column of the table allows a simple relative scaling of those ionization parameters for the case of $L_E \propto E^{-1.3}$.

Line and Continuum Emitting Processes

The most important continuum emitting processes are due to bound-free transitions. The signature of such processes in the optical-UV band are the

Table 1. Various ionization parameters used in the literature

Ionization Parameter	E_1	E_2	Conversion Factors, $L_E \propto E^{-1.3}$
U(hydrogen)	13.6 eV	∞	1000
U(helium)	54.4 eV	∞	165
U(X-ray)	0.1 keV	10 keV	73.3
U(oxygen)	0.54 keV	10 keV	8.2
ξ	13.6 eV	13.6 keV	31.1

bound-free transitions to levels $n = 1$ and $n = 2$ of hydrogen and helium. In the X-ray band the observationally important processes are bound-free transitions to the ground levels of H-like and He-like carbon, nitrogen, oxygen, neon, magnesium and silicon. Free-free emission is important at radio frequencies, in a low temperature gas, and at hard X-ray energies, in a high temperature plasma.

Line emission (bound-bound) transitions dominate the heating-cooling balance in low ionization (U(hydrogen ~ 0.01) photoionized AGN gas. The most intense transitions are due to forbidden lines, in low density gas, and due to recombination and collisionally ionized permitted lines, in high density gas. For highly ionized X-ray gas (U(oxygen) ~ 0.01), most cooling is via bound-free transitions because the lowest accessible levels are at energies that are much larger than kT. This gas is dominated by permitted, intercombination and forbidden recombination lines. Starburst X-ray gas, which is sometimes found in AGN, is characterized by higher temperatures and lower densities. The most intense lines are produced via collisional excitation of $n = 2$ and $n = 3$ levels.

Several additional processes that were not discussed so far result in line emission and absorption. Here we comment on two of those.

Continuum fluorescence: This process, sometimes referred to as "photoexcitation" or "continuum pumping", is the result of populating low lying levels by absorbing the incident continuum in various resonance transitions. This increases the level population and the line emission but the effect on the global energy balance is negligible, especially in low density environment, since heating and cooling almost exactly balance.

The increase or decrease in the observed line intensity, due to this process, is geometry dependent. For example, in a static spherical atmosphere, or in a static spherical thin shell, where the point continuum source is in the center, the process can be considered as pure scattering and emission is completely canceled out by absorption. In a moving spherical atmosphere, the emitted photons are Doppler shifted relative to the gas rest frame, and absorption and emission appear at different energies. This results in both emission and absorption with the same equivalent widths (EWs). In cases of incomplete covering, absorption is stronger than emission (in terms of the measured EW). Finally, in optically thick lines, the absorption and emission optical depths are not necessarily the same which can affect the emitted line intensity.

Continuum fluorescence in the X-ray band can also excite inner shell transitions. In principle, this is similar to absorption by resonance lines except that some of these transitions can leave the ion in an autoionized stage which can affect the ionization balance.

Line fluorescence: Photoionization of K-shell electrons can lead to line fluorescence, following the ejection of an Auger electron. The process is important at X-ray energies where lines of this type (most notably iron Kα lines in the 6.4–7.0 keV energy range) are occasionally seen. The fluorescence efficiency ("fluorescence yield") depends on Z^3 and is roughly 0.3 for iron. Thus, the most abundant elements, like oxygen, do not produce strong fluorescence lines, despite the order of magnitude difference in abundance between iron and oxygen. This difference can partly be compensated by the larger incident flux around 0.54 keV (the K-shell threshold for neutral oxygen), compared with the flux near 7.1 keV (the K-shell threshold of neutral iron).

4.5 Clouds, Condensations, Confinement and Winds

The basic line emitting entity is normally referred to as a *cloud.* The assumption of clouds, or more generally condensations, is reasonable given the observations of galactic HII regions, and the interstellar medium, where condensations are indeed observed. It is also required by line intensity considerations, mainly the observation that the typical observed line widths, e.g. in the BLR, are similar for low and high ionization lines. The assumption of clouds is somewhat problematic since such entities must be either gravitationally bound or else confined by external pressure.

Likely candidates for self gravitating clouds are stars in the central cluster. In particular winds produced by "bloated stars" have been considered as the origin of material in the broad and the narrow line regions. It is interesting to note that "typical" NLR clouds (that are not just theoretical entities since direct observations of nearby AGN show clear condensations in the NLR) contain masses of about one $M_\odot$. Much lower mass clouds, containing as little as 10^{-8} $M_\odot$, have been considered in various BLR models. Such clouds must be confined or else created and destroyed on a very short (the sound speed crossing time) time scale. For the BLR clouds, this time is of order one year.

The most likely confinement mechanisms for clouds that are not self-gravitating are either thermal confinement, by a low density hot inter-cloud medium (HIM), or confinement by magnetic fields. Confinement requires pressure equilibrium and the exact physical conditions are obtained from general considerations of a two-phase medium, similarly to what is done in the ISM [16,17].

Consider a medium where the fractional ionization of the gas and its kinetic temperature are determined solely by the central radiation field. Suppose there are two components at the same location with very different temperatures that are noted here by ‘cold" and "hot". In this case,

$$U_{cold} \propto \frac{L/c}{r^2 N_H(cold)} \ ; \ U_{hot} \propto \frac{L/c}{r^2 N_H(hot)} \ . \tag{49}$$

Dividing by the relevant temperature (T_{cold} or T_{hot}) and noting that all quantities are calculated at the same location we get

$$\frac{U_{cold}}{T_{cold}} = \frac{U_{hot}}{T_{hot}} \propto \frac{P_{rad}}{P_g} \ . \tag{50}$$

At a given distance r, $P_{rad} \propto L/r^2$ and is independent of the gas density. Thus, for all components at that location $\frac{U}{T} \propto P_g^{-1} \propto (N_H T)^{-1}$. This allows several different temperatures and densities for the same value of U/T. For the case considered here there are two possible values of T for the same U/T. For other cases there may well be even more such solutions, all with the same P_g. Such two (or more) components can survive, side by side, in pressure equilibrium, provided they are stable against thermal perturbations. The situation is illustrated in Fig. 5 where we show the famous "S-curve", which is the photoionization equilibrium curve (sometimes referred to as the "stability curve") for such gas.

The upper branches of stability curves like the ones shown here are stable, i.e. a small deviation from the curves results in return to an adjacent stable region. Their temperature is the Compton temperature of the fully ionized gas which is determined solely by Compton heating and cooling and is given by

$$T_C = \frac{h\overline{\nu}}{4k} \ , \tag{51}$$

where $h\overline{\nu}$ is the mean photon energy weighted by the cross section. T_C depends only on the SED and, for AGN, is $\sim 10^7$ K. The lower branch with the positive slope is also stable and is typical of AGN photoionized gas with a characteristic temperature of $\sim 10^4$ K. The intermediate part contains one or more regions with negative slopes. These are unstable regions where small perturbations to the right (the region where $H > C$) or to the left ($H < C$) result in further removal from the stability line. Thus a two, three and even four-component medium can be formed in gas exposed to a typical AGN radiation field. Such localized components, with a well defined temperature, could be described as "clouds" or condensations.

An alternative and perhaps a more likely solution is confinement by magnetic pressure. In this case, $B^2/8\pi \geq N_H kT$ and the required magnetic field is of the order of one gauss for the BLR and much smaller for the NLR.

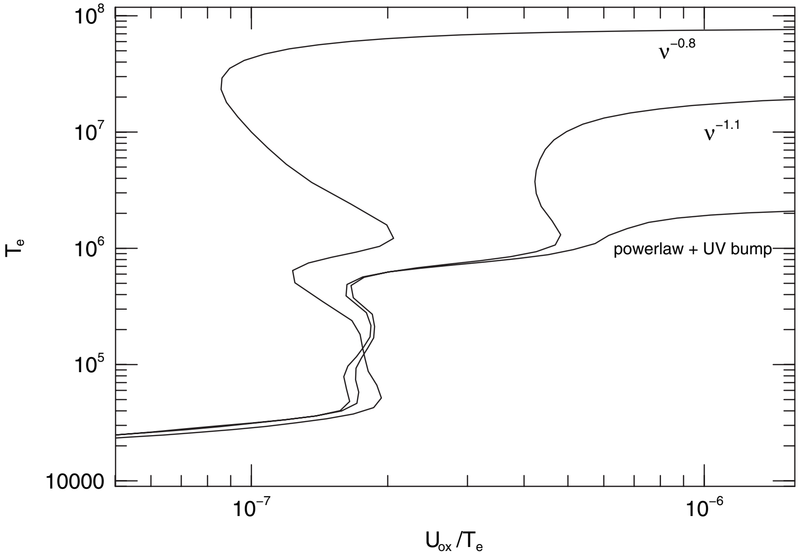

Fig. 5. Stability curves for several assumed SEDs, calculated by the author: 1. $L_E \propto E^{-0.8}$. 2. $L_E \propto E^{-1.1}$. 3. "Typical" AGN SED with a strong blue-bump and an X-ray power-law. Note the several unstable regions (negative slope parts) that are less noticeable in the SED containing a UV bump. Such SED results in lower T_C due to the efficient Compton cooling by the blue-bump photons

Such magnetic fields are likely to occur near the central accretion disk and, perhaps, throughout the emission line region [7, 8]. The "clouds" in this case are likely to have a non spherical shape, perhaps elongated filaments.

One can consider also a range of properties in any given location, in a time dependent situation, where condensations or filaments are constantly produced and destroyed. Such situations are common in the interstellar medium and have also been considered in AGN. The one that attracted most attention is the locally optimally emitting cloud (LOC) model which involves an assumed range of density, column density and covering factor at any given location [18]. The various components are not in pressure equilibrium and there is little work, if any, regarding their formation or stability. However, the resulting spectral properties are quite appealing since they produce a good fit to the observed optical-UV spectrum of many AGN.

Finally we consider gas flowing away from the central disk or from other large mass reservoirs in the nucleus, in a form of a continuous wind. Such flows are characterized by continuous change of density and velocity and are very different from the above considered "clouds". The emission and absorption spectrum of such flows are very different from those produced by clouds mostly because of the different run of opacity and the structure and location of the various ionization fronts. The dynamics and the spectrum of AGN flows are further discussed in Subsect. 4.6.

4.6 The Motion of Ionized Gas

The Equation of Motion

AGN gas is exposed to a strong radiation field that can, under certain conditions, produce large scale flows. Such flows may be continuous, in which case they are classified as pure "winds", or else drive clouds or condensations in a ballistic way (the general term "wind" as used in the literature can include condensations inside a continuous flow). The general form of the equation of motion for a cloud of mass M_c is

$$a(r) = a_{rad}(r) - g(r) - \frac{1}{\rho}\frac{dP}{dr} + f_d/M_c \ , \tag{52}$$

where $g(r)$ is the gravity, $a_{rad}(r)$ the acceleration due to radiation pressure and f_d the drag force. For pure wind flows we neglect the drag force term and the internal radiation pressure. This gives

$$v\frac{dv}{dr} = a_{rad}(r) - g(r) - \frac{1}{\rho}\frac{dP_g}{dr} \ . \tag{53}$$

An additional required equation describes the continuity condition and allows to obtain the density and velocity at every location. For the wind scenario, this is given by

$$\dot{M} = 4\pi\rho r^2 v C_f(r) = const. \ , \tag{54}$$

where $C_f(r)$ is the location dependent covering factor. For ballistic flows $\dot{M}$ needs to be specified more carefully, given the initial conditions, the number of moving clouds and the flow geometry.

The importance of the various terms in the equation of motion can vary a lot depending on the local conditions. Gravity dominates in almost all cases of low ionization, large column density gas. The reason is that the neutral gas at the part of the cloud away from the illuminated surface does not feel the radiation pressure force. Radiation pressure force dominates the motion of ionized, low optical depth gas near a source with a high accretion rate and large η. This is not the case for fully ionized gas where only electron scattering contributes to the pressure. Pressure gradient can be the most important factor when $a_{rad}(r) \ll g(r)$ and there is a high ionization, high temperature gas component [19].

Radiation Pressure

The radiation pressure force at a distance r is obtained by considering all absorption processes and by summing over the contributions to a_{rad} from all ions at that location. The contribution to a_{rad} due to ion X is

$$a_{rad}(r,X) = \frac{N_X}{\rho(r)c}\int_{\nu_X}^{\infty}\frac{L_\nu \chi_\nu e^{-\tau_\nu}}{4\pi r^2}\,d\nu\,, \tag{55}$$

where χ_ν is the total absorption cross section including all bound-bound and bound-free transitions. This should be supplemented by two additional terms that count the free-free and the Compton contributions to a_{rad}.

The relationship between a_{rad} and the level of ionization can be understood by considering the important case where the largest contribution to the radiation pressure force is due to ionization (bound-free transitions). To illustrate this we assume that the mean energy of an ionizing photon is $\overline{h\nu}$ and use the definition of I_X from (37). For ionized gas with $N_e \simeq N_H$ this gives

$$a_{rad}(X) \propto \frac{\overline{h\nu}}{cN_e}N_X I_X\ . \tag{56}$$

In a steady state, photoionization is exactly balanced by radiative recombination, i.e. $N_X I_X = R_X N_{X+1} \propto N_e^2$. Thus a_{rad} is proportional to N_{X+1}. For the specific case of hydrogen, $a_{rad} \propto N_e$.

It is customary to introduce the "force multiplier", $M(r)$, which is the ratio of the total radiation pressure to the Compton radiation pressure

$$M(r) = \frac{a_{rad}(total,r)}{a_{rad}(Compton,r)}\ , \tag{57}$$

where

$$a_{rad}(Compton,r) = \frac{N_e(r)\sigma_T L}{4\pi r^2 \rho(r) c}\,. \tag{58}$$

$M(r)$ is then used to express the radiation pressure force in the equation of motion. Note that by definition $M(r) \geq 1$ and that $M(r) = 1$ for fully ionized gas.

For gas which is not fully ionized, the main contributions to $M(r)$ are from bound-bound and bound-free transitions. The former dominates in those cases where the line optical depths are small. The contribution is much reduced for optically thick lines except for the illuminated surface of the cloud (or wind). The Compton and free-free terms are usually small except for the case where the gas is fully ionized. Figure 6 shows several examples of the various contributions to the force multiplier in a highly ionized AGN gas.

The gravitational and radiation pressure forces can be directly compared by writing the equation of motion in a slightly different form. This is done by noting that the mass of the central BH is proportional to L_{Edd} and the radiation pressure force is proportional to L. Denoting $\Gamma = L/L_{Edd}$ and noting that $L_{Edd} \propto M_{BH}$, we obtain

$$v\frac{dv}{dr} \simeq \frac{\sigma_T L}{m_H c 4\pi r^2}[M(r) - 1/\Gamma] - \frac{1}{\rho}\frac{dP_g}{dr} \tag{59}$$

as the equation of motion for the wind scenario. The term $M(r) - 1/\Gamma$ in this equation gives the relative importance of gravity and radiation pressure.

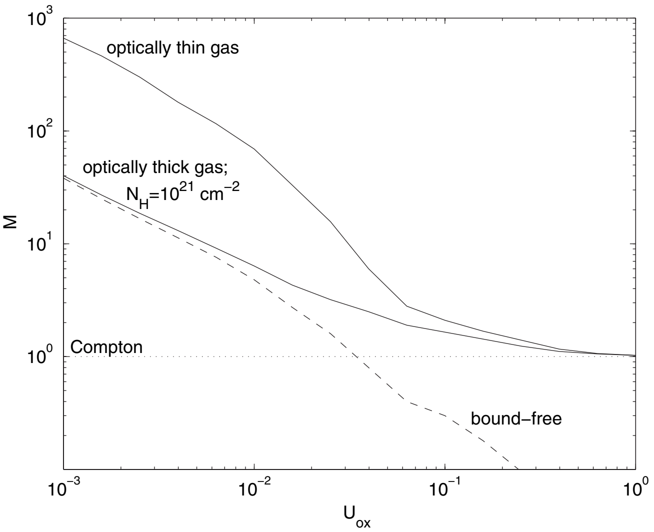

Fig. 6. The force multiplier calculated for a "standard AGN SED" and solar metallicity over a large range in ionization parameter (U(oxygen) in the case shown here). The optically thin case includes free-free, bound-free and bound-bound contribution to M. Bound-bound absorption dominates for all levels of ionization except for almost completely ionized gas. The large column density case is dominated by bound-free absorption since the resonance lines are very optically thick (courtesy of D. Chelouche)

Thus sources of large Γ require smaller force multipliers to drive the wind. E.g. a source with $\Gamma = 0.1$, typical of low and intermediate luminosity AGN, requires $M \geq 10$ in order to drive a wind by radiation pressure force. Such large values of $M(r)$ are normally achieved only in situations where bound-bound contributions to the force multiplier are important, see Fig. 6. (Note that the constant multiplying $M(r)-1/\Gamma$ in (59) is defined to within a factor of $\sim$1.2 since the force multiplier and the Eddington ratio differ by the value of μ, the mean molecular weight per electron. This is the reason for the use of $\simeq$ in (59)).

There are several important scenarios of radiation pressure driven flows that produce a similar asymptotic solution of the form

$$v(r) = v_\infty [1 - r_0/r]^{1/2} \ , \tag{60}$$

where r_0 is the launching distance of the flow and v_∞ is the asymptotic flow velocity. For cases of constant $[M(r) - 1/\Gamma]$ we get

$$v_\infty \simeq v_c \, [M - 1/\Gamma]^{1/2} \ , \tag{61}$$

where v_c is proportional to $[L/r_0]^{1/2}$ and is roughly the escape velocity at the launching point of the flow. The above results are geometry dependent and the velocities under questions can be very different in, e.g., spherical and disk geometries. However, the launching point is a very important parameter in determining the largest obtained velocities.

The equation of motion has been solved, in various works, in an attempt to explain the velocities inferred from observations of UV and X-ray absorption lines. The first seem to indicate very high velocity flows ($v_\infty = few \times 10^4$ km s^{-1}) while the latter are characterized by $v_\infty = few \times 10^2$ km s^{-1}. This can be translated to conditions near the base of the flow. For example, the maximum velocity of AGN flows that are driven by bound-free dominated radiation pressure force is of the same order as the escape velocity at the base of the flow. In such radiation pressure dominated AGN flows, the highest velocity gas must originate in the vicinity of the central accretion disk. The slower X-ray outflows can originate much further out, perhaps outside the BLR. As explained, bound-bound radiation dominated flows can reach higher velocities ($v_c \propto \sqrt{M(r)}$).

5 Main Components of AGN

The AGN phenomenon, while defined as being in the nuclear region, can cover a large range of radii and hence a large range of physical conditions. The physical properties of any gas component are determined by its location (which determine the gravity and the ionizing flux), density (that gives the ionization parameter) and column density (specifies the optical depth and hence the effective radiation pressure force and the self-gravity). Other important factors are the local pressure (confinement), composition (opacity and emissivity) and interaction with the other components (e.g. direct or indirect line of sight to the center). The following is an attempt to specify some of those properties and investigate the observational consequences.

5.1 The Broad Line Region

Consider large column density ($\sim 10^{23}$ cm^{-2}) high density ($\sim 10^{10}$ cm^{-3}) gas clouds situated at a location where $L/4\pi r^2 \simeq 10^9$ erg s^{-1} cm^{-2} (i.e. about 1 pc for a very luminous AGN). Consider also that the clouds can survive over many dynamical times ($t_{dyn} \sim 300$ years in that location) either because they are confined or because they are the extensions of large self-gravitating bodies, like bloated stars. We also assume a global (4π) covering factor of order 0.1.

One immediate consequence of these assumptions is that the clouds are bound, because gravity completely dominates over the radiation pressure force which is inefficient in such large column density material. The clouds will have typical velocities of order 3000 km s^{-1} which will be reflected in the observed widths of the emitted emission lines.

The physical conditions in the region under discussion are such that $U(\text{hydrogen}) \sim 10^{-2}$. This means that only the illuminated surface of the cloud is highly ionized. The most abundant ions in this part are He III, O IV-VI, C III-IV, etc. The strongest predicted emission lines are, therefore, Lyα, C IV λ1549 and O VI λ1035. The density is high but not too high to suppress all semi-forbidden lines. Strong predicted lines of this type are C III] λ1909 and O III] λ1663. As explained, much of these large column density clouds must be partly neutral since only X-ray photons can penetrate beyond an hydrogen column of $\sim 10^{22}\,\text{cm}^{-2}$. These parts will produce strong lines of H I, Mg II and Fe II. The observed EWs of the strongest lines depend on the emissivity and covering factor. For the specified condition they are of order 10–100 Å. The observed absorption lines are extremely weak because of the small covering factor. A region with those observed properties and spectrum would definitely justify the name "broad line region" (BLR).

Can such clouds be confined? Magnetic confinement is a likely possibility since the required magnetic field is small, about 1G. Confinement by HIM is more problematic for various reasons. The required HIM density can be obtained by the known density and temperature of the BLR gas. This gives $N_{HIM} \sim 10^7$ cm^{-3} where we assumed a typical Compton temperature of 10^7 K. Combined with the dimension of the BLR, one can obtain the total Compton optical depth that exceeds unity for the more luminous AGN. The observational consequences are clear: smearing of the central source variations and broad Compton scattering wings for all broad emission lines. This is generally not observed in AGN. There are, however, other processes that can raise the HIM temperature beyond the Compton temperature. Such relativistic HIM may still provide the required confining pressure.

Another interesting issue is the number of BLR clouds. Given the above typical velocity, and the fact that individual confined clouds are likely to emit with typical line width of 10 km s^{-1}, we can determine the number of cloud required to produce the symmetrical smooth profiles observed in many cases. This number is huge, of order 10^{6-8}, which raises serious question about the formation and destruction of such clouds. Alternatively, internal turbulent motion, that exceeds the sound speed by about an order of magnitude, could alleviate some of those difficulties. This needs to be justified observationally as well as theoretically but is still an open possibility.

5.2 The Narrow Line Region

Next we consider smaller column density ($\sim 10^{20-21}$ cm^{-2}) low density ($\sim 10^4$ cm^{-3}) clouds situated at a location where $L/4\pi r^2 \simeq 10^2$ erg s^{-1} cm^{-2} (about 3 kpc for a very luminous AGN). We also assume a smaller covering fraction, of order 0.01–0.1. The ionization parameter for this gas is similar to the one obtained for the BLR and the typical velocity, assuming a bound system, 500 km s^{-1}.

The physical conditions in this larger region are considerably different from those in the BLR, despite the very similar level of ionization. First, the column density in some of the clouds is small enough to be optically thin in the hydrogen Lyman continuum. Under such circumstances radiation pressure force may be important and the line profiles may include a non-gravitational components. Optically thin gas is, on the average, more ionized than optically thick gas of the same ionization parameter. This will show in the spectrum of the gas. Confinement may not be an important issue since the life time of such large clouds is rather long. The EW of the emission lines is considerably smaller than that of the broad lines because of the smaller covering factor and the smaller Lyman continuum opacity.

The observed spectrum of this component includes intense forbidden lines, because of the low densities. This shifts the line cooling balance in such a way that the semi-forbidden and permitted lines are relatively weaker. Another group of lines that are predicted to be intense in the innermost part of this region are coronal lines, produced by fine-structure transitions and observed mostly in the infrared. The region producing such a spectrum qualifies for the name "narrow line region" (NLR).

Dust and Reddening

Dust is likely to be present in AGN emission line regions. Such dust will scatter and absorb some of the ionizing and non-ionizing radiation, and will affect the observed spectrum in various ways. There is clear evidence for dust in the NLR but no indication for dust in the BLR. The possible reasons for this, as well as for the differences in ionization and other physical conditions, are explained below.

If AGN dust is of similar composition to the dust observed in the interstellar medium of our galaxy, we can calculate a dust "sublimation radius", i.e. the minimum radius where dust of a certain composition can survive the central radiation field without evaporating. For graphite particles, this radius is given roughly by

$$R_{sub} \simeq 0.2 L_{46}^{1/2}\,\mathrm{pc}\,, \tag{62}$$

where $L_{46} = L/10^{46}$ erg s^{-1}. Reverberation mapping of the BLR shows that the mean R_{BLR} as determined by the emissivity of the Hβ line is

$$R_{BLR} \simeq 0.4 L_{46}^{0.6\pm0.1}\ . \tag{63}$$

Evidently, $R_{sub} \simeq R_{BLR}$ which suggests that there is a natural boundary to a dust-free BLR. Since dust is such a common component in all known HII regions, it is possible that almost the entire volume of the nucleus, from just outside the BLR to the outer edge of the NLR, is filled with dust.

The presence of dust in the NLR can substantially change the level of ionization. Such dust competes, effectively, with the ionization of the gas

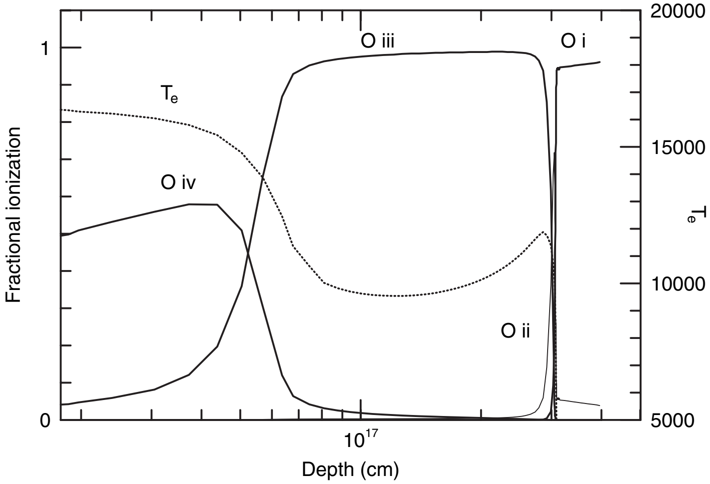

Fig. 7. Calculations by the author showing the ionization and temperature structure of a constant density dust-free gas cloud exposed to a typical AGN ionizing continuum. In this case $N_H = 10^4$ cm^{-3} and U(hydrogen) = 0.03

because of its large absorption cross sections at all energies between about 0.02 µm and 10 µm. The fraction of Lyman continuum photons absorbed by the dust, relative to those absorbed by the gas, is proportional to U(hydrogen) since

$$\frac{N_{dust}}{N_{H^0}} \propto \frac{N_{gas}}{N_{H^0}} \propto \frac{N_{H^+} + N_{H^0}}{N_{H^0}} \propto U(\text{hydrogen})\,. \tag{64}$$

Therefore, dust is more efficient in attenuating the ionizing radiation in highly ionized gas, where it absorbs a larger fraction of the photons capable of ionizing hydrogen. The different ionization structures of dusty and dust-free environments are illustrated in Figs. 7 and 8.

Because of its large cross section, dust is also subjected to large radiation pressure force which is then delivered to the gas through the efficient charge coupling of the two components. In fact, the radiation pressure acceleration due to dust can be orders of magnitude larger than the acceleration given to the gas. This raises the interesting possibility that the internal pressure structure of the cloud is determined by the external radiation field through pressure balance [20]. Since outside such dusty clouds $a_{rad}(r) \propto L/r^2$ and inside the cloud $P_g \propto N_H$, we find that for all such clouds $L/r^2 \propto N_H$. The conclusion is that for dusty ionized gas

$$U \propto L/N_H r^2 = const. \tag{65}$$

This fixes both the absolute value of the ionization parameter and the fact that it is location independent. Recent dusty NLR models based on this idea are quite successful in reproducing the observed NLR spectrum while providing the physical justification for the specific observed value of U(hydrogen).

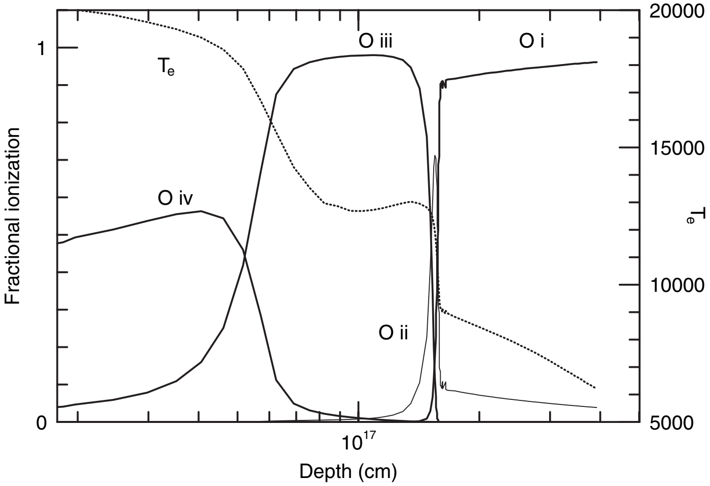

Fig. 8. Similar to Fig. 7 except that galactic type dust with the galactic dust-to-gas ratio is assumed to be present. Note the large reduction in the dimension of the ionized region and the increase in electron temperature mostly because of the depletion of carbon and oxygen that provide the strongest cooling lines

The NLR is, perhaps, the best observed region in active galaxies. We can therefore infer the physical conditions with great accuracy using standard line ratio techniques. We can also deduce the reddening corrections by comparing reddening sensitive line pairs. This may be somewhat tricky regarding internal dust (dust which is mixed in with the ionized gas) because of the source geometry and the distribution of the dust opacity relative to the line formation region.

There are notable exceptions to the dust-regulated constant U model. Some mid-IR emission lines (e.g. [NeVI]7.66 µm) as well as optical lines due to Fe X, Fe XI and Fe XIV, clearly originate in clouds of very high level of ionization that are situated in the general NLR. These *coronal lines* are probably ionized by the same central continuum, but the ionization parameter is clearly much higher. It is possible that the gas producing those lines is density bounded (i.e. of low Lyman continuum optical depth) or is dust-free. In several well studied cases, there are clear signs that this part of the NLR is closer to the central source and is characterized by higher than average density and velocity. Thus, in at least several sources, the NLR is stratified with both density and ionization decreasing outward.

5.3 The Highly Ionized Gas

Next consider the region between the NLR and the BLR, 0.1–10 pc from the center. Assume intermediate to large (10^{21-23} cm^{-2}) column density gas and a density such that the ionization parameter is 10–100 times larger than in the BLR. For this gas $U(\text{oxygen}) \sim 0.02$ thus we expect strong absorption and emission features in the X-ray part of the spectrum. Similar temperatures and levels of ionization can be found in gas with a much lower density which is distributed inside the NLR. A proper name for such a component is the *highly ionized gas* (HIG). In the X-ray literature it is often called "warm absorber" [14]. For scaling purposes we note that the density of a typical HIG cloud is

$$N_{HIG} \simeq 2 \times 10^4 U(\text{oxygen})^{-1} L_{44}(\text{oxygen}) R_{\text{pc}}^{-2} \ \text{cm}^{-3} , \tag{66}$$

and its mass is

$$M_{\text{HIG}} \simeq 10^3 \, N_{22} C_f R_{\text{pc}}^2 \ M_{\odot} , \tag{67}$$

where N_{22} is the hydrogen column density (assuming solar composition) in units of 10^{22} cm^{-2} and $L_{44}(\text{oxygen})$ the 0.54–10 X-ray luminosity in units of 10^{44} erg s^{-1} (for low luminosity AGN, like NGC 5548 and NGC 3783, $L_{44}(\text{oxygen}) \simeq 0.05$ and for high luminosity quasars $L_{44}(\text{oxygen}) \simeq 1$).

Modern X-ray observations of type-I AGN show the clear signature of such gas [21]. The strongest spectral features are numerous absorption lines of the most abundant elements, strong bound-free absorption edges due mostly to O VII and O VIII, and several emission lines. Perhaps the best spectrum of this type (correct to 2004) is a 900 ks spectrum of NGC 3783 obtained by *Chandra* HETG. This spectrum was observed in two states of the source, high and low, that differs by about a factor of 1.5 in flux of the hard X-ray continuum. Part of the low-state spectrum is shown in Fig. 9.

The X-ray spectrum of type-II AGN is completely different, because of the large obscuration in this type of sources. It shows prominent X-ray emission lines that indicate similar properties to the gas producing the strong absorption in type-I AGN. In some of the sources the X-ray emitter is resolved and its dimension can be directly measured. This indicates X-ray emission region of several hundred pc in diameter. It is not yet clear whether the type-I absorbers are of similar dimensions. In fact, some well studied cases indicate much smaller absorption regions.

HIG Outflows

The HIG gas is transparent to the UV and soft X-ray radiation and is thus subjected to a strong radiation pressure force, compared with the opaque BLR gas. This can result in mass outflow from the nucleus. Such outflows have been observed in several well studied cases and it is interesting to consider their effects on the general AGN environment.

The mass outflow rate of the HIG can be inferred from its observed velocity and from the known (or guessed in some cases) location. It is given by

$$\dot{M}_{HIG} = 4\pi r^2 C_f \rho v_{HIG} \epsilon_{HIG} \tag{68}$$

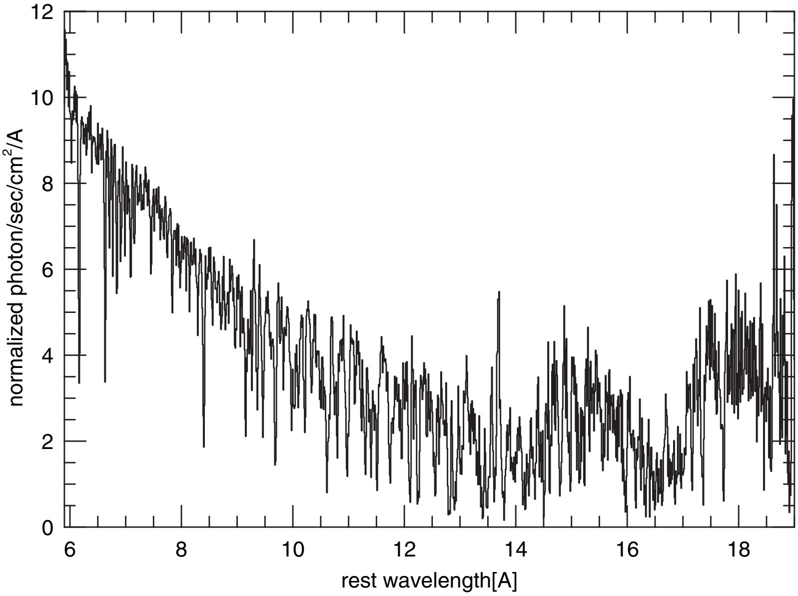

Fig. 9. Part of the low-state spectrum of NGC 3783 observed by *Chandra* in 2002 [22]. Some 60 X-ray lines were identified in this part of the spectrum

where C_f is the covering fraction and ϵ_{HIG} the filling factor of the HIG. The associated dynamical time is r_{HIG}/v_{HIG} which is of order a thousand years for HIG at several pc from the center and outflow velocity of several hundred km s^{-1}. Such outflows can interact with the NLR gas and disturb or modify the gas distribution at that location. It can also reach the ISM and the galactic halo on time scales of several million years with important consequences to the overall cooling and energy output to the inter-galactic medium.

Associated UV Absorbers

Are X-ray absorption lines the only signature of HIG at the above mentioned location? The answer to this question depends on the level of ionization and the column density of the gas [14]. Large columns and low ionization parameters can result in HIG whose less ionized species may well be C IV, N V, O V and O VI. Relatively small column density of such ions along the line of sight can produce strong UV absorption lines with typical widths similar to the widths of the X-ray absorption lines. Interestingly, some of the ions, e.g. O VI, show both UV and X-ray lines which can be compared to test the idea that the UV and X-ray absorption lines originate in the same moving gas component. Figure 11 shows a typical example of UV absorbers in an AGN whose X-ray spectrum shows the clear absorption signature of HIG.

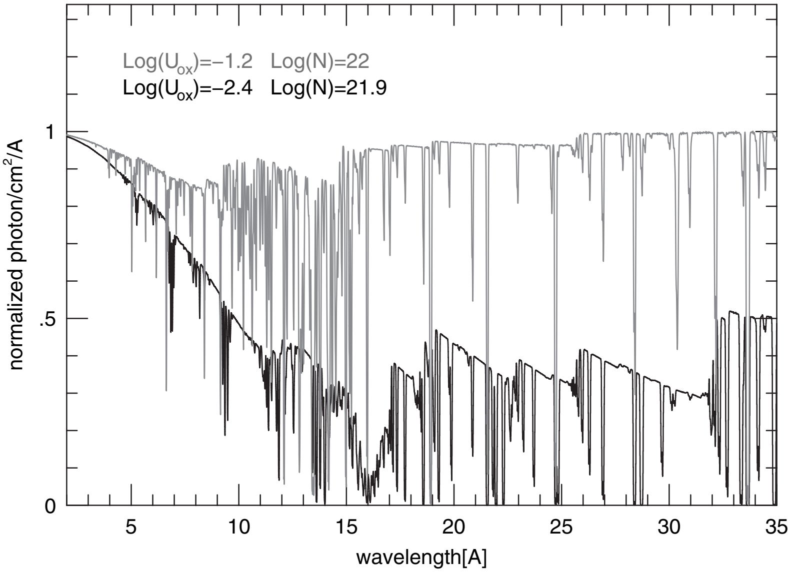

Fig. 10. Two of the three components required to model the low-state spectrum of NGC 3783 shown in Fig. 9 [23]

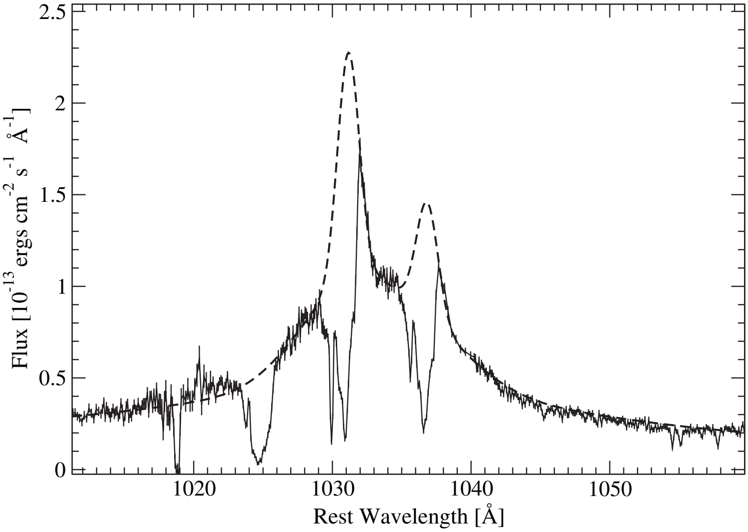

Fig. 11. The FUSE spectrum of the quasar MR 2251-178 showing narrow O VI absorption lines superimposed on emission lines of the same ion [24]. The derived outflow velocities are in the range 300–600 km s^{-1}. The source also shows time-dependent X-ray absorption lines

5.4 The Torus

We next consider a region of size 1–100 pc located around the central BH with density of $10^4 - 10^6$ cm^{-3} and extremely large column density, 10^{25} cm^{-2} and even larger. We assume a flat geometry i.e. a torus or a collection of large column density clouds moving in a plane.

The above structure covers such a large range in radii that we expect a large range in physical conditions. Far from the center, at 10 or even 100 pc, the gas is very optically thick with very low temperature. Only the hard X-ray radiation can penetrate that far and even those photons are limited to a few Compton depths. The conditions must be similar to those in molecular gas in the galaxy. Such regions are likely to contain large amount of dust. Their spectral signature are infrared emission and absorption with strong dust features.

The inner part of such a structure is exposed to the central radiation field. The temperature and ionization of this part depend on the ionization parameter. Using the previously defined L_{44}(oxygen) and assuming a typical X-ray SED we estimate that

$$U(\text{oxygen}) \simeq 0.2 L_{44}(\text{oxygen}) r_{pc}^{-2} N_5^{-1} \ , \tag{69}$$

where $N_5 = N_H/10^5\,\text{cm}^{-3}$ and r_{pc} the distance to the inner wall in pc. Assume for example $L_{44}(\text{oxygen}) = 0.1$ (intermediate luminosity AGN), $r_{pc} = 1$ and $N_5 = 0.1$. This gives $U(\text{oxygen}) = 0.2$. This would produce strong high ionization X-ray lines from the torus cavity. In particular, low, intermediate and high ionization iron K_α lines are expected to show as a clear signature of such a structure. Because of the huge column density, such a wall is a very efficient "X-ray reflector" which will scatter and reflect the incident X-ray continuum radiation. Finally, we note that the inner parts of such a torus are a likely gas reservoir for both X-ray driven winds and accretion onto the central BH.

5.5 Stars and Starburst Regions

We next consider stars and starburst regions at various locations.

Stars in the (hypothetical) nuclear cluster can absorb and reprocess the incident AGN continuum. Regarding main sequence stars, those are not likely to show any spectral signature because of the high density atmospheres and the small cross section (i.e. small covering factor). However, some stars may develop extended winds, or envelops, as a result of the interaction of their atmosphere with the hard, external radiation field. This will lower the density and increase the cross section for absorbing the ionizing radiation. Such extended envelope stars have been named "bloated stars". It has been speculated that star-star collisions may also produce bloated stars.

Bloated stars extended envelopes are similar, in many ways, to the low density gas discussed so far. Close to the center they will show typical BLR spectrum and further away a typical NLR spectrum. Moreover, such winds are likely to have a large density gradient that, far enough from the stars, could be observed as typical HIG. Thus the presence of giant envelope stars can bridge the gap between several observed components of AGN with different densities [25]. This can also provide another reservoir, or source, for the emission line gas.

Starburst regions are normally considered as a different component of AGN. The typical dimension of such regions is several kpc and the typical density $10^0 - 10^3$ cm^{-3}. Such regions contain their own internal source of mechanical energy: supernovae explosions and fast stellar winds. The typical X-ray signature of starburst regions is that of a multi-temperature hot plasma, i.e. strong collisionally excited emission lines. An example is given in Fig. 12 which shows the X-ray spectrum of the typical starburst galaxy M82.

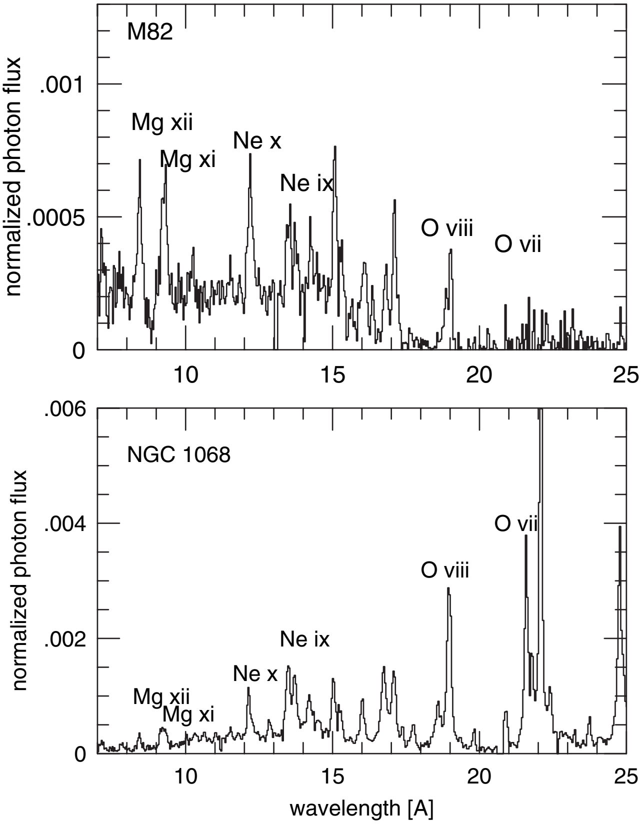

Fig. 12. A comparison of X-ray photoionized gas (NGC 1068, a type-II AGN with an extended X-ray nebulosity) and X-ray collisional gas (M82, a "classical" starburst galaxy). Both spectra were obtained by the RGS on board *XMM-Newton* and were retrieved from the XMM archive

For comparison we also show the X-ray spectrum of the “classical” type-II AGN, NGC 1068, obtained with the same instrument. The two spectra are, indeed, very different and the relative emission line ratios clearly show the different excitation mechanisms.

Close enough to the central source, the starburst excited gas is also exposed to the strong, hard radiation field. Depending on the density and flux, the resulting ionization parameter can be high enough to cause significant additional ionization thus changing the ionization balance and the emergent spectrum. We expect such starburst regions to show a mixed-type spectrum which can differ a lot from the spectrum of starburst regions that are not exposed to a central radiation source.

5.6 The Central Accretion Disk

Finally, consider the central accretion disk itself. This is an object of high density, large column density gas at 10–100 r_g which can be exposed to part of the centrally produced continuum emission. This can result in emission and/or absorption lines, depending on the geometry.

The term “central radiation field” should be defined more carefully since, so far, it has been used to describe the radiation emitted by material in the disk and in its vicinity. The disk itself, i.e. the part emitting the optical-UV radiation, is exposed to only some of the emitted SED, e.g. the X-ray radiation. The exact geometry is, therefore, of paramount importance. Plausible models considered in the literature include the possibility that the disk is illuminated by a central point source, by a double X-ray source (i.e. two point sources above and below the plane of the disk) and by the X-rays produced and scattered by an extended corona [26]. The unknown geometry, plus the fact that the density of the disk depends on the uncertain value of the parameter α, leave a large range of acceptable ionization parameters.

The results of such a scenario are easy to predict. For a low ionization parameter, the main spectral signatures are fluorescence lines of relatively neutral species and a strong reflected X-ray continuum. The strongest lines are due to FeI-XVII, at around 6.4 keV, because of the combination of high abundance and large fluorescence yield for such ions. For higher ionization parameter, the “skin” of the disk is ionized and is a source of high excitation high ionization X-ray lines. The strongest predicted lines are the leading resonance transitions in the H-like and He-like sequences of carbon, nitrogen, oxygen, neon silicon and iron. Such disk lines are predicted to be broadened by the strong gravitational field, and by the large rotational velocity of the gas, if produced close enough to the central BH.

To summarize all these possibilities we show in Table 2 a list of all possible locations considered here with the known or expected density and the corresponding ionization parameters.

Table 2. AGN components

Component	Location	Density	Ionization Parameter
Accretion disk	10^{-3} pc	$\sim10^{15}$ cm^{-3}	U(oxygen) $= 10^{-3} - 10^{-1}$
BLR	0.01–1 pc	$\sim10^{10}$ cm^{-3}	U(hydrogen) $\sim10^{-2}$
Torus	1–10 pc	10^{3-6} cm^{-3}	U(oxygen) $\sim10^{-2}$
HIG	$\sim$1 pc	10^3–10^5 cm^{-3}	U(oxygen) $\sim10^{-2}$
NLR	100–1000 pc	10^{3-5} cm^{-3}	U(hydrogen) $\sim10^{-2}$
Starburst	$\sim$1000 pc	10^0–10^3 cm^{-3}	U(oxygen) $= 10^{-2}$–10^{-4}

5.7 AGN Winds

Can all or most of these components be combined into a simplified model where the various locations, and the different dynamics, form a single, more coherent picture? Such ideas have been proposed in recent years [27]. The general picture is that of a "wind model" that originates from the vicinity of the central BH, perhaps from the surface of the accretion disk. The components are linked to this global picture, some with more justification than others. The more important ingredients of the model are:

Disk wind: Such winds are assumed to be lifted along the magnetic field lines as shown in Fig. 3. They can also be evaporated from the disk upper surface in a vertical direction. The gas gets an additional "kick" at some height above the disk surface, where it is first exposed to part or all of the central radiation. The flow accelerates due to radiation pressure force and moves to large distances. The flow lines can change direction due to the combined effect of all forces and the global velocity field is rather different from a simple spherical outflow.

Broad and narrow absorption lines: These are most naturally explained by the wind model. They are observed to be broad or narrow depending on the location and the line-of-sight to the center. An important factor for disk winds is the amount of spectral shielding since very high velocity outflows require shielding at their base. The spectral appearance of such flows is very sensitive to the observers' line-of-sight and can produce rather different appearances at different viewing angles. Many absorption lines are optically thick yet they are not black at their center since the covering factor along some viewing angles is less than unity.

Broad emission lines: The BLR in the wind scenario is assumed to be made of high density, optically thick condensations (or regions) in the wind. This is a weak point of the model since no calculation can currently produce all the observed properties of this region. Moreover, there are some indications, based on reverberation studies, of Keplerian (random orientation) velocities in the BLR of several well studied objects.

Narrow emission lines: Naive arguments suggest that the outflow can reach the NLR and form larger condensations. There are no calculations that support this claim.

HIG outflows: This is, perhaps, the most successful aspect of the model since highly ionized X-ray gas is easy to produce in many different situations. However, the observed velocities of such gas are an order of magnitude smaller than the escape velocity from the disk.

Mass loss rate: The required mass loss rate in most global wind models considered so far is of the same order as the mass accretion rate to the BH.

Global AGN wind models are only just starting to emerge. It remains to be seen whether the more complete calculations can explain the observed properties of AGN, given such winds.

6 Summary

The basic physics of the central power house in AGN must be related to accretion disks around massive BHs. A well developed theory of such structures, in particular of optically thick geometrically thin accretion disks, helps to understand the optical-UV continuum of AGN. The observed strong X-ray source is likely to originate in a separate component, such as a disk corona. The emergent continuum SED, together with the density, column density and the location of the gas relative to the BH, determine the physical conditions and the emission and absorption spectra of such objects. Examples that were treated here in some detail include gas in the BLR, in the NLR, the HIG, bloated stars, gas in star forming regions and the central accretion disk.

Acknowledgements

This work is supported by the Israel Science Foundation and by the Jack Adler Chair of Extragalactic astronomy.

References

1. Barkana, R., & Loeb, A., 2001, Phys. Rep, 349, 125
2. Krolik, J.H., "Active Galactic Nuclei" (Princeton Series in Astrophysics)
3. Shakura, N., & Sunyaev, R., 1973 (A&A 24, 337)
4. Frank, King & Raine "Accretion Power in Astrophysics" (Cambridge University Press)
5. Blandford, R., 1990 in "Active Galactic Nuclei" (Saas-Fee 1990)
6. Laor, A., & Netzer, H., 1989, MNRAS, 238, 897
7. Konigl, A. & Kartje, J., 1994, ApJ, 434, 446
8. Emmering, R.T., Blandford, R.D., & Shlosman, I. 1992, ApJ, 385, 460

9. Balbus, S.A., & Hawley, J.F., 2003, LNP Vol. 614: Turbulence and Magnetic Fields in Astrophysics, 614, 329
10. Narayan, R. 2005, Astrophys.Space.Science 300, 177
11. Liu, B.F., Mineshige, S., & Ohsuga, K., 2003, ApJ 587, 571
12. Netzer, H. 1990, in “Active Galactic Nuclei” (Saas-Fee 1990)
13. Peterson, B., “Introduction to Active Galactic Nuclei” (Cambridge University Press)
14. Crenshaw, Kraemer & George, I., 2003, AnnRevAstAP, 41, 117
15. Ferland, G., 2003, AnnRevAstAp, 41, 517
16. Krolik, J.H., McKee, C.F., & Tarter, C.B., 1981, ApJ, 249, 422
17. Rees, M.J., Netzer, H., & Ferland, G.J., 1989, ApJ, 347, 640
18. Baldwin, J., Korista, K., & Ferland, G., 1997, ApJ, 487, 555
19. Chelouche, D., & Netzer, H., 2005, ApJ, 625, 95
20. Groves, B.A., Dopita, M.A., & Sutherland, R.S., 2004, ApJS, 153, 9
21. Paerels, F.B.S., & Kahn, S.M., 2003, AnnRevAstAp, 41, 291
22. Kaspi, S., et al. 2002, ApJ, 574, 643
23. Netzer, H., et al. 2003, ApJ, 599, 933
24. Kaspi, S., Netzer, H., Chelouche, D., George, I.M., Nandra, K., & Turner, T.J. 2004, ApJ, 611, 68
25. Alexander, T., & Netzer, H., 1997, MNRAS, 284, 967
26. Fabian, A.C., 2005, Astrpy. SS, 300, 97
27. Elvis, M., 2000, ApJ, 545, 63

Multiwavelength Evidence of the Physical Processes in Radio Jets

D.M. Worrall and M. Birkinshaw

Department of Physics, University of Bristol, Tyndall Ave, Bristol BS8 1TL, UK
`d.worrall@bristol.ac.uk`, `mark.birkinshaw@bris.ac.uk`

Abstract. Over the last few years, high-quality X-ray imaging and spectroscopic data from *Chandra* and XMM-Newton have added greatly to the understanding of the physics of radio jets. Here we describe the current state of knowledge with an emphasis on the underlying physics used to interpret multiwavelength data in terms of physical parameters.

1 Introduction

Jets in active galaxies emit over a wide range of energies from the radio to the γ-ray. Synchrotron radiation and inverse-Compton scattering are the two main radiation processes, with their relative importance depending on observing frequency and location within the jet. The thermally-emitting medium into which the jets propagate plays a major role in the properties of the flow and the appearance of the jets.

In this chapter, we concentrate primarily on the interpretation of observations of resolved jet emission. Much new information about jets has been gained over the past few years. The high spatial resolution of *Chandra* has been key to the study of X-ray jets, and the large throughput of XMM-Newton has assisted studies of the X-ray-emitting environments. While historically the search for optical jet emission has been carried out using ground-based telescopes, a major problem has often been one of low contrast with light from the host galaxy. The sharp focus of the Hubble Space Telescope (HST) helps to overcome this difficulty, and the HST is playing a particularly important role in optical polarization studies. The recent opening of the Very Large Array (VLA) data archive has greatly assisted multiwavelength studies of radio jets. It, together with the Australian Telescope Compact Array (ATCA) in the southern hemisphere, allows multifrequency radio mapping at angular resolutions well matched to *Chandra* and XMM-Newton. The combination of data from the radio to the X-ray has been a key element in advancing our understanding of jets.

Active-galaxy jets provide an exciting laboratory where the interplay of relativistic effects, plasma physics, and radiation mechanisms can be seen. Our purpose in this chapter is to provide an introduction to the field and some pointers to recent research.

D.M. Worrall and M. Birkinshaw: *Multiwavelength Evidence of the Physical Processes in Radio Jets*, Lect. Notes Phys. **693**, 39–76 (2006)
DOI 10.1007/3-540-34621-X_2

2 Radiative Processes

The non-thermal mechanisms of synchrotron radiation and inverse Compton scattering are described extensively in several places, e.g., [14, 49, 81, 96, 111], and descriptions of thermal radiation can be found, e.g., in [81, 111, 114, 123]. In the following sections we give equations important in the interpretation of jet emission in useful forms that are normally independent of the system of units.

2.1 Synchrotron Radiation

The rate of loss of energy of an electron (or positron) of energy E is given by

$$-\,dE/dt = 2\sigma_{\rm T} c \gamma^2 \beta^2 u_{\rm B} \sin^2\alpha \tag{1}$$

where $\sigma_{\rm T}$ is the Thomson cross section, c is the speed of light, γ is the Lorentz factor of the electron ($= E/m_{\rm e}c^2$), β is the speed of the electron in units of c, $u_{\rm B}$ is the energy density in the magnetic field, and α is the pitch angle between the direction of motion and the magnetic field. Averaging over isotropic pitch angles, $P(\alpha)d\alpha = (1/2)\sin\alpha\, d\alpha$,

$$-\,dE/dt = (4/3)\sigma_{\rm T} c \gamma^2 \beta^2 u_{\rm B} \tag{2}$$

The radiative lifetime of the electrons is usually calculated as

$$\tau_{rad} = E/(-dE/dt) \tag{3}$$

which, for energy losses proportional to energy squared, as here, is the time for any given electron to lose half of its energy. High-energy electrons, responsible for high-energy radiation, lose their energy fastest.

The spectral distribution function of synchrotron radiation emitted by monoenergetic electrons of Lorentz factor γ is rather broad. It is usual to define the critical frequency,

$$\nu_{\rm c} = (3/2)\gamma^2 \nu_{\rm g} \sin\alpha \tag{4}$$

where $\nu_{\rm g}$ is the non-relativistic electron gyrofrequency, which is proportional to the magnetic field strength, B. Written in SI units, $\nu_{\rm g} = eB/2\pi m_{\rm e} \approx 30B$ GHz, where B is in units of Tesla. As a rough approximation, something close to $\nu_{\rm c}$ can be used as the frequency of emission, but for the full distribution function it is convenient to define $X = \nu/\nu_{\rm c}$, and then the spectral distribution function depends on frequency through

$$F(\nu, \nu_{\rm c}) = X \int_X^\infty K_{5/3}(\zeta) d\zeta \tag{5}$$

where $K_{5/3}$ is the modified Bessel function of order 5/3. The spectrum peaks at $0.29\nu_{\rm c}$, as shown in Fig. 1. The synchrotron spectrum from a distribution

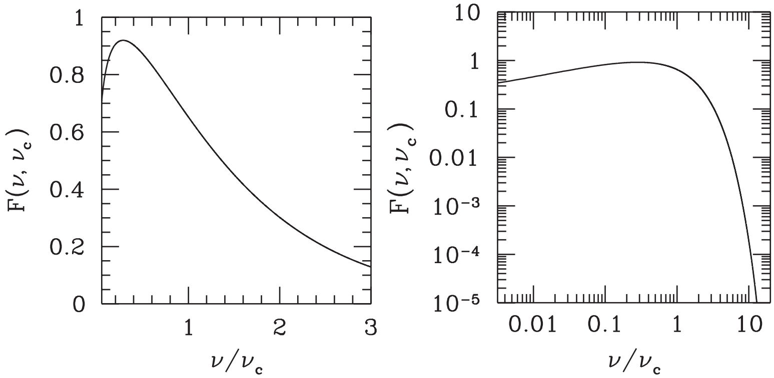

Fig. 1. The spectral distribution function of synchrotron radiation from a mono-energetic electron of Lorentz factor γ, where $\nu_{\rm c}$ is defined in (4). The spectrum is plotted linearly on the left and logarithmically on the right

of electrons with some maximum Lorentz factor, $\gamma_{\rm max}$, will therefore fall exponentially at high frequencies. The gradual turn-down at low frequencies, $\propto \nu^{1/3}$, is not expected to be seen in practice. Instead, for a homogeneous emitting region the low-energy fall off will have a slope $\propto \nu^{2.5}$ due to the source having become optically thick and suffering from synchrotron self-absorption.

The luminosity per unit frequency for a number spectrum of electrons $N(\gamma)d\gamma$ is

$$L_\nu = 2\pi\sqrt{3}\, c\, m_{\rm e}\, r_{\rm e}\, \nu_{\rm g} \sin\alpha \int F(\nu, \nu_{\rm c})\, N(\gamma) d\gamma \tag{6}$$

where $r_{\rm e}$ is the classical electron radius, and this form is valid in all systems of units. For a power-law number distribution of electrons $N(\gamma)\, d\gamma = \kappa\gamma^{-p}\, d\gamma$, an analytical result can be found for frequencies satisfying $\gamma_{\rm min}^2 \nu_{\rm g} \ll \nu \ll \gamma_{\rm max}^2 \nu_{\rm g}$

$$L_\nu = \kappa \nu^{-(p-1)/2} \nu_{\rm g}^{(p+1)/2} \sin\alpha^{(p+1)/2} m_{\rm e}\, c\, r_{\rm e} \frac{3^{p/2}}{p+1} \Gamma\left(\frac{p}{4} + \frac{19}{12}\right) \Gamma\left(\frac{p}{4} - \frac{1}{12}\right) \tag{7}$$

where Γ is the gamma function, and, again this form is valid in all systems of units. If the pitch angle distribution is isotropic, the term in $\sin\alpha$ may be replaced by $\sqrt{\pi}\Gamma[(p+5)/4]/2\Gamma[(p+7)/4]$. The result for more complicated electron spectra must be computed numerically using (6).

It is usual to use the symbol α for the negative exponent of the power-law radiation spectrum ($L_\nu \propto \nu^{-\alpha}$), where $\alpha = (p-1)/2$. Particle acceleration by ultra-relativistic shocks, likely to be important in jets, produces $p \approx 2.2 - 2.3$ [1], and p will be steeper in regions where energy losses are important.

2.2 Inverse Compton Scattering

Relativistic electrons lose energy by scattering photons to higher energy in a process analogous to synchrotron radiation, where virtual photons are replaced with real ones. Providing an electron of mass m_e and Lorentz factor γ is scattering a photon of low enough energy such that

$$\gamma h\nu_\mathrm{o} \ll m_\mathrm{e}c^2 \tag{8}$$

the Thomson cross section, σ_T is applicable, and the rate of loss of energy of the electron in an isotropic radiation field of total energy density u_rad is

$$-\,dE/dt = (4/3)\sigma_\mathrm{T} c\gamma^2\beta^2 u_\mathrm{rad} \tag{9}$$

where other symbols are as defined in Sect. 2.1. The radiative lifetime can be calculated using (3). Where electrons are losing energy both by synchrotron and inverse Compton emission, the ratio of total luminosity in the two emissions is simply $L_\mathrm{iC}/L_\mathrm{syn} = u_\mathrm{rad}/u_\mathrm{B}$.

An exact result [14] exists for the spectral distribution function for a monoenergetic electron of Lorentz factor γ scattering an isotropic radiation field of photons of frequency ν_o up to frequency ν. It is convenient to define $X = \nu/4\gamma^2\nu_\mathrm{o}$, where $(1/4\gamma^2) \le X \le 1$, and then the spectral distribution function is proportional to

$$F(\nu,\nu_\mathrm{o},\gamma) = Xf(X) = X(1 + X - 2X^2 + 2X\ln X) \tag{10}$$

This is plotted in Fig. 2. The luminosity per unit frequency for a number spectrum of electrons $N(\gamma)d\gamma$ and a spectral number per unit volume of photons $n(\nu_\mathrm{o})d\nu_\mathrm{o}$ is

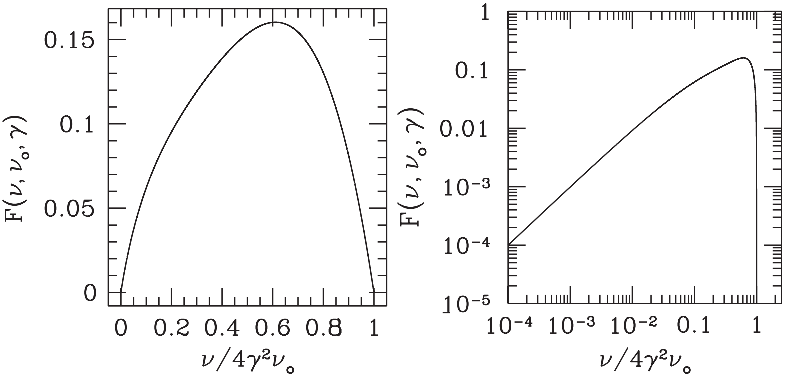

Fig. 2. The spectral distribution function of inverse Compton radiation from a monoenergetic electron of Lorentz factor γ in an isotropic radiation field of frequency ν_o in the Thomson limit (8). The spectrum is plotted linearly on the left and logarithmically on the right

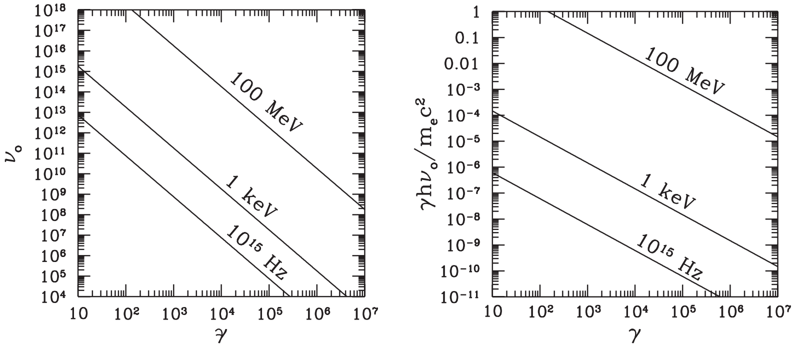

Fig. 3. The combinations of electron Lorentz factors and photon frequencies which will produce photons of typical optical, X-ray and γ-ray energies via inverse Compton scattering of an isotropic photon distribution in the Thomson limit is shown on the left. The Thomson limit only holds if $\gamma h\nu_o/m_ec^2 \ll 1$. The right-hand plot shows this is satisfied for optical and X-ray photons, but for γ-ray and higher energies the Klein-Nishina cross-section begins to become important

$$L_\nu = 3h\,c\,\sigma_\mathrm{T} \int\int n(\nu_\mathrm{o})\,N(\gamma)\,F(\nu,\nu_\mathrm{o},\gamma)\,d\nu_\mathrm{o}\,d\gamma \tag{11}$$

The relationship between the electron power-law index, p, and the spectral index of inverse Compton radiation, α, is the same as for synchrotron radiation. Since the mean value of X,

$$\int Xf(X)dX/\int f(X)dX = 1/3 \tag{12}$$

the mean frequency of photons scattered in an isotropic radiation field is given by

$$\nu = (4/3)\gamma^2\nu_\mathrm{o} \tag{13}$$

Figure 3 plots combinations of γ and ν_o that will produce optical, X-ray, and γ-ray photons according to (13), and shows that in most situations (8) is applicable. As (8) becomes violated, the Klein-Nishina cross section must be used in place of σ_T, as in these situations the electron begins to lose a significant fraction of its energy in a single scattering, e.g., [14].

The photon fields available to the electrons in jets are the synchrotron radiation they have produced, in which case the emission is known as synchrotron self-Compton (SSC) radiation, the cosmic microwave background (CMB) which scales with redshift as $(1+z)^4$, and photons from the active-galaxy nucleus. In practice the photon fields are rarely expected to be isotropic when the effects of the geometry and bulk speed of the jet plasma are taken into account.

2.3 Thermal Radiation and Energy Loss

For hot X-ray emitting gas containing ions of atomic number $Z_{\rm i}$, and where the cooling in line radiation is unimportant, the X-ray emissivity, $\mathcal{E}_\nu$ depends on temperature, T, and electron and ion number densities, $n_{\rm e}$, $n_{\rm i}$, as

$$\mathcal{E}_\nu = \frac{32\pi}{3}\left(\frac{2\pi}{3}\right)^{1/2} Z_{\rm i}^2\, g_{\rm ff}\, n_{\rm i}\, n_{\rm e}\, m_{\rm e}\, c^2\, r_{\rm e}^3 \left(\frac{m_{\rm e}c^2}{kT}\right)^{1/2} e^{-h\nu/kT} \tag{14}$$

for all systems of units, where $r_{\rm e}$ is the classical electron radius. $g_{\rm ff}(\nu, T, Z)$, the free-free Gaunt factor which accounts for quantum-mechanical effects, is of order unity and a weak function of frequency. An approximate form [111] in the case of most X-ray interest, $(kT/{\rm eV}) > 13.6 Z_{\rm i}^2$ and $Z_{\rm i} \le 2$, is

$$g_{\rm ff} = \begin{cases} \left(\frac{3}{\pi}\frac{kT}{h\nu}\right)^{1/2} & h\nu > kT \\ \frac{\sqrt{3}}{\pi}\ln\left(\frac{4}{\zeta}\frac{kT}{h\nu}\right) & h\nu < kT \end{cases} \tag{15}$$

where the constant $\zeta = 1.781$. For heavier ions more complicated forms must be used, and since the heavier ions contribute disproportionately to the X-ray output (because of the $Z_{\rm i}^2$ factor), the calculation of the correct Gaunt factor becomes a computational issue.

In SI units, (14) becomes

$$\mathcal{E}_\nu = 6.8 \times 10^{-51} Z_{\rm i}^2 T^{-1/2} (n_{\rm e}/{\rm m}^{-3})\,(n_{\rm i}/{\rm m}^{-3})\, g_{\rm ff}\, e^{-h\nu/kT} \quad {\rm W\ m^{-3}\ Hz^{-1}} \tag{16}$$

At temperatures below ∼2 keV, line radiation cannot be ignored, and the plasma models incorporated into X-ray spectral-fitting programs such as `XSPEC` [4] can be used to find $\mathcal{E}_\nu$. Examples for plasma of $kT = 0.5$ keV and 5.0 keV are shown in Fig. 4. It is common to define the rate of loss of energy per unit volume per unit frequency per $n_{\rm p}n_{\rm e}$ as $\Lambda(\nu)$, where $n_{\rm p}$ is the proton number density. In `XSPEC`, $\Lambda(\nu)$ is normalized by $(1+z)^2 \int n_{\rm p}n_{\rm e} dV/4\pi D_{\rm L}^2$ in fitting to data, where V is volume and $D_{\rm L}$ is luminosity distance. This allows $n_{\rm p}$ to be determined assuming some geometry for the source, as illustrated in Sect. 4.

The total energy-loss rate per unit mass is given by

$$\epsilon = \frac{X \int \mathcal{E}_\nu d\nu}{m_{\rm H} n_{\rm p}} \tag{17}$$

where X is the mass fraction in hydrogen, which is 0.74 for normal cosmic (i.e., solar) abundances.

If the energy loss is sufficiently rapid, the radiating gas will fall towards the centre of the gravitational potential well as its pressure support fails. The rate of infall is therefore determined by whether the characteristic timescale

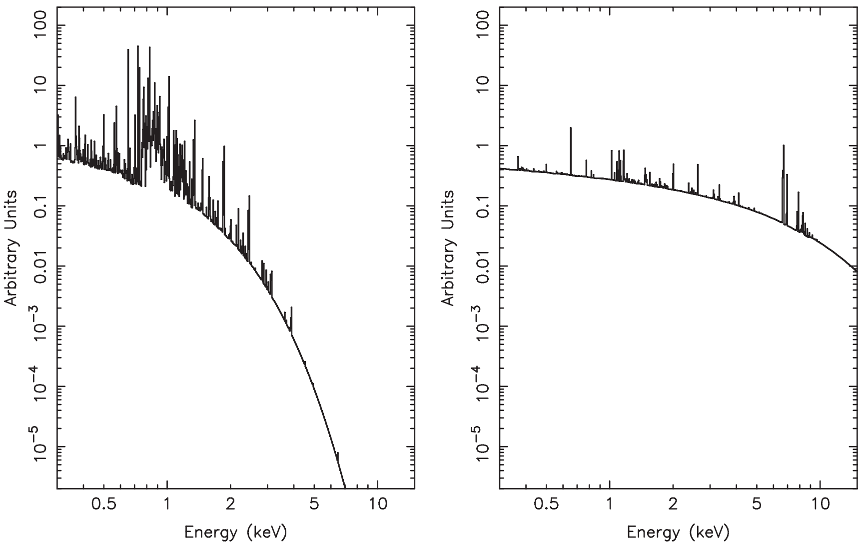

Fig. 4. Emissivity distributions of gas of $kT = 0.5\,\mathrm{keV}$ (*left*) and $kT = 5\,\mathrm{keV}$ (*right*) with normal cosmic abundances from the `APEC` model in `XSPEC`. The line emission, superimposed on the thermal bremsstrahlung continuum, is a more prominent channel for cooling in cool gas. The resolution of these model spectra is higher than that of current X-ray detectors

for radiative energy loss is greater than or less than the timescale for energy input, perhaps from the dissipation of energy by the radio jet.

If the timescale for energy loss is defined by the logarithmic rate of change of temperature in the comoving fluid frame,

$$\tau_{\mathrm{cool}} = -\left(\frac{D \ln T}{Dt}\right)^{-1} \tag{18}$$

where D/Dt is the convective derivative $(\partial/\partial t + \boldsymbol{v} \cdot \boldsymbol{\nabla})$, which describes the rate of change in a frame moving at velocity $\boldsymbol{v}$ with the gas, then the value of τ_{cool} can be obtained from the entropy equation

$$\frac{Ds}{Dt} = -\frac{\epsilon}{T} \tag{19}$$

where $s(T, P)$ is the entropy per unit mass of the gas and P is the pressure. For isobaric (P = constant) losses in a monatomic gas,

$$\tau_{\mathrm{cool}} = \frac{5}{2}\frac{kT}{\epsilon\,\mu m_{\mathrm{H}}} \tag{20}$$

Curves of $\tau_{\mathrm{cool}} n_{\mathrm{p}}$ as a function of kT, calculated using the `APEC` plasma model in `XSPEC`, are shown in Fig. 5. At high temperatures, $\tau_{\mathrm{cool}} \propto T^{1/2}$, but

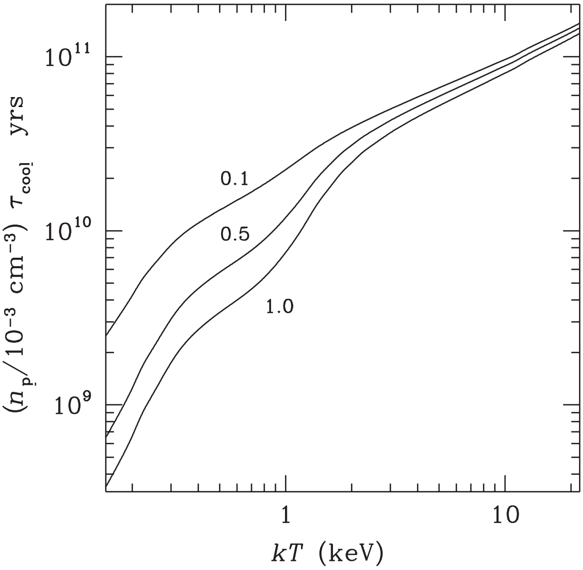

Fig. 5. Curves of $\tau_{\rm cool} n_{\rm p}$ (with $n_{\rm p}$ in units of 10^3 m^{-3} or 10^{-3} cm^{-3}) as a function of kT for three metal abundances with respect to solar: 0.1, 0.5, 1.0

at $T \leq 2$ keV line emission from common metal species in the gas becomes important, and so ϵ and $\tau_{\rm cool}$ become strong functions of kT. The offset between the curves at large kT arises from the increased density of highly-charged ions in the radiating plasma as the metal abundance increases.

3 Effects of Bulk Relativistic Motion

The first evidence for bulk relativistic motion in jets was on pc scales. In the early 1970s, the technique of Very Long Baseline Interferometry (VLBI) became established, and the first detections of apparent superluminal motion of small-scale radio components were being made (see [97] for a short review). After some years of debate, it became widely accepted that the true explanation of the effect was as proposed earlier [107], and that the apparently superluminal velocities resulted from components moving towards the observer at bulk relativistic speeds.

From a consideration of light travel time, it is straightforward to show that a source moving with a relativistic speed βc, bulk Lorentz factor $\Gamma (= 1/\sqrt{1-\beta^2})$, at angle θ towards the observer, has an apparent transverse velocity of

$$v_{\rm app} = \beta c \sin\theta / (1 - \beta\cos\theta) \tag{21}$$

which takes on a maximum value of $\Gamma\beta c$ when $\sin\theta = 1/\Gamma$. Hence large β ($> 1/\sqrt{2}$) and small θ can easily result in $v_{\rm app} > c$.

The bulk relativistic Doppler factor is defined as

$$\delta = 1/\Gamma(1 - \beta \cos\theta) \tag{22}$$

and takes on a large value for large β, small θ. For a spherical blob of emission, the apparent spectral luminosity, L_ν, is increased by $\delta^{3+\alpha}$, where α is the spectral index (Subsect. 2.1), although in most steady jets $\delta^{2+\alpha}$ is more appropriate [103, 116].

A variable source will show intensity changes on timescales shorter by δ than the true value, and beaming effects can have a dramatic effect on model parameters. For example, it was shown that the predicted X-ray flux density from SSC in a compact, spherical, variable radio source must be reduced by a factor of $\delta^{2(3\alpha+5)}$ if variability is used to measure the size, and the observed self-absorption turn-over frequency is used to estimate the magnetic-field strength [85, 86].

Relativistic beaming has a major effect on a jet's appearance, and by the 1980s the first models to unify classes of active galaxies through jet orientation were proposed, e.g., [94, 116], and subsequently further developed to account for multiwavelength properties, e.g., [6, 115, 130]. It is common to assume that a radio source has two oppositely-directed jets that are intrinsically the same, although there is doubt that this need be the case either from jet-production considerations or asymmetries in the environment through which the jets propagate. Under the assumption of intrinsic similarity, the brightness ratio between the approaching and receding jets is given by

$$R_{\rm J} = \left(\frac{1 - \beta\cos\theta}{1 + \beta\cos\theta}\right)^{-(\alpha+2)} \tag{23}$$

This constraint on β and θ is often used in conjunction with a constraint on $v_{\rm app}$ to estimate separately the jet speed and the angle to the line of sight, e.g., [68, 147]. The core dominance, $R_{\rm cd}$, defined as the flux density of beamed emission divided by that of unbeamed extended emission, is also commonly used as an indicator of the angle of a source to the line of sight. When applied to a twin-jet source, e.g., [94],

$$R_{\rm cd} = \frac{R_{{\rm cd}(\cos\theta=0)}}{2}\left[(1 - \beta\cos\theta)^{-(\alpha+2)} + (1 + \beta\cos\theta)^{-(\alpha-2)}\right] \tag{24}$$

Bulk relativistic motion towards the observer has an important effect on inverse-Compton scattering in the case of an external photon field (i.e., not SSC) [34]. In its rest frame an electron now sees a directionally enhanced radiation field and the rate of scatterings increases. If the photon field is isotropic in the observer's frame (such as the CMB), the increase in scattering rate is by a factor of δ. Since lower-energy electrons (of which there are more) are needed to scatter photons of a given rest-frame frequency to the observed frequency, there is a further enhancement by δ^α. This means that the observed luminosity at a given frequency from inverse-Compton scattering of the CMB will be increased relative to the synchrotron luminosity by $\delta^{1+\alpha}$, assuming that the magnetic field and electron spectrum are held constant, or $\delta^{3+\alpha}$ if both the field and electron spectrum are reduced by factors of δ^2 to keep the synchrotron flux constant under the condition of minimum energy in a spherical source (Subsect. 6.1). This can be sufficient to raise the ratio of X-ray to radio emission enough to fit observations of quasar jets (Subsect. 6.3). For a source radiating at minimum energy (Subsect. 6.1), while the X-ray emission from inverse-Compton scattering of CMB photons is increased, the X-ray emission from SSC is reduced. To increase this ratio via the SSC mechanism requires taking the source out of minimum energy through adding relativistic electrons. A beamed source has intrinsically lower total energy content, and, despite the higher jet speed increasing the kinetic energy, it may be in a condition of minimum total (radiative and kinetic) power [47].

4 The External Medium

Often the bulk of the X-ray emission from an active galaxy is not jet related, but rather arises from the hot interstellar, intergalactic, or intracluster medium around the jet. If we assume that this gas is almost at rest in the local gravitational potential well, with little transfer of energy or matter to or from it, then it will be in a state close to hydrostatic equilibrium.

Under these circumstances, the gas will take up a density and temperature profile which depend on the distribution of mass and the thermal history of the gas. The principal governing equation is that of hydrostatic equilibrium, with

$$\boldsymbol{\nabla} P = -\rho \, \boldsymbol{\nabla} \Phi \tag{25}$$

where Φ is the gravitational potential and ρ is the total mass density of the gas. We may assume that the gas has the ideal gas equation of state

$$P = \frac{\rho \, k \, T}{\mu \, m_{\mathrm{H}}} = \frac{n_{\mathrm{p}} \, k \, T}{X \, \mu} \tag{26}$$

to relate P and ρ, where μ is the mass per particle in units of m_{H}, given by

$$\mu = \left(2X + \frac{3}{4} Y + \frac{1}{2} Z \right)^{-1} = 0.60 \tag{27}$$

for a gas with solar abundances ($X = 0.74$, $Y = 0.25$, $Z = 0.01$).

If the mass distribution is spherically symmetrical, then

$$\frac{d\Phi}{dr} = \frac{G \, M_{\mathrm{tot}}(r)}{r^2} \tag{28}$$

where $M_{\mathrm{tot}}(r)$ is the total mass within radius r, and so the mass density and temperature profiles of the gas are related to the distribution of total mass by

$$G\,M_{\mathrm{tot}}(r) = -\frac{kTr}{\mu m_{\mathrm{p}}}\left(\frac{d\ln\rho}{d\ln r} + \frac{d\ln T}{d\ln r}\right)\,. \tag{29}$$

A consistent solution of this equation is obtained by assuming that the gas is isothermal and that the mass distribution follows

$$M_{\mathrm{tot}}(r) = 2\,M_{\mathrm{c}}\frac{r^3}{r_{\mathrm{c}}\,(r^2 + r_{\mathrm{c}}^2)} \tag{30}$$

where r_{c}, the core radius, defines the characteristic scale of the mass distribution and M_{c} is the mass within r_{c}. The density of the gas then follows the so-called isothermal β model [25]

$$\rho = \rho_0\left(1 + \frac{r^2}{r_{\mathrm{c}}^2}\right)^{-\frac{3}{2}\beta} \tag{31}$$

where β is a constant which determines the shape of the gas distribution, and depends on the ratio of a characteristic gravitational potential to the thermal energy per unit mass in the gas

$$\beta = \frac{2}{3}\,\frac{\mu m_{\mathrm{H}}}{kT}\,\frac{GM_{\mathrm{c}}}{r_{\mathrm{c}}}\,. \tag{32}$$

The alternative derivation of (31) given by [25] brings out the interpretation of β in terms of the relative scale heights or, equivalently, the ratio of energy per unit mass in gas and dark matter. ρ_0, r_{c}, and β in (31) are often fitted to provide convenient measures of the gas distribution's mass, scale and shape without considering the detailed properties of the underlying mass distribution (30).

It should be noted that the physical consistency of this much-used model for the gas distribution depends on radial symmetry (a simple distortion into an ellipsoidal model for the gas density ρ would imply a mass distribution which is not necessarily positive everywhere), and on gas at different heights in the atmosphere having come to the same temperature without having necessarily followed the same thermal history. This is problematic given the various origins of gas – some will have fallen into the cluster, some will have been ejected from stars, and some may have been moved by the jet. The solution (31) is not unique in the sense that gas with a different thermal history might follow a significantly different density distribution. Thus, for example, if the gas has the same specific entropy at all heights, and so is marginally convectively unstable, with

$$P \propto \rho^{\gamma} \tag{33}$$

with γ being the usual specific heat ratio (5/3 for a monatomic gas), the mass distribution (30) would lead to a gas density distribution

$$\rho^{\gamma-1} = \rho_0^{\gamma-1}\left(1 - \beta_{\mathrm{A}}\ln\left(1 + \frac{r^2}{r_{\mathrm{c}}^2}\right)\right) \tag{34}$$

where $\beta_{\rm A}$ is a structure constant with a similar meaning to β, and the relation only applies within the radius at which $\rho \rightarrow 0$.

A similar procedure for an NFW mass profile [92] and an isothermal gas leads to a density distribution

$$\rho = \rho_{\rm s}\, 2^{-\alpha} \left(1 + \frac{r}{r_{\rm s}}\right)^{\alpha r_{\rm s}/r} \tag{35}$$

where the new structure constant is

$$\alpha = \frac{1}{\ln 2 - \frac{1}{2}} \frac{\mu m_{\rm H}}{kT} \frac{GM_s}{r_s} \tag{36}$$

(r_s is the scale of the NFW model and M_s is the mass within radius r_s). It can be seen that in this solution $\rho \rightarrow \infty$ as $r \rightarrow 0$, since the NFW density profile has a cusp at $r = 0$.

The total masses for either the NFW profile or (30) diverge as $r \rightarrow \infty$, so both profiles must be truncated at some radius, such as r_{200}, the radius at which the mean enclosed mass density is 200× the critical density of the Universe, $\rho_{\rm crit}$ at redshift z. Care should be taken in using the solutions for ρ to ensure that the fraction of mass in gas does not exceed the cosmological bound (∼17%).

The run of density and temperature in the atmosphere around a jet are measured from the X-ray image and spectrum, where a density model for the gas of the form of (31) or (35) is fitted to the X-ray surface brightness. The X-ray surface brightness at frequency ν at a point offset by r in *projected* distance from the centre of a spherical gas distribution is (in energy per unit time per unit solid angle per unit frequency)

$$\Sigma_\nu(r) d\Omega = (1+z)\, \frac{\int n_{\rm e} n_{\rm p} \Lambda_\nu(T) dV}{4\pi D_{\rm L}^2} \tag{37}$$

where Λ_ν is the emissivity of the gas at frequency ν, and $D_{\rm L}$ is the luminosity distance of the gas. The $(1+z)$ factor takes account of the redshifting of time and frequency in the definition of Σ_ν. The volume element in the integral is

$$dV = dl D_{\rm A}^2 d\Omega \tag{38}$$

where dl is an element of distance along the line of sight, $D_{\rm A}$ is the angular diameter distance, and $d\Omega$ is the element of solid angle, so that the surface brightness becomes

$$\Sigma_\nu(r) = \frac{\int n_{\rm e} n_{\rm p} \Lambda_\nu(T) dl}{4\pi(1+z)^3} \quad . \tag{39}$$

For an isothermal gas with a β-model density distribution

$$\Sigma_\nu(r) = \frac{\Lambda_\nu(T)\, n_{\rm e0}\, n_{\rm p0}}{4\,\pi\,(1+z)^3} \int \left(1 + \frac{r^2 + l^2}{r_{\rm c}^2}\right)^{-3\beta} dl \tag{40}$$

where the integral is taken along the line of sight with limits of ±∞ or to some cut-off radius. In the former case the integral is simple, and the X-ray surface brightness can be written [11]

$$\Sigma_\nu(r) = \frac{\Lambda_\nu(T)\, n_{\mathrm{e}0}\, n_{\mathrm{p}0}\, r_c}{4\,\pi\,(1+z)^3} \left(1+\frac{r^2}{r_{\mathrm{c}}^2}\right)^{\frac{1}{2}-3\beta} \sqrt{\pi}\,\frac{\Gamma(3\beta-\frac{1}{2})}{\Gamma(3\beta)} \tag{41}$$

It is also useful to write down the observed flux density, S_ν, of a gas distribution out to angle θ from the centre, since this is the quantity generally fitted in X-ray spectral analyses. This is

$$S_\nu(\theta) = \int_0^\theta 2\pi\theta\, d\theta \Sigma_\nu(r) \tag{42}$$

$$= \frac{\Lambda_\nu(T)\, n_{\mathrm{e}0}\, n_{\mathrm{p}0}\, \theta_c^3 D_{\mathrm{A}}}{(1+z)^3}\, \frac{\sqrt{\pi}}{4}\, \frac{\Gamma(3\beta-\frac{3}{2})}{\Gamma(3\beta)} \left(1-\left(1+\frac{\theta^2}{\theta_{\mathrm{c}}^2}\right)^{\frac{3}{2}-3\beta}\right) \tag{43}$$

where θ_c is the angular equivalent of the core radius r_c. Now, `XSPEC` (Subsect. 2.3) calculates a normalization, $\mathcal{N}$, for any thermal gas model, where

$$\mathcal{N} = \frac{\int n_{\mathrm{e}} n_{\mathrm{p}} dV}{4\pi D_{\mathrm{A}}^2 (1+z)^2} \equiv \frac{(1+z) S_\nu}{\Lambda_\nu} \tag{44}$$

and this is related to the parameters of the isothermal β model by

$$\mathcal{N} = \frac{n_{\mathrm{e}0}\, n_{\mathrm{p}0}\, \theta_c^3 D_{\mathrm{L}}}{(1+z)^4}\, \frac{\sqrt{\pi}}{4}\, \frac{\Gamma(3\beta-\frac{3}{2})}{\Gamma(3\beta)} \left(1-\left(1+\frac{\theta^2}{\theta_{\mathrm{c}}^2}\right)^{\frac{3}{2}-3\beta}\right) \tag{45}$$

The fitting process in `XSPEC` takes account of the form of Λ_ν and the conversion from energy units (S_ν) to the count rate in energy bins used by X-ray detectors.

This expression shows how to use the normalization $\mathcal{N}$ found by `XSPEC`, or some other code, from the X-ray spectrum within angular radius θ of the centre, and β and r_c as fitted from the radial profile of the X-ray emission, to measure the central electron and proton densities which are related by

$$\frac{n_{\mathrm{e}}}{n_{\mathrm{p}}} = \frac{1+X}{2X} \sim 1.18 \tag{46}$$

The proton density found in this way, and the gas temperature measured from the spectrum, provide crucial information on the gas environments of radio jets and hence the external forces on the jet. The separation of X-rays from the non-thermal processes associated with the radio source and the thermal processes associated with the environment is therefore an important part of establishing the physics of a jet. This is particularly true for low power (FRI) jets, which are in direct contact with the external medium (Subsect. 5.2), but is also important for establishing the balance of static and ram-pressures in powerful sources.

5 Simple Radio-Jet Models

The FRI and FRII classifications, into which extragalactic radio sources were separated by [41], remain in wide usage today. Although based on radio morphology, [41] found a relatively sharp division in radio luminosity, with most sources of total 178-MHz luminosity below 2×10^{25} W Hz^{-1} sr^{-1} being of FRI classification, and the others of FRII. Some dependence of the location of the FRI/FRII boundary on the optical magnitude of the host galaxy has also been found [95], suggesting the effects of environment or the mass of the originating black hole are important in determining the type of radio galaxy formed by an active nucleus. However, the primary cause of the structural differences is believed to be due to the speed of the primary jet fluid, with the beams producing FRII sources being supersonic with respect to the ambient medium, whereas FRI jets are seen as the result of turbulent transonic or subsonic flows. The following sub-sections describe the observational consequences.

5.1 High-Power FRII Jets

The standard model for the expansion of a powerful radio source powered by a jet which is supersonic with respect to the X-ray-emitting interstellar medium (ISM) is illustrated in Fig. 6. The jet terminates at the beam head (in a feature identified as the radio hotspot) where the jet fluid passes through a strong shock to inflate a cocoon of radio-emitting plasma. The energy and momentum fluxes in the flow are normally expected to be sufficient to drive a bow shock into the ambient medium ahead of the jet termination shock. In the rest frame of the bow shock, ambient gas is heated as it crosses the shock to fill a region surrounding the lobe of radio-emitting plasma. Observationally on kpc scales we see the radio emission from well-collimated jets feeding edge-brightened lobes. An example is shown in Fig. 7.

The sound speed in gas of temperature T is

$$c_{\rm s} = \sqrt{\frac{\gamma k T}{\mu m_{\rm H}}} \tag{47}$$

where γ is the ratio of specific heats ($\gamma = 5/3$), k is the Boltzmann constant, $m_{\rm H}$ is the mass of the hydrogen atom, and $\mu m_{\rm H} = 0.6 m_{\rm H}$ is the mass per particle. Under these conditions, $c_{\rm s} \approx 516 (kT/{\rm keV})^{1/2}$ km s^{-1} $\approx 0.54 (kT/{\rm keV})^{1/2}$ kpc Myr^{-1}.

The Mach number of the speed of advance, $v_{\rm adv}$, of the bow shock into the ambient medium is $\mathcal{M} = v_{\rm adv}/c_{\rm s}$, which in convenient units can be expressed as

$$\mathcal{M} \approx 580 (v_{\rm adv}/c)(kT/{\rm keV})^{-1/2} \tag{48}$$

where c is the speed of light. The advance of the bow shock is likely to be slow with respect to the bulk speed of the jet. However, the jets need not

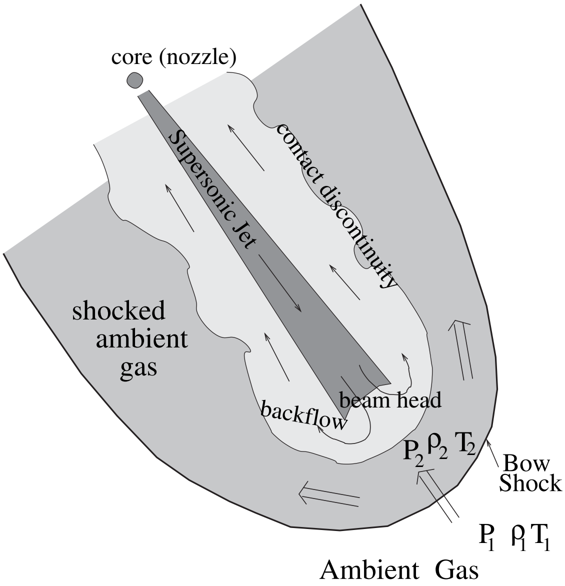

Fig. 6. In the standard model for powerful radio sources, a supersonic jet (*dark grey*) terminates at the beam head, producing a radio hotspot. Provided the shocked radio-emitting fluid forming the radio lobe (*light grey*) has enough internal energy or momentum density to drive a leading bow shock, ambient X-ray-emitting gas will be heated as it crosses the bow shock to fill the medium-grey region

Fig. 7. 8.4 GHz VLA A and B array radio image of an example high-power FRII source, the radio galaxy 3C 220.1. The resolution is 0.3×0.2 arcsec, and the image is from [145]. The angular length of the source is about 28 arcsec which, at the galaxy's redshift of $z = 0.61$, corresponds to 187 kpc ($H_o = 70$ km s^{-1} Mpc^{-1}, $\Omega_{\mathrm{m}} = 0.3$, $\Omega_{\Lambda} = 0.7$)

have bulk relativistic motion (Sect. 3) for the general description of an FRII source to hold.

The state of the X-ray gas close to a radio galaxy with a leading bow shock should reflect the source dynamics, and in addition to the X-ray emission of the ambient medium (Sect. 4), we may also expect to see hotter shocked gas surrounding the radio lobe (Fig. 6). In a simple application of the Rankine-Hugoniot conditions for a strong shock [123], the pressure, density, and temperature, respectively, in the unshocked (subscript 1) and shocked (subscript 2) regions at the head of the bow shock are related by

$$P_2/P_1 = (5\mathcal{M}^2 - 1)/4 \tag{49}$$

$$\rho_2/\rho_1 = 4\mathcal{M}^2/(\mathcal{M}^2 + 3) \tag{50}$$

$$T_2/T_1 = (5\mathcal{M}^2 - 1)(\mathcal{M}^2 + 3)/16\mathcal{M}^2 \tag{51}$$

for a monatomic gas.

Using the above equations, and converting density into an emissivity using `XSPEC`, as discussed in Subsect. 2.3, we find that in the energy band 0.8–2 keV, where *Chandra* and XMM-Newton are most sensitive, for an $\mathcal{M} = 4$ shock, the X-ray emissivity contrast between shocked and unshocked gas is a factor of 3 higher if the ambient gas is at a galaxy temperature of ~0.3 keV than if the external medium has a cluster temperature of ~4 keV.

Complications apply in reality, and in practice these are difficult to treat even with data from observatories as powerful as *Chandra* and XMM-Newton. Firstly, there is observational evidence that in supernova remnants the post-shock electrons are cooler than the ions [66, 106]. Secondly, the simple Rankine-Hugoniot equations do not take into account the fact that the bow shock around a lobe is oblique away from its head, with a consequent change in the jump conditions and the emissivity contrast [135]. However, in Subsect. 6.5, we describe how the above equations can also be applied to an overpressured lobe in the inner structure of the low-power radio galaxy Cen A. The closer such a structure is to a spherical expansion, the more normal the shock will be everywhere and the better the applicability of the above equations.

5.2 Low-Power FRI Jets

The appearance of low-power jets, in FRI radio galaxies, is quite different from that of high-power jets: compare Fig. 7 with Fig. 8. Whereas the high-power jets in FRII radio sources are generally weak features within the well-defined, and much brighter radio lobes, low-power jets are usually of high contrast against the inner radio structures. Low-power jets are usually brightest near the active galactic nucleus, and then fade gradually in brightness at larger distances from the core, although this pattern is often interrupted by patterns of bright knots in all wavebands. FRI jets show a range of morphologies, from almost straight and symmetrical two-sided jets to the bent "head-tail" sources that are characteristic of low-power extended radio sources in clusters.

While it is believed that the same basic mechanism is responsible for the generation of low-power and high-power jets, it appears that low-power

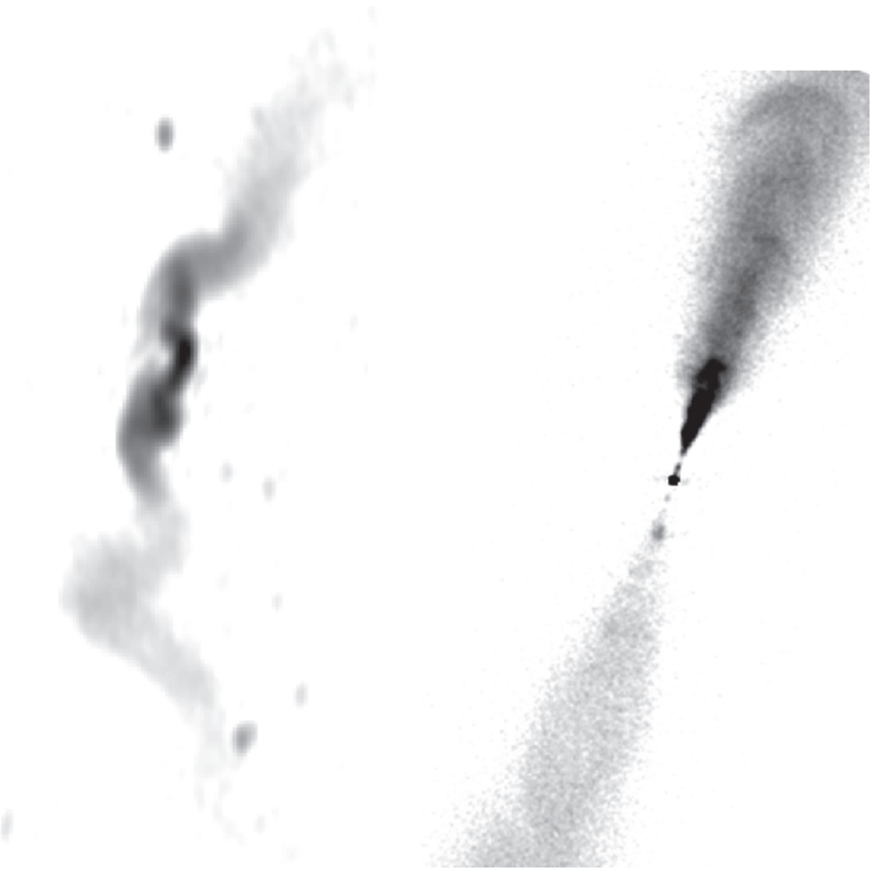

Fig. 8. Radio images of an example low-power FRI source, the radio galaxy 3C 31. The left image shows the large-scale structure. The map [80] has a resolution of 29×52 arcsec and is from a 608 MHz observations using the Westerbork Synthesis Radio Telescope. The angular length is about 40 arcmin which, at the galaxy's redshift of $z = 0.0169$, corresponds to about 830 kpc ($H_o = 70$ km s^{-1} Mpc^{-1}). The right image shows an 8.4-GHz VLA map of the the inner $\sim$18 kpc where the jets are relatively straight. The resolution is 0.25 arcsec, and the data are from [78]

jets dissipate much of their energy without developing a well-defined beam head. This is often taken to indicate that low-power jets are in good contact with the external medium in which they are embedded, so that they share momentum and energy with entrained material as they propagate. The strong velocity shear between a fast-moving (originally relativistic, based on VLBI observations of the cores of low-power radio sources [50]) jet flow and the almost stationary external medium must generate instabilities at the interface [10], and drive the flow into a turbulent state. The physics of the resulting flow is far from clear, although with simplifying assumptions [8] it has been shown that subsonic and transonic turbulent jet flows can reproduce the general trend of radio brightness and jet width for plausible confining atmospheres. More recently, model fits [57] using (41) and (45) to X-ray measurements of the thermal atmosphere of 3C 31, coupled with results from high-dynamic range radio mapping [78] (Fig. 8), have led to a self-consistent model of the flow.

The kinematic model of 3C 31 [79] involves three distinct regions in the jet: an inner ∼1 kpc long section of narrow jet where the flow is fast (relativistic) and the opening angle of the jet is small, a flaring region of ∼2 kpc during which the jet broadens rapidly and brightens in the radio, and an outer region in which the jet expands steadily with a smaller opening angle than in the flaring region. Here the jet decelerates steadily from a moderate Mach number as it picks up mass from the external medium or from the mass lost by stars embedded in the jet [73]. Buoyancy forces are important for much of the outer flow, and the jet becomes sensitive to local changes in the density of the intracluster medium, and hence is liable to deflect from straight-line motion.

While this model provides a good description of 3C 31, which displays a smoothly-varying radio brightness profile and so is amenable to simple fitting for flow speed, and is based simply on the basic conservation laws for mass, momentum, and energy [9], it represents only the overall deceleration of the jet from entrainment. This leaves unanswered important questions about the origins of the bright knots which appear in many jets, and which are interpreted as the sites of strong shocks (which are relativistic, based on proper motion studies), and about the detailed physics responsible for mass entrainment and the deceleration of the jet.

While the knots in jets are sites where intense acceleration of particles to highly relativistic energies can take place, it is clear from the diffuse X-ray and other emission between the knots that there is continuing particle acceleration even in the inter-knot regions. The structures of the acceleration regions can only be studied in detail in the closest radio galaxies, such as Cen A [58] and M 87 [99]. A number of features of the knots have become apparent through detailed multiwavelength studies, of which the most important are the presence of high-speed motions (apparent speeds greater than the speed of light having been noted in M 87), and the distinct locations for the acceleration of particles radiating in different wavebands.

6 The Interpretation of Multiwavelength Data in Terms of Physical Parameters

6.1 Minimum Energy and Magnetic Field Strength

The evidence for magnetic fields in the jets and lobes of active galaxies (Sect. 6.6) confirms the synchrotron origin of the radio emission, which is therefore an inseparable function of the magnetic-field strength and electron spectrum. To progress further it is usual to assume that the source is radiating such that its combined energy in particles and magnetic field is a minimum [20]. In this situation the energy in the magnetic field is ∼3/4 of the energy in the particles, and so this is similar to the condition in which the two are equal and the source is in "equipartition".

We can calculate the minimum-energy magnetic field for a power-law spectrum using the equations in Subsect. 2.1, and for more complicated spectra results can be obtained via numerical integration. The total energy in electrons

$$U_{\mathrm{e}} = \kappa\, m_{\mathrm{e}} c^2 \int_{\gamma_{\min}}^{\gamma_{\max}} \gamma \gamma^{-p} d\gamma = \kappa\, m_{\mathrm{e}} c^2 \int_{\gamma_{\min}}^{\gamma_{\max}} \gamma^{-2\alpha} d\gamma \tag{52}$$

From (7), the synchrotron luminosity per unit frequency is

$$L_\nu \propto \kappa \nu^{-\alpha} B^{\alpha+1} \tag{53}$$

and we can eliminate κ from (52) and (53) to give (for $\alpha \neq 0.5$)

$$U_{\mathrm{e}} \propto L_\nu \nu^\alpha B^{-(\alpha+1)} \left(\gamma_{\max}^{1-2\alpha} - \gamma_{\min}^{1-2\alpha}\right) (1-2\alpha)^{-1} \tag{54}$$

If K is the ratio of energy in heavy non-radiating particles to that in electrons, and η is the filling factor in the emission region of volume V, the energy density in particles can be written as

$$u_{\mathrm{pcl}} = C_1 B^{-(\alpha+1)} \frac{(1+K)}{\eta V} L_\nu \nu^\alpha \left(\gamma_{\max}^{1-2\alpha} - \gamma_{\min}^{1-2\alpha}\right) (1-2\alpha)^{-1} \tag{55}$$

where the constant of proportionality, C_1, is determined from synchrotron theory and can be expressed explicitly using the equations in Subsect. 2.1. The energy density in the magnetic field can be written as

$$u_{\mathrm{B}} = C_2 B^2 \tag{56}$$

where the value of the constant C_2 depends on the system of units. The value of B which minimizes the sum of the two energy densities, B_{me}, is then given by

$$B_{\mathrm{me}} = \left[\frac{(\alpha+1) C_1}{2 C_2} \frac{(1+K)}{\eta V} L_\nu \nu^\alpha \frac{\left(\gamma_{\max}^{1-2\alpha} - \gamma_{\min}^{1-2\alpha}\right)}{(1-2\alpha)} \right]^{1/(\alpha+3)} \tag{57}$$

Since luminosity density and flux density, S_ν, are related by

$$L_\nu = (1+z)^{\alpha-1} S_\nu\, 4\pi D_{\mathrm{L}}^2 \tag{58}$$

where D_{L} is the luminosity density, and the volume of a radio source can be specified [90] in terms of its angular component sizes, θ_{x}, θ_{y} and path length through the source, d, as

$$V = \theta_{\mathrm{x}} \theta_{\mathrm{y}}\, d\, D_{\mathrm{L}}^2 / (1+z)^4 \tag{59}$$

we find

$$B_{\mathrm{me}} = \left[\frac{(\alpha+1)C_1}{2C_2} \frac{(1+K)}{\eta\,\theta_{\mathrm{x}}\theta_{\mathrm{y}} d} 4\pi\, S_\nu\, \nu^\alpha\, (1+z)^{3+\alpha} \frac{\left(\gamma_{\max}^{1-2\alpha} - \gamma_{\min}^{1-2\alpha}\right)}{(1-2\alpha)} \right]^{1/(\alpha+3)} \tag{60}$$

At minimum energy, the total energy density is $u_{\mathrm{me}} = u_{\mathrm{B}}(\alpha+3)/(\alpha+1) = u_{\mathrm{pcl}}(\alpha+3)/2$. Any change in the ratio of u_{B} to u_{pcl} increases the total energy and pressure of the source. Thus if, as seen in the southwest radio lobe of Cen A (Subsect. 6.5), the minimum pressure is below that of the external medium, expansion can immediately be inferred if lobes are to stay inflated for an appreciable time. Using minimum energy, the magnetic fields in radio sources are generally estimated at 1–10 nT (10–100 μGauss).

The dependence of B_{me} on the lower and upper cut-off energies of the electron spectrum are weak, and since α is expected to be $\sim$0.6 from shock acceleration [1], or larger where energy losses have steepened the spectrum, it is the lower energy which has most effect. There is a stronger dependence of B_{me} on $(1+K)$, the ratio of energy in all particles to that in radiating electrons, and η. It is usual to adopt unity for both of these factors, which gives a true minimum value for the energy and pressure [$(\gamma-1)$ times the energy density, where in this context γ is the ratio of specific heats, which is 4/3 for a relativistic gas]. Indirect arguments are then applied to estimate if $(1+K)$ should be larger than unity or η less than unity for a particular source.

In the case of SSC, L_ν must be converted into a spectral number of photons per unit volume, n_ν [i.e., $n(\nu_o)$ in the equations of Subsect. 2.2]. For a uniform spherical source of radius r and volume V, the average value for n_ν is given by

$$n_\nu = \frac{3\,r\,L_\nu}{4\,c\,h\nu\,V} \tag{61}$$

It is instructive to consider the effects of relativistic beaming (Subsect. 3). For the case of a spherical blob, (58) is now the equation for $L_\nu \delta^{(3+\alpha)}$, and the total synchrotron luminosity in the source frame is reduced by a factor of δ^4, with u_{B} and κ, the normalization of the electron spectrum, both reduced by δ^2. SSC in the source frame is reduced by a factor of δ^6, and thus sources which are more beamed have lower SSC to synchrotron flux ratios. [The factors of δ^2 become $\delta^{2(2+\alpha)/(3+\alpha)}$ for the jet case where $\delta^{(2+\alpha)}$ should be used.]

The minimum-energy assumption can be tested by combining measurements of synchrotron and inverse-Compton emission from the same electron population. If the inverse Compton process is responsible for most of the higher-energy radiation that is measured, and the properties of the photon field are well known, the electron population is probed directly (Subsect. 2.2). In combination with measurements of radio synchrotron emission from the same electrons, both the electron density and magnetic field strength can then be estimated, and the minimum-energy assumption can be tested. Since the application of this test requires that the volume and any bulk motion of the emitting plasma be known, the best locations for testing equipartition are the radio hotspots, which are relatively bright and compact, and are thought to arise from sub-relativistic flows at jet termination (but see [46]), and old radio lobes that should be in rough overall pressure balance with the external medium. There is no reason to expect dynamical structures to be at minimum energy.

Chandra has allowed such tests to be made on a significant number of lobes and hotspots, with results generally finding magnetic field strengths within a factor of a few of their minimum energy (equipartition) values for $(1 + K) = \eta = 1$ (e.g., [19, 56]). A study of $\sim$40 hotspot X-ray detections concludes that the most luminous hotspots tend to be in good agreement with minimum-energy magnetic fields, whereas in less-luminous sources the interpretation is complicated by an additional synchrotron component of X-ray emission [59].

6.2 Jet Composition

Much of the detailed physics of jets depends on what they contain. While it is clear from their polarized radiation that jets contain fast charged particles and magnetic fields, it is less clear whether these are primary energy-carrying constituents of the jet or secondary. A number of different possibilities for the energy carriers have been suggested. The most popular are that the jets are primarily composed of electrons and positrons, or of electrons and protons, although electromagnetic (Poynting flux) jets [108] and proton-dominated jets [84, 104] have been discussed.

Since jets are presumed to obtain much of their energy from the infall of matter into a supermassive black hole, it is natural to suppose that electromagnetic radiation would carry much of the energy from the system on the smallest scales, since a plausible mechanism for the extraction of energy from the system is the twisting of magnetic field linked to the accretion disk. [82] describe the processes that might launch such a jet, but fast interactions with the plasma environment and efficient particle acceleration are expected quickly to load the field with matter. In the resulting magnetohydrodynamic flow much of the momentum would be carried by particles, although Poynting flux may carry a significant fraction of the total energy [5].

An electron-positron pair plasma is a natural consequence of the high energy density near the centres of active galaxies, and hence it might be expected that electrons and positrons would be an important, and perhaps dominant, component of the jet outflow. Acceleration of such a plasma by the strong radiation field near the active nucleus is certainly possible [93, 109], although radiation drag is an important limitation on the speeds to which the flow can be accelerated if only electron-positron pairs are present [121].

The importance of electron-positron pairs can be established by assuming that they are the major constituents of the flow, and then testing that the kinetic energy that they carry down the jet is comparable with the radiated power. While an electron-positron jet with arbitrary distribution of Lorentz factors, $N(\gamma)$, can satisfy this condition, an electron-proton jet should not extend to Lorentz factors $\gamma_{\rm min} < 100$, to avoid a high order of inefficiency and depolarization by Faraday rotation (Subsect. 6.6). Another test, based on the assumption that the hotspots at jet termination in the FRII radio galaxy Cygnus A are reverse shocks in the jet fluid [72], suggests that the jet material is an electron-positron fluid at that point.

Further information supporting the interpretation of jets as electron-positron plasmas at their point of injection has been obtained by VLBI polarization studies [134]. Here the detection of circular polarization, interpreted as arising from mode conversion, strongly suggests that, on the pc scale, jets have low $\gamma_{\rm min}$ (< 20) and hence are electron-positron plasmas (Subsect. 6.6), although [110] find that other possibilities (notably that the jet is composed of electrons and protons) are not ruled out since other mechanisms exist for the generation of circular polarization.

There are also grounds for expecting high-power jets to contain a considerable fraction of their mass in protons. Protons outnumber electrons by 100 to 1 in Galactic cosmic rays at energies at which they leak out of the galaxy before significant energy loss [138]. Some jet acceleration mechanisms can accelerate heavy components of the jets as much as, or more than, the lighter components, and, even if the jet is initially light, interactions with the external medium are expected to load it with protons (Subsect. 5.2). However, there is little direct evidence for the proton contents of jets. Polarization studies do not find evidence of embedded thermal material, and protons are relatively inefficient radiators so the synchrotron emission of relativistic protons is probably not detectable (but see [2]). It does seem to be clear that jets are of low density relative to the external medium, on the basis of numerical simulations, since the structures of dense jets are unlike those observed, but there could still be appreciable mass in the flow.

Support for the presence of protons is also found in models for the broad-band spectral energy distributions of emission from some sub-parsec scale quasar jets, where a significant proton contribution is required to boost the kinetic power sufficiently to match the total radiated power [126]. Some radio lobes in low-power radio galaxies, if assumed to be radiating at minimum energy (Subsect. 6.1), would collapse because the X-ray-emitting medium has a higher pressure [32]. Although there are several ways of boosting the internal pressure in such a situation, magnetic dominance would make the sources unusual, electron dominance can be ruled out by constraints on inverse-Compton scattering of the CMB, and non-relativistic protons are disfavoured on grounds of Faraday rotation (Subsect. 6.6), leaving a relativistic proton component most likely. In contrast, there are other cases where the radio lobes would be overpressured with respect to the ambient X-ray-emitting medium if a relativistic proton contribution were included [56]. Although this in itself is not a major difficulty, since there are open issues concerning lobe expansion (Subsect. 6.5), these are lobes for which the magnetic field can be measured (Subsect. 6.1) and the sources are close to minimum energy when only the radiating electrons are considered.

It has been known for some time that the minimum pressure in low-power jets (calculated assuming an electron-positron plasma) is typically below that of the external X-ray-emitting medium, e.g., [42, 71, 91, 140]. However, this cannot be used simply to infer that the jets are launched as an electron-proton plasma to give them the extra required pressure. Low-power jets are believed to slow down to sub-relativistic speeds via entrainment of thermal material (Subsect. 5.2), and even the most detailed hydrodynamical modeling, such as that which has been applied to 3C 31 by [79], does not decide the issue of primary jet content.

On balance it appears that the dominant energy-carrying constituents of jets are electrons and positrons, but that relativistic protons may also be important, particularly in momentum transport. However, the situation remains unclear, and there are degeneracies between the measurement of jet speed and jet composition which render this conclusion tentative.

6.3 Jet Speed

The resolved X-ray jets (excluding hotspots) that *Chandra* detects in powerful radio sources are mostly in quasars, e.g., [28, 87, 112, 113, 117, 118, 119, 120], with the bright radio sources Pictor A and Cygnus A [136, 137] being exceptions. The quasar X-ray jet emission is one-sided, on the same side as the brighter radio jet, implying that relativistic beaming is important. Two-sided X-ray emission, such as that in the quasar 3C 9 [40] most likely does not imply the presence of counter-jet emission, but rather the presence of the more isotropic lobe, hotspot, or cluster-related emission expected at some level in all sources and detected in many, e.g., [56, 145].

Currently there are several tens of quasar X-ray jet detections. They have mostly been found through targeted programs to observe bright, prominent, one-sided radio jets. In most cases there has been no pre-existing reported optical jet detection, but there has been reasonable success from follow-up work. The level of many such detections lies below an interpolation between the radio and X-ray spectra, suggesting that synchrotron emission from a single power-law distribution of electrons is not responsible for all the emission, e.g., [113, 117], although it has been pointed out that since high-energy electrons lose energy less efficiently via inverse Compton scattering in the Klein-Nishina regime, the electron spectrum should harden at high energies and in some cases may produce a synchrotron spectrum that can match observations [35].

In order to avoid the total energy in particles and magnetic field being orders of magnitude above its minimum value, as would arise from a simple SSC explanation, the most widely favored model for the X-ray emission in these cases is that the X-rays are produced by inverse Compton scattering of CMB photons by the electrons in a fast jet (Sect. 3) that sees boosted CMB radiation and emits beamed X-rays in the observer's frame [27, 127]. The model can produce sufficient X-rays with the jet at minimum energy, but only if the bulk motion is highly relativistic (bulk Lorentz factor, $\Gamma \approx 5-20$) and the jet at a small angle to the line of sight. Although such a speed and angle are supported on the small scale by VLBI measurements, at least for the source which has guided this work, PKS 0637–752 [83, 117], the jet must remain highly relativistic hundreds of kpc from the core (after projection is taken into account) for the X-rays to be produced by this mechanism. This conclusion, based on multiwavelength data, has been something of a surprise, since earlier statistical studies of the structures of powerful radio sources suggested that jet velocities average only about $0.7c$ at distances of tens of kpc from the core [54, 133].

A difficulty with the interpretation of the X-rays from radio jets as due to inverse-Compton scattering of CMB photons by a fast jet is that the observed gradients in X-ray surface brightness at the edge of the knots are sharp [128]. Since the X-rays are generated from low-energy electrons (as the CMB photons are boosted in the electron rest frame), the lifetimes of the electrons are long and it is difficult to see how X-ray knots in high-power jets can have a steeper gradient (as in PKS 0637–752, [117]) than radio knots, which are produced by electrons of similar or higher energy. Although [128] suggest that the effects of strong clumping in the jets may resolve this issue, a fast jet and the proposed mechanism is then no longer required, since such clumping would increase the SSC yield for a slow jet at minimum energy [117].

6.4 Particle Acceleration in Jets

The electrons that emit synchrotron radiation in the radio, optical, and X-ray wavebands are all ultrarelativistic (with Lorentz factors $> 10^3$). The existence of large numbers of such relativistic electrons depends on their acceleration to high energies *locally* within the jet, or hot-spot, since their lifetimes against synchrotron losses (Subsect. 2.1) are usually less than the minimum transport times from the active nuclei. (This may not be the case if proton synchrotron radiation is important [2].) Particle acceleration is generally discussed for the cases of a particle interacting with a distributed population of plasma waves or magnetohydrodynamic turbulence, or shock acceleration. Reviews of these processes may be found in [13, 36, 37] and elsewhere.

Resolved X-ray jets in active galaxies with *low* radio power are detected with *Chandra* in sources covering the whole range of orientation suggested by unified schemes, suggesting that beaming is less important than in their more powerful counterparts. The several tens of detected sources range from beamed jets in BL Lac objects [12, 101] to two-sided jets in radio galaxies [29, 58], with most X-ray jets corresponding to the brighter radio jet, e.g., [55, 57, 60, 61, 88, 144, 146]. Several of the observations have been targeted at sources already known to have optical jets, from ground-based work or HST. However, it's easier [144] to detect X-ray jets in modest *Chandra* exposures than to detect optical jets in HST snapshot surveys, because there is generally better contrast with galaxy emission in the X-ray band than in the optical.

Inverse Compton models for any reasonable photon field suggest an uncomfortably large departure from a minimum-energy magnetic field in most low-power X-ray jets, e.g., [55]. Synchrotron emission from a single electron population, usually with a broken power law, is the model of choice to fit the radio, optical, and X-ray flux densities and the relatively steep X-ray spectra, e.g., [16, 55]. X-ray synchrotron emission requires TeV-energy electrons which lose energy so fast that they must be accelerated *in situ*.

The above arguments, applied to the bright northeast jet of the nearest radio galaxy, Cen A, find in favor of X-ray synchrotron emission [74], and the proximity of Cen A allows its acceleration sites to be probed in the greatest possible detail. Unfortunately the dramatic dust lane spanning the galaxy masks any optical jet emission. Proper motion of order 0.5c, observed both in the diffuse emission and some knots of Cen A's radio jet, is indicative of bulk motion rather than pattern speed [58]. Since Cen A has a strong jet to counter-jet asymmetry, (23) then suggests that the jet is at a small angle to the line of sight. Since this contradicts the evidence based on parsec-scale properties, and other considerations, that the jet is at about 50 degrees to the line of sight, Cen A appears to be a case where intrinsic effects render (23) inapplicable. Some of the bright X-ray knots have only weak radio emission with no indication of proper motion, but with the radio emission brightening down the jet in the direction away from the nucleus. While the radio association confirms that these X-ray knots are indeed jet related, the emission profiles are not what are expected from a simple toy model where the electrons are accelerated and then advect down the jet, with the X-ray emitting electrons losing energy faster than the radio-emitting electrons. Instead, it is proposed [58] that there are obstacles in the jet (gas clouds or high-mass-loss stars). Both radio and X-ray-emitting electrons are accelerated in the standing shock of this obstacle, and a wake downstream causes further acceleration of the low-energy, radio-emitting, electrons. The resulting radio-X-ray offsets, averaged over several knots, could give the radio-X-ray offsets seen in more distant jets, e.g., [55].

The X-ray and radio emission in hot-spots are also offset in some cases [56], presumably also because of the different locations of acceleration of the particles at the different energies probed in these wavebands, if the emission is all synchrotron. However, it is expected that much of the X-ray emission should be of inverse-Compton origin, and then the offsets become harder to understand, although [46] have described a model in which decelerations of the jet plasma near a hot-spot can generate X-ray enhancements and small offsets.

Optical polarization might help us to learn more about the acceleration processes in low-power jets. In M 87 there is evidence for strong shock acceleration at the base of bright emitting regions, in compressed transverse magnetic fields [99]. A knot in the jet of M 87 has been observed to vary in the X-ray and optical on the timescale of months, consistent with shock acceleration, expansion, and energy losses [62, 100].

6.5 Pressure and Confinement

It is interesting to compare the minimum pressure (Subsect. 6.1) in radio lobes with that of the external X-ray-emitting environment. Over-pressure in the lobe would imply an expansion, which may be supersonic and should involve significant heating of the external gas. Under-pressure suggests either that the lobe is undergoing collapse (this should be rare given their prevalence) or that there is an additional component of pressure that may be in the form of relativistic protons. FRII sources are generally at high redshift where observations lack sufficient sensitivity and resolution to draw strong conclusions. However, notwithstanding the fact that there is no reason to expect dynamical structures to be at minimum energy, where tests are possible it tends to be confirmed (Subsect. 6.1), and it appears that with this assumption rough pressure balance prevails, e.g., [7, 33, 56].

The situation in low-power radio galaxies is more complicated, because the medium plays an important role in the deceleration of the jets, such that they share momentum and energy with entrained material. However, these sources have the advantage of being closer, and Subsect. 5.2 describes how the pressure profile predicted by the mass-entrainment model for deceleration of the jet in 3C 31 gives an excellent match to observations. The outer radio structures of FRIs may sometimes be buoyant, e.g., [143], and in other cases show evidence of having done significant work on the gas in pushing it aside [15, 89], or responsibility for lifting gas in hot bubbles, e.g., [31], expected to result in eventual heating, e.g., [105]. Such heating would help to explain the weakness or absence of lines from gas cooling below 1 keV in the densest central regions of galaxy and cluster atmospheres, e.g., [102]. A statistical study shows that atmospheres containing radio sources tend to be hotter than those without [32].

One place where heating is definitely expected is from gas crossing the supersonically advancing bow shock of an expanding lobe (Subsect. 5.1). It is possible to interpret X-ray cavities coincident with the inner parts of the radio lobes of Cygnus A as due to an emissivity contrast between bow-shock heated gas outside the lobes and the more easily detected ambient cluster medium [24], although the parameters of the shock are not effectively constrained by the data. More recent *Chandra* observations of Cygnus A find gas at the

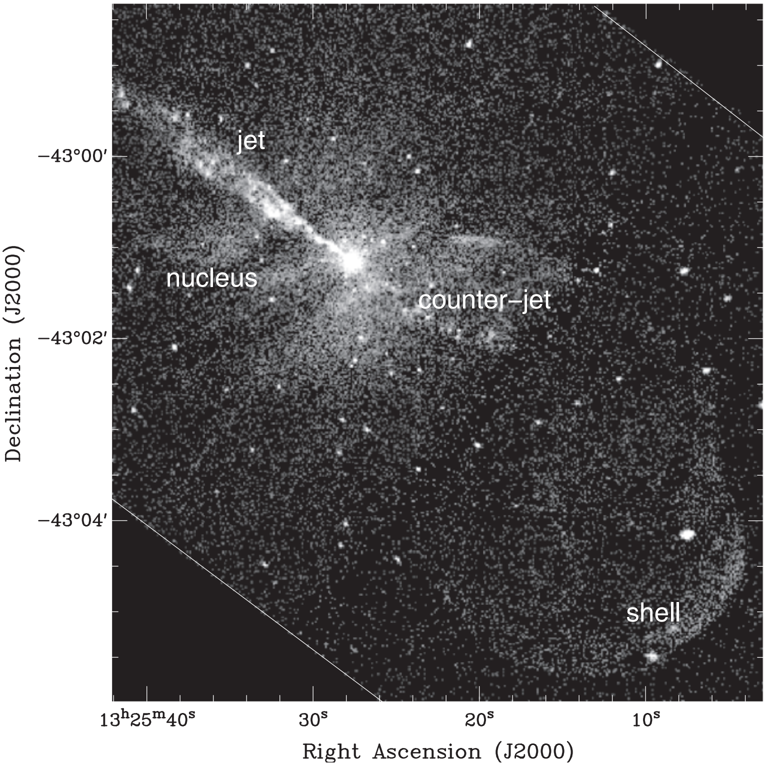

Fig. 9. 0.5–5 keV X-ray image of Cen A from a 50 ks *Chandra* exposure using the ACIS-S instrument. Pixel size is 1 arcsec. Data are exposure-corrected. Many point sources associated with Cen A's host galaxy NGC 5128 are seen, together with a diffuse background of X-ray-emitting gas of $kT \approx 0.3$ keV, and labelled X-ray structures that are in whole or part related to radio structures (see the text and [38, 58, 74, 75] for more details)

sides of the lobes to have $kT \sim 6$ keV, slightly hotter than the value of 5 keV from ambient medium at the same cluster radius, possibly indicating cooling after bow-shock heating, but again the data do not usefully constrain model parameters [122].

The first and best example of a shell of heated gas which can reasonably be attributed to supersonic expansion is not in an FRII radio source, but in Cen A (Fig. 9). High-quality *Chandra* and XMM-Newton data [74] provide the temperature and density constraints needed to test the model and measure the supersonic advance speed of the bow shock responsible for the heating. This source is an excellent example where much of the theory outlined earlier in this chapter can be applied.

Cen A is our nearest radio galaxy, at a distance of 3.4 Mpc [67] so that 1 arcsec corresponds to ~17 pc. The full extent of Cen A's radio emission

Table 1. Physical parameters of the gas in various regions of Cen A

Structure	kT (keV)	$n_{\rm p}$ ($\rm m^{-3}$)	Pressure (Pa)†	0.4–2 keV Relative Emissivity, $\mathcal{E}$
ISM (measured)	0.29	1700	1.8×10^{-13}‡	1
Behind bow shock (inferred)	6.8	6530	2.1×10^{-11}	13
Shell (measured)	2.9	20000	2.1×10^{-11}	127

† 1 Pascal = 10 dyn $\rm cm^{-2}$

‡ incorrectly reported in Table 5 of [75]

ISM and Shell pressures are thermal only. Inferred pressure includes ram pressure of $\rho_1 v_{\rm adv}^2(\mathcal{M}^2+3)/4\mathcal{M}^2$.

The minimum-energy pressure in the radio lobe is $\sim 1.4 \times 10^{-12}$ Pa.

covers several degrees on the sky [69]. Within this lies a sub-galaxy-sized double-lobed inner structure [21] with a predominantly one-sided jet to the northeast and a weak counter-jet to the southwest [58], embedded in a radio lobe with pressure 1.4×10^{-12} Pa or more, greater than the pressure in the ambient ISM ($\sim 1.8 \times 10^{-13}$ Pa; Table 1), and so which should be surrounded by a shock. Around this southwest lobe there is a shell of X-ray emitting gas which appears to have the geometry of the shocked ambient gas in Fig. 6 [75]. Although the capped lobe is around the weak counterjet, so it is not evident that the lobe is being thrust forward supersonically with respect to the external interstellar medium (ISM) by the momentum flux of an active jet, the current high internal pressure in the radio lobe ensures its strong expansion.

The temperature, proton density and pressure of the ambient ISM and the X-ray shell, taken from [75] are given in Table 1. The ambient medium is measured to have $n_{\rm p} \sim 1.7 \times 10^3$ $\rm m^{-3}$ and $kT = 0.29$ keV, whereas the shell is ten times hotter, at $kT = 2.9$ keV, and twelve times denser, with $n_{\rm p} \sim 2 \times 10^4$ $\rm m^{-3}$. From (51) and (50), we see that temperature and density measurements for both the ambient medium and the shocked gas directly test shock heating, since only two of the four parameters are required to measure the Mach number, and the other two test the model.

The most straightforward application of the equations finds that the densities and temperatures are not self-consistent. The shell's density and temperature are wrong for gas directly in contact with the bow shock. However, we can find a Mach number consistent with shocking the gas to a temperature and density such that the combined thermal and ram pressure is in pressure equilibrium with the thermal pressure of the detected shell: $\mathcal{M} = 8.5, v_{\rm adv} \approx 2400$ km $\rm s^{-1}$. The post-shock temperature is $kT_2 \sim 6.8$ keV. The 6.8 keV gas flows back from the shock, into the X-ray-detected shell at 2.9 keV. The characteristics of this undetected hotter gas are given in Table 1. In this table we also quote estimates of the relative X-ray emissivity (per unit volume) of gas in the different structures over the 0.4–2 keV energy band, where the *Chandra* response peaks and is relatively flat. The gas directly behind the bow shock has a predicted emissivity that is an order of magnitude fainter than that in the shell, accounting for its non-detection in our measurements.

The radiative timescale for material in the shell is $\sim 2 \times 10^9$ yrs (Fig. 5), which is large compared with the lobe expansion time ($< 2.4 \times 10^6$ years), so the material in the shell behaves as an adiabatic gas [3]. The shell is overpressured compared with the minimum-energy pressure in the radio lobe (in magnetic field and radiating electrons) by a factor of ~ 10. If we assume minimum energy in the lobe, despite it being a dynamical structure, and that the shell has reached equilibrium [but note that the sound-crossing time in the shell (thickness ~ 0.3 kpc, $c_{\mathrm{s}} \sim 9 \times 10^{-7}$ kpc yr^{-1}) is about 15 per cent of the maximum time we estimate it has taken the lobe to reach its current size], the shell's overpressure relative to the radio lobe could be balanced by the ram pressure from internal motions in the lobe for a moderate relativistic proton loading. Finally, the shell's kinetic energy is ~ 5 times its thermal energy, and exceeds the thermal energy of the ISM within 15 kpc of the centre of the galaxy. As the shell dissipates, most of the kinetic energy should ultimately be converted into heat and this will have a major effect on Cen A's ISM, providing distributed heating.

6.6 Magnetic Field Structures

Jets contain significant magnetic fields. The orientation of these fields is displayed by the linear polarization that they show in their radio synchrotron emission, and for many years detailed maps of the radio polarizations of jets have been used to infer the magnetic field geometries both on kpc scales, e.g., for NGC 6251 [98] and 3C 31 [43], and pc scales, e.g., for BL Lac objects [26].

The fractional linear polarization expected from an optically-thin synchrotron-emitting plasma is

$$\pi_{\mathrm{L}} = \frac{\alpha + 1}{\alpha + \frac{5}{3}} \tag{62}$$

where α is the radio spectral index, and it is assumed that the frequency of observation satisfies $\gamma_{\mathrm{min}}^2 \nu_{\mathrm{g}} \ll \nu \ll \gamma_{\mathrm{max}}^2 \nu_{\mathrm{g}}$, as for (7). The electric field of the emitted radiation is perpendicular to the magnetic field within which the emitting electron population lies, so that an observation of the plane of polarization of the radio emission can be used to infer the geometry of the magnetic field within the radio jet.

In mapping the direction of the projected magnetic field, account has to be taken of the effect of Faraday rotation of the plane of polarization by mixed thermal material and fields within and around the radio jets. For a foreground plane slab of plasma, at a wavelength of λ, the Faraday rotation is an angle

$$\theta_{\mathrm{F}} = \mathrm{RM}\,\lambda^2 \tag{63}$$

where the rotation measure along path length L is

$$\mathrm{RM} = \frac{1}{c}\int_0^L n_{\mathrm{e,th}}\, r_{\mathrm{e}}\, \nu_{\mathrm{g}}\, \cos\psi\, dz \tag{64}$$

where $n_{\mathrm{e,th}}$ is the thermal electron density, r_{e} is the classical electron radius, ν_{g} is the non-relativistic gyrofrequency, and ψ is the angle of the magnetic field to the line of sight. In SI units,

$$\mathrm{RM} = 8100 \int_0^L n_{\mathrm{e,th}}\, B_{\|}\, dz \quad \mathrm{rad\ m}^{-2} \tag{65}$$

for a magnetic field with line-of-sight component $B_{\|}$ (in Tesla), where $n_{\mathrm{e,th}}$ is in m^{-3} and the thickness L is in pc. If the thermal material is mixed with the radio jet's plasma, an effect half this size is expected. If the magnetic field and thermal material have a complicated structure, different dependencies of θ_{F} on λ are possible. Faraday rotation is measured by multi-frequency radio mapping, and uncertainties in the intrinsic plane of polarization of the radio emission are minimized by working at high radio frequencies.

Since the presence of an appreciable density of thermal material around the jets is revealed in X-rays (Sect. 4), a significant Faraday rotation is possible if this medium contains a magnetic field, and some evidence for such Faraday rotation has been seen, e.g., in Hercules A [63], with the brighter, and presumably Lorentz-boosted, jet often showing a lower Faraday rotation than the opposite jet, as might be expected from simple geometrical considerations [45, 77].

A more important consequence of the significant linear polarizations measured for radio jets, however, is that those jets cannot contain large amounts of thermal material. If they did, then the change in θ_{F} between the front and back sides of the jets would lead to linear polarizations far less than the observed values, which are sometimes near the theoretical maximum given by (62). The limits on Faraday depolarization thus lead us to conclude that the X-rays from radio jets and lobes cannot come from thermal material mixed with the jets.

Polarization mapping also reveals that the magnetic fields in radio jets are relatively well ordered. In many jets of both high and low radio power the magnetic fields are longitudinal for the first few kpc of the length of the jet, then, in lower-power jets, the magnetic field becomes perpendicular to the jet axis at larger distances from the core, often after a significant change in the jet's width [17]. Variations on this pattern are seen where bright emission knots exist in the jet. These knots appear to be associated with strong shocks, which compress the magnetic field strongly into a transverse pattern at the upstream (core-side) of the knot, after which the magnetic field pattern becomes complicated. A particularly good example [99] is M 87, where the use of HST to map the optical polarization provides additional information, since the lifetimes of the electrons emitting optical synchrotron emission are far shorter than those of radio-emitting electrons. Significant differences between the polarization structures seen in the optical and radio in M 87 suggest that the sites of acceleration are different for different electron energies, with the strongest shocks, that provide acceleration to the highest energies, appearing in the most central parts of the jet.

In the central parts of radio sources, VLBI observations have been able to detect both linear and circular polarization. A distinct difference is seen in the polarizations of the pc-scale (VLBI) jets of BL Lac objects and radio galaxies or quasars. Generally, the magnetic field orientation in the cores of the highest-power sources is along the jet [26], but in lower-power sources, such as BL Lac objects, regions of significant transverse field are seen, and change on short timescales [44]. Again this is interpreted as the effect of shock structures moving along the jets, compressing the magnetic field as they pass, although helical patterns in the flow may affect the apparent field pattern, as may the effects of relativistic aberration. The lower linear polarizations often seen in pc-scale jets may be due to depolarization caused by the superposition of many small-scale structures, with different field orientations, within the resolution of the observations, or perhaps to significant structure in the nuclear gas near the jet.

Circular polarization is generally undetectable in large-scale jets, but has been mapped in pc-scale jets using VLBI [64, 65], where it is interpreted as arising from mode conversion – the conversion of linear polarization to circular polarization by Faraday rotation in the jet plasma, although this is not certain [110]. If the X-ray emission from powerful jets on the large scale is interpreted as inverse-Compton scattering of CMB photons in a fast jet, then the population of relativistic electrons has to extend down to low energies (Lorentz factors $\gamma_{\min} = 10 - 20$, [27, 127]). On the other hand, the detection of significant polarization in VLBI jets implies that there should be few low-energy protons and electrons (which would cause excessive internal Faraday rotation), and hence that the jets should be composed of an electron-positron plasma. However, the presence of an electron-positron plasma in the VLBI jet does not necessarily imply its presence on the largest scales, where the jet may have become loaded with material gathered from the ambient medium or the radio lobe.

6.7 Cores: The Inner Jets

Radio jets extend into the cores of active galaxies, where they are faster and more compact. Special-relativistic effects then cause their brightness and variability time scales to be strong functions of jet orientation. As a result of synchrotron self-absorption, it is important to use VLBI techniques at high radio or mm frequencies to see close to the bases of jets.

The inner jets are difficult to distinguish at non-radio wavelengths because of their closeness to the central engine, which is bright at infrared to X-ray energies, and because of orientation-dependent absorption in the optical to soft-X-ray bands from gas and dust structures. These gas and dust structures are only sometimes detected, but their presence is often inferred so that models which unify various classes of active galaxy might work, e.g., [6]. They affect the central engine the most. Thus, the first problem in using multiwavelength data to gain additional insight into radio jets on small scales is to separate out the jet emission. For no waveband does telescope spatial resolution match that of radio VLBI, and so only the tool of spectral separation is usually available.

In the most core-dominated radio sources, such as bright variable BL Lac objects and quasars (often classed together as blazars), jet emission appears dominant at all energies, sometimes up to TeV. The multi-wavelength spectral energy distributions and variability time scales are used to probe the beaming parameters and the physical properties of the emitting regions, e.g., [48, 76, 125]. Correlated flares are sometimes measured across wavebands, giving support to the present of a dominant spatial region of emission, e.g., [124, 131], but otherwise uncertainties of size scales, geometries, and parameters for the competing processes of energy loss and acceleration, often force the adoption of oversimplified or poorly-constrained models for individual jets.

In the quasar population in general, there is good evidence that in the X-ray an inner jet dominates the emission from core-dominated quasars, but not lobe-dominated quasars, e.g., [18, 70, 141, 142, 148]. Radio galaxies are of particular interest, because here emission from the central engine is weak. In these sources a correlation of core soft-X-ray and radio emission, lost in lobe-dominated quasars, re-appears [22, 39, 52, 139], suggesting that jet-related X-ray emission is again dominant. In low-power radio galaxies, an optical core is often seen with HST, and is interpreted as synchrotron emission from a similar small-scale emitting region [23, 30, 53, 132]. However, at higher X-ray energies, a number of radio galaxies show clear evidence of a hard, highly absorbed continuum, sometimes accompanied by Fe-line emission, e.g., [51, 129]. Here the decreasing jet output is leading to the central engine becoming increasingly dominant. Both X-ray components can sometimes be distinguished in the same spectrum, e.g., [33, 38].

The richness of high-energy structure in larger-scale radio jets has been revealed because it is well resolved with HST and *Chandra*. Our knowledge of the inner jets is limited by the confusion of components that has now been lifted for the jets discussed earlier in this chapter.

7 Conclusion

The new results on radio jets which have resulted from complementary X-ray and optical observations have brought some surprises. Firstly, synchrotron X-ray jets are common in low-power sources, which implies that the intrinsic electron spectrum continues to TeV energies, and requires substantial *in situ* particle acceleration. Secondly, the detection of many quasar X-ray jets, most commonly interpreted as due to beamed CMB photons, would suggest that highly relativistic bulk flows exist far from the cores, contradicting earlier statistical studies of radio sources [54, 133]. Jet theory has had some pleasing successes, such as the agreement of the X-ray pressure profile with the prediction from a hydrodynamical model for 3C 31 [79].

There is still much observational work to be done. Firstly, there is considerable bias in the jets which have been observed in the X-ray, and we need observations of unbiased samples over broader luminosity and redshift ranges, together with observations of the X-ray-emitting medium through which the jets propagate. The measurements should elucidate the relative importance of synchrotron emission, inverse Compton scattering, and relativistic beaming, refine our knowledge of the source energetics, and improve constraints on jet composition and speed. Secondly, we need more deep X-ray observations (and refined theory) to understand jet-lobe/intercluster medium interactions. Finally, to study acceleration sites and processes, deeper and more detailed multiwavelength mapping, spectroscopy, and temporal monitoring is required. In combination with multifrequency polarization measurements, such data could map the spatial distributions and follow the acceleration of the electrons responsible for the radiation in the radio to X-ray bands.

It should, however, be acknowledged that much basic physics of jets is still not well understood, e.g., the origin of the magnetic field, the method of jet production and collimation, and the effects of turbulence. Much theoretical work in these areas is necessary to relate measurements to the processes occurring in the fascinating jet flows that are observed.

References

1. A. Achterberg, Y.A. Gallant, J.G. Kirk, A.W. Guthmann: MNRAS **328**, 393 (2001)
2. F.A. Aharonian: MNRAS **322**, 215 (2002)
3. P. Alexander: MNRAS **335**, 610 (2002)
4. K. Arnaud, B. Dorman:
   `XSPEC`, An X-ray Spectral Fitting Package, version 11.3
   `http://heasarc.gsfc.nasa.gov/docs/software/lheasoft/xanadu/xspec`
   (2003)
5. S. Appl, M. Camenzind: A&A **274**, 699 (1993)
6. P.D. Barthel: ApJ **336**, 606 (1989)

7. E. Belsole, D.M. Worrall, M.J. Hardcastle, M. Birkinshaw, C.R. Lawrence: MNRAS **352**, 924 (2004)
8. G.V. Bicknell: ApJ **286**, 68 (1984)
9. G.V. Bicknell: ApJ **422**, 542 (1994)
10. M. Birkinshaw: The stability of jets. In *Beams and Jets in Astrophysics*, ed. by P.A. Hughes (Cambridge University Press), pp 278–341 (1991)
11. M. Birkinshaw, D.M. Worrall: ApJ **412**, 568 (1993)
12. M. Birkinshaw, D.M. Worrall, M.J. Hardcastle: MNRAS **335**, 142 (2002)
13. R.D. Blandford, D. Eichler: Phys. Rep. **154**, 1 (1987)
14. G.R. Blumenthal, R.J. Gould: Rev. Mod. Phys. **42**, 237 (1970)
15. H. Böhringer, W. Voges, A.C. Fabian, A.C. Edge, D.M. Neumann: MNRAS **264**, L25 (1993)
16. H. Böhringer, E. Belsole, J. Kennea et al.: A&A **365**, L181 (2001)
17. A. Bridle, R.A. Perley: ARAA **22**, 319 (1984)
18. I.W.A. Browne, D.W. Murphy: MNRAS **226**, 601 (1987)
19. G. Brunetti, M. Bondi, A. Comastri, G. Setti: A&A **381**, 795 (2002)
20. G.R. Burbidge: ApJ **124**, 416 (1956)
21. J.O. Burns, E.D. Feigelson, E.J. Schreier: ApJ **273**, 128 (1983)
22. C.M. Canosa, D.M. Worrall, M.J. Hardcastle, M. Birkinshaw: MNRAS **310**, 30 (1999)
23. A. Capetti, A. Celotti, M. Chiaberge, H.R. de Ruiter, R. Fanti, R. Morganti, P. Parma: A&A **383**, 104 (2002)
24. C.L. Carilli, R.A. Perley, D.E. Harris: MNRAS **270**, 173 (1994)
25. A. Cavaliere, R. Fusco-Femiano: A&A **70**, 677 (1978)
26. T.V. Cawthorne, D.C. Gabuzda: MNRAS **278**, 861 (1996)
27. A. Celotti, G. Ghisellini, M. Chiaberge: MNRAS **321**, L1 (2001)
28. G. Chartas, D.M. Worrall, M. Birkinshaw et al.: ApJ **542**, 655 (2000)
29. M. Chiaberge, R. Gilli, F.D. Macchetto, W.B. Sparks, A. Capetti: ApJ **582**, 645 (2003)
30. M. Chiaberge, A. Capetti, A. Celotti: A&A **349**, 77 (1999)
31. E. Churazov, M. Brüggen, C.R. Kaiser, H. Böhringer, W. Forman: ApJ **554**, 261 (2001)
32. J.H. Croston, M.J. Hardcastle, M. Birkinshaw, D.M. Worrall: MNRAS **346**, 1041 (2003)
33. J.H. Croston, M. Birkinshaw, M.J. Hardcastle, D.M. Worrall: MNRAS **353**, 879 (2004)
34. C.D. Dermer: ApJ **446**, L63 (1995)
35. C.D. Dermer, A.M. Atoyan: ApJ **568**, L81 (2002)
36. L.O'C. Drury: Rep. Prog. Phys. **46**, 973 (1983)
37. J.A. Eilek, P.A. Hughes: Particle acceleration and magnetic field evolution. In *Beams and Jets in Astrophysics*, ed. by P.A. Hughes (Cambridge University Press), pp 428–483 (1991)
38. D.A. Evans, R.P. Kraft, D.M. Worrall, M.J. Hardcastle, C. Jones, W.R. Forman, S.S. Murray: ApJ **612**, 786 (2004)
39. G. Fabbiano, L. Miller, G. Trinchieri, M. Longair, M. Elvis: MNRAS **277**, 115 (1984)
40. A.C. Fabian, A. Celotti, R.M. Johnstone: MNRAS **338**, L7 (2003)
41. B.L. Fanaroff, J.M. Riley: MNRAS **167**, 31P (1974)
42. L. Feretti, R. Fanti, P. Parma, S. Massaglia, E. Trussoni, W. Brinkmann: A&A **298**, 699 (1995)

43. E.B. Fomalont, A.H. Bridle, A.G. Willis, R.A. Perley: ApJ **237**, 418 (1980)
44. D.C. Gabuzda, T.V. Cawthorne: MNRAS **319**, 1056 (2000)
45. S.T. Garrington, J.P. Leahy, R.G. Conway, R.A. Laing: Nature **331**, 147 (1988)
46. M. Georganopoulos, D. Kazanas: ApJ **589**, L5 (2003)
47. G. Ghisellini, A. Celotti: MNRAS **327**, 739 (2001)
48. G. Ghisellini, A. Celotti, G. Fissati, L. Maraschi, A. Comastri: MNRAS **301**, 451 (1998)
49. V.L. Ginzburg, S.I. Syrovatski: The origin of cosmic rays. Pergamon Press, Oxford (1964)
50. G. Giovannini, W.D. Cotton, L. Feretti, L. Lara, T. Venturi: ApJ **552**, 508 (2001)
51. M. Gliozzi, R.M. Sambruna, W.N. Brandt: A&A **408**, 949 (2003)
52. M.J. Hardcastle, D.M. Worrall: MNRAS **309**, 969 (1999)
53. M.J. Hardcastle, D.M. Worrall: MNRAS **314**, 359 (2000)
54. M.J. Hardcastle, P. Alexander, G.G. Pooley, J.M. Riley: MNRAS **304**, 135 (1999)
55. M.J. Hardcastle, M. Birkinshaw, D.M. Worrall: MNRAS **326**, 1499 (2001)
56. M.J. Hardcastle, M. Birkinshaw, R.A. Cameron, D.E. Harris, L.W. Looney, D.M. Worrall: ApJ **581**, 948 (2002)
57. M.J. Hardcastle, D.M. Worrall, M. Birkinshaw, R.A. Laing, A.H. Bridle: MNRAS **334**, 182 (2002)
58. M.J. Hardcastle, D.M. Worrall, R.P. Kraft, W.R. Forman, C. Jones, S.S. Murray: ApJ **593**, 169 (2003)
59. M.J. Hardcastle, D.E. Harris, D.M. Worrall, M. Birkinshaw: ApJ **612**, 729 (2004)
60. D.E. Harris, H. Krawczynski, G.B. Taylor: ApJ **578**, 60 (2002)
61. D.E. Harris, A. Finoguenov, A.H. Bridle, M.J. Hardcastle, R.A. Laing: ApJ **580**, 110 (2002)
62. D.E. Harris, J.A. Biretta, W. Junor, E.S. Perlman, W.B. Sparks, A.S. Wilson: ApJ **586**, L41 (2003)
63. N.A.B. Gizani, J.P. Leahy: MNRAS **342**, 399 (2003)
64. D.C. Homan, J.F.C. Wardle: AJ **118**, 1942 (1999)
65. D.C. Homan, J.F.C. Wardle: ApJ **602**, L13 (2004)
66. U. Hwang, A. Decourchelle, S.S. Holt, R. Petre: ApJ **581**, 1101 (2002)
67. F.P. Israel: A&A Rev **8**, 237 (1998)
68. D.L. Jones: ApJ **309**, L5 (1986)
69. N. Junkes, R.F. Haynes, J.I. Harnett, D.L. Jauncey: A&A **269**, 29 (1993)
70. A. Kembhavi: MNRAS **264**, 683 (1993)
71. N.E.B. Killeen, G.V. Bicknell, R.D. Ekers: ApJ **325**, 180 (1988)
72. M. Kino, F. Takahara: MNRAS **349**, 336 (2004)
73. A. Komissarov: MNRAS **269**, 394 (1994)
74. R.P. Kraft, W.R. Forman, C. Jones, S.S. Murray, M.J. Hardcastle, D.M. Worrall: ApJ **569**, 54 (2002)
75. R.P. Kraft, S.E. Vázquez, W.R. Forman, C. Jones, S.S. Murray, M.J. Hardcastle, D.M. Worrall, E. Churazov: ApJ **592**, 129 (2003)
76. H. Krawczynski, R. Sambruna, A. Kohnle et al.: ApJ **559**, 187 (2001)
77. R.A. Laing: Nature **331**, 149 (1988)
78. R.A. Laing, A.H. Bridle: MNRAS **336**, 328 (2002)

79. R.A. Laing, A.H. Bridle: MNRAS **336**, 1161 (2002)
80. J.P. Leahy: Atlas of radio sources. `http://www.jb.man.ac.uk/atlas/` (1998)
81. M.S. Longair: High Energy Astrophysics. 2nd ed. Cambridge University Press (1994)
82. R.V.E. Lovelace, H. Li, A.V. Koldoba, G.V. Ustyugova, M.M. Romanova: ApJ **572**, 445 (2002)
83. J.E.J. Lovell, S.J. Tingay, B.G. Piner et al.: VSOP and ATCA observations of PKS 0637–752. In *Astrophysical Phenomena Revealed by Space VLBI*, ed. by H. Hirabayashi, P.G. Edwards, D.W. Murphy (ISAS, Japan), pp 215–218 (2000)
84. K. Mannheim, P.L. Biermann: A&A **221**, 211 (1989)
85. A.P. Marscher: ApJ **264**, 296 (1983)
86. A.P. Marscher, F.E. Marshall, R.F. Mushotzky, W.A. Dent, T.J. Valonek, M.F. Hartman: ApJ **233**, 498 (1979)
87. H.L. Marshall, D.E. Harris, J.P. Grimes et al.: ApJ **549**, L167 (2001)
88. H.L. Marshall, B.P. Miller, D.S. Davis, E.S. Perlman, M. Wise, C.R. Canizares, D.E. Harris: ApJ **564**, 683 (2002)
89. B.R. McNamara, M. Wise, P.E.J. Nulsen et al.: ApJ **534**, L135 (2000)
90. G. Miley: ARAA **18**, 165 (1980)
91. R. Morganti, R. Fanti, I.M. Gioia, D.E. Harris, P. Parma, H. de Ruiter: A&A **189**, 11 (1988)
92. J.F. Navarro, C.S. Frenk, S.D.M. White: MNRAS **275**, 720 (1995)
93. S.L. O'Dell: ApJ **243** L147 (1981)
94. M.J.L. Orr, I.W.A. Browne: MNRAS **200**, 1067 (1982)
95. F.N. Owen, M.J. Ledlow: The FRI/II break and the bivariate luminosity function in Abell clusters of galaxies. In *The Physics of Active Galaxies*, ed. by G.V. Bicknell, M.A. Dopita, P.J. Quinn (ASP Conf. Proc 54), pp 319–323 (1994)
96. A.G. Pacholczyk: Radio Astrophysics. W.H. Freeman and Co., San Francisco (1970)
97. T.J. Pearson, J.A. Zensus: Superluminal Radio Sources: Introduction. In *Superluminal Radio Sources*, ed. by J.A. Zensus, T.J. Pearson (Cambridge University Press), pp 1–11 (1987)
98. R.A. Perley, A.H. Bridle, A.G. Willis: ApJS **54**, 291 (1984)
99. E.S. Perlman, J.A. Biretta, F. Zhou, W.B. Sparks, F.D. Macchetto: AJ **117**, 2185 (1999)
100. E.S. Perlman, D.E. Harris, J.A. Biretta, W.B. Sparks, F.D. Macchetto: ApJ **599**, L65 (2003)
101. J.E. Pesce, R.M. Sambruna, F. Tavecchio, L. Maraschi, C.C. Cheung, C.M. Urry, R. Scarpa: ApJ **556**, L79 (2001)
102. J.R. Peterson, F.B.S. Paerels, J.S. Kaastra et al.: A&A **365**, L104 (2001)
103. E.S. Phinney: Central Radio Sources. In *Astrophysics of active galaxies and quasi-stellar object*, (University Science Books, Mill Valley CA), pp 453–496 (1985)
104. R.J. Protheroe, A.-C. Donea, A. Reimer: ApPh **19**, 599 (2003)
105. V. Quilis, R.G. Bower, M.L. Balogh: MNRAS **328**, 1091 (2001)
106. C.E. Rakowski, P. Ghavamian, J.P. Hughes: ApJ **590**, 846 (2003)
107. M.J. Rees: Nature **211**, 468 (1966)
108. M.J. Rees: Nature **229**, 312 and erratum 510 (1971)

109. N. Renaud, G. Henri: MNRAS **300**, 1047 (1998)
110. M. Ruszkowski, M.C. Begelman: ApJ **573**, 485 (2002)
111. G.R. Rybicki, A.P. Lightman: Radiative processes in astrophysics. Wiley, New York (1979)
112. R.M. Sambruna, C.M. Urry, F. Tavecchio, L. Maraschi, R. Scarpa, G. Chartas, T. Muxlow: ApJ **549**, L161 (2001)
113. R.M. Sambruna, L. Maraschi, F. Tavecchio, C.M. Urry, C.C. Cheung, G. Chartas, R. Scarpa, J.K. Gambill: ApJ **571**, 206 (2002)
114. C.L. Sarazin: Rev. Mod. Phys. **58**, 1 (1986)
115. P.A.G. Scheuer: Tests of beaming models. In *Superluminal Radio Sources*, ed. by J.A. Zensus, T.J. Pearson, (Cambridge University Press), pp 1–113 (1987)
116. P.A.G. Scheuer, A.C.S. Readhead: Nature **277**, 182 (1979)
117. D.A. Schwartz, H.L. Marshall, J.E.J. Lovell et al.: ApJ **540**, L69 (2000)
118. A. Siemiginowska, J. Bechtold, T.L. Aldcroft, M. Elvis, D.E. Harris, A. Dobrinska: ApJ **570**, 543 (2002)
119. A. Siemiginowska, R.K. Smith, T.L. Aldcroft, D.A. Schwartz, F. Paerels, A.O. Petric: ApJ **598**, L15 (2003)
120. A. Siemiginowska, C. Stanghellini, G. Brunetti et al.: ApJ **595**, 643 (2003)
121. M. Sikora, H. Sol, M.C. Begelman, G.M. Madejski: MNRAS **280**, 781 (1996)
122. D.A. Smith, A.S. Wilson, K.A. Arnaud, Y. Terashima, A.J. Young: ApJ **565**, 195 (2002)
123. L. Spitzer: Physical Processes in the Interstellar Medium. Wiley, New York (1978)
124. T. Takahashi, J. Kataoka, G. Madejski et al.: ApJ **542**, 535 (2000)
125. G. Tagliaferri, M. Ravasio, G. Ghisellini et al.: A&A **400**, 477 (2003)
126. F. Tavecchio, L. Maraschi, G. Ghisellini et al.: ApJ **543**, 535 (2000)
127. F. Tavecchio, L. Maraschi, R.M. Sambruna, C.M. Urry: ApJ **544**, L23 (2000)
128. F. Tavecchio, G. Ghisellini, A. Celotti: A&A **403**, 83 (2003)
129. S. Ueno, K. Koyama, M. Nishida, S. Yamauchi, M.J. Ward: ApJ **431**, L1 (1994)
130. C.M. Urry, P. Padovani: PASP **107**, 803 (1995)
131. C.M. Urry, A. Treves, L. Maraschi: ApJ **486**, 799 (1997)
132. G.A. Verdoes Kleijn, S.A. Baum, P.R. de Zeeuw, C.P. O'Dea: AJ **123**, 1334 (2002)
133. J.F.C. Wardle, S.E. Aaron: MNRAS **286**, 425 (1997)
134. J.F.C. Wardle, D.C. Homan, R. Ojha, D.H. Roberts: Nature **395**, 457 (1998)
135. A.G. Williams: Numerical Simulations of Radio Source Structure. In *Beams and Jets in Astrophysics*, ed. by P.A. Hughes (Cambridge University Press), pp 342–378 (1991)
136. A.S. Wilson, A.J. Young, P.L. Shopbell: ApJ **547**, 740 (2001)
137. A.S. Wilson, A.J. Young, P.L. Shopbell: Chandra X-ray Observations of Cygnus A. and Pictor A. In: *Particles and Fields in Radio Galaxies*, ASP Conf. Ser. Vol. 250, ed. by R.A. Laing, K.M. Blundell (ASP, San Francisco), pp 213–223 (2001)
138. A.W. Wolfendale: Introductory cosmic rays. In: *Origin of Cosmic Rays*, NATO Advanced Study Institute Series, ed. by J.L. Osborne, A.W. Wolfendale (Reidel, Dordrecht), pp 1–12 (1975)
139. D.M. Worrall, M. Birkinshaw: ApJ **427**, 134 (1994)
140. D.M. Worrall, M. Birkinshaw: ApJ **530**, 719 (2000)

141. D.M. Worrall, P. Giommi, H. Tananbaum, G. Zamorani: ApJ **313**, 596 (1987)
142. D.M. Worrall, C.R. Lawrence, T.J. Pearson, A.C.S. Readhead: ApJ **420**, L17 (1994)
143. D.M. Worrall, M. Birkinshaw, R.A. Cameron: ApJ **449**, 93 (1995)
144. D.M. Worrall, M. Birkinshaw, M.J. Hardcastle: MNRAS **326**, L7 (2001)
145. D.M. Worrall, M. Birkinshaw, M.J. Hardcastle, C.R. Lawrence: MNRAS **326**, 1127 (2001)
146. D.M. Worrall, M. Birkinshaw, M.J. Hardcastle: MNRAS **343**, L73 (2003)
147. D.M. Worrall, M.J. Hardcastle, T.J. Pearson, A.C.S. Readhead: MNRAS **347**, 632 (2004)
148. G. Zamorani: X-ray properties of extragalactic compact objects. In: *VLBI and compact radio sources*, IAU Symp. 110 ed. by R. Fanti, K. Kellermann, G. Setti (Reidel, Dordrecht), pp 85–94 (1984)

The Broad-Line Region in Active Galactic Nuclei

B.M. Peterson

Department of Astronomy, The Ohio State University, 140 West 18th Avenue,
Columbus, OH 43210, USA
`peterson@astronomy.ohio-state.edu`

Abstract. We review the basic observed and inferred properties of the broad emission-line region in AGNs, as well as the basics of the reverberation-mapping technique that can be used to determine the size and structure of the region. We argue that the current best evidence points to a multi-component line-emitting region, with a disk-like structure, possibly an extension of the accretion disk itself, and a disk wind being strong candidates for the origin of the broad-line emission.

1 Introduction

Some 40 years after the discovery of quasars and 60 years after the publication of C.K. Seyfert's initial observations of high central surface brightness galaxies, we are finally quite certain that active galactic nuclei (AGNs) are powered by accretion onto supermassive collapsed objects. While many important details remain poorly understood, the black-hole/accretion-disk paradigm is now reasonably secure. In contrast, however, we still have no self-consistent models of the nuclear regions that produce (a) the broad emission lines that are so prominent in the UV/optical spectra of AGNs and (b) the strong absorption features seen in the X-ray/UV spectra. Given the proximity of these regions to the central engine, it seems likely that these are some manifestation of the accretion process and related outflow; there is a good deal of empirical evidence that connects these spectral features to disk-related outflows. Unfortunately, solid information about the broad-line region (BLR) is hard to come by: the BLR is spatially unresolved in even the nearest AGNs and the information in line profiles, sampling only one of six dimensions in phase space, is highly ambiguous.

In this review, we will concentrate on a few things that we can infer about the nature of the BLR, and reverberation mapping, arguably the most promising technique for exploring the nature of the BLR. The scope of this review is limited to a few basic topics and represents, of course, the author's own highly biased personal view.

B.M. Peterson: *The Broad-Line Region in Active Galactic Nuclei*, Lect. Notes Phys. **693**, 77–100 (2006)
DOI 10.1007/3-540-34621-X_3

2 Basic Inferences about the Line-Emitting Gas

The relative strengths of the various emission lines in AGN spectra are very similar to those emitted by a wide variety of astrophysical plasmas: to some low order of approximation, the emission lines of planetary nebulae, H II regions, supernova remnants, and AGN emission-line regions are all quite similar. The reason is simple: photoionization equilibrium of all these gases is achieved at the same temperature, $T \approx 10^4$ K, e.g., [20, 21, 23, 28].

Photoionization equilibrium is attained when the rate of photoionization is balanced by the rate of recombination. The conditions where this occurs can be expressed in terms of an "ionization parameter"

$$U = \frac{Q_{\rm ion}({\rm H})}{4\pi r^2 c n_e}\ , \tag{1}$$

where $Q_{\rm ion}({\rm H})$ is the number of H-ionizing photons produced per second by the central source, i.e.,

$$Q_{\rm ion}({\rm H}) = \int \frac{L_\nu}{h\nu}\, d\nu\ , \tag{2}$$

and the integral is over ionizing photon energies and L_ν is the specific luminosity of the ionizing source. The ionization parameter is thus the ratio of the density of ionizing photons $Q_{\rm ion}/4\pi r^2 c$, reflecting ionization rate, divided by the electron density n_e, reflecting recombination rate, at the face of a cloud exposed to the radiation. Over a rather broad range in U, photoionization equilibrium is realized at electron temperatures in the range 10,000 to 20,000 K. As a result, the line spectra emitted by such gases are very similar, since variations in elemental abundances are also relatively minor. There are, of course, more subtle spectral differences among these various 10^4 K gases which are attributable to differences in gas density n_e, the input ionizing spectrum, the gas dynamics, and elemental abundances (which appear to be approximately solar or slightly enhanced in AGN emission-line regions) [14].

Since at least the work of Khachikian and Weedman in the mid-1970s, it has been recognized that AGN emission-line spectra are kinematically composite. The "narrow components" have Doppler widths usually less than around 500 $\rm km\,s^{-1}$; these emission lines arise in relatively low-density ($n_e \approx 10^3\ \rm cm^{-3}$) gas that is spatially extended – indeed, the narrow-line region (NLR) is at least partially resolved in some of the nearest AGNs. In contrast, the "broad components" have Doppler widths in the range $\sim$1000 to 25,000 $\rm km\,s^{-1}$ and arise in gas of fairly high density by nebular standards (i.e., $n_e > 10^9\ \rm cm^{-3}$), as determined from the weakness of certain metastable and forbidden lines that are relatively more prominent in lower-density gases.

Indeed, the gas density leads us to an important distinction that is not obvious based on line widths alone: the widest narrow lines and the narrowest broad lines have similar Doppler widths, around 1000 $\rm km^{\,-1}$ (though not in the same object: in general, there is a correlation between the widths of the narrow and broad lines in AGN spectra). However, even the subclass of broad-line objects known as narrow-line Seyfert 1 (NLS1) galaxies still have distinct "broader" and "narrower" features, with the difference between them being the absence of the "broader" component in the forbidden or semiforbidden lines, i.e., from the high-density gas in which these transitions are collisionally suppressed.

Temperatures of order 10^4 K correspond to thermal line widths of order $10\,\mathrm{km\,s^{-1}}$, so it is clear that in both the NLR and the BLR, the gas moves supersonically. This in turn suggests some kind of organized flow, or a system of discrete "clouds." The especially large Doppler widths of the broad lines immediately suggest that they may arise in a deep gravitational potential, which makes the broad lines especially valuable as probes of the central source.

Compared to the NLR, the actual amount of emission-line gas required to produce the broad emission lines can be quite modest as line emission is very efficient in high-density gases (the emissivity per unit volume is proportional to n_e^2). We also observe that the emission-line fluxes vary with the continuum flux, but with a short time delay. From this we infer that the broad-line emitting gas is photoionized and optically thick to ionizing continuum radiation. Moreover, we conclude that the BLR must be fairly small from light-travel time arguments. Conversely, the narrow lines do *not* vary on short time scales; the region is spatially extended, geometrically diluting the variable continuum signal over a large volume, and the recombination timescale is very long, further smearing out any temporal variations.

The kinematics of the line-emitting gas remains problematic. In the case of the NLR, the narrow-emission line widths are correlated with the stellar bulge velocity dispersions, albeit with considerable scatter in the correlation. Nevertheless, this suggests that gravitation provides an important component. However, there is also considerable evidence for interaction between the narrow lines and radio-emitting plasma that is being ejected from the nucleus.

How the BLR gas is moving, whether in infall, outflow, or orbital motion, remains unknown. Emission-line profiles alone only weakly constrain the possibilities because there are a wide variety of profoundly different kinematic models that yield similar line profiles. Broad-line profiles are distinctly non-Gaussian, and have sometimes been described as "logarithmic," i.e., the flux at some displacement $\Delta\lambda$ from line center is proportional to $-\ln \Delta\lambda$, for $\Delta\lambda$ not too close to line center. In many cases, however, the line profiles have some structure, variously described as "bumps" or "humps" or, in other cases as "asymmetric wings" or "shelves." Such features can be either prominent or subtle, and they can change over long time scales. Indeed, the persistence of features in profiles over longer than dynamical time scales $\tau_{\mathrm{dyn}} \approx R/\Delta V$, where R is the size of the region and ΔV represents a typical velocity, strongly suggests that the BLR has some kind of regular structure

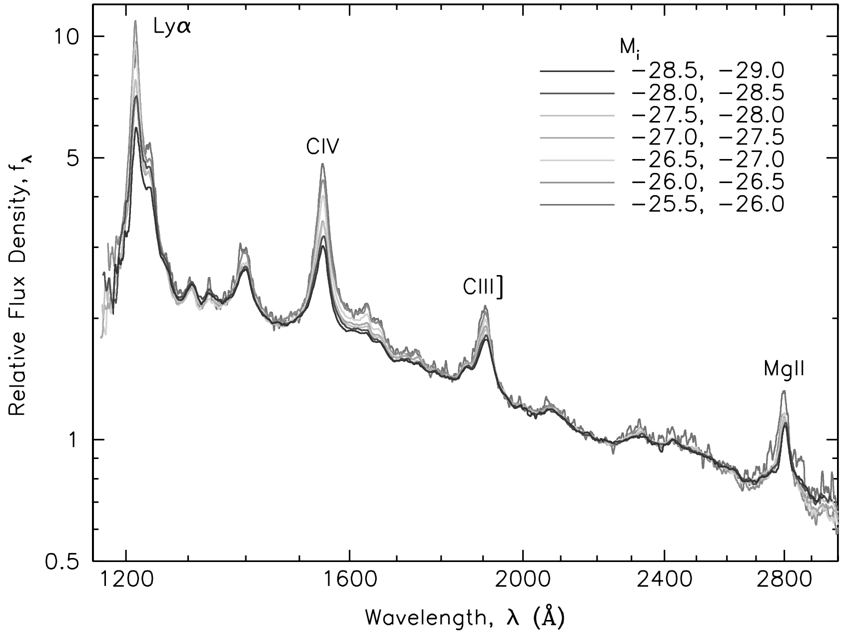

Fig. 1. Composite spectra from the Sloan Digital Sky Survey, binned by luminosity (M_i). Note the similarity of the spectra, with the exception of the "Baldwin Effect" in C IV λ1549; i.e., relative to the continuum and the other lines, C IV is stronger in lower-luminosity objects. From Vanden Berk et al., 2004, in "AGN Physics with the Sloan Digital Sky Survey" ed. G.T. Richards and P.B. Hall (San Francisco: Astronomical Society of the Pacific)

or symmetry, though variability studies, as discussed below, show that the gas is not in simple spherical outflow (like a supernovae explosion) or infall (i.e., Bondi accretion).

The average line spectra of AGNs are very similar over a wide range of luminosity, as shown in Fig. 1. This suggests that in addition to temperature, particle densities and ionization parameters are quite similar. An important exception to this statement is the behavior of the C IV λ1549 emission line; relative to the continuum, C IV is weaker in more luminous objects (i.e., its equivalent width decreases with luminosity), a well-known anticorrelation known as the Baldwin Effect. There are other obvious differences among AGNs, and many of the more subtle differences are correlated with one another; these correlated properties stand out prominently in "principal component analysis." The strongest set of correlations, known as "Eigenvector 1," reveals that strong optical Fe II emission is negatively correlated with the width of the Hβ broad component and the strength of the narrow [O III] λ5007 line, e.g., [5]. It has been speculated by many investigators that Eigenvector 1 measures the Eddington ratio $\dot{M}/\dot{M}_{\rm Edd}$, i.e., the accretion rate relative to the rate required to produce the Eddington luminosity; in terms of the observables, this is essentially the luminosity to mass ratio.

3 Broad Emission-Line Variability and Theory of Reverberation Mapping

As noted above, the broad emission-line fluxes are observed to vary in response to continuum variations, with a short (usually days to weeks for typical Seyfert 1 galaxies) time delay that is attributable to the light travel time across the BLR, i.e., $\tau_{\rm lt} = R/c$. This affords a potentially powerful tool; by closely examining the velocity-resolved response of the broad lines to continuum variations, the kinematics and geometry of the BLR can be determined, or at least highly constrained. Moreover, while the observations have to be done with great attention to quality and homogeneity, they are not otherwise a serious observational challenge since they involve prominent emission lines in the UV/optical spectra of relatively bright AGNs.

3.1 Basic Assumptions

The reverberation method as it is currently applied to AGNs rests on some simple assumptions, e.g., [4, 22, 24], which to some extent can be justified *ex post facto*:

1. *The continuum originates in a single central source.* The continuum source in typical Seyfert galaxies (10^{13} – 10^{14} cm) is much smaller than the BLR, which reverberation reveals to be $\sim 10^{16}$ cm in these objects. Note in particular that it is not necessary to assume that the radiation from the central source is emitted isotropically.
2. *The light travel time across the BLR is the most important time scale.* Specifically, we assume that that cloud response time is instantaneous, which is a good assumption given the high densities of the BLR gas. The response time is given by the recombination time scale $\tau_{\rm rec} = (n_e \alpha_{\rm B})^{-1} \approx 0.1 (10^{10}\,{\rm cm}^{-3}/n_e)$ hours, where $\alpha_{\rm B}$ is the hydrogen case B recombination coefficient and n_e is the particle density. It is also assumed that the BLR structure is stable over the duration of the reverberation experiment, which is typically no longer than several months for Seyfert galaxies, but can be much longer for high-luminosity quasars. For typical low-luminosity reverberation-mapped AGNs (i.e., Seyfert galaxies), $\tau_{\rm dyn} \approx$ 3–5 years.
3. *There is a simple, though not necessarily linear, relationship between the observed continuum and the ionizing continuum.* The key element is that the ionizing continuum and observable (either satellite UV or optical) continuum vary in phase. In only two well-studied cases, NGC 7469 and Akn 564, has this assumption been critically tested. In both cases, time lags between the continuum variations at the shortest UV wavelengths

and those at the longest optical wavelengths are of order ∼1 day. In terms of amplitude of variation, the variations in the UV are generally much stronger, even after taking dilution of the optical continuum into account.

Under these simple assumptions, a simple linearized model to describe the emission-line light curve as a function of line-of-sight (LOS) velocity V can be written as

$$\Delta L(t,V) = \int_0^\infty \Psi(\tau,V)\,\Delta C(t-\tau)\,d\tau\ , \tag{3}$$

where $\Delta C(t)$ and $\Delta L(t,V)$ are the differences from the mean continuum and mean LOS velocity-resolved emission-line light curves, respectively, τ is the time delay, and $\Psi(\tau,V)$ is the "transfer function" or, more descriptively, the "velocity-delay map." The velocity-delay map is thus the response of an emission line to an instantaneous or δ-function continuum outburst at some time t_0, i.e., $C(t) = \delta(t-t_0)$. The goal of reverberation mapping is to recover the velocity-delay map from the observables, $\Delta C(t)$ and $\Delta L(t,V)$, and thus infer the structure and geometry of the line-emitting region. This represents a classical inversion problem in physics, and solution by Fourier methods immediately suggests itself. However, solution by Fourier transforms is successful only in the case of large amounts of data and low noise, neither condition which can be easily realized in astronomical observations of faint sources. A less ambitious goal is to determine the "delay map" $\Psi(\tau) = \int \Psi(\tau,V)\,dV$ by using the integrated emission-line light curve. In practice, unambiguous determination of delay maps has also been illusive. In most cases, we have to settle for an estimate of the mean time scale for response, which we obtain by cross-correlating the continuum and emission-line light curves; the mean delay time, or "lag," is usually a good estimate of the centroid of the delay map.

3.2 Isodelay Surfaces

As noted above, the velocity-delay map represents the response of the BLR to a δ-function outburst, as seen by a distant observer. At the moment the observer detects the outburst, he will also detect the response of the emission-line gas along his line of sight to the continuum source; the total path followed by the ionizing photons that leave the central source and are replaced by emission-line photons when they encounter the BLR gas is the same path followed by the unabsorbed photons that are eventually detected by the observer. There is thus no time delay for the response of gas along the line of sight to the continuum source. For all other points, however, there will be a time delay τ between detection of the continuum outburst and the line response because of the increased path length.

Suppose for simplicity that the BLR clouds are all in a circular Keplerian orbit of radius r around the central black-hole/accretion-disk structure. At

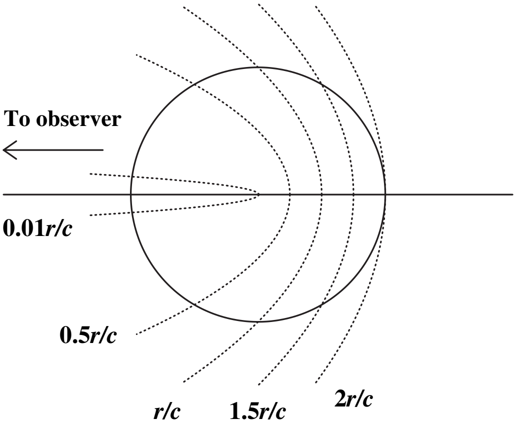

Fig. 2. For illustrative purposes, BLR clouds are supposed to be on circular Keplerian orbits of radius r and the continuum source is a point at the center of the ring. Following a continuum outburst, at any given time the observer far to the left sees the response of clouds along a surface of constant time delay, or isodelay surface. Here we show five isodelay surfaces, each one labeled with the time delay (in units of the shell radius r) we would observe relative to the continuum source. Points along the line of sight to the observer are seen to respond with zero time delay. The farthest point on the shell responds with a time delay $2r/c$

any given delay τ we observe the response of all points on a surface of constant delay, or "isodelay surface," which must be a paraboloid of revolution around the line of sight. This is illustrated in Fig. 2. which shows that, compared to the signal from the central source, the signal from anywhere else on the ring is delayed by light-travel time effects.

In the upper panel of Fig. 3, we define a system of polar coordinates measured from our line of sight to the continuum source. For clouds at position (r, θ), we see that the observer detects a time delay

$$\tau = (1 + \cos\theta)r/c. \tag{4}$$

Thus, at any time delay τ relative to the detection of the outburst, the distant observer sees the response of all gas clouds that intersect with the isodelay surface defined by (4).

3.3 Velocity-Delay Maps

We are now prepared to construct a simple velocity-delay map from first principles. We adopt as a toy model the edge-on (inclination $i = 90°$) ring system shown in the upper panel of Fig. 3 and consider now how the points on the ring translate from (r, θ) measured in configuration space to (V, τ) measured in velocity-delay space. This is illustrated in the lower panel of

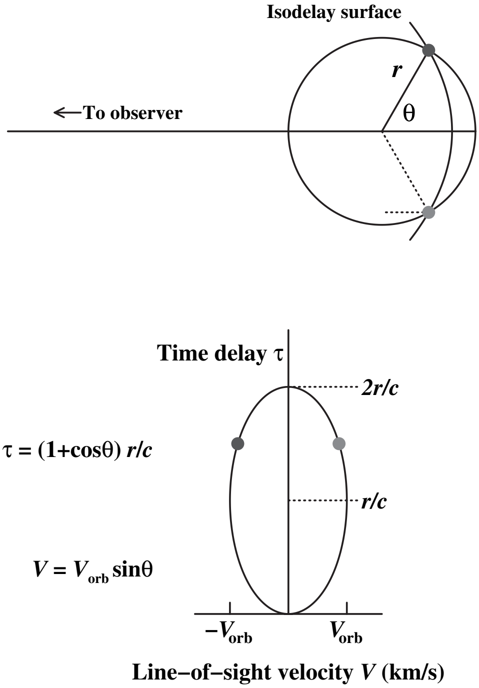

Fig. 3. The upper diagram shows a ring-like BLR with the observer to the left. The upper diagram shows a ring (or cross-section of a thin shell) that contains line-emitting clouds, as in Fig. 2. An isodelay surface for an arbitrary time is given; the intersection of this surface and the ring shows the clouds that are observed to be responding at this particular time. The *dotted line* shows the additional light-travel time, relative to light from the continuum source, that signals reprocessed by the cloud into emission-line photons will incur (4). In the lower diagram, we project the ring of clouds onto the LOS velocity/time-delay (V, τ) plane, assuming that the emission-line clouds in the upper diagram are orbiting in a clockwise direction (so that the cloud upper part of the orbit in the upper diagram is blueshifted and is represented on the left side of the lower diagram)

Fig. 3, where we have identified two of the emission-line clouds that lie on a particular isodelay surface. These points obviously project to $\tau = (1 + \cos\theta)r/c$ and LOS velocities $V = V_{\rm orb}\sin\theta$, where $V_{\rm orb}$ is the circular orbit speed. It is thus easy to see that an edge-on ring in configuration space projects to an ellipse in velocity-delay space, with semi-axes $V_{\rm orb}$ and r/c. It is also trivial to see that decreasing the inclination of the ring from 90° will decrease both axes by a factor of $\sin i$; as i approaches 0, the velocity-delay ellipse contracts towards a single point at $V = 0$ (since the orbital motion is now in the plane of the sky) and $\tau = r/c$ (since all points on the ring are now equidistant from the observer).

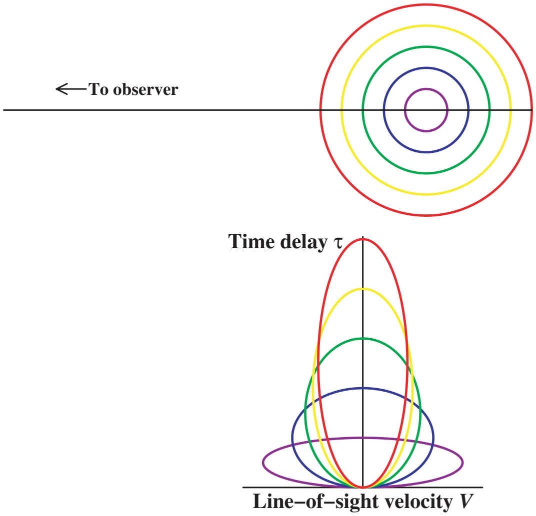

Fig. 4. Same information as in Fig. 3, except that here, the response is given for a multiple-ring system for clouds in circular Keplerian orbits. Each orbit projects to an ellipse in velocity–delay space. Larger orbits project to larger ranges in time delay ($2r/c$) and smaller LOS velocities ($V \propto r^{-1/2}$)

Once we have a velocity-delay map for a ring at arbitrary inclination, it is simple to take the next step to a Keplerian disk, which is essentially a sequence of rings with orbital velocities decreasing like $V_{\rm orb} \propto r^{-1/2}$. Such a sequence of rings is shown in Fig. 4.

Generalization to a disk (a series of rings), a thin shell (a series of rings of fixed r and varying i), or a thick shell (a series of thin shells of varying r) is trivial because of the linear nature of (3). All simple geometries dominated by Keplerian motion will show in the velocity-delay map the same characteristic "taper" with increasing time delay as seen in Fig. 4.

The ultimate goal of reverberation mapping is to recover the velocity-delay map from the observations. Unfortunately, in no case to date has this been possible, though in fairness it must be noted that recovery of a velocity-delay map has not been the principal goal of any experiment that has been carried out as designed; virtually all previous reverberation mapping programs have been designed to measure only the mean response time of emission lines to continuum variations. Even this comparatively modest goal has led to a number of important results, as described below.

4 Reverberation Results

4.1 Size of the BLR and Scaling with Luminosity

Emission-line time delays, or lags, have been measured for 35 AGNs, in some cases for multiple emission lines. The highest ionization emission lines are found to respond most rapidly to continuum variations, demonstrating clearly that there is ionization stratification within the BLR. Moreover, the highest ionization lines tend to be *broader* than the lower ionization lines, and indeed a plot of line width versus time delay shows that the lag τ varies with line width ΔV as $\tau \propto \Delta V^{-2}$, the virial relationship expected if the dynamics of the BLR is dominated by the gravitational potential of the central source. We show the results for the well-studied case of NGC 5548 in Fig. 5, [26].

As noted earlier, AGN spectra are remarkably similar over a broad range of luminosity. We might thus infer that, to some low order of approximation, their BLRs have similar physical conditions. Specifically, referring back to (1) and (2), we might conclude that U and n_e are about the same in all AGNs, and from that we could easily reach the conclusion that there should be a very simple relationship between the size of the BLR and the AGN continuum luminosity [16], (since $L \propto Q$) of the simple form

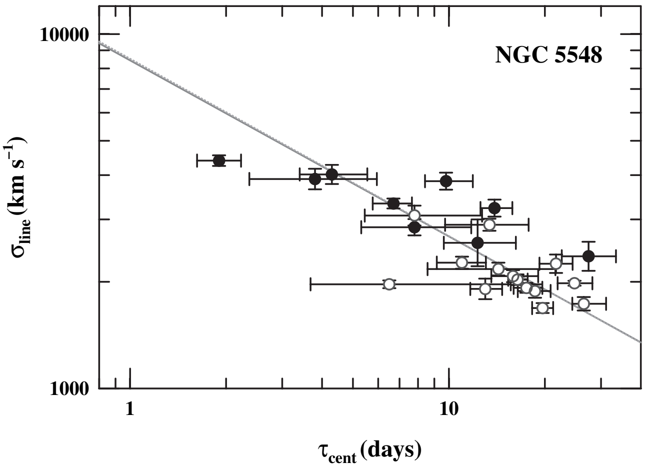

Fig. 5. Line width, as characterized by the line dispersion (second moment of the profile), versus time delay, as measured by the cross-correlation function centroid $\tau_{\rm cent}$, for multiple lines and multiple epochs for NGC 5548. The line measurements are for the variable part of the line only, as isolated by forming a root-mean-square (rms) spectrum from the many spectra obtained in the variability monitoring program. The *solid line* is the best fit to the data, and the *dotted line* is a forced fit to slope −1/2, the virial slope; in this particular case, the two lines are indistinguishable. The *open circles* are measurements of Hβ for 14 different years. The *filled circles* represent all of the other lines. From [25]

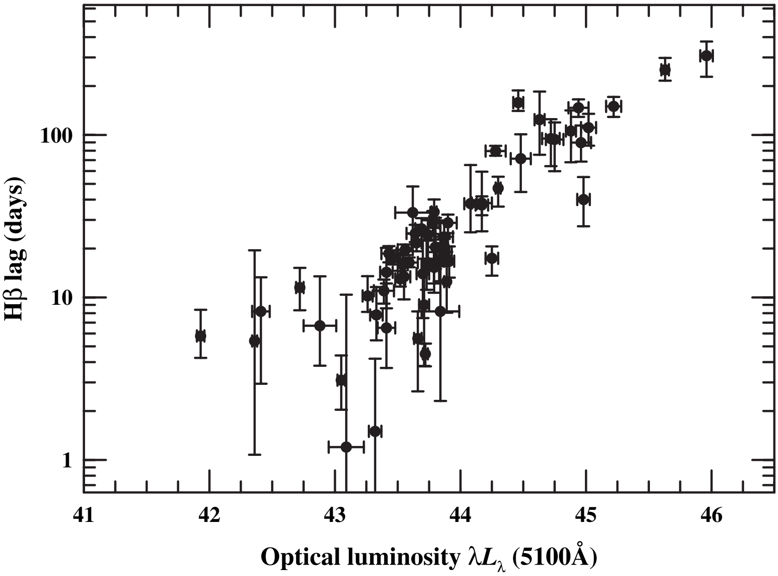

Fig. 6. Measured time delays for Hβ versus optical luminosity for over 30 reverberation-mapped AGNs, including multiple measurements of some sources. The relationship between BLR size and luminosity can be fitted with $r(\mathrm{H}\beta) \propto L^{0.6\pm0.1}$. Based on data from [25]

$$r \propto L^{1/2} . \tag{5}$$

In Fig. 6, we plot the size of the broad Hβ-emitting region as a function of the optical luminosity at 5100 Å for over 30 AGNs, some represented by multiple measurements at different epochs. The observed slope of this correlation, $\tau(\mathrm{H}\beta) \propto L^{0.6\pm0.1}$ is surprisingly consistent with the very naive prediction of (5) over four orders of magnitude in luminosity and time delays ranging from a few days to hundreds of days.

In the case of the particularly well-observed Seyfert 1 galaxy NGC 5548, the Hβ response time has been measured for 14 individual observing seasons, yielding measured lags ranging from as short as 6 days to as long as 26 days, depending on the mean continuum luminosity. The relationship between Hβ lag and continuum luminosity is shown in Fig. 7. The best-fit slope to this correlation is $\tau(\mathrm{H}\beta) \propto L^{0.9}$, significantly steeper than the result obtained by comparing different objects. The explanation for this, however, seems to be simple: the shape of the continuum changes as AGNs vary. As the continuum source gets brighter, it also gets *harder*, i.e., the amplitude of variability is larger at shorter wavelengths. Comparison of the ultraviolet (1350 Å) and optical (5100 Å) continuum fluxes in NGC 5548 shows that $L_{\mathrm{opt}} \propto L_{\mathrm{UV}}^{0.56}$, and since the size of the line-emitting region is controlled by the H-ionizing continuum ($\lambda < 912$ Å), the UV flux should be a much better measure of the

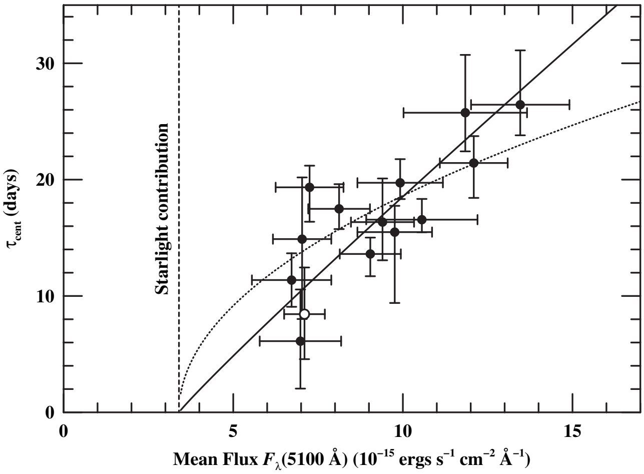

Fig. 7. Measured time delays for Hβ in NGC 5548 versus optical continuum flux for 14 independent experiments. The vertical line shows the constant stellar contribution to the measured continuum flux. The best-fit slope to this relationship is shown as a solid line $\tau(\mathrm{H}\beta) \propto F_\lambda^{0.9}$ and the dotted line shows the naive prediction $\tau(\mathrm{H}\beta) \propto F_\lambda^{1/2}$. From [27]

ionizing flux than the optical flux. Thus, we find that

$$\tau(\mathrm{H}\beta) \propto L_{\mathrm{opt}}^{0.9} \propto (L_{\mathrm{UV}}^{0.56})^{0.9} \propto L_{\mathrm{UV}}^{0.5} , \tag{6}$$

again, consistent with the naive prediction.

These results lead us immediately to some uncomfortable questions. What fine-tunes the size of the BLR? How does the BLR know precisely where to be? Why are the ionization parameter and electron density the same for all AGNs?

At least a partial answer is provided by considering the result shown in Fig. 7. What is obvious from this diagram is that the Hβ-emitting gas of the BLR is *not* changing its location from year to year as the continuum changes; it is obviously not moving fast enough to do this. What we are forced to conclude is that gas is *everywhere* in the line-emitting region (in this case, at least between ∼6 and ∼26 light days of the central source) and that what is actually changing with time, or more properly with luminosity, is the distance from the continuum source at which the physical conditions are optimal for emission in the Hβ line. This is often referred to as the "locally optimally-emitting cloud (LOC)" model, e.g., [3]. The basic idea is that line-emitting clouds of gas of various density are distributed throughout the nuclear region. Emission in a particular line comes predominantly from clouds with optimal conditions for that line, and this location can vary as the continuum flux changes. An example of a grid of emission line strengths in an LOC photoionization equilibrium model is shown in Fig. 8.

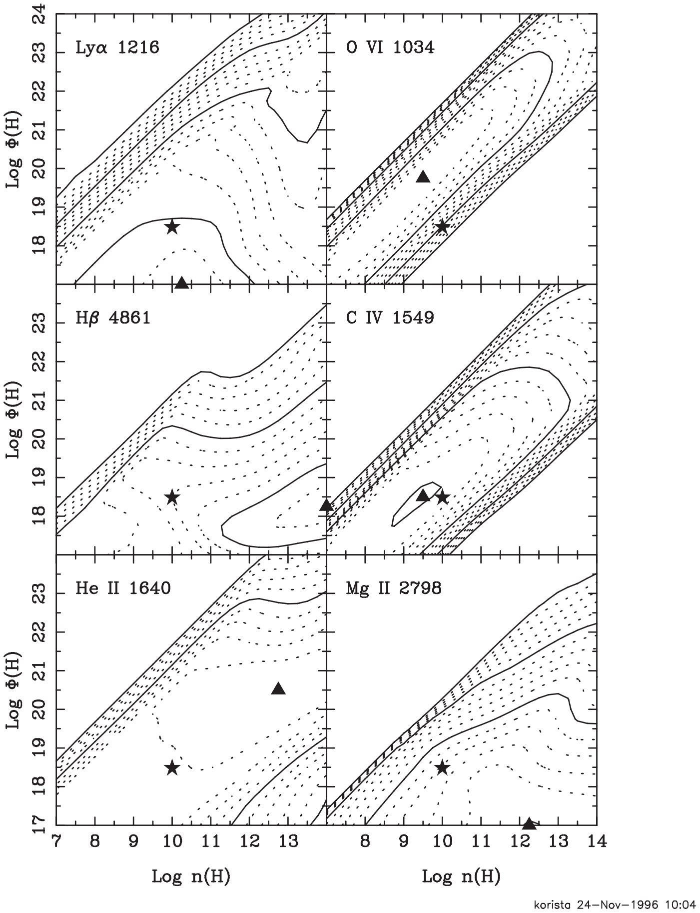

Fig. 8. Contours of constant equivalent width for a grid of photoionization equilibrium models, as a function of input ionizing flux ($\Phi(\mathrm{H}) = Q_{\mathrm{ion}}/4\pi r^2$) and particle density $n(H)$ for strong emission lines in AGN spectra. Dotted contours are separated by 0.2 dex and solid contours by 1 dex. The solid star is a reference point to "standard BLR parameters," i.e., the best single-zone model. The triangle shows the location of the peak equivalent width for each line. From [17]

4.2 Reverberation and AGN Black Hole Masses

As shown in Fig. 5, for AGNs with multiple reverberation measurements, there is virial relationship between line width and lag. This is strong evidence that the broad lines can be used to measure the mass of the central black hole via the virial theorem,

$$M_{\mathrm{BH}} = \frac{f c\tau \Delta V^2}{G}, \tag{7}$$

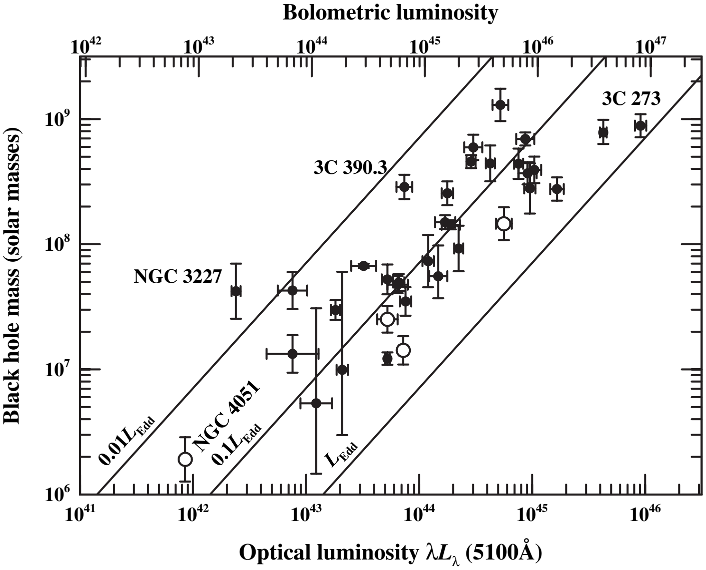

Fig. 9. The mass–luminosity relationship for reverberation-mapped AGNs. The luminosity scale on the lower x-axis is $\log \lambda L_\lambda$ in units of ergs s^{-1}. The upper x-axis shows the bolometric luminosity assuming that $L_{\mathrm{bol}} \approx 9\lambda L_\lambda(5100\,\text{Å})$. The *diagonal lines* show the Eddington limit L_{Edd}, $0.1L_{\mathrm{Edd}}$, and $0.01L_{\mathrm{Edd}}$. The *open circles* represent NLS1s. From [25]

where f is a factor of order unity that depends on the unknown geometry and kinematics of the BLR. That these virial masses are valid at some level is demonstrated by the fact that a plot of the reverberation-based masses versus the stellar bulge velocity dispersion σ_* is consistent with the same $M_{\mathrm{BH}} - \sigma_*$ relationship seen in quiescent galaxies based on masses measured primarily by stellar or gas dynamics, e.g., [12, 13]. Indeed, a mean value for the scaling factor f in (7) can be obtained by scaling the AGN $M_{\mathrm{BH}} - \sigma_*$ relationship to that for quiescent galaxies. Unfortunately, however, most dynamical models of the BLR give similar values of f, so this affords no useful constraint on the BLR structure and kinematics.

Figure 9 shows the relationship between black hole mass and luminosity for the 35 reverberation-mapped AGNs. All of the AGNs are sub-Eddington, contrary to some earlier estimates, and based on an admittedly nominal bolometric correction, the accretion rates are typically about 10% of the Eddington rate, but with considerable scatter about this value. The scatter in this mass-luminosity relationship is apparently real; the NLS1s occupy the right-hand edge of the envelope, as expected, and the one reverberation-mapped AGN with very broad double-peaked emission lines (see below), 3C 390.3, is on the left-hand side of the envelope, again as expected. Of course, the scatter in the relationship is not attributable only to differences in accretion rate, but other effects such as inclination and obscuration (e.g., the luminosity of NGC 3227 is probably underestimated as it appears to have heavy internal absorption).

5 What is the BLR?

The fundamental question that we still have not yet addressed is what is the nature of the BLR? What is the origin of the gas that gives rise to the emission lines, and how is it related to the accretion flow, if at all? A number of different scenarios have been proposed, and we discuss some of these briefly.

5.1 "Cloud" Models

Our first notions of what the BLR might look like was based on observations of Galactic nebulae, especially the Crab Nebula system of "clouds" or "filaments," partly because of the obvious merits of such an interpretation (e.g., we noted earlier that supersonic motions of BLR gas argue for such a system or some kind of large-scale flow), but probably also because the first astrophysicists to work on the problem had previously worked on nebular physics in Galactic sources.

If we suppose that the BLR is comprised of some large number N_c of identical line-emitting clouds of radius R_c, the BLR covering factor (i.e., fraction of the sky covered by BLR clouds, as seen by the central source) will be proportional to $N_c R_c^2$. The covering factor is determined by estimating the fraction of ionizing continuum photons absorbed by BLR clouds and reprocessed into emission lines, and is constrained to be of order $\sim$10% by the equivalent widths of the emission lines, notably Lyα. An independent constraint is given by the total line luminosity, which is proportional to $N_c R_c^3$, i.e., the total volume of line-emitting material. By combining these two independent constraints on $N_c R_c^2$ and $N_c R_c^3$, we can independently solve for N_c and R_c. For a typical Seyfert galaxy like NGC 5548, we find $N_c \approx 10^7$ and $R_c \approx 10^{13}\,\mathrm{cm}$. Furthermore, we can combine the size of the cloud with the putative particle density of $n_e \approx 10^{10}\,\mathrm{cm}^{-3}$, to get a cloud column density of $N_\mathrm{H} \approx 10^{23}\,\mathrm{cm}^{-2}$. Interestingly and ultimately coincidentally, this is the same order of magnitude as the first measurements of the column densities of "warm absorbers" detected in the X-ray spectra of AGNs, and it was natural to ascribe this absorption to BLR clouds along the line of sight to the very small X-ray continuum source. Finally, given the number, size, and density of the clouds, it is trivial to compute a mass for the entire BLR, which works out to be $\sim 1\,M_\odot$.

Early views on cloud dynamics were also influenced by observations of supernova remnants, the only other astrophysical environment in which line-emitting gas moves at velocities higher than $1000\,\mathrm{km\,s}^{-1}$. Either ballastic outflow (as in supernovae) or radiation pressure driven outflow can produce logarithmic line profiles. A preference for outflow models was at least in some part driven by concerns that a gravitationally bound BLR seemed quite implausible. Prior to the advent of reverberation mapping, there was no way to determine the size of the BLR other than photoionization equilibrium modeling. Sizes predicted by single-zone (i.e., all lines produced in approximately the observed ratios by a single representative cloud) models were about an order of magnitude too large, thus leading to mass estimates that were far too large to be consistent with observed nuclear stellar kinematics.

The argument on the number of emission-line clouds is, however, flawed. The fact that the emission lines vary in response to continuum variations argues that the clouds are optically thick, in which case the emitting volume of a cloud is proportional to $R_c^2 R_{\rm S}$ where $R_{\rm S}$ is the depth of the ionized layer of the cloud (i.e., the Strömgren depth), rather than R_c^3, which removes the independent constraints on N_c and R_c. However, there is a second argument that also argues for a large number of clouds, namely the smoothness of the emission-line profiles, as shown in Fig. 10. Here the argument is that a collection of clouds with a velocity dispersion of a few thousand $\mathrm{km\,s^{-1}}$, but each emitting lines of thermal widths of 10 $\mathrm{km\,s^{-1}}$, ought to produce a rather "grainy" composite line profile on account of statistical fluctuations in the number of clouds at different line-of-sight velocities. The lack of grainy structure at high spectral resolution and high signal-to-noise ratios argues either that the number of clouds must be very large or the BLR gas is some kind of continuous flow rather than in the form of discrete clouds. Even an extreme case, NGC 4395, the least luminous known Seyfert 1, has characteristically smooth line profiles. But the BLR in this source is expected to be so small that the number of individual clouds could not exceed a few thousand. This essentially leaves us only with models that involve some kind of supersonic flow, unless there is another significant source of microturbulence within the broad-line clouds.

Another interesting consequence of variability observations is the realization that there must be a large reservoir of gas throughout the BLR, but at any particular time we are detecting primarily the gas that is emitting most efficiently [2]. Thus the total mass of gas in the BLR is much larger than computed above. Estimates of the BLR mass run as high as 10^3–$10^4\,M_\odot$.

5.2 "Bloated Stars" Model

An early suggestion for the origin of the broad emission lines was gas provided by stars in the nucleus, e.g., [1]. This solves a number of problems, such as fuel supply and cloud stability, but encounters other problems, notably that, except in the case of giant stars, which are relatively rare, the stellar surface gravity is too high to easily remove gas. Another problem is whether or not one can fit an adequate number of stars into the BLR, i.e., a variation on the "number of clouds" problem referred to above.

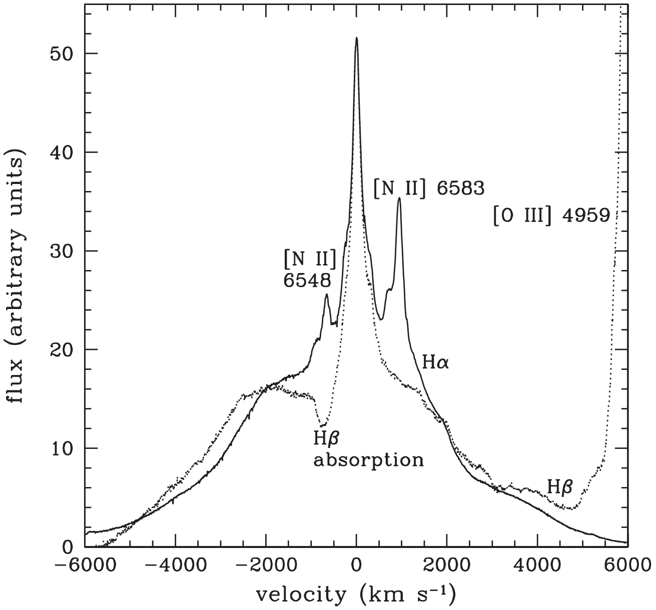

Fig. 10. Keck high-resolution spectra of the Hα and Hβ emission lines in NGC 4151. Some contaminating narrow-line features are labeled. While there is structure in the line profiles on scales of hundreds of $\mathrm{km\,s^{-1}}$, on thermal scales the profiles are very smooth. From Arav et al., 1998, MNRAS, 297, 990

5.3 Double-Peaked Emission Lines

A relatively small subset of AGNs have very broad double-peaked Balmer line profiles, as shown in Fig. 11. Double-peaked profiles are characteristic of rotating disks; such profiles are commonly observed in accretion disks in Galactic binaries. Sometimes double-peaked profiles appear in the *variable* part of an AGN spectrum, i.e., the line profile that appears in difference spectra (i.e., high-state spectrum minus low-state spectrum) or in rms spectra.

Extensive efforts to model such disks indicate that they must be fairly complex. Profile variations show that the disks are clearly not axisymmetric, and they sometimes have large-scale structural or thermal irregularities. Sources with double-peaked emission lines are certainly important to our understanding of AGNs, as these are the only sources where we can identify a plausible disk structure. Whether or not such disk structures are present and merely hidden or disguised in other AGNs remains an open question. Clearly, however, disk-like structures cannot explain everything about the BLR.

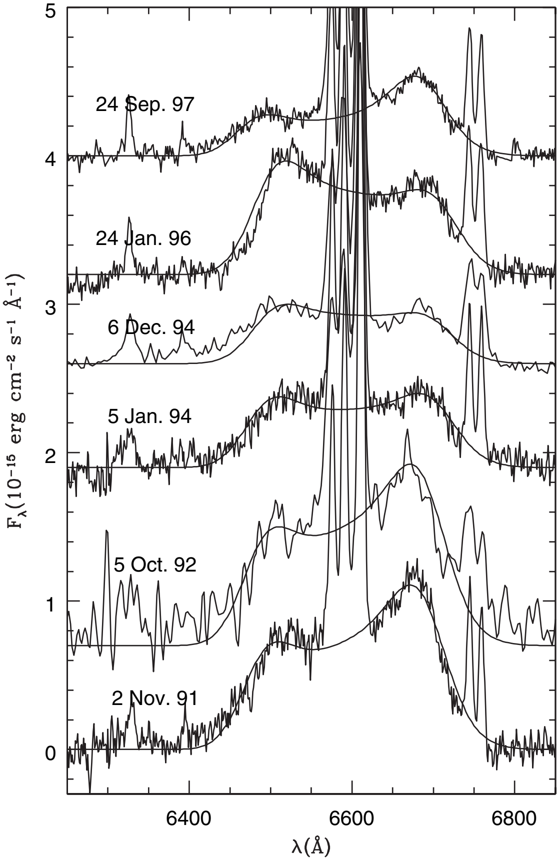

Fig. 11. The variable double-peaked Hα line profile in NGC 1097, with best-fit elliptical disk models shown as solid lines. From [29]

5.4 Disk Winds

There is increasing evidence for the existence of disk winds in AGNs, e.g., [6, 10, 19]. Disk winds are observed in both Galactic binaries and in young stellar objects, and may be a feature common to accretion disks on all scales. There has been much theoretical investigation in this area, but it is not clear whether AGN disk winds ought to be driven radiatively or hydromagnetically, or perhaps by some hybrid.

It may well be that what we think of as the BLR is itself a composite. Photoionization equilibrium modeling has suggested that there must be two distinct regions, one highly ionized and of moderate optical depth and another of moderate ionization and greater optical depth, but of similar physical scale and distance from the central source, e.g., [8]. The latter, which is largely responsible for the Balmer-line emission, has been often identified with the disk itself; and indeed the double-peaked Balmer lines in some sources support this. The higher ionization lines, however, may represent a disk wind component. Observational evidence for this includes:

1. clear blueward asymmetries in the higher-ionization lines in narrow-line Seyfert 1 galaxies, e.g., [18];
2. the peaks of high ionization lines tend to be blueshifted relative to systemic, and the maximum observed blueshift increases with source luminosity, e.g., [11];
3. in radio-loud quasars, the width of the bases of the C IV lines are larger in edge-on sources than in face-on sources, implying that the wind has a strong radial component to it, e.g., [30].

The disk wind model also affords a possible connection to the outflows detected in AGNs of all luminosity. Strong absorption features, generally blueshifted relative the systemic velocities, are ubiquitous features of the X-ray and UV spectra of AGNs, e.g., [9]. In high-luminosity quasars, the outflows are manifest as "broad absorption-line (BAL)" quasars. In lower-luminosity Seyfert galaxies, the features are weaker, but may still cover a fairly large velocity range, but with multiple discrete components as opposed to more-or-less continuous absorption troughs as in BAL quasars. In Seyfert galaxies, the amount of kinetic energy in the absorbing region can easily be of the same order as the radiative energy; in any given case, this calculation is subject to uncertainties due to unknown covering factors, which are quite reasonably assumed to be high because strong absorption features are so common in these sources.

6 What will it Take to Map the Broad-Line Region?

The BLR has been studied extensively for over 30 years. Despite this long history, new surprises seem to emerge regularly. An important recent step has been the Sloan Digital Sky Survey, which is providing extremely large samples of AGNs for statistical study.

Probably the most obvious area of untapped potential remains reverberation mapping. Emission-line response times have been measured for only 35 AGNs; the response of the high-ionization UV lines has been measured in only in a bare handful of these cases, and only in the case of NGC 5548 have measurements been made on multiple occasions.

Most importantly, however, is that *no reliable velocity-delay map has been obtained for any source.* There have been attempts to recover velocity-delay maps from reverberation data, but in every case to date, the data have been of insufficient quality or quantity for this analysis. However, as a result of these pioneering observational studies, we understand the characteristics of AGN continuum and emission-line variability well enough that we can carry out fairly realistic simulations to determine what kind of data are necessary to recover a velocity-delay map. Extensive numerical simulations, e.g., [15], have shown the typical data requirements for a bright Seyfert galaxy (with specific numbers appropriate for the case of NGC 5548) are:

1. *High time resolution* ($\Delta t \approx 0.2$–1.0 day). The time sampling of the light curves translates directly into the time resolution of the velocity-delay maps. Sampling as crude as $\sim$1 day provides around 4–20 resolution elements through the line response for a typical Seyfert galaxy, depending on the particular emission-line and on the luminosity of the source.
2. *Long duration* (several months). A rule of thumb in time series is that the duration of the experiment should be at least three times as long as the longest time scale phenomenon to be addressed. This suggests a minimum duration of some 60–80 days, again depending on the emission line and continuum luminosity. What simulations show is that the main requirement for detecting a clear reverberation signal is a major "event" in the driving light curve, i.e., an outburst, "drop-out," or change in the sign of the derivative. One needs to detect a strong signal and then follow it as it propagates through the entire BLR. Simulations based on the behavior of NGC 5548 suggest that a 200-day program is necessary for a high probability of success, but success can occur with programs as short as 75 days or so, if one is lucky.
3. *Moderate spectral resolution* ($\sim$600 km s^{-1}). Spectral resolution translates directly into velocity-resolution in the velocity-delay map. Resolution of $R > 500$ typical yields some 10 or so resolution elements across a typical broad emission line. Given the lack of profile structure at high spectral resolution, resolution approaching the thermal line width (i.e., $R \approx 30,000$) does not seem necessary.
4. *High homogeneity and signal-to-noise ratio ($S/N \approx 100$).* A high level of homogeneity and high signal-to-noise ratios per datum are required in order to detect the continuum variations, usually a few percent or less between observations, and the responding signal in the emission lines. At least high *relative* photometric accuracy is required; this is a major challenge to ground-based observations where one has little control over the point-spread function, which ultimately limits the photometric accuracy. For this reason and that some of the most responsive emission line are in the UV, space-based observations are preferred for obtaining velocity-delay maps.

Detailed simulations show that *Hubble Space Telescope* can carry out such a project, but practically speaking, only for a limited number of emission lines. To get a complete map of the BLR, it is desirable to use as many lines as possible; this is especially important if the BLR is, as suggested earlier, a composite structure such as a disk plus a disk wind. A space mission concept, called *Kronos*, has been designed to obtain spectroscopic time series data on accretion-driven sources, with the specific application of obtaining velocity-delay maps of the BLR as a major goal. The key to *Kronos* science is long on-target times at high time resolution, enabled by long on-target times as a consequence of its high-Earth orbit ($P \approx 14\,\text{days}$).

Figure 12 shows an example of a reverberation-mapping simulation. The BLR model in this case is an inclined ($i = 45°$) Keplerian disk with a somewhat arbitrary spiral-arm structure. The emission-line response for each of four strong lines is calculated from a photoionization equilibrium model based on NGC 5548. The recovered velocity-delay maps are based on the quality and quantity of data expected from the *Kronos* spectrometers.

7 Some Obvious Questions to Keep in Mind

Any explanation for the origin of the BLR must provide answers to a number of fairly obvious questions:

1. Stellar-mass black hole systems (microquasars) do not have (obvious) BLRs. Why are they different from AGNs? While Galactic binaries have accretion disks and jets, they do not have anything like the BLR. This may, of course, be related to the fact that accretion disks around stellar-mass black holes are considerably hotter than those around supermassive black holes, as the highest disk temperature is inversely proportional to the mass in simple thin-disk models.
2. Some AGNs may lack BLRs, i.e., they are true type 2 AGNs. Whereas many type 2 AGNs have “hidden” BLRs that are detected at extremely high signal-to-noise ratios or in polarized (scattered) light, there is evidence suggesting that some Type 2 objects have no broad lines at all. Why is this? What are the physical differences between AGNs with BLRs and those without?
3. Broad lines widths are in the range 1000 to 25,000 $\text{km}\,\text{s}^{-1}$. What imposes these limits? And, a key question closely related to this, what does spectrum of an AGN at $i = 0°$ look like? If not orbital motion, what causes the line broadening in this extreme case?
4. Where does the energy to power the emission lines come from? The BLR energy budget problem is still unsolved. The observed AGN continuum is neither luminous enough nor hard enough to account for the broad lines. Does the BLR see a different continuum than we do, or is there an energy source we have not yet recognized?

8 Summary

Despite years of observational and theoretical effort, we cannot yet say much about the BLR with any certainty. We know that it is dense by nebular standards and hot ($T \approx 10^4\,\text{K}$), and that it is probably some supersonic, fairly continuous flow. The dynamics of the BLR are unknown, though it

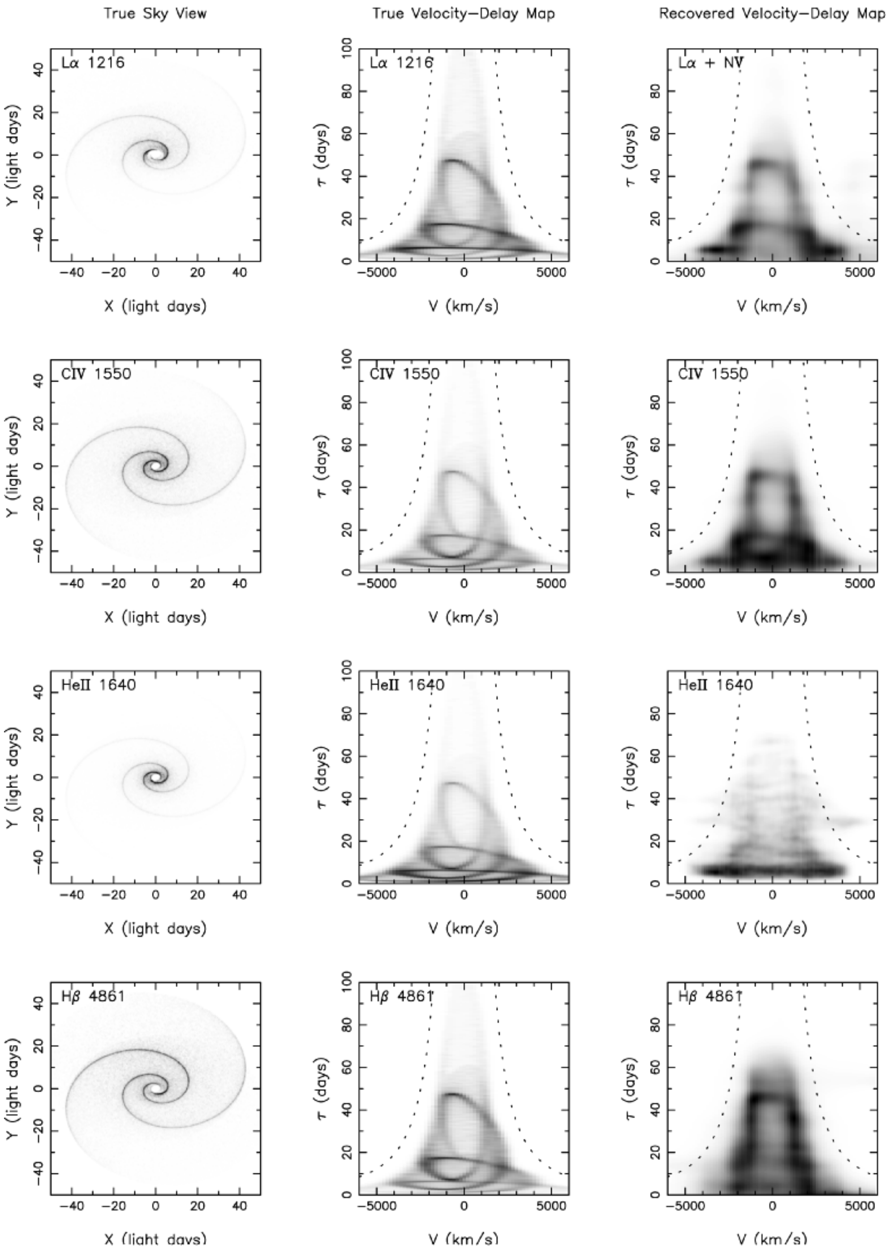

Fig. 12. Reverberation mapping simulations. From left to right, the three columns show, for four emission lines, (**a**) a map of a BLR with a spiral arm structure in configuration space, (**b**) a corresponding velocity-delay map, and (**c**) a velocity-delay map recovered from simulated data. From top to bottom, the rows represent results for Ly$\alpha\,\lambda$1216, C IV λ1550, He II λ1640, and H$\beta\,\lambda$4861. From [15]

is clear that the gravitational acceleration caused by the central source is sufficiently dominant that we can use the BLR size and velocity dispersion to estimate the central black hole masses with an accuracy of about a factor of three. We know that it has a stratified ionization structure, and may have multiple kinematic components; currently strong candidates include both disk winds and emission from the extended parts of the accretion disk itself, with the strength of different components varying from object to object and from emission line to emission line. We know that the size of the BLR scales simply with luminosity, both globally from object-to-object and in a single source as the mean luminosity varies with time. The latter in particular argues that there is considerably more mass involved in the BLR than presumed by naive models. We have argued that the best way to determine the structure and kinematics of the BLR is by obtaining velocity-delay maps for multiple emission lines by reverberation mapping, and that the data requirements, while somewhat daunting, are realizable with existing technology.

Acknowledgements

I am grateful for support of this work by the US National Science Foundation through grant AST-0205964 to The Ohio State University. I would like to thank my collaborators and Nahum Arav, Kirk Korista, Thaisa Storchi-Bergmann, and Dan Vanden Berk for assistance with figures. I also thank the organizers for their fine hospitality in Chile. I apologize for the incomplete nature of the reference list; the few selected references are intended to provide the reader an entry point to the extensive literature on the broad-line region rather than a comprehensive bibliography.

References

1. T. Alexander, H. Netzer: MNRAS, **284**, 967 (1997)
2. J.A. Baldwin, G. Ferland, K.T. Korista, F. Hamann, M. Dietrich: ApJ, **582**, 590 (2003)
3. J.A. Baldwin, G. Ferland, K.T. Korista, D. Verner: ApJL, **455**, L119 (1995)
4. R.D. Blandford, C.F. McKee: ApJ, **255**, 419 (1982)
5. T.A. Boroson, R.F. Green: ApJS, **80**, 109 (1992)
6. M. Bottorff, K.T. Korista, I. Shlosman, R.D. Blandford: ApJ, **479**, 200 (1997)
7. S. Collier et al.: ApJ, **500**, 162 (1998)
8. S. Collin-Souffrin, M. Joly, D. Pequignot, S. Dumont: A&A, 166, 27 (1986)
9. D.M. Crenshaw, S.B. Kraemer, I.M. George: ARAA, **41**, 117 (2003)
10. M. Elvis: ApJ, **545**, 63 (2000)
11. B. Espey: Velocity Diagnostics of the Broad Emission-Line Region of Active Galactic Nuclei. In: *Emission Lines in Active Galaxies: New Methods and Techniques*, ed. by B.M. Peterson, F.-Z. Cheng, A.S. Wilson (Astronomical Society of the Pacific 1997) pp 175–178

12. L. Ferrarese, R.W. Pogge, B.M. Peterson, D. Merritt, A. Wandel, C.L. Joseph: ApJ, **555**, L79 (2001)
13. K. Gebhardt et al.: ApJ, **543**, L5 (2000)
14. F. Hamann, G.J. Ferland: ARAA, **37**, 487 (1999)
15. K. Horne, B.M. Peterson, S. Collier, H. Netzer: PASP, **116**, 465 (2004)
16. S. Kaspi, P.S. Smith, H. Netzer, D. Maoz, B.T. Jannuzi, U. Giveon: ApJ, **533**, 631 (2000)
17. K.T. Korista, J.A. Baldwin, G. Ferland, D. Verner: ApJS, **108**, 401 (1997)
18. K.M. Leighly: *ASCA* (and *HST*) Observations of NLS1s. In:*New Astronomy Reviews*, ed. by Th. Boller, W.N. Brandt, K.M. Leighly, M.J. Ward (Elsevier: Amsterdam 2000) pp 395–402
19. N. Murray, J. Chiang: ApJ, **474**, 91
20. H. Netzer: AGN Emission Lines. In:*Active Galactic Nuclei*, ed. by T.J.-L. Courvoisier, M. Mayor (Springer-Verlag, Berlin Heidelberg New York 1990) pp 57–160
21. H. Netzer: Physical Processes in Starburst and Active Galaxies. In: *The Starburst–AGN Connection*, ed. by I. Aretxaga, D. Kunth, R. Mújica (World Scientific: Singapore 2001) pp 117–165
22. B.M. Peterson: PASP, **105**, 247 (1993)
23. B.M. Peterson, *An Introduction to Active Galactic Nuclei*, (Cambridge University Press: Cambridge 1997)
24. B.M. Peterson: Variability of Active Galactic Nuclei. In:*The Starburst–AGN Connection*, ed. by I. Aretxaga, D. Kunth, R. Mújica (World Scientific: Singapore 2001) pp 3–67
25. B.M. Peterson, L. Ferrarese, K.M. Gilbert, S. Kaspi, M.A. Malkan, D. Maoz, D. Merritt, H. Netzer, C.A. Onken, R.W. Pogge, M. Vestergaard, A. Wandel: ApJ, **613**, 682 (2004)
26. B.M. Peterson, A. Wandel: ApJL, **521**, L95 (1993)
27. B.M. Peterson et al.: ApJ., **581**, 197 (2002)
28. J.W. Sulentic, P. Marziani, D. Dultzin-Hacyan: ARAA, **38**, 521 (2000)
29. T. Storchi-Bergmann, R. Nemmen da Silva, M. Eracleous, J.P. Halpern, A.S. Wilson, A.V. Filippenko, M.T. Ruiz, R.C. Smith, N.M. Nagar: ApJ, **598**, 956 (2003)
30. M. Vestergaard, B.J. Wilkes, P.D. Barthel: ApJ, **538**, 103 (2000)

Cold Gas Near Active Galactic Nuclei

J.F. Gallimore[1], M. Elitzur[2], and S.A. Baum[3]

[1] Department of Physics and Astronomy, Bucknell University, Moore Avenue, Lewisburg, PA 17837, USA
`jgallimo@bucknell.edu`

[2] Department of Physics and Astronomy, University of Kentucky, 600 Rose Street, Lexington, KY 40506-0055, USA
`moshe@uky.edu`

[3] Center for Imaging Science, Rochester Institute of Technology, 54 Lomb Memorial Drive, Rochester, NY 14623, USA
`baum@cis.rit.edu`

1 Background

Active galactic nuclei (AGNs) are identified mainly by the properties of visibly hot gas located within the nucleus (e.g., line emitting regions, UV – X-ray emission from the accretion disk) or without (e.g., radio lobes of relativistic plasma). Colder gas and dust however play an important role in the standard model for AGNs. Firstly, there are arguments that the (relatively) cold ISM from either the host galaxy or a cannibalized companion ultimately provides the fuel for the AGN. Secondly, apropos of AGN unifying schemes [1], the cold, potentially fueling gas can obscure the hot, innermost regions, at least along some sight-lines, affecting the apparent properties of the AGN.

The fueling argument stems mainly from considerations for QSOs and radio galaxies. For example, a luminous ($10^{13} L_\odot$) QSO needs to accrete of order 1 $M_\odot$ yr^{-1} to support its bolometric luminosity, and it further requires a reservoir of at least 10^8 $M_\odot$ of fuel to support that luminosity over its lifetime. For comparison, this minimum reservoir mass is about 1–10% of the mass of cold (<1000 K) atomic and molecular gas within a "typical" spiral galaxy. Cold gas also offers an advantage towards accretion over hotter phases of the ISM, simply because the colder gas cannot support itself against collapse into an accretion disk. Based on spherical (Bondi) accretion arguments, the temperature of the fueling gas must be less than or of order 1000 K in order to funnel gas efficiently into an accretion disk. For example, [2] provides a straightforward review of these principles.

There is now plenty of indirect evidence for cold absorbing gas near AGNs. Arguably the most remarkable evidence comes from polarimetry studies of Seyfert galaxies. Antonucci & Miller [3] demonstrated that the optical spectrum of the Seyfert galaxy NGC 1068 changed character when viewed in polarized light. NGC 1068 is classified as a type 2 Seyfert galaxy, meaning that its optical spectrum exhibits equally broad permitted and forbidden emission

J.F. Gallimore et al.: *Cold Gas Near Active Galactic Nuclei*, Lect. Notes Phys. **693**, 101–120 (2006)
DOI 10.1007/3-540-34621-X_4

lines. In polarized light, however, broader wings appear on the profiles of the permitted lines, which is a defining characteristic of type 1 Seyfert galaxies. Their explanation was that NGC 1068 is intrinsically a type 1 Seyfert, but the “broad line region” (BLR) responsible for the broad permitted line wings is obscured by foreground dust. The “narrow line region” (NLR) is much larger than the BLR and is viewed without obscuration. Emission from the BLR however scatters off of electrons or dust grains in the NLR; scattering polarizes light, and so the BLR appears only in the reflected, or polarized, spectrum. Spectropolarimetry has since turned up hidden BLRs in a handful of other Seyfert 2’s, promoting the unifying scheme that perhaps all Seyfert galaxies are intrinsically the same, but type 2 Seyferts are viewed through some cold, dusty medium surrounding the BLR.

X-ray spectroscopy broadly supports this view. The soft (low energy) X-ray spectra of Seyfert 2 nuclei appear suppressed compared to Seyfert 1 nuclei. This soft X-ray suppression is well fit by models involving a Seyfert 1 spectrum that is absorbed by cold, foreground gas (in X-ray astronomy, cold usually means $T_e < 10^6\,\mathrm{K}$). The characteristic column densities are of order $N_H = 10^{22}$ cm^{-2} up to $N_H > 10^{25}$, at which point the obscuring medium is opaque to Thomson scattering and even hard (high energy) X-ray emission is suppressed. Seyfert 2 galaxies also show high equivalent width Fe K-shell emission lines near 6.4 keV. The observed continuum is too weak to produce the emission line, and so the interpretation is that the AGN, which provides the requisite X-ray continuum to produce the Fe K-shell lines, is obscured along our sight-line.

The morphology of Seyfert NLRs also provides evidence of a dusty obscuring medium surrounding the AGN. Many Seyfert nuclei are near enough to us that we can resolve the structure of the NLR using narrow-band imaging filters; the Hubble Space Telescope has produced spectacular images of Seyfert NLRs. Resolved NLRs are commonly asymmetric and elongated, usually aligning near the radio source axis. One interpretation is that the obscuring medium shadows the nuclear continuum source and the ionizing radiation can escape only along an unobscured axis. Circumnuclear clouds along the unobscured axis are ionized by the incident continuum and appear as the resolved NLR. These extended NLRs are often called “ionization cones,” although this is a misnomer since NLRs rarely appear precisely conical (see the review by Fosbury in this volume). The intrinsic structure of gas in the NLR, which may itself be affected by the radio jet or winds, the relative orientation of the obscuring medium and the circumnuclear gas, and large-scale obscuration can all compete to produce patterns more complicated than simple bi-cones.

Putting the pieces together, the unifying model for the narrow-line AGN vs. broad-line AGN dichotomy is a supermassive black hole surrounded by a hot accretion disk, which is in turn surrounded by a dusty obscuring medium. For a galaxy to be classified as a Seyfert (or other emission-line AGN), the NLR must not be completely obscured, which means that the projected size of the obscuring medium must be smaller than the projected size of the NLR, typically less than a few tens of parsecs. On the other hand, for a galaxy to appear as a Seyfert 2, the BLR must be completely obscured, which means that the obscuring medium must be larger than a few light days – light weeks, the characteristic scale of Seyfert BLRs and the nuclear continuum source (see the review by Peterson, this volume). The simplest model, then, is a parsec-scale disk of cold, dusty material surrounding the AGN. If we view the disk relatively edge-on, the BLR is hidden from view and we classify the galaxy as a Seyfert 2; otherwise, the BLR and AGN are not significantly obscured and the galaxy is classified as a Seyfert 1.

The relative fraction of Seyfert 2s vs. Seyfert 1s gives the covering fraction of the obscuring disk, which is the fraction of the sky viewed by the AGN that is blocked by the disk. Because of selection effects in samples of Seyfert galaxies the average covering fraction is uncertain but probably exceeds 50%. Such a high covering fraction implies that the scale height of the obscuring disk is comparable to the inner radius. For this reason, the obscuring disk is argued to be a fat disk, or "torus." Such a torus may not be dynamically stable in a steady state, and Pringle and others have argued that the obscuring geometry might instead be a geometrically thin but strongly warped disk. For the purpose of brevity, we will simply call the obscuring medium the "torus."

The physical conditions within the torus are straightforward to constrain and have been modeled in some detail. The sublimation temperature sets an upper limit to the dust temperature, $T_d < 1000 - 1500$ K. Ward et al. [4] found that the near–mid infrared spectrum of Seyfert galaxies is compatible with a combination of a QSO-like spectrum plus thermal radiation from hot dust grains; this result argues in favor of the view that dusty ISM exists as close as possible to the AGN. The opacity through the torus is sufficiently high that soft X-rays and UV radiation cannot penetrate very deeply into the obscuring clouds, and so the clouds are probably largely molecular with maximum temperatures of order a few $\times 100\,\mathrm{K}$. On the other hand, hard X-rays can travel far into the obscuring clouds, producing a significant column of warm ($T \sim 8000\,\mathrm{K}$) atomic gas.

The dust sublimation radius sets the innermost scale of the dusty medium: $r_{\rm min} \sim 0.1 L_{44}^{1/2}$ pc, where L_{44} is the AGN luminosity in units of 10^{44} ergs s^{-1}, typical for a Seyfert nucleus. Because of the relatively low temperature and compact size, mapping the obscuring torus of Seyfert galaxies has been and will continue to be difficult. Looking first at the compact size, 1 parsec at the distance of the nearest Seyfert galaxies subtends only $\lesssim$15 mas. Such fine angular resolution is currently only achievable with interferometric techniques. Turning to the temperature, low temperature generally means low surface brightness, and it is difficult to detect 100 K gas in emission on such small scales. To illustrate the problem, we consider the detection of thermally emitting molecular gas in a parsec-scale torus at the distance of NGC 1068. The Planck function sets an upper limit on the flux density for thermal line or continuum radiation for a given gas or dust temperature. Assuming $T \sim 100\,\mathrm{K}$ molecular gas distributed in a parsec-diameter source, thermal emission at millimeter wavelengths produces flux densities of order 0.1 milli-Jansky (mJy; 1 mJy $= 10^{-26}$ erg s^{-1} cm^{-2} Hz^{-1}). For comparison, the characteristic line sensitivity of current millimeter interferometers is of order a few mJy.

Because of these observational difficulties described above, the basic properties of the torus have been difficult to measure directly. Currently, the best direct measurements have come from centimeter-wave aperture synthesis measurements, which provide the requisite angular resolution, and infrared measurements, which measure the peak energy output from the torus. This review will cover these two broad categories of observations. Firstly reviewed are millimeter-wave (and complementary near-infrared) and centimeter-wave radio observations at that probe the molecular and atomic gas components via thermal emission, absorption lines, free-free (continuous) absorption, and maser (non-thermal) emission. We next review infrared measurements and torus models that constrain the properties of the torus. We conclude with an outlook for future observations, including the potential impact of future millimeter-wave and infrared interferometers towards *imaging* the torus.

2 Radio Observations

2.1 Thermal Emission Lines: CO & H_2

There are two classical, direct tracers of molecular gas in galaxies, namely, CO emission at millimeter wavelengths (usually the $1\rightarrow0$ and the $2\rightarrow1$ transitions at 2.6 mm and 1.3 mm wavelengths, respectively), and the H_2 $S(1)$ emission line at $2.12\,\mu\mathrm{m}$, which traces warmer ($\sim$1000 K) molecular gas. Current mm-wave interferometers can achieve subarcsecond resolution, corresponding to linear scales of 50–100 pc in the nearest Seyfert galaxies. There are as yet no systematic surveys of active nuclei at subarcsecond resolution, but a few key targets have been studied in considerable detail. NGC 1068 is easily the best studied to date. There, mm-wave observations have resolved the central CO emission into a 3 kpc radius starburst “ring” (actually, a pair of tightly wound spiral arms seen in radio continuum and millimeter line emission), linear structures aligned with the stellar bar, and a classical “twin peaks” structure at $\sim$100 pc radius which is also detected in H_2 $S(1)$ emission [5, 6], and HCN emission [7, 8].

The central “twin peaks” structure of NGC 1068 has gained much attention owing to the possibility that it may contribute significantly to the nuclear obscuration. Looking at the kinematics, the twin CO/H_2 peaks have been interpreted variously as either the loci of orbit crossings within the kpc-scale stellar bar (the classical “twin peaks” interpretation), an inclined, rotating nuclear disk [9], or a heavily warped disk that may mark the outer extent of the torus proper [10]. Interestingly, the twin CO/H_2 peaks orient roughly perpendicular to the [OIII] cone, motivating the torus interpretation.

On the other hand, it is not clear whether the "twin peaks" structure is unusual, and there remains no direct evidence of a physical connection to either a parsec-scale torus (if one is needed) or the central accretion disk. Such a distribution of molecular gas is generally expected to arise and is observed to occur in barred spiral galaxies, and one could argue that the molecular gas distribution of NGC 1068 (and other Seyferts) is not really all that different from that of the Milky Way and other "normal" spiral galaxies [11] (see also the review by Jogee in this volume). The present and forthcoming BIMA SONG [12] and NuGA [14] surveys will provide a useful database to interpret more generally the 100 pc–kpc scale molecular gas of Seyfert galaxies and address such issues as obscuration from larger scale distributions of cold gas.

2.2 Absorption Lines

Because of sensitivity limitations, it is currently very difficult to detect millimeter-wave or centimeter-wave line emission from thermally emitting gas associated with a parsec-scale obscuring torus. The detection of absorption lines, however, is primarily limited by the brightness of a background source. Radio galaxies and many Seyfert galaxies harbor bright radio continuum sources suitable as background sources for detecting thermal line absorption [13]. Using aperture synthesis (interferometric) techniques, the location and distribution of atomic and molecular gas can be mapped against extended radio sources. Of course, absorption line studies are limited in that AGN radio sources are rarely so friendly as to provide a uniform background; instead, near the AGN, the radio sources are usually jet-like, meaning that absorption can be detected only along a narrow axis. On the other hand, localized absorption detected against a radio jet still provides useful constraints on the extent of the absorption line gas.

The λ21 cm transition of the ground state of neutral hydrogen is a potential tracer of the warm atomic component of the torus. Because the transition occurs in the Rayleigh-Jeans limit, the line opacity, $\tau(v)$, is related simply to the ratio of the atomic hydrogen column density, N_H and the excitation temperature, T_S, commonly called the "spin temperature" because it is a statistical measure of the excitation of the electron spin in the ground state. For the high gas densities expected in the torus, the spin temperature should be very close to the gas temperature. For reference, the atomic hydrogen column density is given by

$$N_H = 1.456 \times 10^{22} \left(\frac{T_S}{8000\mathrm{K}} \right) \int \tau\left(v\right)\, dv\ \mathrm{cm}^{-2}, \tag{1}$$

where the integral is carried out over the line profile as a function of Doppler velocity v in km s^{-1}. A bright Seyfert nucleus might produce of order 100 mJy background continuum, and short integrations with the Very Large Array (VLA) provide sensitivities of order 1 mJy in a 20 km s^{-1} channel. Therefore, the VLA is rapidly sensitive to $N_H \sim 3 \times 10^{21}\,\mathrm{cm}^{-2}$.

Gallimore et al. [15] reported a systematic imaging survey of 21 cm absorption in Seyfert galaxies. The surprising result was that 21 cm absorption generally failed to detect absorption that could be uniquely identified with a compact, parsec-scale torus. The material was commonly detected at positions offset from the radio nucleus, and the 21 cm columns showed no correlation with measured X-ray absorption columns. The 21 cm absorption columns showed better correlation with the inclination of the host galaxy, although, interestingly, the inferred 21 cm columns are on average greater in Seyfert galaxies than they are through our own Galaxy. Kinematic models demonstrated that the 21 cm absorption usually traces 100 pc scale disks participating in normal galactic rotation. The ISM is predominantly molecular in the inner kiloparsec of non-active spiral galaxies. The warm atomic disks detected in Seyfert galaxies might therefore arise from an ambient, "normal" molecular cloud population (discussed above) that has been heated by the AGN.

There is no sign of 21 cm absorption from an obscuring torus even in those Seyfert 2 galaxies found by polarimetry to harbor a hidden Seyfert 1 nucleus. Mrk 348 is a dramatic example, because it harbors a bright, variable VLBI radio source that must be located very near the central engine [16], and yet no 21 cm absorption is detected. Either the obscuring torus is entirely molecular, which might occur if the atomic medium is swept away by radiation pressure and subsequently ionized (see Sect. 3.2); the spin temperatures are out of thermal equilibrium, which is unlikely given the high gas densities required for the torus; or very little 21 cm continuum radiation is produced by the central engine or makes it to the torus. We will return to this latter point in the discussion of free-free absorption, below.

There are however interesting exceptions to the 21 cm absorption studies. Mundell et al. [17, 18] observed NGC 4151, an intermediate Seyfert 1, with the long baseline MERLIN and VLBA telescopes. They found that the 21 cm absorption is confined to within ~10 pc of a single compact radio knot centrally located within a 240 pc long radio jet (Fig. 1). The measured opacity and centroid velocities vary over the absorbed radio source implying a clumpy atomic medium. Such absorbing clouds would have been easily detected against neighboring radio components, and so the simplest interpretation is that the 21 cm absorption arises from a clumpy, parsec-scale disk surrounding the central engine. Recently, Beswick, Pedlar, & McDonald [19] showed that the Seyfert nucleus of NGC 7674 shows a similar geometry, with the absorption localized to the inner hundred parsecs.

Another peculiar case is found in IC 5063, a Seyfert 2 galaxy. Morganti et al. [20] found high velocity 21 cm absorption against a kiloparsec-scale radio jet arising from the nucleus. The absorption line is broad, spanning

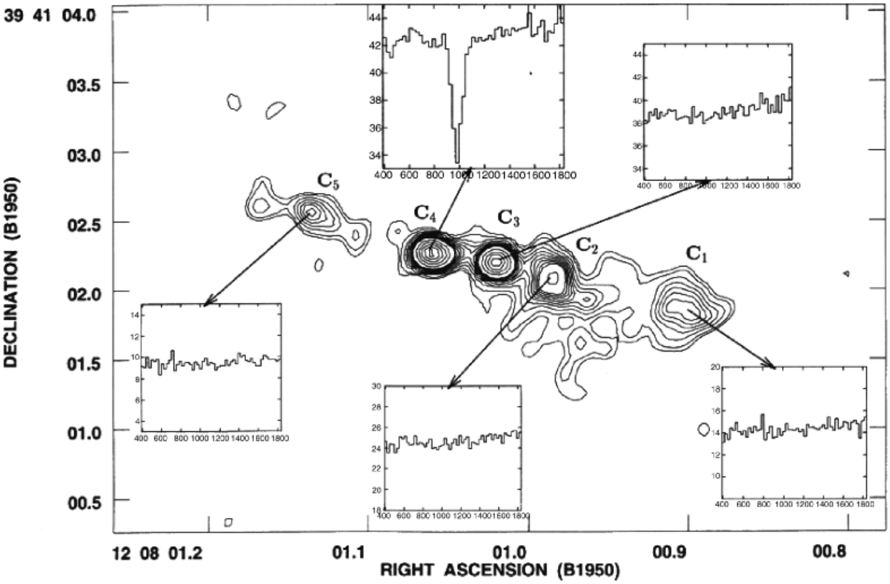

Fig. 1. MERLIN 21 cm absorption spectra (insets) of the brighter radio continuum sources (contours) of the jet of NGC 4151, from [17]. The absorption is localized to the C_4 source, which constrains the size of the 21 cm absorbing region to <10 pc along the jet. See [18] for higher resolution VLBA measurements

700 km s^{-1}, and is blueshifted. Since the gas is seen in absorption, it must be located on the near side of the radio source, and the neutral gas is probably participating in a rapid outflow.

The case for compact, 21 cm absorbing gas is clearer in radio-loud sources, particularly those with compact jets. Recently, Pihlström, Conway, & Vermeulen [21] demonstrated that the 21 cm absorption column is anti-correlated with the size of the (compact) radio source in a sample of GPS and CSS sources. Since these objects commonly resolve into compact, double-lobed structures with little radio continuum emission from the AGN proper, the simplest interpretation is that the neutral gas density increases towards the center of these objects. A handful of such sources have been studied in detail with VLBI arrays, and the 21 cm absorption commonly resolves over the span of the compact radio jet. The radio source 1946 + 708 is a dramatic example [22] (see Fig. 2): there, the absorption column and line-width increases towards the center of the jet. Were a kiloparsec-scale disk responsible for the absorption, projection onto the parsec-scale nuclear source should show no such strong concentration of absorption line properties on sub-parsec scales. The 21 cm absorption must arise from a parsec-scale distribution of cold, neutral gas surrounding the central engine, most likely in the form of a disk.

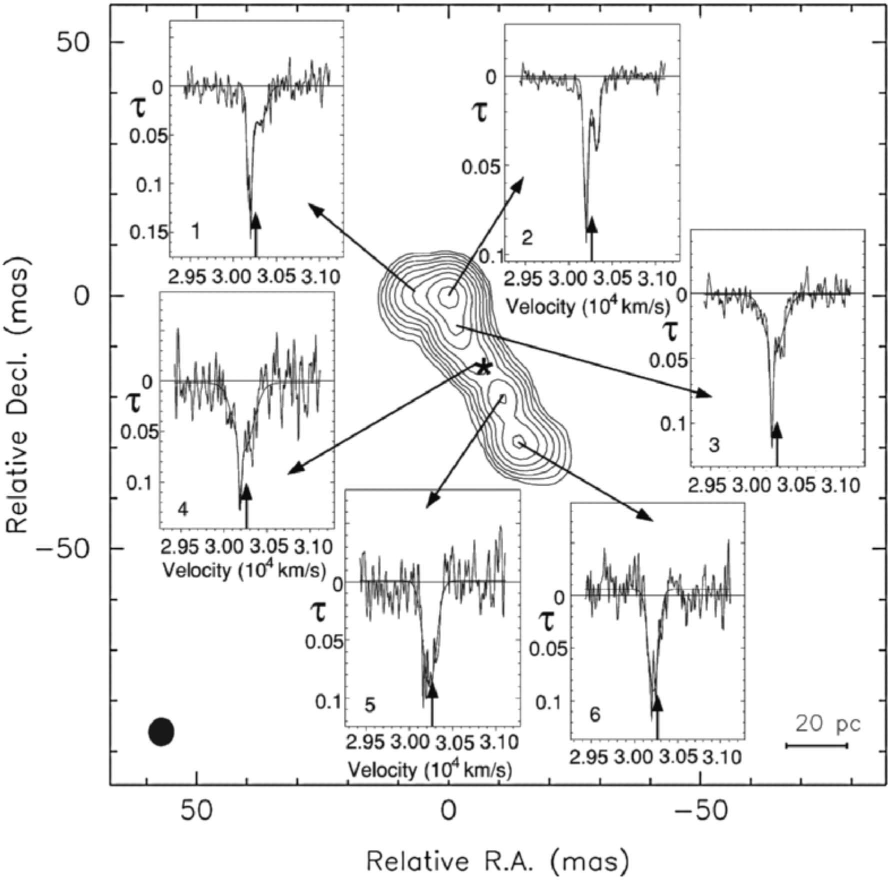

Fig. 2. VLBA 21 cm absorption spectra (insets) of the CSO radio source 1948 + 708 (contours), from [22]

Similar evidence for parsec–hundred parsec scale disks has been found in NGC 4261 [23], Cygnus A [24], Hydra A [25], & 4C31.04 [26], among others.

In contrast to these results, van Gorkom et al. [27] and Morganti et al. [28] demonstrated that 21 cm absorption is relatively uncommon in powerful radio galaxies. Unlike Seyfert galaxies, the nuclear radio sources of radio galaxies are commonly very bright at 20 cm, and so free-free absorption of the background continuum cannot explain the lack of detections. Instead, any compact disk surrounding the AGN must have a low covering fraction, perhaps because any such disk is geometrically thin. This result is consistent with evidence from other wavebands that no geometrically thick torus is necessary to explain the observed properties of radio galaxies.

2.3 Free-Free Absorption

The molecular clouds in a parsec-scale torus will be exposed to intense ionizing radiation. The result is an abundance of free electrons in the ionized skin and warm atomic layer that can absorb centimeter-wave radiation by the inverse bremsstrahlung process, which, in astronomy, is more commonly referred to as free-free absorption. The general expression for free-free opacity is fairly complicated but can be simplified at specific wavebands by applying a suitable approximation for the Gaunt factor. At centimeter wavelengths, the following expression is accurate to about 10%,

$$\tau_{ff}\left(\nu\right) = 0.08235\ \nu_G^{-2.1} T_e^{-1.35} \int n_e^2\, dl,$$

where ν_G is the continuum frequency in GHz, T_e is the electron temperature in K, and $\int n_e^2\, dl$ is called the emission measure, with the electron density n_e in cm^{-3} and path-length integral $\int dl$ in parsecs.

The free-free opacity has a strong inverse frequency dependence resulting in low frequency cutoffs in radio continuum spectra. Free-free absorption through NLR clouds should produce cutoffs below 1 GHz. In contrast, models for the torus predict cutoffs as high as 10 GHz because of the large (predicted) emission measure. Detecting free-free absorption has however proven tricky. Synchrotron self-absorption also produces a low frequency cutoff, and so it can be difficult to distinguish between the two absorption mechanisms in bright radio cores. The radio sources in Seyfert galaxies are often so compact, usually no larger than a few arcseconds, that it is difficult to resolve a free-free absorbed core from surrounding steep spectrum, optically thin synchrotron emission.

In spite of these difficulties, recent long-baseline aperture synthesis measurements have been demonstrating that free-free absorption may commonly occur in Seyfert nuclei. Pedlar et al. [29] found that the (near-)nuclear radio component of NGC 4151, the one showing localized 21 cm absorption, is also marked by a low frequency cutoff between 0.4 GHz and 1.6 GHz. The average brightness temperature of the compact radio source is simply too low for synchrotron self-absorption to have much impact; free-free absorption is the most likely cause of the low frequency cutoff. The emission measure requirements are modest, and NLR clouds might be sufficiently opaque at these low frequencies to affect the spectrum. On the other hand, the free-free absorption is localized (in projection) to the same region as the neutral hydrogen absorber, and it seems more likely then that the free-free absorber is associated with the neutral disk or torus.

Ulvestad et al. [16] presented indirect evidence for free-free absorption in their study of the Seyfert galaxies Mrk 231 (a Seyfert 1 nucleus) and Mrk 348 (Seyfert 2). They argue that, in both sources, the parsec-scale jets appear one-sided. The proper motions of the individual radio components are subrelativistic, however, and therefore relativistic de-boosting cannot suppress the receding jet component. Free-free absorption through a parsec-scale disk may instead hide the receding jet. The cutoff frequency would have to exceed the observing frequency, 15 GHz in this case, implying the high emission measures thought to be present in parsec-scale obscuring disks.

VLBA observations [30] have revealed low frequency cutoffs in the nuclear radio source in NGC 1068 and in the two radio sources associated with the "binary AGN" NGC 6240. In NGC 1068, the cutoff frequency is between 1.4 and 5 GHz, and in NGC 6240 the cutoff frequency for both radio nuclei occurs between 1.6 and 2.2 GHz. The argument applied to NGC 4151 similarly applies here: the average brightness temperatures are generally too low for synchrotron self absorption to apply, and so free-free absorption likely dominates the opacity. There is also evidence that, above 5 GHz, the radio continuum emission from the nuclear radio source of NGC 1068 may be dominated by free-free *emission* from the inner regions of a circumnuclear maser disk [31, 32].

Turning to radio-loud AGNs: we described above how 21 cm absorption is more commonly detected in compact, radio-loud AGNs (e.g., CSS, CSO, and GPS sources) compared to larger, classical radio galaxies. This general picture also *appears* to apply to detections of free-free absorption. For example, Bicknell, Dopita, & O'Dea [33] demonstrated that free-free absorption through a jet cocoon may be the origin of the high cutoff frequencies of GPS sources. The best (current) example for free-free absorption in a compact, circumnuclear disk is again the CSO $1946 + 708$ [22], in which a region of low-frequency attenuation is localized to the central 10 pc. The cutoff frequency for synchrotron self-absorption falls below the observed GHz-frequency range, and so free-free absorption is the most likely explanation. Free-free absorption in radio-loud AGNs is currently an active research area of radio astronomy [25], and the next few years should see a dramatic improvement in our understanding of this phenomenon.

2.4 Masers

H_2O masers have provided unusual insight to the circumnuclear environment of low luminosity AGNs like Seyfert and LINER nuclei. Luminous extragalactic H_2O masers were discovered in the late 1970's, and it was quickly recognized that they were more commonly found in active spirals than in other types of galaxies. Braatz et al. [35] produced the most complete search for H_2O masers in galaxies to date, and they found that, to the sensitivity limitations of the radio telescopes used, roughly 7% of active spirals harbor a luminous H_2O maser source. Early VLA measurements by Claussen & Lo [36] and references therein) demonstrated that the masers are usually concentrated at the center of the galaxy, all but proving that the masers are somehow connected to the AGN. Because the emission mechanism is non-thermal and self-amplifying, masers greatly exceed the thermal limit and produce high brightness temperatures. The rest frequency of the H_2O maser transition is near 22 GHz within a standard radio astronomy waveband, and, owing to the high brightness temperatures, the distribution and kinematics of individual maser beams, or "spots," can be mapped using VLBI techniques.

The landmark extragalactic H_2O maser source has been NGC 4258 [37, 38, 39]. The integrated maser spectrum is rich with a group of bright, blended lines near the systemic velocity of the host galaxy and also a number of high-velocity, "satellite" lines offset by roughly ± 1000 km s^{-1} from systemic. The grouping of masers in velocity probably arises from a radiative transfer effect; masers will be most efficiently amplified along pencil beams through the disk with the smallest spread of Doppler velocities. Clumping and the effects of cloud-cloud alignments probably further enhance the emission at particular velocities. Clumpiness appears to be an essential ingredient in understanding the production and behavior, particularly variability, of disk masers in general [40].

The near-systemic lines drift towards higher Doppler velocities with time but the satellite lines remain fixed in velocity. It was recognized by several groups that this pattern of radial acceleration is consistent with centripetal acceleration of an edge-on, rotating ring of molecular gas. Follow-up VLBA observations by Miyoshi et al. [38] showed that the maser spots trace a linear structure spanning $\sim$0.5 pc. The position-velocity diagram of the maser spots precisely matches the profile expected from an edge-on disk undergoing Keplerian rotation around a central mass concentration of $3.5 \times 10^7\, M_\odot$ within the inner 0.13 pc (the inner radius of the maser spots). Based on detailed kinematic models of the maser spot distribution and Doppler velocities, the disk must be geometrically thin, gently warped , and viewed nearly edge-on. Proper motion studies later confirmed that the maser spots undergo rotation.

The maser disk of NGC 4258 probably contributes little to any nuclear obscuration because it is geometrically thin. The VLBA observations also detected a parsec-scale radio jet located at the center of the maser disk and oriented along the symmetry axis of the disk. This arrangement is exactly that expected for an accretion disk and jet in the AGN paradigm! Motivated by this result, Neufeld and Maloney [41, 42] demonstrated how X-rays arising from the inner accretion disk could heat molecular gas in an outer, molecular region of the disk and naturally produce H_2O masers. Disk warping is a key element of the model, because it allows the outer disk to be heated obliquely over a large range of radii and correspondingly spreads out the distribution of masers. Clumpiness may also contribute, because X-rays can travel farther in a clumpy disk, and, as a result, the radial distribution of maser spots increases.

NGC 1068 similarly shows a disk-like group of masers associated with its nuclear radio source [43, 34, 44] (see Fig. 3). Like NGC 4258, the disk appears to be geometrically thin, and so probably contributes little to the nuclear obscuration. There are otherwise a few interesting differences. For instance, the maser disk is larger than that of NGC 4258, probably owing to the greater

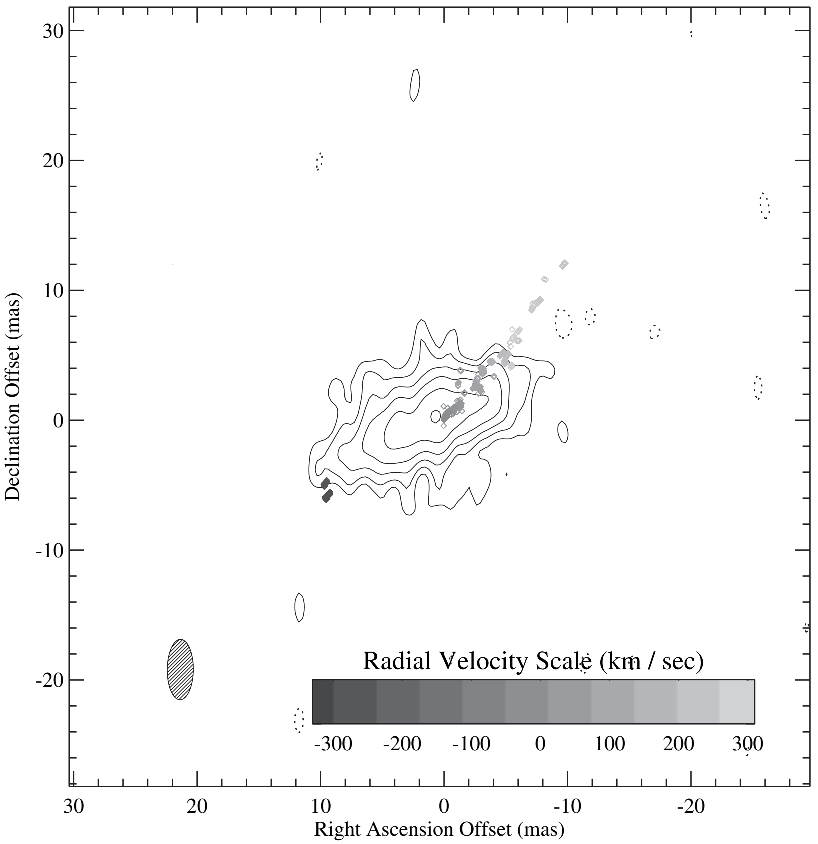

Fig. 3. Positions of the nuclear H_2O maser spots (*open circles*) relative to the nuclear radio continuum source S1 (*contours*) of NGC 1068. The maser spots are greyscaled based on their Doppler velocity relative to systemic according to the velocity scalebar located at the bottom of the figure. The linear scale is 10 mas = 0.72 pc. The maser spot positions were derived from Greenhill & Gwinn [34], and the VLBA continuum image and maser astrometry are from new work by Gallimore, Baum, & O'Dea [32]. The characteristic brightness temperatures of the 8.4 GHz continuum source range from a few $\times 10^5$ K to $\sim 5 \times 10^6$ K. The VLBA continuum observations resolve out the extended jet emission, but MERLIN and VLA measurements show that the local jet axis runs nearly north-south along PA 12°. Both the S1 radio continuum source and the H_2O masers probably trace emission from the hot and cold phases, respectively, of an edge-on, geometrically thin disk

X-ray luminosity of the (hidden) AGN. There are also weak OH masers that appear to be associated with the H_2O maser disk. The rotation velocity falls off more slowly than Keplerian, and some have argued that the maser disk must be partly self-gravitating. The maser spots align along a position angle of ∼45° from the local jet axis, suggesting a strong warp between the inner accretion disk and the outer maser disk. The nuclear masers of NGC 1068 also show coherent brightness variations over time scales of a few months to a few years, which is much shorter than the dynamical time scale of the disk. The likeliest explanation is that the masers are responding to variations of the luminosity or spectrum of the hidden AGN; this result provides support for the picture that the masers are powered by X-ray heating and not local effects such as shocks. Supporting this picture, NGC 1068 showed a remarkable flare that appeared simultaneously on opposite sides of the maser disk. Explaining the simultaneity of the flare requires both variable input from the central engine (probably a flare of the central X-ray source) and clumpiness to isolate the flare to narrow ranges in velocity.

After NGC 4258 and NGC 1068, there are few extragalactic maser sources that resolve unambiguously into disk-like structures. Arguably the next best cases at the time of this writing are the Circinus galaxy [45] and NGC 3079 [46]. Other maser sources resolve into structures that better align *with* the jet axis, and may trace ambient ISM that has been shock-heated by interaction with the radio jet. NGC 1052 is a standout example [47]. In both Circinus [45] and NGC 1068 [43, 44] there is also evidence for masers associated with an outflow or interaction with the jet apart from the disk-like masers. The outlook for finding additional maser disks is positive, however, as single dish surveys continue to turn up new candidates. The spectra of many of these candidates show the three groups of maser lines as expected for maser disks. The only difficulty is that, since extragalactic maser sources are rare, new candidates are turning up mainly in distant objects and are correspondingly faint, which will make it difficult to study them with current VLBI capabilities.

3 Infrared SEDs and Modeling

3.1 Overview

An obscuring dusty torus should reradiate in the IR the fraction of nuclear luminosity it absorbs, and AGN continua indeed show significant IR emission, as is evident from Fig. 4. The composite data for type 1 sources shows at wavelengths $\lesssim 1\,\mu\mathrm{m}$ the power law emission expected from the AGN accretion disk. The excess radiation at longer wavelengths can be attributed to reprocessing by the dusty torus [49]. The right panel shows data for some representative type 2 sources. In accordance with unified schemes, the spectral shape of the IR excess at $\gtrsim 2\,\mu\mathrm{m}$ in type 1 objects is very similar to the emission from typical type-2 sources, where the AGN is blocked and only the torus IR radiation is detected. Yet in spite of this basic agreement, the IR observations present some rather puzzling features:

1. X-ray observations reveal a large range of torus column densities, $N_{\mathrm{H}} \sim 10^{22} - 10^{25}\ \mathrm{cm}^{-2}$. In spite of the huge range of corresponding optical

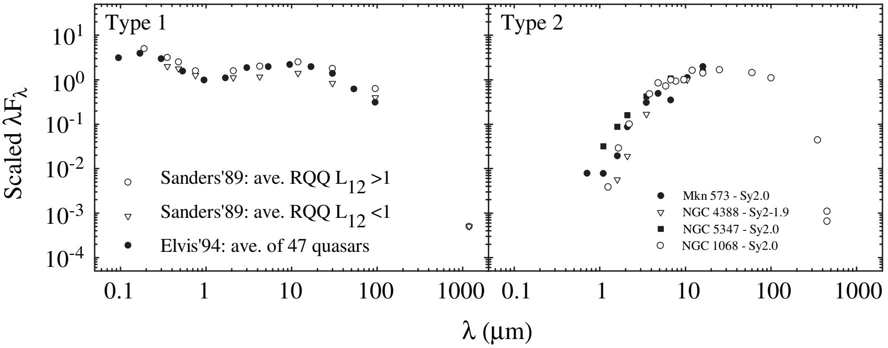

Fig. 4. AGN spectral energy distributions (SEDs). *Left*: Type 1 composite SED from various compilations, normalized to unity at 1 μm. *Right*: Some representative type 2 sources. Data from [48], normalized at 10 μm

depths, the observed spectral energy distributions (SEDs) show only moderate variation in the infrared.

2. Type 2 sources display the typical 10 μm dust silicate feature in absorption, but the absorption depth is never large and is not well correlated with the column density inferred from X-ray observations.
3. The 10 μm feature is absent altogether in type 1 sources, even though the dust is presumably the same in both types of AGN.

Krolik & Begelman [50, 51] present a number of arguments that the obscuring tori comprise a large number of individually very optically thick, dusty molecular clouds. But because of the difficulties in modeling a clumpy medium, early modeling calculations approximated the density distribution with a uniform one instead. Pier & Krolik [52, 53] were the first to explore dust radiative transfer in a toroidal geometry. Ignoring scattering, their model results still produced reasonable overall agreement with observations. However, a major shortcoming was an enormous silicate absorption feature at high equatorial optical depths, contrary to observations. In addition, the silicate feature has never been observed in emission in either type 1 or type 2 sources but the models managed to suppress it only in a narrow range of torus parameters and viewing angles. Finally, the spectral range of the model infrared emission was not as broad as observed.

The problems persisted in subsequent calculations that employed variations of the geometry. Granato & Danese [54] and Granato et al. [55] used an extended flared disk with homogeneous dust distribution. Their models produced an overall agreement with observations, the main shortcoming was a strong 10 μm emission feature in pole-on viewing from the disk hot surface layer. In addition, the moderate optical depths ($A_V \sim 6 - 60$) in their calculations are not in accord with the X-ray observations of large obscuring columns, corresponding to A_V of at least a few hundreds. Efstathiou &

Rowan-Robinson [56] advocated a tapered disk geometry and were able to minimize the silicate feature in face-on viewing and reproduce the observed shapes of the IR continuum in selected sources. However, their model near-infrared slopes are too steep compared to those observed in type 2 sources [48], and in many cases the silicate absorption feature is too deep.

3.2 Clumpy Tori

The recent inclusion of clumpiness in torus models resolves most of the difficulties [57, 58]. The fundamental differences between clumpy and continuous density distributions stem from the fact that radiation can propagate freely between different regions of an optically thick medium when it is clumpy, but not otherwise.

In the clumpy models, the torus inner boundary is determined by dust sublimation at $R_\mathrm{i} = 0.9\,\mathrm{pc} \times L_{12}^{1/2}$, where we have re-normalized the AGN luminosity to $L_{12} = L/10^{12} L_\odot$; the corresponding angular scale is 20 mas for a $10^{12} L_\odot$ source at a distance of 10 Mpc. This relation holds for optically thick clouds of standard Galactic interstellar dust with a sublimation temperature of 1,500 K. According to unified schemes, viewing the same torus from the directions shown in Fig. 5 should give rise to the SEDs observed in type 1 and type 2 sources. From comparison with overall properties of observed SEDs, especially the CfA sample of Seyfert galaxies [48], the data can be described adequately with standard dust and the following torus properties:

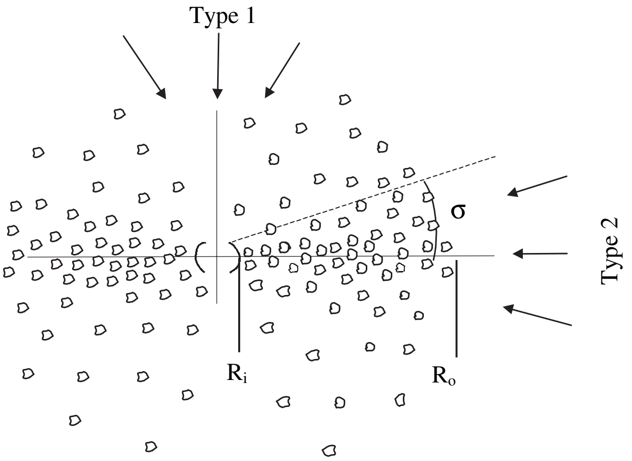

Fig. 5. Geometry for modeling AGN clumpy tori between radii R_i and R_o. The average number of clouds along a radial ray decreases with inclination from the equatorial plane, with the distribution width characterized by σ

- Gaussian variation for the cloud angular distributions. The average number of clouds along a radial ray at angle β from the equatorial plane is $\mathcal{N}(\beta) = \mathcal{N}_0 e^{-(\beta/\sigma)^2}$ with $\mathcal{N}_0 = 5$–10 and $\sigma = 45^\circ \pm 15^\circ$.
- The mean free path between clouds varies radially as a power law r^q with $q = 1$–2.
- Torus outer radius $R_\mathrm{o} = (10\mathrm{pc}–100\mathrm{pc}) \times L_{12}^{1/2}$. While this range ensures proper reproduction of observed SEDs, pinning down the actual size within this range is surprisingly difficult, even in high-resolution imaging, because of the rapid decline of brightness with distance from the AGN. The decline becomes more moderate as the wavelength increases, and imaging in the far-infrared with ALMA holds the promise of distinguishing a 30pc torus from a 100pc one in the foreseeable future.
- Optical depth at visual $\tau_\mathrm{V} \geq 40$ for individual clouds. The torus emission varies only slightly when τ_V increases from 40 to 100, and hardly at all during further increase. The reason is that each cloud is heated from outside so only its surface is heated significantly when τ_V is large, further increase of τ_V only adds cool material and the emission saturates (similar to standard black-body emission).

The first two puzzling features of IR emission from AGN are simply hallmarks of clumpy dust distributions, independent of the distribution geometry. The SED would show a moderate range and the 10 μm absorption would never be deep even if the cloud configurations were spherical or irregular rather than toroidal. Only the complete absence of the 10 μm feature from type 1 spectra is a consequence of the toroidal geometry of the cloud distribution.

The torus clumpiness has some additional important consequences. IR flux is collected from the entire observed area and thus is determined by the average number of clouds along radial rays through the torus. In contrast, X-ray attenuation is controlled by the clouds along one particular ray, the line of sight to the AGN, and their number can differ substantially from the average. For example, in a Poisson distribution with $\mathcal{N}_0 = 5$ more than 80% of the paths will have a number of clouds different from 5, and the probability to encounter just 1 cloud or as many as 9 is a full 20% of the probability to encounter the average 5. Two sources with the same average $\mathcal{N}_0$ will have an identical IR appearance, yet their X-ray absorbing columns could differ by an order of magnitude. Therefore, type 2 sources with the same IR properties are expected to display a large scatter in X-ray properties, as observed.

Type 1 are distinguished from type 2 sources by the visible AGN contribution they display. In a clumpy torus with Gaussian angular distribution, the AGN visibility becomes an issue of probability. In spite of the strong variation with viewing angle, there is always a finite probability for an unperturbed line of sight to the AGN even in the torus equatorial plane. Such a source would be classified as type 1, in contrast with the strict interpretation of unified schemes. A few recorded transitions between type 1 and 2 may correspond to transits of obscuring clouds across the line of sight to the

AGN. A recent study by Risaliti, Elvis & Nicastro [59] finds X-ray variability on all available time scales – from months to 20 years, with the latter simply marking the longest time span currently in the data archives. They conclude that the X-rays are absorbed by clouds belonging to two populations. The first one includes clouds at very short distances ($\lesssim 10^{17}$ cm) from the AGN. These are dust free as they are located within the dust sublimation radius. The second involves clouds at ∼5–10 pc, presumably in the dusty torus.

Dust emission provides a snapshot of the torus that carries no information about its dynamics. Only line observations can probe the kinematics of clouds in the obscuring torus. The properties of individual clouds can be deduced from their volume filling factor, which is not constrained by the SED modeling. A reasonable, though not unique, assumption is a 10% filling factor constant throughout the torus. For a model with $q = 1$, $\mathcal{N}_0 = 5$, $\tau_V = 100$, $L_{12} = 1$ and $R_o = 100\,\mathrm{pc}$ this implies clouds that vary from size of ∼0.1 pc and gas density $\sim 3 \times 10^5\,\mathrm{cm}^{-3}$ in the torus inner regions to size ∼10 pc with density $\sim 3 \times 10^3\,\mathrm{cm}^{-3}$ at the outer edge. These resemble ordinary molecular clouds, thus the cloud motions can be mapped in molecular line observations. The intercloud space may be devoid of gas because of radiation pressure by the intense AGN field.

4 Future Observations

The simultaneous and conflicting need for fine angular resolution and sensitivity to low surface brightness emission has made it difficult to study the cold gas near AGNs in detail. This situation is soon to change with the advent of the Atacama Large Millimeter Array (ALMA) telescope, which will provide sensitive line and continuum observations at millimeter $-100\,\mu\mathrm{m}$ wavelengths, and infrared interferometers such as the Keck interferometer and the Very Large Telescope Interferometer (VLTI), which has already proven successful at single baseline measurements of dust emission from AGNs. Based on the current specifications, ALMA will provide ∼50 K line brightness temperature sensitivity and an angular resolution of $\sim 0.2''$ at CO $1 \rightarrow 0$, sensitive to molecular gas on ∼10 pc scales in the nearest Seyfert galaxies. The prospect for imaging cold dust is equally encouraging: the continuum sensitivity at $\lambda 400\,\mu\mathrm{m}$ is of order a few degrees K at an angular resolution of ∼20 mas, or roughly 2 pc in the nearest Seyferts.

HST/NICMOS imaging [60] and ground-based adaptive optics and speckle techniques [61, 62, 63], and references therein) have usefully constrained the size of the nuclear infrared source, presumably the inner regions of the torus, to parsec-scales, but the resolution of these instruments and techniques is still too coarse to image the inner regions of the obscuring medium. The VLTI has however recently proven successful at probing the geometry on parsec–subparsec scales. Using the MIDI instrument, Jaffe et al. [64] have measured two single-baseline fringes on NGC 1068 over the wavelength range 8–13.5 μm.

The resolution is ∼10 mas, or ∼0.7 pc, effective for measuring the geometry of the putative parsec-scale obscuring torus. Although the two baselines do not suffice to make an image of the warm, dusty medium, they at least constrain the scale of the mid-infrared emission to ∼2–3 pc, and the fringes are consistent with a thick disk viewed edge-on, all but proving the existence of the parsec-scale obscuring medium.

Wittkowski et al. [65] have similarly measured the $2\,\mu m$ fringes on NGC 1068 using VLTI/VINCI/MACAO. They find a very compact source, only ∼5 mas (∼0.4 pc) along PA 44.9°, comparable in size to the maser disk, and they argue that they are actually resolving out much of the parsec-scale emission that is only marginally resolved by adaptive optics and speckle measurements! Instead, VINCI probably recovers emission from torus substructure, or perhaps structure from the maser / accretion disk. The prospects are very exciting: multiple baseline measurements will ultimately contribute to a few mas resolution image, at which point we may finally confront the current torus models with *imaging* information in addition to infrared SEDs.

References

1. Antonucci, R. 1993, ARAA, 31, 473
2. Gunn, J.E. 1979, in Active Galactic Nuclei, eds. C. Hazard, & S. Mitton (Cambridge: CUP), 213
3. Antonucci, R.R.J., & Miller, J.S. 1985, ApJ, 297, 621
4. Ward, M., Elvis, M., Fabbiano, G., Carleton, N.P., Willner, S.P., & Lawrence, A. 1987, ApJ, 315, 74
5. Alloin, D. et al. 2001, A&A, 269, L33
6. Galliano, E., & Alloin, D. 2002, A&A 393, 43
7. Tacconi, L. et al. 1994, ApJ, 426, L77
8. Helfer, T., & Blitz, L. 1995, 419, 86
9. Galliano, E., Alloin, D., Granato, G.L., Villar-Martin, M. 2003, A&A, 412, 615
10. Schinnerer, E. et al. 2000, ApJ, 533, 826
11. Helfer, T., & Blitz, L. 1993, ApJ, 419, 86
12. Helfer, T. et al. 2003, ApJS, 145, 259 (BIMA SONG)
13. Dickey, J.M. 1986, ApJ, 300, 190
14. Garcia-Burillo, S. et al. 2003, in Active Galactic Nuclei: from Central Engine to Host Galaxy, eds. S. Collin, F. Combes, & I. Shlosman (San Francisco: ASP), 423
15. Gallimore, J.F., Baum, S.A., O'Dea, C.P., Pedlar, A., & Brinks, E. 1999, ApJ, 524, 684
16. Ulvestad, J.S., Wrobel, J.M., Roy, A.L., Wilson, A.S., Falcke, H., & Krichbaum, T.P. 1999, ApJ, 517, L81
17. Mundell, C.G., Pedlar, A., Baum, S.A., O'Dea, C.P., Gallimore, J.F., & Brinks, E. 1995, MNRAS, 272, 355
18. Mundell, C.G., Wrobel, J.M., Pedlar, A., & Gallimore, J.F. 2003, ApJ, 583, 192
19. Beswick, R.J., Pedlar, A., & McDonald, A.R. 2002, MNRAS, 335, 1091

20. Morganti, R., Oosterloo, T., & Tsvetanov, Z. 1998, AJ, 115, 915
21. Pihlström, Y.M., Conway, J.E., & Vermeulen, R.C. 2003, A&A, 404, 871
22. Peck, A.B., & Taylor, G.B. 2001, ApJ, 554, L147
23. van Langevelde, H.J., Pihlström, Y.M., Conway, J.E., Jaffe, W., & Schilizzi, R.T. 2000, A&A, 354, L45
24. Conway, J.E. 1998, in Radio emission from Galactic and Extragalactic Compact Sources, eds. J.A. Zensus, G.B. Taylor, & J.M. Wrobel (San Francisco: ASP), 231
25. Taylor, G.B. 1996, ApJ, 470, 394
26. Conway, J.E. 1996, in Extragalactic Radio Sources, eds. R.D. Ekers, C. Fanti, & L. Padrielli (Dordrecht: Kluwer), 92
27. van Gorkom, J.H., Knapp, G.R., Ekers, R.D., Ekers, D.D., Laing, R.A., & Polk, K.S. 1989, AJ, 97, 708
28. Morganti, R., Oosterloo, T.A., Tadhunter, C.N., van Moorsel, G., Killeen, N., & Wills, K.A. 2001, MNRAS, 323, 331
29. Pedlar, A., Fernandez, B., Hamilton, N.G., Redman, M.P., & Dewdney, P.E. 1998, MNRAS, 300, 1071
30. Gallimore, J.F., & Beswick, R. 2004, AJ, 127, 239
31. Gallimore, J.F., Baum, S.A., & O'Dea C.P. 1997, Nature, 388, 852
32. Gallimore, J.F., Baum, S.A., & O'Dea C.P. 1997, ApJ, 613, 794
33. Bicknell, G.V., Dopita, M.A., O'Dea, C.P. 1997, ApJ, 485, 112
34. Greenhill, L.J., & Gwinn, C.R. 1997, Ap&SS, 248, 261
35. Braatz, J.A., Wilson, A.S., & Henkel, C. 1996, ApJS, 106, 51
36. Claussen, M.J., & Lo, K.-Y. 1986, ApJ, 308, 592
37. Watson, W.D., & Wallin, B.K. 1994, ApJ, 432, L35
38. Miyoshi, M. et al. 1995, Nature, 373, 127
39. Herrnstein, J.R. et al. 1999, Nature, 400, 539
40. Kartje, J.F., Königl, A., & Elitzur, M. 1999, ApJ, 513, 180
41. Neufeld, D.A., Maloney, P.R., & Conger, S. 1994, ApJ, 436, L127
42. Neufeld, D.A., & Maloney, P.R. 1995, ApJ, 447, L17
43. Gallimore, J.F., Baum, S.A., O'Dea, C.P., Brinks, E., & Pedlar, A. 1996, ApJ, 462, 740
44. Gallimore, J.F. et al. 2001, ApJ, 556, 694
45. Greenhill, L.J. et al. 2003, ApJ, 590, 162
46. Trotter, A.S., Greenhill, L.J., Moran, J.M., Reid, M.J., Irwin, J.A., & Lo, K.-Y. 1998, ApJ, 495, 740
47. Claussen, M.J., Diamond, P.J., Braatz, J.A., Wilson, A.S., & Henkel, C. 1998, ApJ, 500, L129
48. Alonso-Herrero, A. et al. 2003, AJ, 126, 81
49. Barvainis, R. 1987, ApJ, 320, 57
50. Krolik, J.H., & Begelman, M.C. 1986, ApJ, 308, L55
51. Krolik, J.H., & Begelman, M.C. 1988, ApJ, 329, 702
52. Pier, E.A., & Krolik, J.H. 1992, ApJ, 401, 99
53. Pier, E.A., & Krolik, J.H. 1993, ApJ, 418, 673
54. Granato, G.L., & Danese, L. 1994, MNRAS, 268, 253
55. Granato, G.L., Danese, L., & Franceschini, A. 1997, ApJ, 486, 147
56. Efstathiou, A., & Rowan-Robinson, M. 1995, MNRAS, 273, 649
57. Nenkova, M., Ivezić, Z., & Elitzur, M. 2002, ApJ, 570, L9
58. Nenkova, M., & Ivezić, Z., & Elitzur, M. 2004, in prep.

59. Risaliti, G., Elvis, M., & Nicastro, F. 2002, ApJ, 571, 234
60. Thompson, R.I., & Corbin, M. 1999, Ap&SS, 266, 79
61. Rouan, D. et al. 1998, A&A, 339, 687
62. Weinberger, A.J., Neugebauer, G., & Matthews, K. 1999, AJ, 117, 2748
63. Marco, O., & Alloin, D. 2000, A&A, 353, 465
64. Jaffe, W. et al. 2004, Nature, 429, 47
65. Wittkowski, M. et al. 2004, A&A, 418, L39

AGN Beyond the 100pc Scale

R.A.E. Fosbury

Space Telescope – European Coordinating Facility, Garching bei
München, Germany
`rfosbury@eso.org`

1 Introduction

Galaxies and their nuclear SuperMassive Black Holes (SMBH) appear to be intimately related components of the same fundamental formation and evolutionary process [3], [7]. When the host galaxy contrives to feed its nuclear monster, it unleashes a torrential energy output that can far outshine the gentler and, apart from the occasional supernova, more constant shining of the stellar populations. The observed proportionality between the masses of the black hole and the galaxy bulge (see Chap. 8 of this volume) suggests a close coupling – a negative feedback loop – that limits SMBH growth. One of our more important goals is to understand the nature of this mechanism. To do this, we must look at the galaxy and its larger environment, not at the nuclear regions alone.

A brightly shining AGN does us a useful service by illuminating its host galaxy with an ionizing radiation field. The "fluorescent emission" that this produces can be studied to tell us about some rather fundamental galaxy properties such as, via emission line kinematics, the mass and, via line ratios and ionization modelling, the chemical composition of the interstellar medium (ISM). This is particularly pertinent to the studies of galaxies at early epochs (high redshift) where the hosts are exceedingly faint in the absence of such excitation. Such investigations are crucial for our picture of the assembly of massive galaxies and the chemical enrichment that is an integral part of the process.

With a luminosity that can far exceed that of its host, an AGN will not only illuminate its surroundings, it will push material around. Radiation pressure will affect dust clouds, ambient gas will be shocked in the presence of the particle jets, flows will be driven and the ISM will be generally mixed and redistributed. All these processes have to be recognised and understood if we are to generate a complete picture.

We know that an active galaxy possesses an intrinsic axial symmetry due to the spin of the SMBH. This symmetry leaves its imprint on the host in the form of an anisotropic illuminating radiation field and the effects of the jets that initially trace the AGN spin axis. This has the important consequence that active galaxies appear quite different when viewed from large and small angles to the AGN spin axis. The unravelling of these orientation effects, and

R.A.E. Fosbury: *AGN Beyond the 100pc Scale*, Lect. Notes Phys. **693**, 121–141 (2006)
DOI 10.1007/3-540-34621-X_5

the realisation that the purely empirical AGN classification schemes had to be recrafted, was the work of the last two or three decades. The orientation-based unification of AGN (type 1 viewed ∼ pole-on; type 2 ∼ equator-on) is now well-established [5] and we have moved on to the next, and intrinsically more interesting, task of delineating the more subtle evolutionary processes.

Finally, and this is perhaps the most fundamental question of all, we want to see the role of the AGN in the energy balance of the Universe as a whole. What fraction of the overall photon energy production comes from gravity (matter collapse onto SMBH) and what fraction from nucleosynthesis (element transmutation in stars) and how does this ratio change with cosmic epoch [19]? The observations tell us that the ratio has been not so enormously different from unity. Is this another cosmic coincidence or is it an inevitable consequence of the laws of nature? To answer this question, we need to find all the AGN in the Universe and then to measure their bolometric luminosities. The host galaxies can serve as the "bolometers" that measure the output of the type 2 objects whose AGN we do not see directly.

The aim of this chapter is not to give an exhaustive review of "extranuclear" activity. Rather, by choosing some illustrative examples, we can construct a catalogue of the more important AGN–host galaxy interactions and so equip ourselves to attack the unsolved fundamental questions concerning the relationship between the compact structures at the nucleus and their large-scale galactic surroundings.

2 AGN–host Interactions

The host galaxy "feels" the presence of its nuclear SMBH through the gravitational effect of its mass, through the ionizing and non-ionizing radiation field emitted by the AGN and by the particle flows in the jets and radiation pressure driven winds with any associated magnetic fields. In this chapter, our concern lies principally with the radiation field and with the AGN-related flows.

2.1 The AGN Radiation Field

The primary radiation source associated with the SMBH is assumed to be more-or-less isotropic. However, beyond the scale of 100pc, this intrinsic isotropy may have been destroyed by absorption, most notably by the presence of an optically thick equatorial torus surrounding the black hole (see Chap. 4). Such a structure can produce sharply defined shadows and it is thought to be responsible for most, if not all, of the "ionization cones" first seen in the emission line images of Seyfert galaxies [13]. A schematic illustration of the origin of this form of isotropy is shown in Fig. 1.

As their name implies, the ionization cones are delineated by the shadowing of the AGN radiation below the Lyman limit at 912Å. The optical

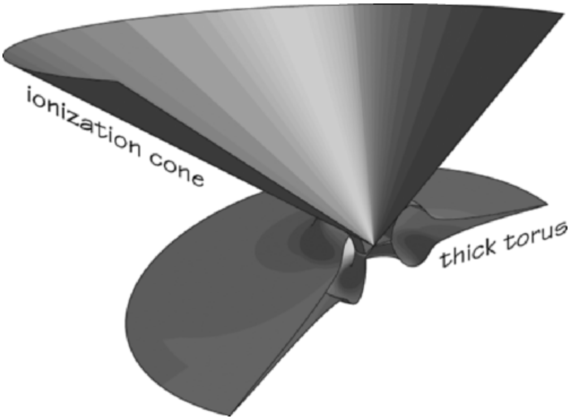

Fig. 1. A sketch of the optically thick nuclear torus whose shadow produces a sharply defined "ionization cone" of the type that is commonly seen in Seyfert 2 galaxies

properties of the torus will clearly vary with wavelength and it is expected to become transparent, in some objects, in the hard X-ray band and also somewhere in the mid-far infrared. In an attempt to illustrate how the apparent isotropy varies with wavelength, Fig. 2 gives a schematic view in four

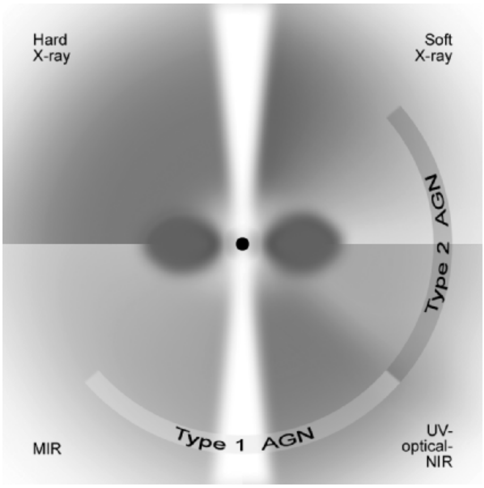

Fig. 2. A schematic representation of the apparent anisotropy of the AGN radiation field in different wavebands. The torus is essentially opaque from the MIR through to the soft X-ray but can become transparent outside this range. This pattern is useful for identifying type 2 AGN in multiband surveys. The spin axis has been left clear to indicate the possible presence of a particle jet and its associated, beamed radiation field

wavebands from hard X-rays to the mid-infrared (MIR). Such a concept is useful when designing strategies for finding type 2 AGN in multiband surveys.

The radiation field affects the material within these ionization cones in a number of ways, each of which produces characteristic observable phenomena. The photoionization process results in a spatially extended emission line region which is variously known as the Extended Emission Line Region (EELR) or the (Extended) Narrow Line Region (ENLR). The term "narrow" is used to contrast this region with the much smaller – and much less massive – Broad Line Region (BLR) surrounding the nucleus. The emission line widths actually extend over a broad range of values from the relatively quiescent interstellar medium of the galaxy which happens to be illuminated by the nucleus, to material that may be shocked by AGN-driven winds or jets. In many objects, there appears to be a correlation between line width and ionization state that may reflect either differences in density or the presence of alternative ionization mechanisms [10]. A notable characteristic of the NLR is the wide range of ionization levels present in the gas. This is a product of the shape of the ionizing Spectral Energy Distribution (SED) emitted by the AGN which extends from the Lyman limit up to very high energies and is markedly different in its effects from a black body-like stellar ionizing spectrum. The high energy photons can penetrate deep into clouds that are optically thick to the Lyman continuum where they partially ionize and heat a predominantly neutral gas, resulting in the coexistence of lines from [O I], [N I], [S II] etc. with lines from much more highly ionized species including, sometimes, even the "Coronal" lines such as [Fe X] [9].

The presence of dust within the ionization cones has a number of consequences. AGN photons can ionize dust grains, donating high kinetic energy electrons to the gas and so affecting the thermal balance. The photons can also efficiently transfer momentum to grains and thus, by dragging the gas along with them, produce radiation pressure driven winds. The presence of these winds can be inferred from the emission line shifts and widths. Another consequence of dust is the result of its high cross section for scattering. Dust clouds can act as "mirrors" that allow us to see the AGN along indirect sightlines in sources where the AGN is obscured. The presence of scattered light can be most directly inferred by measuring linear polarization. In the axisymmetric bi-conical geometry of an AGN, a net linear polarization – with the $\boldsymbol{E}$-vector perpendicular to the cone axis – can remain even after the spatial averaging by the polarimeter of unresolved sources. Hot electrons can also Thompson-scatter and, hence, polarize, AGN radiation although electrons are far less efficient scatters than dust grains, and clouds with sufficient Thompson optical depth appear quite rare away from the immediate nuclear environment. Electron-scattered spectra will be significantly smeared in wavelength due to high Doppler velocities in a hot, highly ionized plasma. The presence of broad-line (type 1) AGN is, indeed, demonstrated in polarized light in many type 2 objects. The energy absorbed – rather than scattered – by the dust will be re-radiated somewhere in the mid-far infrared (MIR, FIR), depending on the equilibrium temperature attained by the dust grains in a given AGN radiation field and at a given radial distance [15]. In general, such AGN-heated dust will compete with starburst-heated dust to dominate the FIR emission from an active galaxy.

2.2 Winds and Jets

The process of accretion onto the SMBH results in the generation of substantial amounts of kinetic energy. This is transported both in the form of highly collimated jets (see Chap. 2), seen predominantly in radio-loud sources, and as more isotropic winds. The evidence from some (radio quiet) Broad Absorption Line (BAL) QSOs suggests that fast winds can carry a substantial fraction of the AGN power, perhaps even approaching that of its radiative output. While the AGN radiation field produces its dominant effects on the ISM of the host galaxy, the winds and jets probably transport their energy to larger scales and dissipate it in the tenuous – and already hot – intergalactic medium (IGM). The passage of this kinetic energy through the galaxy can, however, influence the way the ISM is distributed and, by driving shocks, increase the gas density in some regions. Both the shocks themselves, and the density changes they produce, can influence the line emission from the gas in the galaxy.

Powerful radio galaxies such as *Cygnus A* possess energetic jets that inflate huge cocoons which are overpressured with respect to the surrounding IGM and expand supersonically into it. More generally, however, the jet-inflated cocoons behave as buoyant bubbles and dissipate energy rather gently. Whilst most of the power in the jets is in the form of bulk kinetic energy, the inner regions, which are moving relativistically in some powerful radio sources, can produce a Doppler boosted photon "beam" with a rather small opening angle (see Fig. 3). For observers within the radiation pattern of such a beam, the source will appear as a *blazar* or a *BL Lac* object. The SED of such a beam will differ from that of the more isotropic AGN radiation field and, although it may carry only a small fraction of the bolometric luminosity of the AGN, does illuminate regions along the radio axis with a high photon density. If the beams illuminate gas clouds, the ionizing and other effects on them can be substantial. The high photon densities can be maintained in the beam to very large radial distances and so can, in principle, be responsible for ionizing gas even well beyond the extent of the radio lobes.

The reaction of an AGN, in the form of its radiative and kinetic output, to the rate and form of accretion onto the SMBH constitutes a potential "feedback loop" that can limit growth and, perhaps, establish the apparent observed relationship between the black hole mass and that of the host galaxy bulge (see Chap. 8).

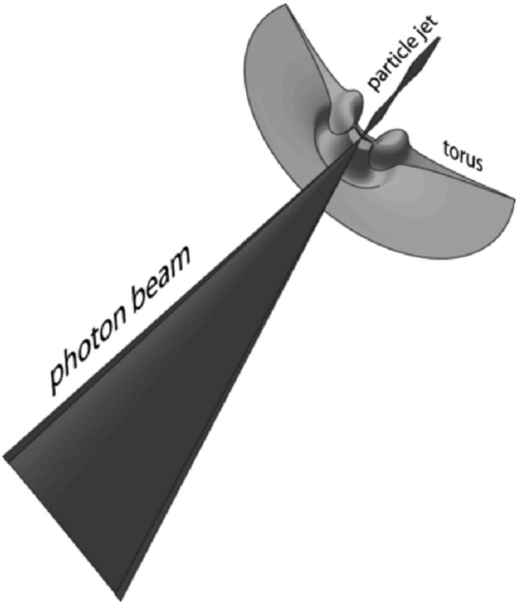

Fig. 3. A sketch showing the relation of the particle jet and associated photon beam to the nuclear torus in a radio-loud AGN. The jet and the beam are both, of course, bi-directional and have only been separated in the diagram for clarity

3 Observational Techniques and Examples

Most observational work on the AGN – host galaxy connection has been carried out on the type 2 sources – Seyfert 2 rather than Seyfert 1 and radio galaxies rather than quasars – for the simple reason that a direct view of the AGN in type 1 objects results in an excessive "glare" that renders it difficult to see faint extra-nuclear structures. The presence of thick obscuration close to the nucleus and in our line-of-sight in type 2 sources results in a very effective "natural coronograph".

The first indications of the spatial scale of the interaction between an AGN and its host galaxy was the discovery of the EELR which, especially in the radio-loud sources, could extend tens, sometimes hundreds, of kiloparsecs from the nucleus. This realisation came as a result of instrumental developments in the early 1970's. While there had been some work with photographic long-slit spectrographs, the introduction of panoramic electronic detectors such as the Image Photon Counting System (IPCS) and, later, CCDs, made it possible to do quantitative, spatially resolved studies of emission lines in the optical spectrum. Before that, most AGN spectroscopy had been carried out with 1-dimensional spectral scanners that gave little – though not zero – information about the extent of the emission line regions. The general picture that emerged was that the ISM of the host galaxy was being photoionized by the hard AGN radiation field [12]. The careful correlation of high resolution radio continuum maps made with interferometric arrays, notably in Cambridge, Westerbork and the Very Large Array (VLA) in New Mexico, suggested, however, that the jets were influencing the distribution and possibly also the excitation of the gas. It was learned that it was not only the radio-loud AGN that had bi-symmetric jets. The much less radio luminous Seyfert galaxies also contained radio nuclei, often with extended structures exhibiting axial symmetry.

Work on the spatial correlations continued with attempts to establish physical connections between the line emitting gas (at $T_e \sim 10^4$K) and the relativistic plasma responsible for the non-thermal radio emission. While there appeared to be an approximate pressure balance between the two components, the interactions were likely to be far from an equilibrium situation. Indeed, later work has shown that the optically emitting gas can have a significantly higher pressure than the surroundings. The effect of shocks, driven by the jets, was not obvious. While shocks are clearly capable of ionizing gas – either collisionally or, in faster cases, radiatively – diagnosing the relative importance of these processes and those produced by the powerful AGN radiation field is complicated and often ambiguous.

Work on the relationship between the radio axis and the symmetry axis (if it has one) of the host galaxy did not produce immediately explicable results. The spiral galaxy hosts of Seyfert galaxies have rotation axes that show no apparent correlation with their AGN axis for the late-type hosts, but a tendency for alignments in the earlier-type hosts. Such work, however, is difficult since jets emerging within the plane of the disc will produce different effects to those emerging close to the pole. The clearest association exists between the cone axis and the elongation of the nuclear radio source – often double or triple – in Seyfert galaxies. In the elliptical galaxy hosts of the radio-loud sources, the situation is less clear since identifying a relevant symmetry axis of a possibly triaxial ellipsoid is both observationally difficult and interpretationally ambiguous. Where a small scale nuclear gas/dust disc can be seen, however, the jet appears to be perpendicular to it.

When observations of radio galaxies at redshifts greater than about 0.6 became routine in the mid 1980's, it was realised that the optical images – corresponding to restframe blue or ultraviolet light – were elongated and shared an axis, although not necessarily a detailed spatial correspondence, with the double-lobed radio source. This became known as the "alignment effect" and it spawned a great variety of ideas for its physical interpretation [1, 6]. Observations of these sources in the near-infrared (NIR) – corresponding more closely to the restframe optical – showed a weaker effect, indicating that it was the blue/UV light that is aligned. Spatially resolved spectroscopy of these distant objects is very demanding with 4m-class telescopes and it was not possible to decide between explanations based on some reprocessing of AGN power transmitted along the axis or the possibility of star formation

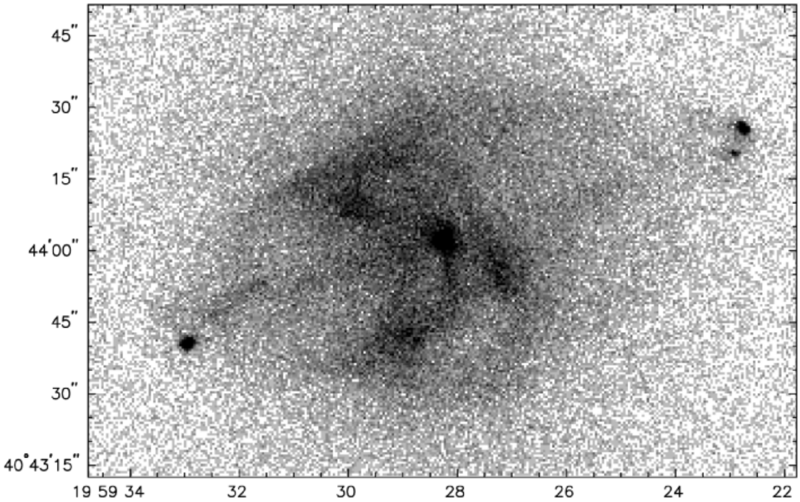

Fig. 4. An X-ray image of *Cygnus A* with the CHANDRA ACIS instrument. (Courtesy A. Wilson et al., NASA/UMD)

triggered by the interaction between the jet and the ISM. At least for the higher redshift sources, inverse Compton scattering of microwave background photons by the relativistic electrons in the radio jets was a possible contributor. Using 4m telescopes, the presence of a linear polarization signal in some sources with an $\boldsymbol{E}$-vector perpendicular to the radio axis suggested the presence of scattering of light from the AGN but it was not until the availability of the much larger 8–10m telescopes that the details of the scattering process became clear and it was shown that this was the dominant, although not necessarily sole, contributor to the alignment effect in the powerful radio sources [15].

The most recent contribution to the study has come from high spatial resolution X-ray maps of the nucleus, the jets and radio hot-spots and the surrounding cluster environment. Detailed studies of *Cygnus A* with Chandra show what can be achieved (Figs. 4, 5) [17]. The principal components are:

- thermal X-ray emission from the intracluster gas,
- a limb-brightened cavity containing the relativistic gas fed by the jets,
- synchrotron self-Compton emission from the radio hot-spots,
- X-ray emission from the jets themselves,
- soft emission from an extended nuclear source and
- hard emission from an unresolved nuclear source.

The extended nuclear source, on a kiloparsec scale, appears to result from the Thompson scattering of nuclear X-ray photons in a very highly (photo)ionized gas extending along the radio axis. The Thompson optical depth, while it can explain the X-rays, is too small to produce the extended

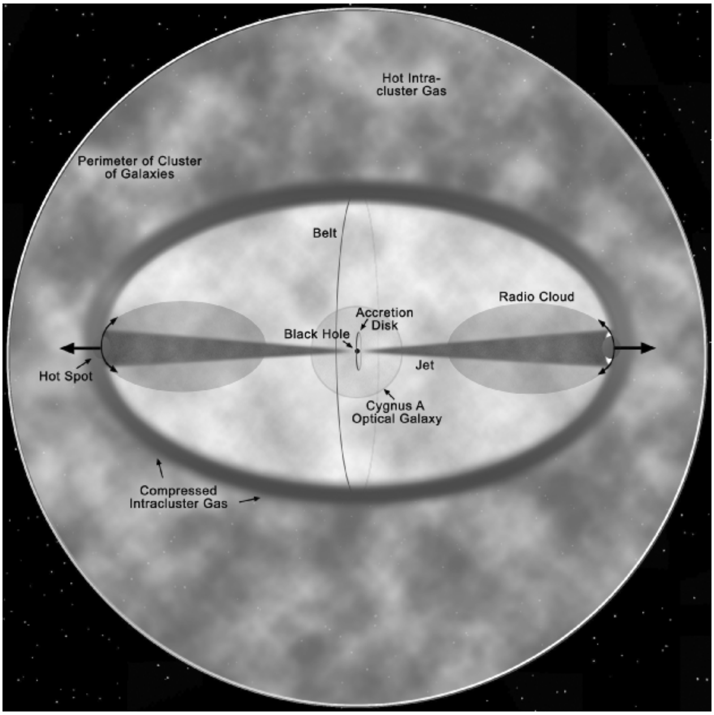

Fig. 5. An X-ray cartoon of *Cygnus A* that illustrates the interactions between the jets and the surrounding intracluster medium. (Courtesy CHANDRA X-ray Observatory, NASA/CXC/SAO)

optical polarization which must, therefore result from dust scattering. The spatial distribution of this soft X-ray component appears to match very closely the emission from the extended coronal lines of [Fe X], [Fe XI], [Ar XI] and [S XII]. This is strong evidence in favour of the photoionization model for the production of the coronal line spectra in AGN in gas with a very high ionization parameter (U = ionizing photon density/particle density).

3.1 Local AGN

Work on the local Seyfert galaxies has concentrated on the nature of the ionization cones (for an example see Fig. 6), their association with the nuclear radio source and their orientation with respect to the host galaxy. Two- and three-dimensional spectroscopy has allowed the study of the ionization state, the kinematic properties and, to some extent, the chemical abundances of the ionized gas. The cones can be seen even in projection against the disc

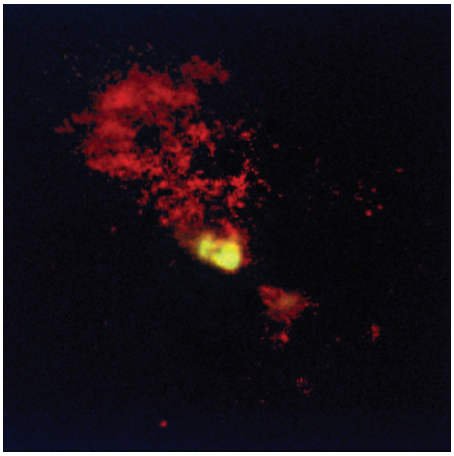

Fig. 6. An early (WFPC) image taken with HST of the ionization cone in the Seyfert galaxy NGC 5728 (Courtesy A.S. Wilson, STScI/NASA). This was taken with the PC before correction of Hubble's spherical aberration. It is in the light of Hα and [O III] and the structure has an overall extent of some 1.8 kpc. The apex of each cone indicates, presumably, the site of the obscured AGN

by virtue of their high ionization state and strong line emission (Fig. 7). By taking account of projection effects, either statistically or by direct inference in individual cases, it is possible to infer the true cone opening angle. This, in turn, can be compared with the value implied by the type 1 to type 2 object ratio in samples selected on the basis of an assumed isotropic property.

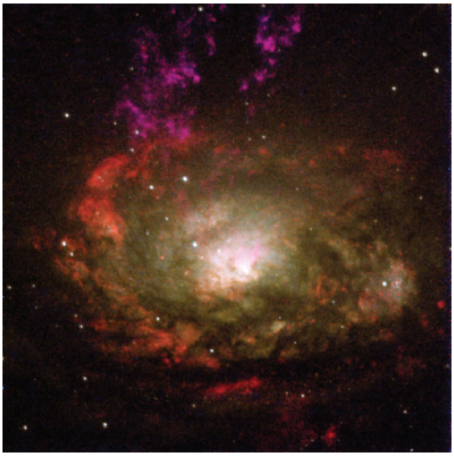

Fig. 7. An HST WFPC 2 image of the Circinus galaxy in two narrow (Hα and [O III]) and two broad (NIR and green) filters. The nearside ionization cone can be seen projected against the disk by virtue of its strong hydrogen and oxygen line emission. (Courtesy A.S. Wilson et al., NASA)

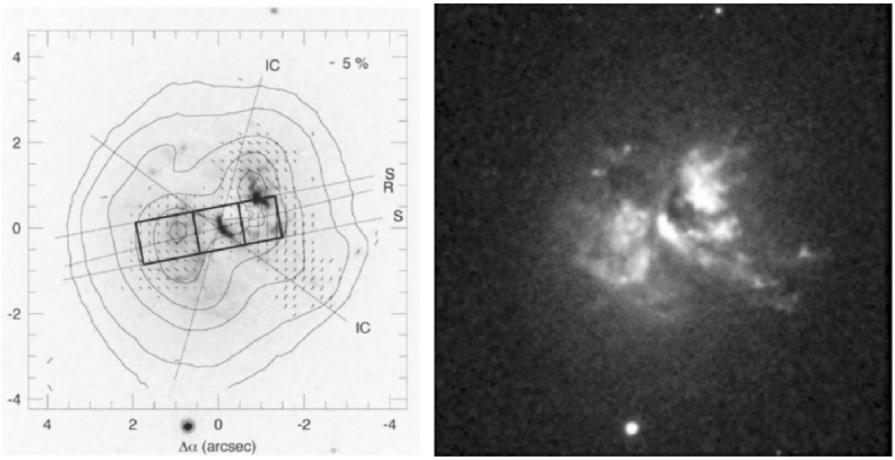

Fig. 8. The powerful radio galaxy *Cygnus A*. The image on the right is a combination of B, V, R and I broad band and [O III] and Hα narrow band exposures from the HST WFPC 2 archive with the addition of a B-band exposure with Keck. The B-band contour map on the left is from the Keck LRISp polarimeter showing the measured $\boldsymbol{E}$-vectors and the location of the apertures used for spectropolarimetry (a negative version of the colour image is shown as an underlay). The position of the radio axis (R) and the outline of the ionization cone (IC) are shown. The equatorial ring on the southeast side consists of starlight and is unpolarized

Cygnus A is also a good relatively low redshift ($z = 0.056$) example of the range of phenomena that can be studied in a powerful radio galaxy. This famous object did, however, strongly resist attempts to fit it comfortably into the emerging orientation-based unification paradigm. There were several unsuccessful attempts to measure the linear optical polarization that might have been expected in a type 2 object harbouring a powerful quasar. It took the application of the Keck polarimeter and a careful subtraction of the elliptical galaxy starlight to reveal the dust scattering (Fig. 8 – left) [8]. The task was rendered more difficult by the relatively high Galactic extinction in the direction of Cygnus.

Images obtained from the ground had never had quite the spatial resolution necessary to make sense of the rather complex optical source structure. It was the assembly of the Hubble images into a colour composite that showed clearly the ionization cones and the ~4.5 kpc diameter ring of blue stars. The combination of large telescope spectropolarimetry and HST imaging allowed, for the first time in this type of object, the direct identification of spatial features with spectral components having a particular polarization pattern and colour (Fig. 9).

The interaction between a powerful radio jet and a substantial obstacle, such as a dwarf satellite galaxy, is a rather rare phenomenon (fortunately for people who live on planets in dwarf galaxies!) that can produce some spec-

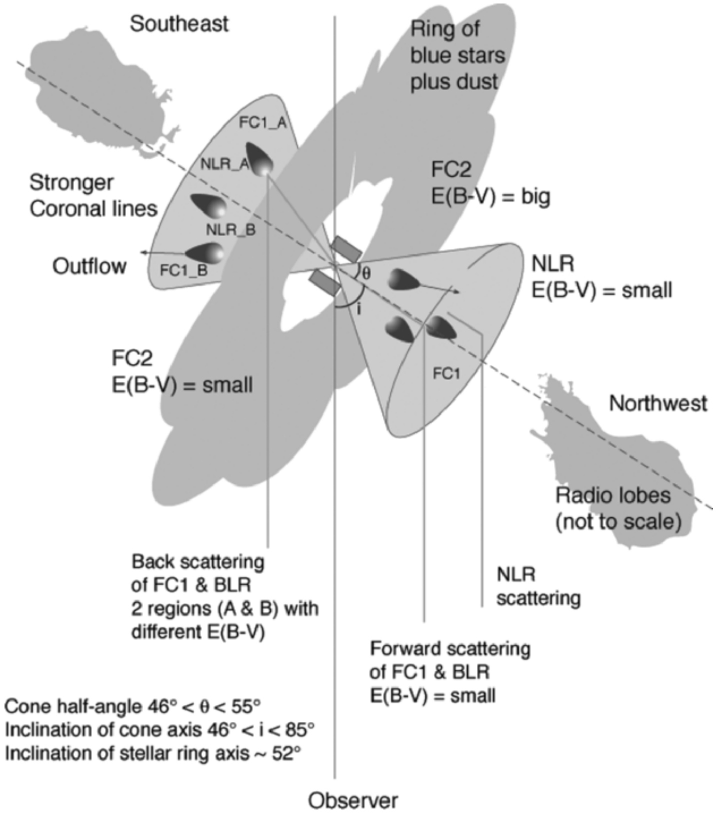

Fig. 9. A cartoon of *Cygnus A* which illustrates the various components, their positions and orientations. The letters “FC” followd by a number represent various so-called “featureless continua” that had been seen in the spectropolarimetry as polarized or unpolarized sources

tacular results. One of the best examples of this occurring is in the relatively nearby radio galaxy PKS B2152-699 ($z = 0.0282$). Figure 10 shows an optical (HST) and radio (ATCA) image of the galaxy with a “shredded” gaseous cloud (HIC) some 8 kpc out along the radio axis: actually along the extension of the axis defined by the VLBI jet on a scale of a few tens of mas [4]. Optical spectroscopy has shown this cloud to have a very highly excited emission line spectrum, including a strong [Fe X] coronal line. It also emits (or scatters) a very blue, polarized optical/near-UV continuum. In addition to a radio component slightly offset from the optical cloud, there is an associated X-ray source seen with both ROSAT and Chandra. Detailed, multiwavelength studies of sources such as this give us an excellent opportunity to measure the properties and makeup of the particle jet and its associated photon beam.

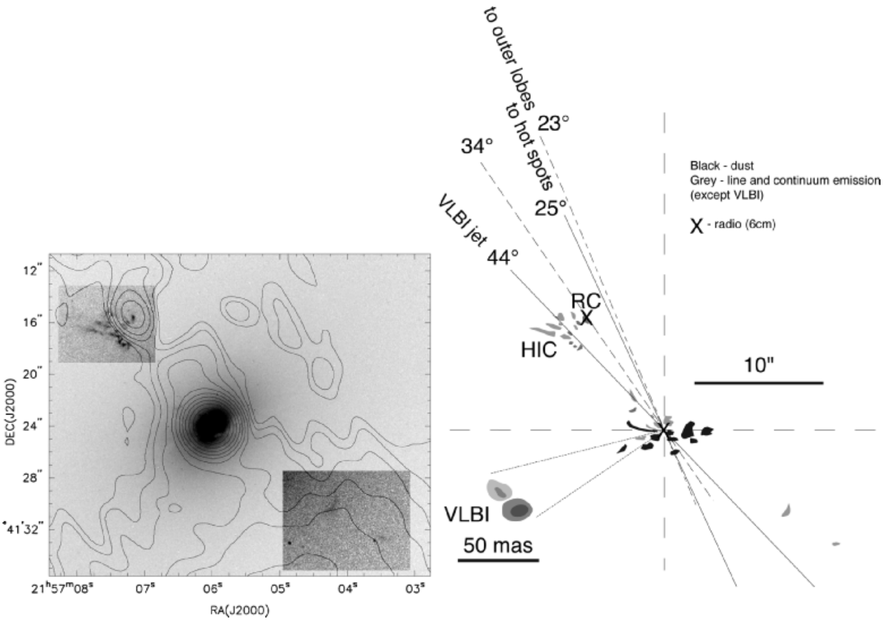

Fig. 10. An optical image (greyscale, HST WFPC 2) and a radio map (contours, ATCA) of the radio galaxy PKS B2152-699 showing the bright jet-cloud interaction (HIC) to the NE and a fainter counter feature to the SW. The associated cartoon illustrates the location and the orientation of the various observed components with "RC" marking the offset position of the 6cm radio source associated with the HIC

3.2 Distant AGN

The space density of powerful AGN was highest during the so-called "Quasar Epoch" at redshifts around 2–3, corresponding to lookback times of around 10 billion years. This was the time when the most massive galaxies were forming in environments that were to become the rich clusters we see today. The presence of luminous active nuclei in these galaxies has a significant effect on their surroundings and, fortunately for us, "illuminates" structures and processes that we should otherwise find it very difficult to study. The ionization cones represent the ISM of the host galaxy ionized by the hard AGN spectrum and, with a suitable understanding of the photoionization physics, allows the study of the chemical composition and the kinematic state of the gas. A *disadvantage* of the presence of the AGN is, however, the extent to which its effects can dominate the other source of luminosity, i.e., the stellar population. Efforts to look at the stellar evolutionary properties of those galaxies that contain a currently active AGN are seriously hampered in the restframe ultraviolet and optical wavebands where the young populations should be most accessible. Studies of more evolved populations in the NIR are, in principle, less disturbed but do require the ability to observe at wavelengths inaccessible from the ground.

The most detailed studies of AGN host galaxies in the Quasar Epoch have been those of the powerful radio galaxies since these are our only currently complete samples of high redshift type 2 objects. Radio-quiet type 2 sources are beginning to be found now as a result of deep surveys in (hard-ish) X-ray bands and, most recently, in the mid-infrared. While we recognise that the radio-loud sources do not represent the entire AGN host population, we do believe that they mark the most massive galaxies in the Universe.

At a redshift of around 2.5, the restframe ultraviolet spectrum (from Lyman-α to C III] λ1909) and beyond is accessible to optical instruments while the NIR instruments nicely accommodate, within the J, H and K-bands, the important optical emission lines from [O II] λ3727 to Hα, [N II] and [S II] in the red. An example of such a source, observed in the optical with the Keck II instrument LRIS and in the NIR with ISAAC on the VLT, is shown in Fig. 11. While this object is extreme in the sense that it appears to be completely dominated by reflected light from the obscured quasar, it does caution us about the difficulty of accessing the underlying stellar population.

The detailed nature of the scattering process that we see at work in an object like this and producing a high redshift example of the alignment effect has been the subject of much discussion. The fact that the scattering appears to be almost perfectly grey, with the reflected continuum and broad emission lines being so well represented by a bright quasar reduced in intensity by around 3 magnitudes, suggests Thompson scattering as the responsible mechanism. It can be shown, however, that a dusty medium consisting of optically thick clumps embedded in a thin intraclump medium, makes an efficient grey scatterer without broadening spectral features to the extent that would occur after scattering by hot electrons [14, 15, 18]. The physics of this process is rather simple and worth understanding because of its general applicability in complex geometries.

Assume that the AGN is surrounded by such a clumpy, dusty medium but completely obscured from our direct view. An emergent photon is scattered into our line of sight with a probability determined by the scattering optical depth along its original path. On its route to the observer, the photon suffers the possibility of being either scattered out of our line of sight or of being absorbed by a dust grain. The scattered source is consequently attenuated by an extinction optical depth, giving an emergent luminosity

$$L_{obs} \propto L_{AGN}(\lambda) \times \tau_{scat}(\lambda) \times e^{-\tau_{ext}(\lambda)}$$

where $L_{AGN}(\lambda)$ is the quasar luminosity. This function is a maximum when $\tau_{scat}(\lambda) \simeq \tau_{ext} \simeq 1$. A geometrically complex scattering medium acts, therefore as a self-regulating "mirror" that reflects a wavelength-independent fraction of the source, i.e., as a grey scatterer. Galactic dust has extinction and scattering cross sections that are proportional to one another throughout the

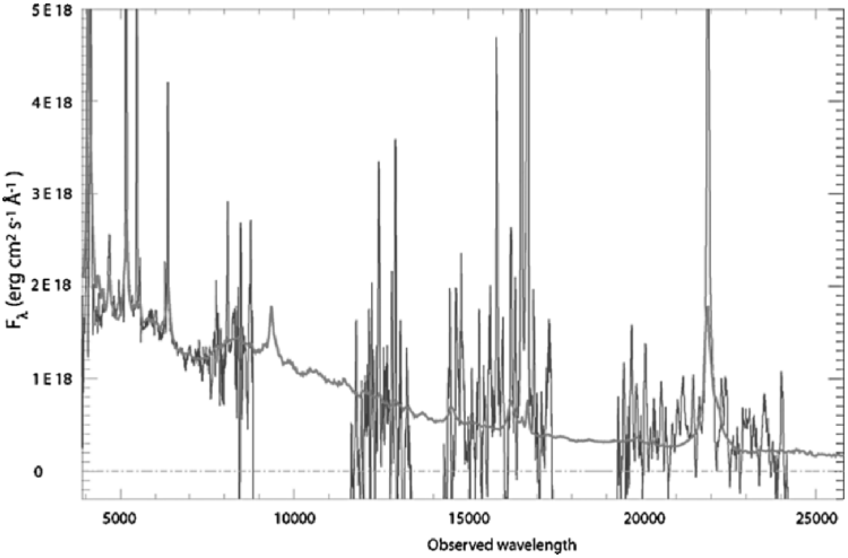

Fig. 11. The ultraviolet–optical SED of the radio galaxy TXS 0211-122, $z = 2.340$ measured with the Keck (optical) and VLT (NIR) telescopes. The red continuous line shows a template radio-loud quasar spectrum scattered by a clumpy dust distribution within the illuminated "ionization cones" in the host galaxy. In this powerful source, the entire UV-optical continuum can be ascribed to the scattered AGN light which completely dominates the host galaxy starlight. The narrow emission lines – which are relatively much stronger than in the scattered quasar light – originate from the photoionized ISM of the host. The shortest wavelength emission line is Lyman-α and the redmost line is Hα. (Courtesy J. Vernet et al.)

UV-optical spectrum except close to the resonance at 2200Å. This has the result that, while the overall scattered spectrum is grey, a 2200Å feature is imprinted on the emergent spectrum. This can be seen as the bump around an observed wavelength of 8000Å in the observed and modelled spectra in Fig. 11.

Spectropolarimetry of the powerful $z \sim 2.5$ radio galaxies shows spatially-averaged fractional polarizations from zero (≤ 2 or 3%) up to 25% in the continuum longward of Lyman-α (see Fig. 12 for an example of a high polarization source). This suggests the presence of an unpolarized continuum source in some of the objects in addition to the dust-scattered quasar. It is natural to assume that this is starlight – although it is very difficult to demonstrate this conclusively by searching for UV absorption lines from hot star photospheres. Some photospheric lines have been reported but these are not always seen in the spectra of high redshift radio source hosts even with long exposures on the biggest telescopes [2].

Evidence for high rates of star formation in these massive galaxies probably comes more directly from observations of the heat radiated by the large

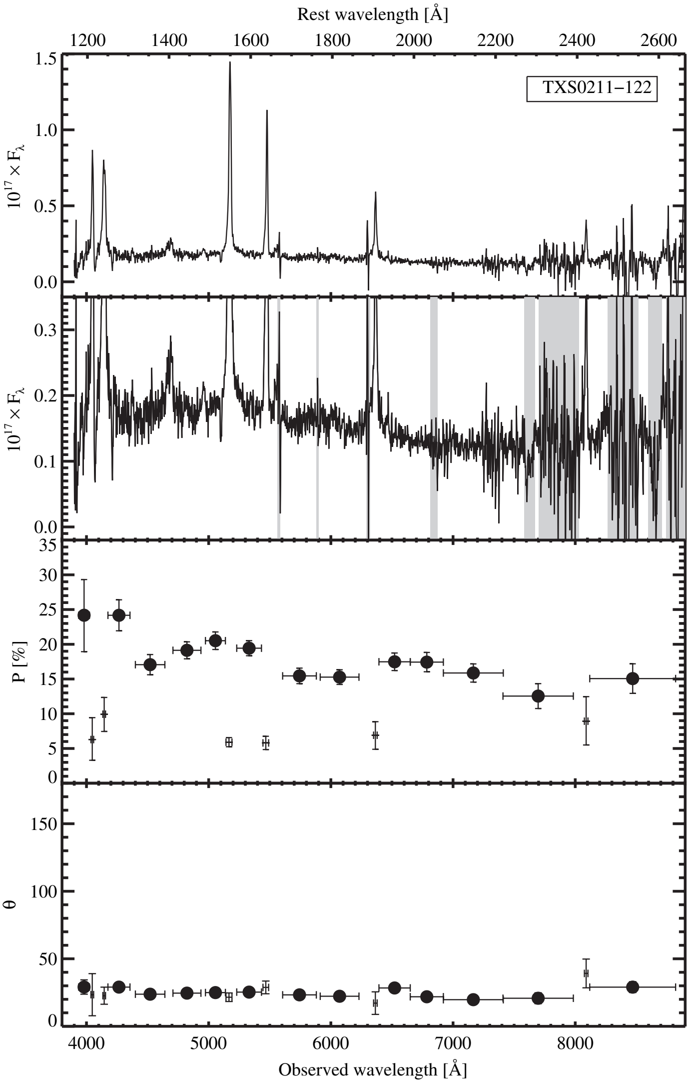

Fig. 12. Optical spectropolarimetry of TXS 0211-122 obtained with the Keck LRISp instrument. The top two panels show differently scaled versions of the total flux. The third panel shows the fractional polarization in both line and continuum wavelength bins. The final (bottom) panel shows the position angle of the $\boldsymbol{E}$-vector in the same wavelength bins: for comparison, the PA of the major axis of the restframe UV image of this source is 122° which is geometrically consistent with the scattering origin of the alignment effect. The degree of polarization is also consistent with dust (single-)scattering models within ionization cones where the integration over scattering angle dilutes the net polarization. Note the low fractional polarization in the narrow bands centred on the strong, narrow emission lines. (Courtesy J. Vernet et al.)

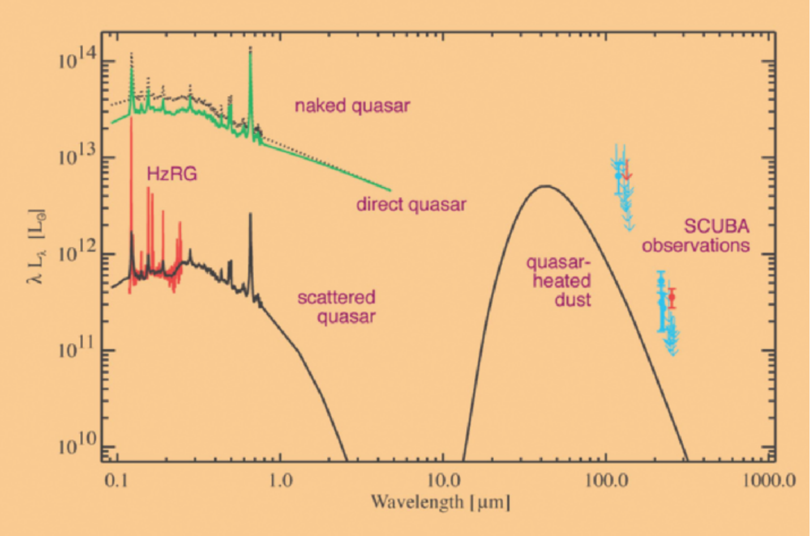

Fig. 13. A diagram representing the scattering, absorption and thermal emission balance for the dust content of the ionization cones (*only*) in an type 2 active galaxy. Starting with the naked quasar SED in the upper left corner, the lower green line represents the escaping flux after some modest dust extinction when seen close to the AGN axis (as a type 1 object). The lower left black line represents the quasar light scattered out of the ionization cone, by the same dust that produces the extinction for a type 1 observer, towards an observer who would see this as a type 2 object. The red spectrum is an observation of one of the redshift 2.5 radio galaxies. The FIR SED (black curve on the right) is the thermal emission from the dust that is causing the extinction and scattering within the cones. The red sub-mm points are observations of the source represented by the red spectrum and the blue observations represent other high-z radio galaxies. This analysis suggests that the bulk of the cool dust thermal emission from AGN like this is coming from dust heated by star formation rather than being reprocessed quasar power. (Courtesy J. Vernet et al.)

scale dust distribution. Figure 13 suggests that the bulk of FIR/sub-mm thermal emission observed from these massive objects is unlikely to come from the AGN-heated dust within the ionization cones. It is more likely to come, instead, from dust-obscured star formation which must be occurring in these objects at this epoch. The hotter AGN-heated dust in the nuclear regions (torus) will radiate at much shorter wavelengths in the MIR.

Another observational characteristic of these sources is the presence of huge gaseous halos seen in the Lyman-α line as emission that is sometimes overlayed with spatially extended narrow absorption. These halos can be seen on scales of 100 kpc or more and have luminosities of 10^{43-44} erg s^{-1} [16], [11]. Figure 14 shows long-slit spectra of TXS 0211-122 aligned with its HST image and VLA radio map. Other UV lines are spatially extended but usually

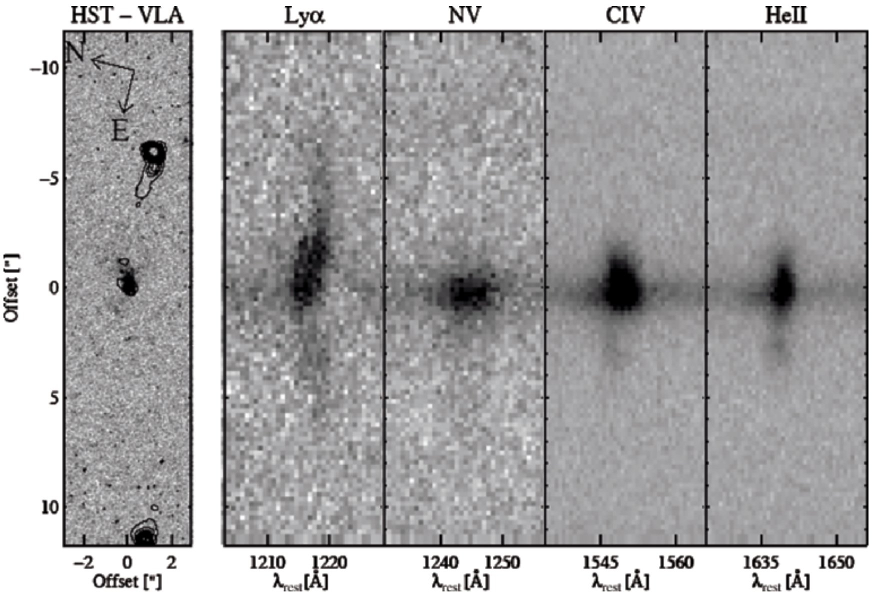

Fig. 14. The emission line halo of TXS 0211-122 from Keck long-slit spectroscopy. The halo can be seen in all the strong UV emission lines but is most extended in Lyman-α. the spectra are plotted aligned and on the same spatial scale as the restframe UV and radio images. Some of these halos can be detected even beyond the extent of the radio lobes. (Courtesy M. Villar-Martín et al.)

not to the extent of Lyman-α. Kinematically, these halos appear to divide into two zones, one that is highly disturbed and probably associated with the passage of the jet and another that is more quiescent and may represent the remnants of the gaseous reservoir from which the galaxy is forming.

The line-emitting photoionized gas in the AGN host galaxies can provide a measure of the chemical composition of the ISM and so provide an important indicator of the galaxy's evolutionary history. Early studies of the EELR in low redshift, radio-loud AGN hosts showed that the extended gas generally has approximately Solar composition with remarkably little variation from object to object. Such apparent uniformity suggests that the current state of the extended ISM in these massive galaxies is the result of a steady integration of star formation and gas mixing taking place over a long period of time. The ability, now provided by the large groundbased telescopes, to extend these studies to earlier epochs, offers the exciting opportunity to witness the build up of the metallicity during the assembly of the massive galaxies.

Since the optical spectrum gives us access to only the restframe UV of high redshift (within the "quasar epoch") sources, it is necessary either to develop metallicity measures using these UV lines or to observe in the NIR where the "traditional" techniques developed for the optical spectrum can be employed. Both of these approaches are starting to be employed now but they still provide significant observational challenges. The UV spectrum is dominated by resonance lines of hydrogen, nitrogen, carbon and magnesium that are easy to measure but difficult to interpret since they are subject to optical depth effects and absorption by dust. There are reasonably prominent intercombination (semi-forbidden) lines of C, N, O and Si but these, while being easier to interpret, do require long integrations with the largest telescopes to measure with any precision. In the restframe optical band, we are on more familiar territory with the traditional favourite forbidden spectra of N, O, Ne and S referenced to the Balmer series of H and the He II line at 4686Å. These lines come, however, at the cost of the bright NIR sky and the atmospheric absorption between the J, H and K-bands.

Analyses are carried out with the aid of photoionization models and some knowledge of the shape of the ionizing SED from the AGN. The results are generally presented and interpreted by plotting the line-ratio diagnostic diagrams that have become familiar in the field. Models are run with sequences of ionization parameter U, gas metallicity and ionizing SED shape (e.g., power law slope) and compared with the observed points in different diagrams that have different dependencies on the input parameters. Ratios are chosen to be as independent as possible of the effects such as dust reddening. As usual, difficulties arise in the detail: the effects of dust on resonance line transfer means that gas clouds look very different when seen from their illuminated and shadowed sides. Also, gas clouds that are transparent to the ionizing continua ("matter bounded") and opaque ("ionization bounded") can produce different spectra for a given set of basic input parameters. However, there are clear trends apparent in the observational data that currently challenge our interpretational ability. An example is the C IV, N V, He II diagram that has been used for a decade or more to interpret the BLR of quasars. This is shown for the NLR of radio galaxies in Fig. 15. While diagrams such as these suggest a broader range in gas metallicity than seen within quasar hosts at the current epoch, there is still a need for confirmation using the forbidden line ratio diagrams that require NIR spectroscopy at the higher redshifts.

4 What Next?

The deep surveys currently being carried out from the X-ray to the radio, including the important space-infrared capabilities, offer us the first opportunity for a rather complete census of both type 1 and type 2 AGN extending to the highest redshifts. These sources will act as "markers" for elucidating the relationship between star and SMBH formation in galaxies and the nature of the feedback process responsible for the close observed relationship between them. The host galaxies serve as "bolometers" that allow luminosity estimates even for the obscured type 2 sources, enabling us to complete

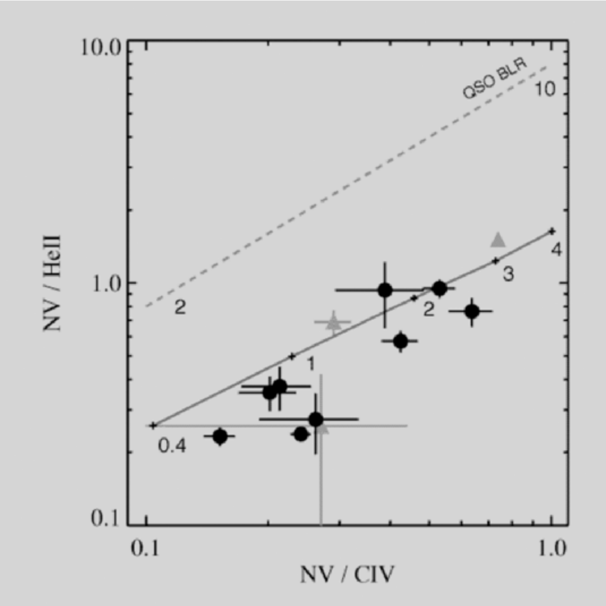

Fig. 15. An ultraviolet line ratio diagram plotted for the NLR radio galaxies at $2 \leq z \leq 3$. The red plain line shows a sequence of photoionization models as a function of metallicity in Solar units. The black (*dots*) and blue (*triangles*) points are (similar) sources from different observed samples. The dashed green line shows the locus of points established for quasar BLR – which of course have a much higher particle density – again labelled with metallicity in Solar units. (Courtesy J. Vernet et al.)

the inventory of gravitational and nucleosynthetic photon production in the Universe.

Further detailed studies of individual objects will yield more information about the chemical evolution of massive galaxies at early epochs and, by discrimination between reprocessed AGN light and starlight, map their stellar evolutionary histories. While observations of AGN hosts at high redshift are observationally challenging, the AGN themselves are bright and their effects on the surrounding medium can be reasonably prominent and accessible to a broad range of observational techniques.

5 Further Reading

The ESO conference proceedings: "Extranuclear activity in galaxies", Garching, Germany, 1989; E.J.A. Meurs and R.A.E. Fosbury (eds.) will give a general introduction to the field.

Acknowledgments

While an article such as this, in not attempting to comprehensively review the field, represents a personal view, I should like to single out and thank several colleagues who have worked with me over the years in this area of astronomy. The work on radio galaxy polarimetry was started with Sperello di Serego Alighieri and Clive Tadhunter and was given a huge boost by Marshall Cohen who used significant amounts of his Keck observing time to help us study sources at $z \sim 2.5$. Luc Binette, Andy Robinson, Montse Villar-Martín and Joël Vernet have worked extensively on data acquisition, interpretation and modelling. Finally, I thank the organisers of the meeting in Santiago for giving me both the opportunity of presenting this work and also of assimilating the other overviews.

References

1. Chambers, K.C., Miley, G.K., & van Breugel, W. 1987, Nature, 329, 604
2. Chambers, K.C., Miley, G.K., & van Breugel, W.J.M. 1990, ApJ, 363, 21
3. Ferrarese, L., & Merritt, D. 2000, ApJ, 539, L9
4. Fosbury, R.A.E., Morganti, R., Wilson, W., Ekers, R.D., di Serego Alighieri, S., & Tadhunter, C.N. 1998, MNRAS, 296, 701
5. Krolik, J.H. 1999, Active galactic nuclei : from the central black hole to the galactic environment/Julian H. Krolik. Princeton, N.J.: Princeton University Press, c1999
6. McCarthy, P.J., van Breugel, W., Spinrad, H., & Djorgovski, S. 1987, ApJ, 321, L29
7. McLure, R.J., & Dunlop, J.S. 2002, MNRAS, 331, 795
8. Ogle, P.M., Cohen, M.H., Miller, J.S., Tran, H.D., Fosbury, R.A.E., & Goodrich, R.W. 1997, ApJ, 482, L37
9. Osterbrock, D.E. 1989, Research supported by the University of California, John Simon Guggenheim Memorial Foundation, University of Minnesota, et al. Mill Valley, CA, University Science Books, 1989, 422 p.
10. Pelat, D., Alloin, D., & Fosbury, R.A.E. 1981, MNRAS, 195, 787
11. Reuland, M. et al. 2003, ApJ, 592, 755
12. Robinson, A., Binette, L., Fosbury, R.A.E., & Tadhunter, C.N. 1987, MNRAS, 227, 97
13. Tadhunter, C., & Tsvetanov, Z. 1989, Nature, 341, 422
14. Városi, F., & Dwek, E. 1999, ApJ, 523, 265
15. Vernet, J., Fosbury, R.A.E., Villar-Martín, M., Cohen, M.H., Cimatti, A., di Serego Alighieri, S., & Goodrich, R.W. 2001, A&A, 370, 407
16. Villar-Martín, M., Vernet, J., di Serego Alighieri, S., Fosbury, R., Humphrey, A., Pentericci, L., & Cohen, M. 2003, New Astronomy Review, 47, 291
17. Wilson, A.S., Young, A.J., & Shopbell, P.L. 2000, ApJ, 544, L27
18. Witt, A.N., & Gordon, K.D. 1996, ApJ, 463, 681
19. Zurita Heras, J.A., Türler, M., & Courvoisier, T.J.-L. 2003, A&A, 411, 71

The Fueling and Evolution of AGN: Internal and External Triggers

S. Jogee

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore,
MD 21218, USA
`jogee@stsci.edu`

1 Introduction

The quest for a coherent picture of nuclear activity has witnessed giant leaps in the last decades. Four decades ago, the idea was put forward that accretion of matter onto a massive compact object or a supermassive black hole (SMBH) of mass $>10^6$ $M_\odot$ could power very luminous active galactic nuclei (AGN), in particular, quasi-stellar objects (QSOs) (Lynden-Bell 1969; Soltan 1982; Rees 1984). In the last decade, dynamical evidence increasingly suggests that SMBH pervade the centers of most massive galaxies (Sect. 2 and references therein). The challenge has now shifted towards probing the fueling and evolution of AGN over a wide range of cosmic lookback times, and elucidating how they relate to their host galaxies in both the local and cosmological context.

In this review, I will focus on the fueling and evolution of AGN under the influence of internal and external triggers. In the nature versus nurture paradigm, I use the term internal triggers to refer to intrinsic properties of host galaxies (e.g., morphological or Hubble type, color, and non-axisymmetric features such as large-scale bars and nuclear bars) while external triggers refer to factors such as environment and interactions. The distinction is an over-simplification as many of the so called intrinsic properties of galaxies can be induced or dissolved under the influence of external triggers. Connections will be explored between the nuclear and larger-scale properties of AGN, both locally and at intermediate redshifts. One of the driving objectives is to understand why not all relatively massive galaxies show signs of AGN activity (via high-excitation optical lines or X-ray emission) despite mounting dynamical evidence that they harbor SMBHs. The most daunting challenge in fueling AGN is arguably the angular momentum problem (Subsect. 3.2). Even matter located at a radius of a few hundred pc must lose more than 99.99% of its specific angular momentum before it is fit for consumption by a BH.

The sequence of this review is as follows. Section 2 briefly addresses BH demographics and the BH-bulge-halo correlations. Section 3 sets the stage for the rest of this paper by providing an overview of central issues in the

S. Jogee: *The Fueling and Evolution of AGN: Internal and External Triggers*, Lect. Notes Phys. **693**, 143–183 (2006)
DOI 10.1007/3-540-34621-X_6

fueling of AGN and circumnuclear starbursts. In particular, I review mass accretion rates, angular momentum requirements, the effectiveness of different fueling mechanisms, and the growth and mass density of BHs at different epochs. These central issues in Sect. 3 are attacked in more detail in Sects. 4–9 which describe different fueling mechanisms including mergers and interactions (Sect. 5), large-scale bars (Sect. 6), nuclear bars (Sect. 7), nuclear spirals (Sect. 8), and processes relevant on hundred pc to sub-pc scales (Sect. 9). I conclude with a summary and future perspectives in Sect. 10. Complementary reviews on mass transfer and central activity in galaxies include those by Shlosman (2003), Combes (2003), Knapen (2004), and Martini (2004).

2 BH Demographics and BH-Bulge-Halo Correlations

2.1 Measurement of BH Masses

The term SMBHs refers to BHs having masses $M_{\rm bh} > 10^6$ M$_\odot$, in contrast to intermediate mass BHs (IMBHs) with $M_{\rm bh} \sim 10^2$–10^6 M$_\odot$, and stellar mass BHs. Properties of SMBHs are generally studied through accretion signatures of BHs or their gravitational influence. The strongest dynamical evidence for SMBHs are in our Galaxy and in NGC 4258. In these systems, the large central densities inferred within a small resolved radius can be accounted for by a SMBH, but not by other possibilities such as collections of compact objects, star clusters, or exotic particles. In our Galaxy, proper motion measurements set stringent constraints on the central potential (Schödel et al. 2003; Ghez et al. 2003; Genzel et al. 2000), yielding $M_{\rm bh} \sim 3$–4×10^6 M$_\odot$. In NGC 5248, VLBA maser observations reveal Keplerian motions implying $M_{\rm bh} \sim 3.9 \times 10^7$ M$_\odot$ (Miyoshi et al. 95).

In the last decade, high resolution gas and stellar dynamical measurements from ground-based (e.g., Kormendy & Richstone 1995) and *HST* observations (e.g., Harms et al. 1994; Ferrarese et al. 1996; van der Marel & van den Bosch 1998; Ferrarese & Ford 99; Gebhardt et al. 2000) have provided compelling evidence that several tens of galaxies host massive central dark objects (CDOs) which are likely to be SMBHs. The more reliable dynamical measurements tend to be from observations which resolve the radius of influence ($R_{\rm g-bh}$) within which the gravitational force of the BH exceeds that of nearby stars with velocity dispersion σ, namely,

$$R_{\rm g-bh} = \frac{GM_{\rm bh}}{\sigma^2} = 11.2\ {\rm pc} \left(\frac{M_{\rm bh}}{10^8\ {\rm M}_\odot}\right) \left(\frac{\sigma}{200\ {\rm km\ s}^{-1}}\right)^{-2} \qquad (1)$$

However, the scales probed by these measurements are still several 10^5–10^6 times the Schwarzschild radius ($R_{\rm s-bh}$) of the BH, namely,

$$R_{\rm s-bh} = \frac{2GM_{\rm bh}}{c^2} = 5 \times 10^{-4}\ {\rm pc} \left(\frac{M_{\rm bh}}{10^8\ {\rm M}_\odot}\right) \qquad (2)$$

The majority of the afore mentioned reliable measurements target ellipticals and a few early-type (Sa-Sbc) spirals with central $\sigma < 60$ km s^{-1}, and probe BH masses in the range 10^7–10^9 M$_\odot$. Conversely, measuring BH masses in late-type spirals and dwarf galaxies poses many challenges, and there are no firm measurements of BH masses below 10^6 M$_\odot$. However, theoretical models and a mounting body of observational evidence put the existence of IMBHs on a relatively firm footing (see review by van der Marel 2003). The first challenge in measuring the masses of IMBHs is that the gravitational radii of such BHs are typically too small to be easily resolved even with *HST*. A second complication is that late-type spirals and dwarf galaxies which might harbor such BHs also tend to host a bright 10^6–10^7 M$_\odot$ stellar cluster (Boker et al. 1999) whose dynamical effect can mask that of the BH. A 10^4–10^5 M$_\odot$ BH (Filipenko & Ho 2003) has been invoked in the Sm dwarf NGC 4395 which hosts the nearest and lowest luminosity Seyfert 1 nucleus. Upper limits on BH masses are reported in several systems, e.g., 10^6–10^7 M$_\odot$ for six dwarf ellipticals in Virgo (Geha, Guhathakurta, & van der Marel 2002), 5×10^5 M$_\odot$ for the Scd spiral IC342 (Boker et al.1999). Gebhardt, Rich, & Ho (2002) infer the presence of an IMBH with a mass of a few $\times$ 10^4 M$_\odot$ in one of the most massive stellar clusters (G1) in M31, but an alternative interpretation of the dataset has been presented by Baumgardt et al. (2003). A tantalizing dark central mass concentration of a few $\times 10^3$ M$_\odot$ (Gerssen et al. 2003) is reported in the globular cluster M15 from *HST* data, but it remains unclear whether it is an IMBH. *Chandra* observations of ultraluminous X-ray sources also suggest the presence of IMBHs (Clobert & Miller 2004 and references therein).

At many levels, measuring BH masses in local AGN such as Seyferts and LINERS is more challenging than corresponding measurements in massive quiescent galaxies. The bright non-thermal active nucleus in Seyfert galaxies can drown the spectroscopic features from which dynamical measurements are made. Consequently, BH masses in local AGN are commonly mapped with alternative techniques such as reverberation mapping (Blandford & McKee 1982; Peterson 1993; see Peterson these proceedings) where one estimates the virial mass inside the broad-line region (BLR) by combining the velocity of the BLR with an estimate of the size of the BLR based on time delay measurements. Reverberation mapping can typically probe scales $\sim$600 $R_{\rm s-bh}$ and has yielded BH masses for several tens of AGN (Peterson 1993; Wandel, Peterson, & Malkan 1999; Kaspi 2000). Earlier controversies existed on the reliability of the method due to purported systematic differences in the BH-to-bulge mass ratio between AGN with reverberation mapping data and quiescent galaxies or QSOs. However, recent work (e.g., Ferrarese et al. 2001) claims that for AGN with accurate measurements of stellar velocity dispersions, the reverberation masses agree with the BH mass determined from the tight $M_{\rm bh}$–σ relation (Subsect. 2.2) which is derived from quiescent galaxies.

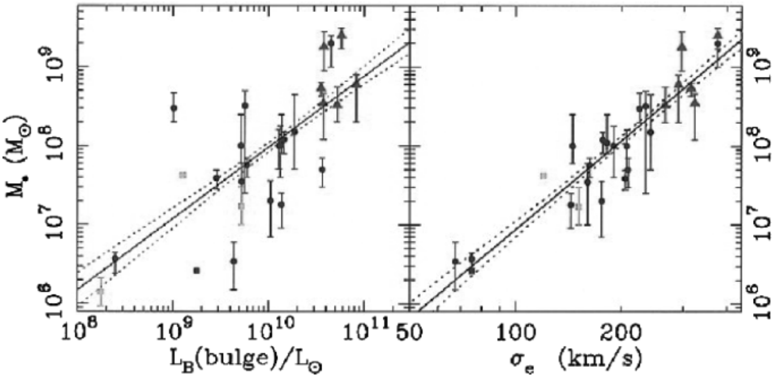

Fig. 1. Correlation between central BH mass and circumnuclear velocity dispersion – Black hole mass versus bulge luminosity (*left*) and the luminosity-weighted aperture dispersion within the effective radius (*right*). Green squares denote galaxies with maser detections, red triangles are from gas kinematics, and blue circles are from stellar kinematics. Solid and dotted lines are the best-fit correlations and their 68% confidence bands. (From Gebhardt et al. 2000)

2.2 Relationship of the Central BH to the Bulge and Dark Halo

A tight correlation has been reported between the mass of a central BH and the stellar velocity dispersion (σ) of the host galaxy's bulge, as shown on Fig. 1 (Ferrarese & Merritt 2000; Gebhardt et al. 2000):

$$M_{\mathrm{h}} = \alpha \left(\frac{\sigma}{200\ \mathrm{km\ s^{-1}}} \right)^{\beta}\ \mathrm{M}_{\odot} \tag{3}$$

where $\alpha = (1.7 \pm 0.3) \times 10^8$, $\beta = (4.8 \pm 0.5)$ (Ferrarese & Merritt 2000), and $\alpha = (1.2 \pm 0.2) \times 10^8$, $\beta = (3.8 \pm 0.3)$ (Gebhardt et al. 2000). Tremaine et al. (2002) assign the range in quoted values for β to systematic differences in velocity dispersions used by different groups. The M_{bh}–σ relation reported originally in the literature (Gebhardt et al. 2000; Ferrarese & Merritt 2000; Tremaine et al. 2002) is primarily based on local early-type galaxies (E/SOs) and a handful of spirals Sb–Sbc, and it primarily samples quiescent BHs with masses in the range a few × (10^7–10^9) $\mathrm{M}_{\odot}$. This relation was subsequently found to also hold in AGN hosts (Ferrarese et al. 2001), and in bright QSOs out to $z \sim 3$ with estimated BH masses of up to 10^{10} $\mathrm{M}_{\odot}$ (Shields et al. 2003). *This suggests that active and quiescent BHs bear a common relationship to the surrounding triaxial component of their host galaxies over a wide range of cosmic epochs and BH masses* (10^6–10^{10} $\mathrm{M}_{\odot}$).

Numerous variants of the M_{bh}–σ relation have been proposed. While earlier correlations between the mass of CDOs/SMBHs and the bulge luminosity (L_{bulge}) had significant scatter (e.g., Kormendy & Richstone 1995), recent work (Häring & Rix 2004) based on improved BH and bulge masses yield a very tight M_{bh}–M_{bulge} relation. Graham et al. (2001) find a correlation between the light concentration of galaxies and the mass of their SMBHs, and claim this relation is as tight as the M_{bh}–σ relation. Grogin et al. (2004) have searched for signs of this correlation at $z \sim 0.4$–1.3 in a comparative study of structural parameters among 34000 galaxies in the GOODS fields, including 350 X-ray selected AGN hosts in the overlapping *Chandra* Deep Fields. Compared to the inactive galaxies, the AGN hosts have significantly enhanced concentration indices throughout the entire redshift range, as measured in rest frame B-band for a volume-limited sample to $M_{\mathrm{B}} < -19.5$ (and to $L(2\text{–}8\ \mathrm{keV}) > 10^{42}$ for the AGN). Finally, Ferrarese (2002) shows that the M_{bh}–σ relation translates to a relation between the mass of the BH and that of the dark matter (DM) halo (M_{dm})

$$M_{\mathrm{h}} \ = 10^7\ \mathrm{M}_{\odot} \left(\frac{M_{\mathrm{dm}}}{10^{12}\ \mathrm{M}_{\odot}} \right)^{1.65} \tag{4}$$

if one assumes that σ correlates with the circular speed V_{c} which bears an intimate relation to the DM halo within the standard ΛCDM paradigm.

A plethora of theoretical studies have explored the growth of BHs and the possible origin of a fundamental M_{bh}–σ relation (e.g., Haehnelt & Kauffmann 2000; Adams, Graff, & Richstone 2001; Burkert & Silk 2001; Di Matteo, Croft, Springel, & Hernquist 2003; Bromm & Loeb 2003; Wyithe & Loeb 2003; El-Zant et al. 2003). According to Haehnelt & Kauffmann (2000), hierarchical galaxy formation models where bulges and SMBHs both form during major mergers produce a M_{bh}–σ correlation. Star-formation (SF) regulated growth of BHs in protogalactic spheroids has been proposed by Burkert & Silk (2001) and Di Matteo et al. (2003). In many of these models, black hole growth stops because of the competition with SF and, in particular, feedback, both of which determine the gas fraction available for accretion. According to Wyithe & Loeb (2003), a tight M_{bh}–σ relation naturally results from hierarchical ΛCDM merging models where SMBHs in galaxy centers undergo self-regulated growth within galaxy haloes until they unbind the galactic gas that feeds them. El-Zant et al. (2003) have suggested that the BH–bulge–DM halo correlation can be understood within the framework of galactic structures growing within flat-core, mildly triaxial haloes.

3 Central Issues in Fueling AGN and Starbursts

I present here an overview of several central issues that are relevant for understanding the fueling of AGN and circumnuclear starbursts.

3.1 Mass Accretion Rates

For a standard BH accretion disk with an efficiency ϵ of conversion between matter and energy, the radiated bolometric luminosity L_{bol} is related to the

Table 1. Typical $L_{\rm bol}$ and $\dot{M}_{\rm bh}$ for QSOs and local AGN

Type of AGN (1)	$L_{\rm bol}$[a] (ergs s^{-1}) (2)	Typical $L_{\rm bol}$ (ergs s^{-1}) (3)	Typical $\dot{M}_{\rm bh}$[b] (M$_\odot$ yr^{-1}) (4)
QSOs	10^{46}–10^{48}	10^{47}–10^{48}	10–100
Seyferts	10^{40}–10^{45}	10^{43}–10^{44}	10^{-3}–10^{-2}
LINERs	10^{39}–$10^{43.5}$	10^{41}–10^{42}	10^{-5}–10^{-4}

Notes to Table – a. The full range in bolometric luminosity ($L_{\rm bol}$) for Seyfert and LINERS is taken from Ho, Filippenko, & Sargent 1997a, while for QSOs different sources in the literature are used; b. The typical $\dot{M}_{\rm bh}$ in column (4) is derived from the typical $L_{\rm bol}$ in column (3) assuming a standard radiative efficiency $\epsilon \sim 0.1$

mass accretion rate ($\dot{M}_{\rm bh}$) at the last stable orbit of a BH by

$$\dot{M}_{\rm bh} = 0.15 \ {\rm M}_\odot \ {\rm yr}^{-1} \left(\frac{\epsilon}{0.1}\right) \left(\frac{L_{\rm bol}}{10^{45} \ {\rm ergs \ s}^{-1}}\right) \tag{5}$$

Table 1 shows typical observed bolometric luminosities and inferred mass accretion rates for QSOs and local AGN (Seyfert, LINERS) assuming a standard radiative efficiency $\epsilon \sim 0.1$. The standard value of $\epsilon \sim 0.1$ applies if the gravitational binding energy liberated by the accreting gas at the last stable orbit of the BH is radiated with an efficiency of $\sim 0.1 \ c^2$. In practice, the radiative efficiency depends on the nature of the accretion disk and gas accretion flows. For instance, thin-disk accretion onto a Kerr BH can lead to a radiative efficiency $\epsilon \sim 0.2$. It has been suggested that the most luminous quasars at high redshift may have grown with $\epsilon \sim 0.2$, or alternatively that they have a super-Eddington luminosity (Yu & Tremaine 2002). Conversely, in certain popular models of gas accretion flows such as adiabatic inflow-outflow solutions (ADIOS; Blandford & Begelman 1999) and convection-dominated accretion flows (CDAF: Narayan et al. 2000) only a small fraction of the matter which accretes at the outer boundary of the flow contributes to the mass accretion rate at the BH due to turbulence and strong mass loss. This leads to an effective radiation efficiency $\ll 0.1$ when applied to the mass accretion rate at the outer boundary of the accretion flow. Thus, within the CDAF and ADIOS paradigms, the gas inflow rates that must be supplied on scales of tens of pc may be much larger than those quoted in Table 1, even for low luminosity Seyferts.

3.2 The Angular Momentum Problem

The most important challenge in fueling AGN is the angular momentum problem rather than the amount of fuel *per se.* The angular momentum per unit

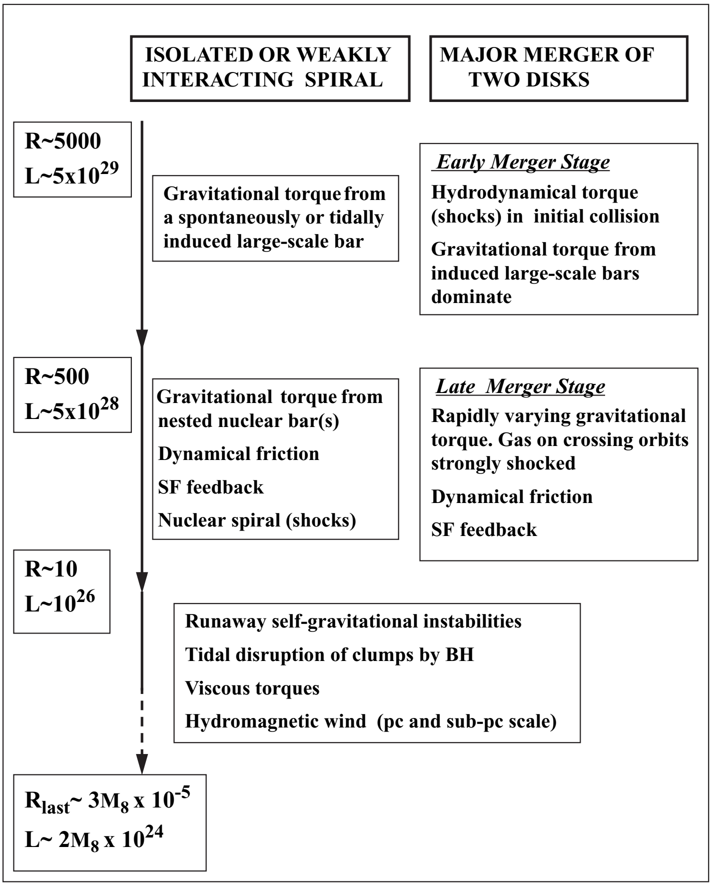

Fig. 2. The angular momentum problem in the fueling of AGN and starbursts: The specific angular momentum (L) of gas located at a radius (R) of several kpc must be reduced by more than 10^4 before it is fit for consumption at the last stable orbit (R_{last}) of a BH. In contrast, powerful starbursts can be more easily triggered via gravitational torques which build large gas densities on circumnuclear (R = 500 pc) scales. This figure schematically illustrates some mechanisms that can reduce L and drive gas inflow on various spatial scales in a relatively quiescent galaxy *(left)* and in a major merger *(right)*. R is in pc, L is in units of cm^2 s^{-1}, and a ($M_8 \times 10^8$ M$_\odot$) BH is assumed. See text for details

mass or specific angular momentum $L = r \times v$ of fuel at the last stable radius of a BH of mass ($M_8 \times 10^8$ M$_\odot$) is several times $10^{24} M_8$ cm^2 s^{-1}. In contrast, matter (star or gas) rotating in a spiral or elliptical galaxy at a radius of 10 kpc has a specific angular momentum of several times $10^{29} M_8$ cm^2 s^{-1}. This is illustrated in Fig. 2 assuming typical galactic rotation velocities. Thus, *the specific angular momentum of matter located at a radius of a few kpc must be* *reduced by more than 10^4 before it is fit for consumption by a BH.* Searching for mechanisms which can achieve this miraculous reduction of angular momentum is one of the driving objectives of AGN research. Even at a radius of 200 pc, L is still a factor of 1000 too large, and the angular momentum barrier is a more daunting challenge than the amount of gas. For instance, in the case of a Seyfert with an accretion rate of $\sim 10^{-2}$ $M_\odot$ yr^{-1} and a duty cycle of 10^8 years, a gas cloud of 10^6 $M_\odot$ may provide adequate fuel. Such clouds are certainly common within the inner 200 pc radius of spiral galaxies, but we yet have to understand what physical processes are able to squeeze their angular momentum out by more than 99.99%. The BH is analogous to an exigent dieter who has a plentiful supply of rich food, but can only consume 99.9% fat-free items!

3.3 Dominant Fueling Mechanisms on Different Scales

Gravitational torques, dynamical friction, viscous torques, and hydrodynamical torques (shocks) are some of the mechanisms which remove angular momentum from the dissipative gas component and channel it to small scales, thereby helping to fuel central starbursts and massive BHs. These different fueling mechanisms assume a different relative importance at different radii in a galaxy, and also, when dealing with a strongly interacting galaxy versus an isolated one. I will review these different mechanisms in detail from an observational and theoretical perspective in Sects. 4–9, but here I discuss a few key concepts and provide a schematic overview in Fig. 2.

Gravitational torques operate on a timescale ($t_{\rm gra}$) comparable to the orbital timescale and provide, therefore, the most efficient way of reducing angular momentum on large to intermediate scales (tens of kpc – a few 100 pc). This can be seen by comparing $t_{\rm gra}$ with the typical timescales on which dynamical friction ($t_{\rm df}$) and viscous torques ($t_{\rm vis}$) operate for a cloud of mass M (Table 2). Dynamical friction on a clump of mass M and speed v at a radius R operates on a timescale which is $\propto (R^2\ v/M\ \ln\Lambda)$, where $\ln\Lambda$ is the Coulomb logarithm (Binney & Tremaine 1987). For a 10^7 $M_\odot$ gas cloud at a kpc radius in a disk galaxy, $t_{\rm df}$ is an order of magnitude larger than $t_{\rm gra}$ (Table 2). However, for massive gas clumps at low radii, dynamical friction becomes increasingly important: it can drive a 10^8 $M_\odot$ cloud from $R \sim 200$ pc down to $R \sim 10$ pc within a few times 10^7 yrs (Sect. 9).

Table 2. Gravitational Torques, Dynamical Friction, and Viscous Torques

R (pc) (1)	M ($M_\odot$) (2)	$t_{\rm gra}$ (Myr) (3)	$t_{\rm df}$ (Myr) (4)	$t_{\rm visc}$ (Myr) (5)
1000	1e7	20	1020	1000
200	1e7	4	60	–

In an isolated galaxy (Fig. 2), gravitational torques are exerted by non-axisymmetric features such as large-scale (Sect. 6) and nuclear (Sect. 7) bars. While a large-scale bar efficiently drives gas from the outer disk into the inner kpc, the bar-driven gas flow slows or even stalls as it crosses the inner Lindblad resonance (ILR) for reasons described in Sect. 6. At this stage, the gas piles up typically at a radius of several 100 pc where powerful starbursts are commonly observed (Sect. 6; Fig. 6). However, gas on these scales has a specific angular momentum that is still more than 1000 times too high for it to be digestible by a BH. If a nuclear bar (Sect. 7) is present, it can break the status quo and torque gas from the ILR region of the large-scale bar down to tens of pc. In addition, if massive gas clumps exist in the inner few 100 pc, dynamical friction can drive them down to tens of pc (Sect. 9). Finally, feedback from SF (e.g., shocks from supernovae) can remove energy and angular momentum (Sect. 9) from a small fraction of the circumnuclear gas. On scales of tens of pc, the tidal torque from the BH itself can disrupt gas clumps and stellar clusters, possibly into an accretion disk (Sect. 9). Subsequently, on pc and sub-pc scales, viscous torques and hydromagnetic outflows in AGN (Sect. 9) may become important.

Simulations suggest that induced large-scale stellar bars remain the main driver of gas inflows down to scales of a few 100 pc, even in the case of interacting galaxies (Fig. 2), namely in many minor mergers (Subsect. 5.2) and during the *early* stages of major (1:1) and intermediate mass ratio (1:3) interactions (Subsect. 5.1). Just like in the case of an isolated barred galaxy, gas inflows driven by an induced bar also slow down near the ILR. However, the *final stages* of a major or intermediate mass ratio merger bring in very different elements. As violent relaxation starts, gas experiences strongly-varying gravitational torques, and if it is on interacting and crossing orbits, it also suffers strong shocks (Subsect. 5.1; Fig. 2). Thus, in the final merger stages, gas loses angular momentum and large gas inflows ($\gg 1$ $M_\odot$ yr^{-1}) down to small scales can result, provided the earlier episodes of SF have not depleted most of the circumnuclear gas already (Subsect. 5.1).

3.4 Census and Growth Epoch of BHs

Table 3 compares the BH mass density ($\rho_{\mathrm{bh-qso}}$) accreted during the optically bright QSO phases ($z = 0.2$–5) to the BH mass density in present-day galaxies (both active and inactive). Yu & Tremaine (2002) find $\rho_{\mathrm{bh-qso}} \sim (2.5 \pm 0.4) \times 10^5$ $(h_0/65)^2$ $M_\odot$ Mpc^{-3} using the extrapolated QSO luminosity function from the 2dF redshift survey and a radiative efficiency of 0.1. Similar values have been reported by others including Wyithe & Loeb (2003), Ferrarese (2002b), and Chokshi & Turner (1992). This value of $\rho_{\mathrm{bh-qso}}$ is a lower limit to the total BH mass density we expect to be in place by $z = 0.2$ since it does not incorporate optically obscured QSOs and any build-up of the BH mass occurring outside the QSO phase. However, it is probably not far off, since the BH mass density from X-ray AGN counts at $z > 0.2$ ($\rho_{\mathrm{bh-xray}}$)

Table 3. Census of BH Mass density

	BH Mass Density [10^5 M$_\odot$ Mpc^{-3}]
$\rho_{\mathrm{bh-QSO}}$ accreted during optical QSO phase ($z = 0.2$–5)	2–4[a,b,c,d]
$\rho_{\mathrm{bh-Xray}}$ from X-ray background ($z > 0.2$)	2–5[e,f]
$\rho_{\mathrm{bh-local}}$ in local early-type galaxies ($z < 0.1$)	2–6[a,b,g]
$\rho_{\mathrm{bh-Sy}}$ in local Seyferts	<0.5[c]

References in table – a. Yu & Tremaine 2002; b. Wyithe & Loeb 2003; c. Ferrarese 2002; d. Chokshi & Turner 1992; e. Cowie & Barger 2004; f. Fabian & Iwasawa 1999; g. Merritt & Ferrarese 2001.

is estimated to be 2–5 × 10^5 M$_\odot$ Mpc^{-3} (Cowie & Barger 2004; Fabian & Iwasawa 1999; Table 3).

In the local Universe, the BH mass density in early-type galaxies at $z < 0.1$ is estimated to be $(2.5 \pm 0.4) \times 10^5$ $(h_0/65)^2$ M$_\odot$ Mpc^{-3}, based on the measured velocity dispersion of early-type galaxies in the Sloan Digital Sky Survey and the M_{bh}–σ relation (Yu & Tremaine 2002). However, rough estimates of the BH mass density in local *active* Seyfert 1 and 2 galaxies yield significantly lower values (Ferrarese 2002; Padovani et al. 1990; Table 3).

In summary, the census of BH mass density (Table 3) suggests that accretion with a standard radiation efficiency of 0.1 during the quasar era can readily account for the BH mass density found in local ($z < 0.1$) early-type galaxies. Only a small fraction of this local BH mass density appears to be currently active as Seyfert galaxies and the inferred mass accretion rates in such cases are typically 10^3 times lower than in QSOs. This suggests that there *is no significant growth of BHs in the present epoch compared to the quasar era.* Thus, we should bear in mind that local AGN (Seyferts) with current low levels of BH growth may well differ from luminous QSOs near $z \sim 2.5$ in one or more of the following characteristics: *the nature of the dominant fueling mechanism, the amount of cold gas reservoir, and the nature of the host galaxy.* For instance, tidal interactions and minor or major mergers may have been much more important in the quasar era and early epochs of galaxy growth than they are in activating present-day Seyfert galaxies.

3.5 The Starburst–AGN Connection

While I discuss the fueling of both AGN and starbursts in this review, I will not explicitly address the starburst–AGN connection. I only mention here that this connection can be *circumstantial, influential, or causal.* A *circumstantial* connection refers to the fact that starburst and AGN activity can both manifest in the same system simply because they are affected by a common element such as a rich supply of gas, or an external trigger (e.g., an interaction). Examples include the ULIRG–QSO connection (Sanders et al. 1988), evolutionary scenarios for Seyfert 2 (e.g., Storchi-Bergmann et al.

2001), and perhaps the blue color of AGN hosts described in Sect. 4. An *influential* connection is one where the AGN and starbursts may contaminate each other's *observed* properties. Examples include the starburst affecting the featureless continuum and line ratios of Seyferts (Cid-Fernandes et al. 2001), or washing out the hard accretion disk spectrum. A *causal* connection is a more fundamental connection where the starburst causes the AGN or vice versa. One example is the evolution of a dense stellar cluster into a BH (Norman & Scoville 1988).

3.6 A Note of Caution on Empirical Correlations

There exists many contradictory reports in the literature of correlations or lack thereof between starburst/AGN activity and host galaxy properties (e.g., Hubble types, bar fraction, nuclear bar fraction), or external triggers (e.g., presence of companions, morphological signs of interactions/mergers). Many caveats conspire towards this dismal state of affairs and should be avoided: *(1)* Many early studies fail to adopt the key practice of having a large control sample which is matched to the active sample or to the starburst sample in terms of relevant parameters such as distance, morphological types, luminosities, inclinations, and environments. *(2)* The classification of morphological features such as bars and Hubble types is still often made from optical catalogs (e.g., the Third Reference Catalogue (RC3); de Vaucouleurs et al. 1991) and suffer from subjectivity, low spatial resolution, and contamination by dust. It is better to use a quantitative method (e.g., ellipse fits) for characterizing bars and apply it to near-infrared (NIR) rather than optical images. The former are less affected by extinction and typically yield a bar fraction which is higher by 20–30 % (e.g., Knapen et al. 2000; Eskridge et al. 2002). *(3)* Cross comparisons of discrepant results are often difficult because they are based on inhomogeneous samples drawn from different local AGN catalogs that have limited overlap and different biases. For instance, optically selected magnitude-limited samples may be biased against faint nuclei embedded in bright galaxies. UV-based catalogs may favor blue Seyfert 1 and quasars. Commonly used catalogs include the Veron-Cetty & Veron Catalog of Seyferts and LINERS, the optically selected CfA sample of 48 Seyferts (Huchra & Burg 1992), the Palomar Optical Spectroscopic Survey (POSS; Ho et al. 1997a) of 486 emission line nuclei geared towards low luminosity HII and AGN nuclei, and the extended 12 μm Galaxy Sample (E12GS) of 891 galaxies (Hunt & Malkan 1999). *(4)* Nuclear types (HII, LINER, Seyferts) listed in literature databases such as NED often show significant discrepancies from recent careful spectroscopic classifications (e.g., Ho et al. 1997a). In Sects. 4–9, I will focus on studies which tend to avoid these caveats or alternatively qualify the caveats as they arise.

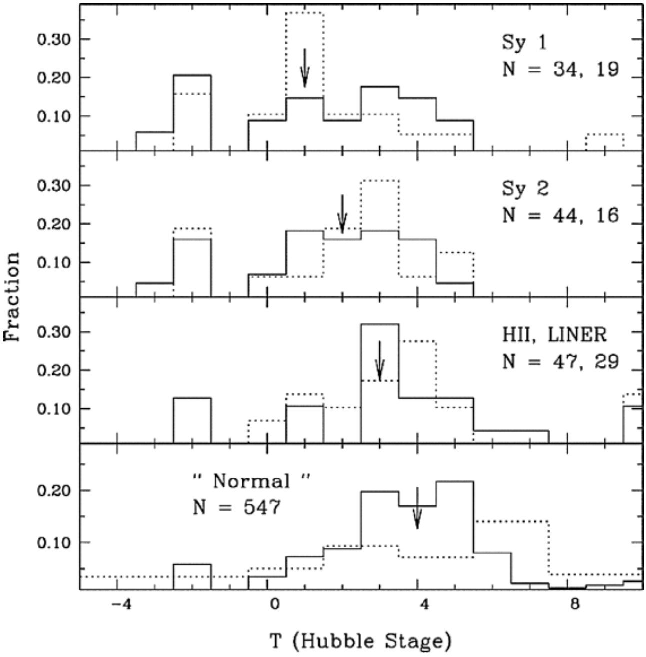

Fig. 3. Distributions of the Hubble types in AGN (Seyferts, LINERS), HII and normal nuclei: The *solid line* in all four panels represents data from the 12 µm sample (E12GS; Hunt & Malkan 1999). The *dotted line* represents data taken from E12GS or other AGN catalogs. The panels represent *(from top to bottom)* (1) The Sy1 sample from E12GS (*solid line*) and the CfA (*dotted line*) sample; (2) The Sy1 sample from E12GS (*solid*) and the CfA (*dotted*) sample; (3) The HII (*solid*) and LINER (*dotted*) samples from E12GS; (4) The normal galaxies from E12GS (*solid*) and the Uppsala General Catalog (*dotted*) as tabulated by Roberts & Haynes (1994). The numbers in each panel refer to the number of objects represented by the solid and dotted histograms. The data have been binned in terms of RC3 Hubble types as follows: S0a and earlier ($T > 0$); Sa, Sab ($0 < T < 2$); Sb, Sbc ($2 < T < 4$); Sc, Scd ($4 < T < 6$); Sd and later ($T > 6$). The vertical arrows mark subsample medians, calculated with a type index resolution of unity. (Figure is from Hunt & Malkan 1999)

4 Hubble Type and Colors of AGN Hosts

Do local Seyferts and central starbursts reside preferentially in certain type of galaxies? Using the 12 µm sample (E12GS) and the CfA sample of Seyferts, Hunt & Malkan (1999; Fig. 3) report that *Sy 1 and Sy 2 nuclei tend to reside primarily in early-type (E–Sbc) galaxies.* The Hubble type quoted here is the RC3 Hubble index based on visual classifications of optical images. A similar result on Seyferts is reported by Ho et al. (1997a) from the POSS optical spectroscopic survey which tends to include lower luminosity galaxies and has a median extinction-corrected Hα luminosity of only 2×10^{39} erg s^{-1}. These findings on Seyferts are consistent with earlier less comprehensive studies (e.g., Hummel et al. 1990; Terlevich, Melnick, & Moles 1987; Balick & Heckman 1982). HII host galaxies tend to have later Hubble types than Seyferts according to both E12GS (Hunt & Malkan 1999) and POSS (Ho et al. 1997a), but the mean value of the Hubble type varies in the surveys, possibly due to luminosity differences.

Properties of AGN hosts in the redshift range 0.5–2.5 are particularly interesting as the optical QSO activity peaks at $z \sim 2.5$. Keen insights are stemming from two large panchromatic *HST* surveys, the Galaxy Evolution from Morphology and SEDs (GEMS; Rix et al. 2004) and the Great Observatories Origins Deep Surveys (GOODS; Giavalisco et al. 2004). A GEMS study of 15 AGN which have $M_{\rm B} \simeq -23$ and are in the redshift range $0.5 < z < 1.1$ where comparable data for control inactive galaxies exist, report that 80% of the AGN hosts are early-type (bulge-dominated) compared to only 20% that are disk-dominated (Sanchez et al. 2004). The high rest-frame B-band concentration indices of the GOODS AGN at $z \sim 0.4$–1.3 (Grogin et al. 2004; Subsect. 2.2) also support the interpretation that these systems are predominantly bulge-dominated.

Furthermore, Sanchez et al. (2004) report that a much larger fraction (70%) of the early-type AGN hosts at $0.5 < z < 1.1$ show blue global rest-frame U-V colors, compared to inactive early-type galaxies in this redshift and luminosity range. These global blue colors are consistent with the presence of young stellar populations over large regions of the AGN host galaxies. The trend of enhanced blue colors in AGN hosts at $0.5 < z < 1.1$ seems to hold both at higher and lower redshifts. SDSS spectra of local $z < 0.2$ Sy2 galaxies show a significant contribution from young stellar populations, and this trend is strongly correlated with nuclear luminosity (Kauffman et al. 2003). At higher redshifts ($1.8 < z < 2.75$), Jahnke et al. (2004) find that the host galaxies of 9 moderately bright ($M_{\rm B} \simeq -23$) AGN in the GEMS survey have rest-frame UV colors that are considerably bluer than expected from an old population of stars. Unfortunately, for these 9 distant AGN the detection images are not deep enough to constrain the morphology. Earlier studies of a handful of luminous QSO at $z > 2$ also reported very UV-luminous hosts (e.g., Lehnert et al. 1992; Hutchings et al. 2002).

One possible interpretation of the enhanced global blue colors exhibited by AGN hosts is that the mechanism which ignites the central BH in these galaxies also triggers global SF. The fact that SF is triggered not only in the nuclear region, but over an extended (several kpc) region, would tend to exclude major (1:1) mergers and favor weaker (e.g., minor (1:10) or intermediate mass-ratio (3:1)) mergers/interactions where the gas has a larger L, and typically settles in extended inner disks during simulations (see Sect. 5). In fact, only 3/15 of the AGN hosts in the Sanchez et al. (2004) study show signs of strong disturbances. In the same vein, the GOODS AGN study (Grogin et al. 2004; Subsect. 2.2) reports no significant difference between the rest-frame B asymmetry index of active and inactive galaxies over $z \sim 0.4$–1.3. This suggests that AGN do not preferentially occur in major mergers over this redshift range.

5 Interactions and AGN/Starburst Activity

5.1 Basic Physics of Major Mergers

The term "major merger" usually refers to the merger of two disk galaxies with a mass ratio of order 1:1. Simulations of major mergers (e.g., Negroponte & White 1983; Noguchi 1988; Barnes & Hernquist 1991; Heller & Shlosman 1994; Mihos & Hernquist 96; Struck 1997) show that they generate large gas inflows into the inner kpc and could plausibly trigger intense starbursts and AGN activity. The full parameter space controlling the outcome of major mergers has not yet been fully explored, but I summarize here (see also Fig. 2) some salient general findings:

1. Not all speeds, energies, angular momenta, and orientations are equally effective in inducing large gas inflows, rapid mergers, and disruptions during a major merger. For instance, while all bound orbits will eventually lead to mergers, low angular momentum and low energy orbits will lead to more rapid mergers. Prograde mergers, where the spin and orbital angular momenta are aligned, occur faster than retrograde mergers, lead to more violent disruption, and excite larger non-circular motions (e.g., Binney & Tremaine 1987).
2. Hydrodynamical torques (shocks) tend to be important in the initial collision when they add spin angular momentum to the gas in both disks (Mihos & Hernquist 1996; Barnes & Hernquist 1996), but gravitational torques dominate thereafter.
3. In the *early stages* of the merger, gas inflows primarily result from *gravitational torques exerted by a stellar bar* induced in the disk of the two galaxies (e.g., Mihos & Hernquist 1996; Heller & Shlosman 1994; Noguchi 1988, Sellwood 1988; Hernquist 1989). The torque from the bar is significantly larger than the gravitational torque exerted by the galaxies on each other. Thus, early gas inflows are primarily the result of gas response in a barred potential as outlined in Sect. 6. Mihos & Hernquist (1996) find that the strength of the bar induced decreases as the bulge-to-disk (B/D) ratio of the disk increases. This behavior is consistent with the idea that a dynamically hot bulge component stabilizes a disk against a bar mode.
4. In the *final stages* of the merger, as the galaxies merge and undergo violent relaxation, the gas experiences rapidly varying gravitational torques as well as shocks on interacting orbits (e.g., Mihos & Hernquist 1996), and therefore loses energy and angular momentum. Large gas inflows

($\gg 1\mathrm{M}_\odot\ \mathrm{yr}^{-1}$) can result in the late merger stages, provided the episodes of SF triggered by earlier inflows have not depleted most of the gas already. In simulations with highly gas-rich disks (Heller & Shlosman 1994), large gas concentrations build up in the inner kpc leading to the formation of self-gravitating gas clumps which are driven to smaller scales by dynamical friction from the stellar background (see also Sect. 9). Gaseous nuclear bars form via gravitational instabilities and drive further inflow.

5. After violent relaxation, the end-product of a major merger tends to have an $\mathrm{r}^{1/4}$ de Vaucouleurs-type stellar profile and boxy isophotes similar to many luminous elliptical galaxies. In the case of intermediate mass ratio (e.g., 1:3) mergers, the gas has a larger specific angular momentum and tends to settle into an extended inner disk (Naab & Burkert 2001) rather than being as centrally concentrated as in a 1:1 merger. The stellar component has an $\mathrm{r}^{1/4}$ profile, disky isophotes, and isotropic velocity dispersion similar to lower luminosity disky ellipticals (Naab & Burkert 2001).
6. Future work has yet to fully explore the parameter space which can influence a merger such as the energy and orbital angular momentum of the initial orbits, the relative alignment of the spin and orbital angular momentum, the orbital geometry, the gas content, and the mass ratios. Another key step is to realistically incorporate in simulations feedback effects from SF and thermal cooling effects which lead to a multi-phase ISM (e.g., Struck 1997; Wada & Norman 2001). Shocks from SNe can dissipate orbital energy and transfer angular momentum outwards. Major heating processes associated with SF such as stellar winds, SNe, and UV photoheating are sources of angular momentum, leading to fountain flows and starburst-driven winds (Jogee, Kenney, & Smith 1998; Heckman et al. 1990).

5.2 Basic Physics of Minor Mergers

The term “minor merger” usually refers to the merger between a large disk galaxy and a satellite with a mass ratio of order 1:10. Such mergers are believed to be very common (Ostriker & Tremaine 1975) and have been the subject of numerous simulations (e.g., Hernquist & Mihos 1995; Mihos et al. 1995; Quinn, Hernquist, & Fullagar 1993; Walker, Mihos, & Hernquist 1996). A few of the general principles underlying this class of mergers are outlined below.

1. As it moves through the dark matter halo, the satellite experiences dynamical friction and sinks rapidly towards the main disk on a timescale $t_{\mathrm{df}} \propto (R^2\ v/M\ \ln\Lambda)$, where M is the mass of the satellite, v is its speed, R is the galactocentric radius, and $\ln\Lambda$ is the Coulomb logarithm (Binney & Tremaine 1987). For typical values of these parameters, the dynamical friction timescale is comparable to a few orbital periods or a few Gyr. The

orbital angular momentum of the satellite is converted into spin angular momentum for the disk and halo of the main galaxy.

2. The satellite first tends to sink in the disk of the primary where it drives warps before it sinks towards the central regions (Quinn et al. 93; Hernquist & Mihos 1995). Depending on the shape of the halo, these warps can persist for a significant fraction of time before they are washed out by phase mixing (Quinn et al. 93; Dubinski 1994).
3. The satellite exerts tidal torques on the main stellar disk and *induces in it large amplitude non-axisymmetries* (Hernquist & Mihos 1995; Mihos et al. 1995; Quinn et al. 1993) such as stellar spirals and bars. Since gas is collisional and dissipative, the gas response leads the stellar response (see Sect. 6) and gas in the disk is gravitationally torqued towards the central few hundred pc. In effect, *the gas in the main disk experiences a much larger gravitational torque from stars in the disk than from the satellite.* In simulations by Hernquist & Mihos (1995), a large fraction of the gas in the main disk can be driven into the central regions in this way.
4. The satellite may have a significant fraction of its material tidally stripped before it sinks towards the inner part of the main disk.
5. Overall, the minor merger leads to large gas concentrations in the inner few 100 pc which may be relevant for fueling starbursts and AGN. It also heats the inner parts of disks vertically and increases the disk thickness and velocity dispersion of stars. Some authors (e.g., Quinn et al. 1993; Walker et al. 1996) have even suggested that minor mergers may be the origin of thick disks.

5.3 Correlations Between Interactions and Starbursts

Colorful examples abound of starbursts and AGN activity occurring in interacting or merging galaxies. However, statistically significant correlations between central activity and signs of morphological disturbance have only been found in the case of highly luminous starbursts or AGN. Signs of strong tidal interactions and mergers are ubiquitous in systems where the SR rate (SFR) is estimated to be $\geq 10\ \mathrm{M}_\odot\ \mathrm{yr}^{-1}$ such as ultra-luminous infrared galaxies (ULIRGs; Veilleux, Kim, & Sanders 2001; Sanders & Mirabel 1996), and the brightest Arp galaxies (Hummel et al. 1990; Kennicutt et al. 1987). In the local Universe, more than 95% of ULIRGS in the 1 Jy IRAS sample show optical and NIR morphological signatures of a strong interaction or merger in the form of tidal tails, bridges, double nuclei, and overlapping disks (Veilleux et al. 2001). In a study based on panchromatic *HST* GOODS data, optically-selected bright starburst galaxies at $z \sim 1$ show a larger frequency ($\sim 50\%$) of disturbed and asymmetric morphologies, compared to 13% and 27%, respectively, in control samples of early-type and late-type galaxies (Jogee et al. 2003; Mobasher et al. 2004).

While the most extreme *individual* starbursts (in terms of luminosity or luminosity per unit mass of gas) may be triggered by a major or intermediate mass ratio interaction, it is important to bear in mind that the *cumulative* SFR density at a given cosmic epoch may be dominated by SF emanating from a large number of relatively undisturbed galaxies rather than from SF in major mergers. This is a particularly viable possibility out to intermediate redshifts ($z \sim 1$) where major mergers are relatively rare. In fact, preliminary findings (Wolf et al. in prep.) from a study of ~ 1400 galaxies at $z \sim 0.7$ in the GEMS survey suggest that while galaxies with strongly disturbed rest-frame optical morphologies are amongst the most UV-luminous candidates, they only make up a small fraction of the UV luminosity density (uncorrected for extinction) at $z \sim 0.7$. While this result on UV luminosity density does not directly translate to SFR density due to potential extinction effects, it does highlight the need for further studies on how mergers/interactions impact central activity.

5.4 Correlations Between Interactions and AGN

An excess of companions in local Seyferts has been reported in early papers (e.g., Dahari 1984; Keel et al. 1985), but more recent studies with large or/and well-matched control samples show no strong correlations (e.g., Schmitt 2001; Laurikainen & Salo 1995). Furthermore, in a study of 69 galaxies belonging to 31 Hickson compact groups (CGs), Shimada et al. (2000) find no difference between the frequency of HII or AGN nuclei in Hickson CGs and field galaxies.

At intermediate redshifts ($z \sim 0.4$–1.3), studies based on GOODS data (Grogin et al. 2004; see Subsect. 2.2) report no significant difference between the rest-frame B asymmetry index of AGN hosts and inactive galaxies. This suggests that over this redshift range, AGN hosts are not preferentially major mergers. Similarly, in the GEMS sample of moderately bright ($M_{\mathrm{B}} \simeq -23$) AGN at $0.5 < z < 1.1$ studied by Sanchez et al. (2004; see Sect. 4), only 3/15 of the AGN hosts signs of strong disturbances, and at most three others are associated with nearby potentially interacting companions.

In fact, statistically significant correlations between AGN activity and signs of strong interactions are reported only in systems with high mass accretion rates ($\geq 10\ \mathrm{M}_{\odot}\ \mathrm{yr}^{-1}$) such as very luminous or radio-loud QSOs (Disney at al. 1995; Bahcall et al. 1997; Kirhakos et al. 1999), and FR-II radio galaxies (Hutchings 1987, Yates et al. 1989). I discuss below possible reasons as to why a correlation is only seen at the very high luminosity end:

- It is possible that the large mass accretion rates ($\geq 10\ \mathrm{M}_{\odot}\ \mathrm{yr}^{-1}$) on small scales required by very high luminosity QSOs are realized in nature primarily during violent processes such as some classes of major/ intermediate mass-ratio interactions. In particular, the last phases of a major merger produce rapidly varying gravitational torques and strong shocks on intersecting orbits (see Subsect. 5.1) that are expected to generate large inflow rates on small scales.

- Conversely, in order to fuel Seyferts over their nominal duty cycles (10^8 years) with a mass inflow rates ≤ 0.01 M$_\odot$ yr^{-1}), any one of the many 10^6 M$_\odot$ clouds commonly present within the inner 200 pc radius is adequate, provided that some fueling mechanism drains its angular momentum by more than 99.99% (Fig. 2). Given that the required fueling mass is 10^{-3}–10^{-2} of the amount of gas typically found in the inner few 100 pc, it is well possible that *localized* low-energy processes (e.g., SNe shocks on ambient clouds) may be adequate for reducing its angular momentum. In other words, interactions and other fueling mechanisms which affect the *bulk* of the gas may not be necessary for fueling Seyferts.

6 Large-Scale Bars in Starbursts/AGN Hosts

6.1 Bar-Driven Gas Inflow: Theory and Observations

Bars are ubiquitous in most (>70%) local spiral galaxies (Grosbøl et al. 2002; Eskridge et al. 2002), and recent work (Jogee et al. 2004a,b; Sheth et al. 03) suggest they may be quite abundant out to $z \sim 1$. Bars drive the dynamical evolution of disk galaxies by exerting gravitational torques which redistribute mass and angular momentum. In fact, large gas inflows into the inner few hundred pc of a disk galaxy primarily result from gravitational torques exerted by a stellar bar. This is true not only in the case of an isolated barred galaxy, but also for some classes of minor mergers (Subsect. 5.2; Quinn et al. 1993; Hernquist & Mihos 1995; Mihos et al. 1995), most intermediate 1:3 mass ratio mergers (Naab & Burkert 2001), and the early phases of most major mergers (Subsect. 5.1; Noguchi 1988, Sellwood 1988; Hernquist 1989; Heller & Shlosman 1994; Mihos & Hernquist 1996). I will, therefore, devote a fair share of this review to a discussion of the basic principles of bar-driven gas inflow, what it can achieve in the context of fueling starbursts and AGN, and related observations.

A barred potential is made up of different families of periodic stellar orbits which conserve the Jacobi integral (E_J), a combination of energy and angular momentum (e.g., Binney & Tremaine 1987). The most important families are those oriented parallel to the bar major axis (x_1 orbits; Contopoulos & Papayannopoulos 1980) and minor axis (x_2 orbits). The x_1 family is the main family supporting the bar and extends between the center and the corotation resonance (CR) of the bar. The inner Lindblad resonances demarcate the transition region where the dominant family of periodic stellar orbits changes from x_1 to x_2. Thus, the x_2 family appears between the center and the ILR if a single ILR exists, and between the inner ILR (IILR) and the outer ILR (OILR) if two ILRs exist. Gas tries to follow these orbits, but due to its collisional and dissipative nature, it cannot remain on periodic orbits which cross. Instead, the gas-populated orbits change their orientation only

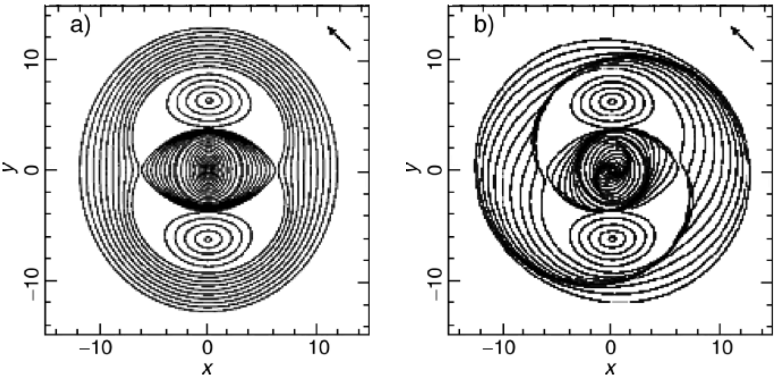

Fig. 4. Stars and gas in a barred potential: *Left:* The periodic *stellar* orbits in the potential of a bar (oriented horizontally here) change orientation by $\pi/2$ at each resonance. The x_1 family of periodic stellar orbits is elongated along the bar major axis and supports the bar. The x_2 orbits are elongated along the bar minor axis and and exist inside the ILR or between the ILRs. *Right:* In contrast, gas being collisional and dissipative, follows orbits which change their orientation only gradually, leading to spiral-shaped gas streamlines. (Figures are adapted from Buta & Combes 1996)

gradually due to shocks induced by the finite gas pressure, leading to spiral-shaped gas streamlines, offset with respect to the stars (Fig. 4). The negative torque exerted by the stars between CR and the OILR causes the gas to lose angular momentum and flow towards the inner kpc.

There is mounting observational evidence for bar-driven gas inflow into the inner kpc. Moderate ($4''$) resolution CO($J = 1$–0) interferometric surveys of molecular gas in the circumnuclear region of nearby spirals show that the molecular gas central concentration (Fig. 5) is on average higher in barred than in unbarred galaxies (Sakamoto et al. 1999). Barred galaxies also show shallower metallicity gradients across their galactic disks than unbarred ones (Martin & Roy 1994; Vila-Costas & Edmunds 1992). Observations of cold or ionized gas velocity fields show evidence for shocks and non-circular motions along the large-scale stellar bar (Quillen et al. 1995; Benedict, Smith, & Kenney 1996; Regan, Vogel, & Teuben 1997; Jogee 1999; Jogee, Scoville, & Kenney 2004c). Bar-driven gas inflow rates into the inner kpc have been estimated only in the case of a few strong bars and range from 1 to 4 $M_\odot$ yr^{-1} (Quillen et al. 1995; Regan et al. 1997; Laine, Shlosman, & Heller 1998).

What happens once gas reaches the central kpc region of barred galaxies? High resolution (1–$2''$; 100–200 pc) observations of molecular gas and SF in the central few kpc of eleven barred galaxies (Jogee et al. 2004c; Fig. 6) reveal a variety of gas distributions and a large range in SFR per unit mass of molecular gas (SFR/$M_{\rm H2}$). These differences in gas distributions and SF efficiencies can be partly understood in terms of different stages of bar-driven

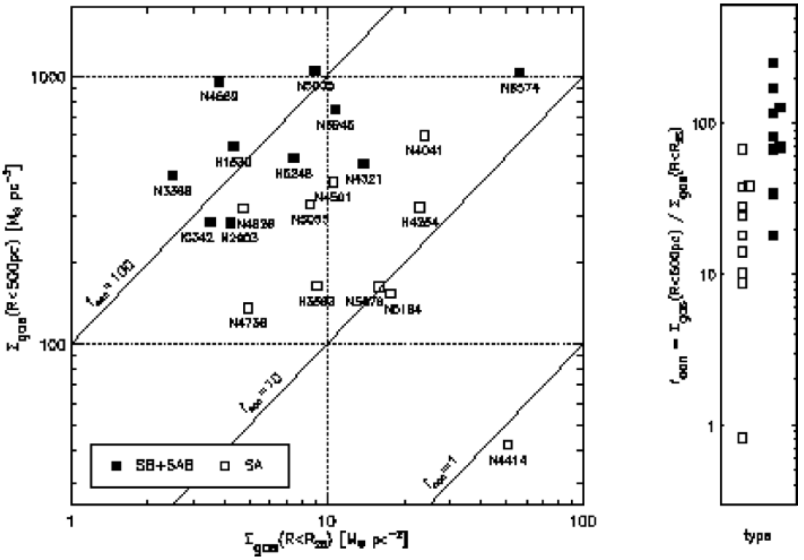

Fig. 5. Molecular gas central concentrations in barred galaxies: *Left:* Comparison of the molecular gas surface densities averaged within the central 500 pc radius and over the optical de Vaucouleurs radius (R_{25}). The spatial resolution of the data is $\sim$ 4″. The ratio of these two quantities is a measure of the molecular gas central concentration ($f_{\rm con}$) within the central kpc. *Right:* On average the molecular gas central concentration ($f_{\rm con}$) within the central kpc is higher in barred galaxies (filled squares) than in unbarred galaxies (open squares). (Figure is adapted from Sakamoto et al. 1999)

inflow and the existence of a critical density for the onset of SF (Jogee et al. 2004c). Some systems appear in the early stages of bar-driven inflow where a large fraction of the gas is still extended along the large-scale bar, shows large non-circular kinematics and does not form stars efficiently. Other galaxies have developed large gas surface densities (600–3500 M$_\odot$ pc^{-2}; Fig. 7) inside the OILR of the bar. Intense starbursts appear to be triggered only once gas densities approach or exceed the Toomre critical density for the onset of gravitational instabilities (Elmegreen 1994; Jogee et al. 2004c).

While the large-scale bar efficiently drives gas from the outer disk into the inner kpc, theory suggests that the radial inflow of gas slows as it crosses the ILR because shocks associated with the large-scale bar weaken, the gravitational potential becomes more axisymmetric, and gravitational torques on the gas in the vicinity of ILRs weaken or even reverse (e.g., Schwarz 1984; Combes & Gerin 1985; Shlosman et al. 1989; Athanassoula 1992). In support of this picture, gas rings near the ILRs have also been reported in many individual galaxies (e.g., Kenney et al. 1992; Knapen et al. 1995; Jogee 1999; Jogee et al. 2001). Figure 8 illustrates the results from high resolution observations of gas in the centers of barred galaxies (Jogee et al. 2004c). In the

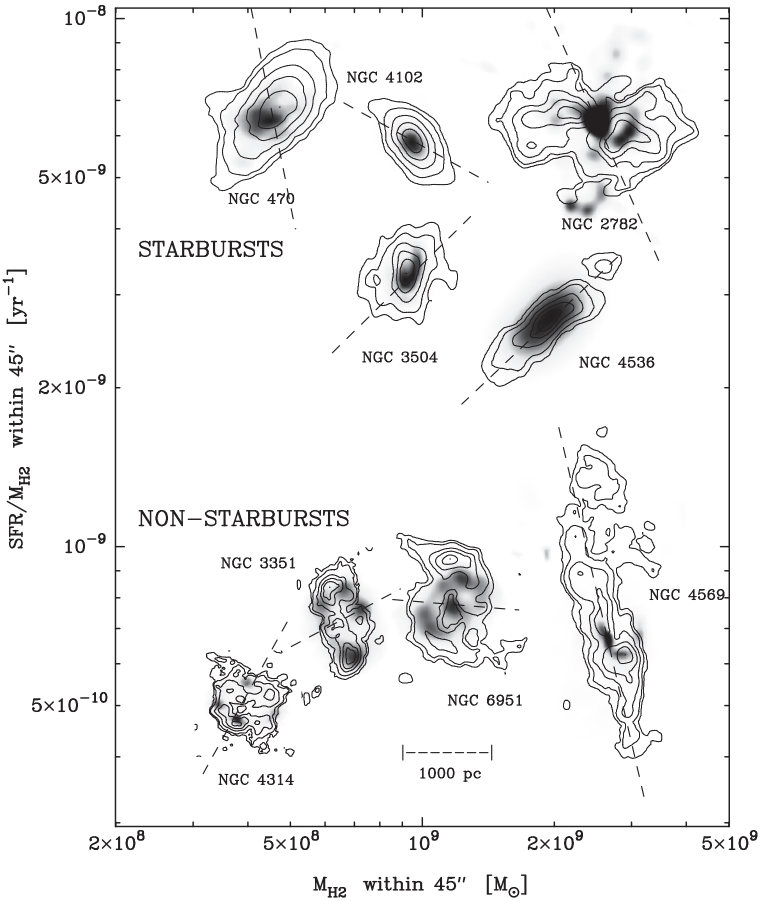

Fig. 6. High resolution (2″ or 100–200 pc) observations of molecular gas and star formation in the inner few kpc of barred galaxies: CO (1–0) intensity (contours) maps of the circumnuclear region of barred galaxies are overlaid on 1.5 or 4.9 GHz radio continuum map or Hα map (greyscale). The dotted line is the P.A. of the large-scale stellar bar/oval. A wide range in gas distributions and in SFR per unit mass of molecular gas (SFR/$M_{\rm H2}$) is present. Systems with high/low circumnuclear SFR/$M_{\rm H2}$ are denoted as starbursts/non-starbursts. Part of the range in SFR/$M_{\rm H2}$ and the variety of gas distributions can be understood in terms of different stages of bar-driven inflow and the critical density for the onset of SF. (From Jogee et al. 2004c; see text for details)

seven barred galaxies shown, typically the OILR radius is $\geq$ 500 pc while the IILR radius is $\leq$ 200–300 pc. It should be noted that even with these high resolution (100 pc) data, it remains difficult to assess the presence/response of gas *inside* the IILR since the latter radius is comparable to the spatial resolution.

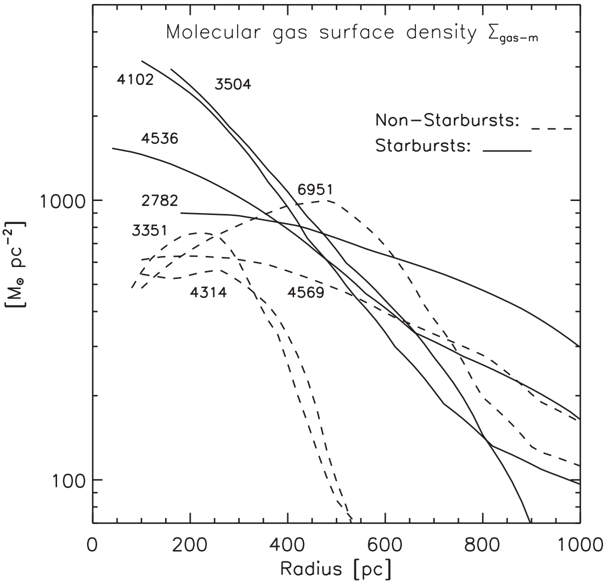

Fig. 7. Molecular gas surface densities in the inner kpc of barred galaxies: The azimuthally averaged molecular gas surface density derived from the high resolution (2″ or 100–200 pc) CO (J = 1–0) maps in Fig. 6 are plotted. Quantities are plotted starting at a radius ≥ half the size of the synthesized beam. Most of the starbursts have developed larger molecular gas surface densities (1000–3500 $M_\odot$ pc^{-2}) in the inner 500 pc radius than the non-starbursts for a given CO-to-H_2 conversion factor. (From Jogee et al. 2004c)

The question of whether bars are long-lived or whether they dissolve and reform recurrently over a Hubble time is highly controversial at the present time. Early studies (Hasan & Norman 1990; Friedli & Benz 1993; Norman, Sellwood, & Hasan 1996, but see also Bournaud & Combes 2002; Shen & Sellwood 2004) proposed that once a large central mass concentration (CMC) builds up in the inner 100 pc of a galaxy, it will destroy or weaken the bar due to the development of chaotic orbits and reduction of bar-supporting orbits. In some scenarios, bars even cause their own demise and self-destruction by building up such CMCs via gas inflows. The disk left behind after a bar is destroyed is dynamically hot and does not reform a new bar unless it is cooled significantly. Recently, Bournaud & Combes (2002) suggested that bars are destroyed primarily due to the reciprocal torques of gas on the stars in the bar (rather than by the CMC *per se*). Within this framework, bars can dissolve and reform recurrently if the galaxy accretes sufficient cold gas over a Hubble time. In fact, Shen & Sellwood (2004) find that in purely stellar N-body simulations, the bar is quite robust to CMCs. Ongoing studies (Jogee et al. 2004a,b) of the bar properties and CMCs of galaxies over lookback

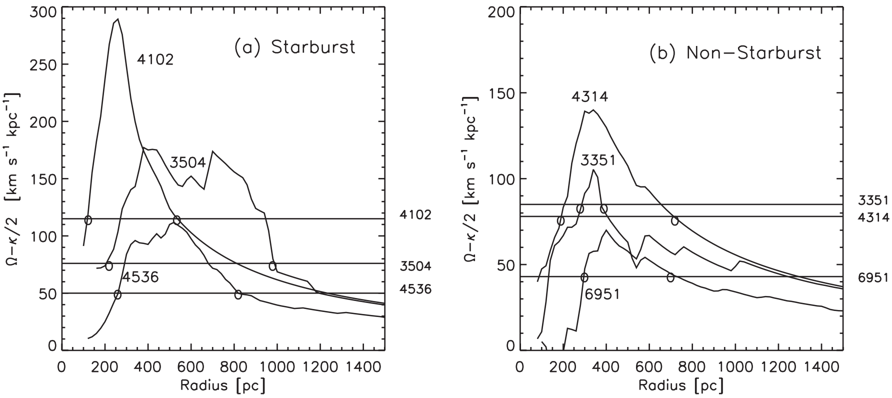

Fig. 8. Location of Inner Lindblad resonances in barred galaxies: $[\Omega - \kappa/2]$ is plotted against radius for the circumnuclear region of the barred galaxies shown in Fig. 6. Under the epicyclic approximation valid for weak bars, the intersection of $[\Omega - \kappa/2]$ with Ω_{p} defines the locations of the ILRs of the large-scale stellar bar. The upper limit on the bar pattern speed (Ω_{p}) is drawn as a horizontal line and is estimated by assuming that the corotation resonance is at or beyond the bar end. Values range from 43 to 110 km s^{-1} kpc^{-1}. Typically, the OILR radius is $\geq$ 500 pc while the IILR radius is $\leq$200–300 pc. Thus, in these systems, the large molecular gas densities (Fig. 5) have developed inside the OILR of the bar/oval. (From Jogee et al. 2004c)

times of 9 Gyr (out to $z \sim 1$), based on the GEMS survey, will help provide discriminant tests on the evolution and lifetime of bars out to $z \sim 1$.

6.2 Correlations Between Large-Scale Bars and Starbursts

Analyses of the extended 12 μm Galaxy sample (E12GS) show that the fraction of large-scale bars is larger in starburst nuclei (82–85%) than in normal (61–68%) ones (Fig. 9; Hunt & Malkan 1999). The term "normal" here denotes quiescent nuclei which do not have HII/starburst, LINER, Sy 1, or Sy 2 signatures. Similar correlations between large-scale bars and circumnuclear starbursts were previously reported by studies with smaller or/and less complete samples. For instance, Hawarden et al. (1986) found that a large fraction of SB and SAB galaxies show an enhanced IRAS 25 μm flux, which they assigned to circumnuclear rings of SF present in the central 20″. Earlier work (e.g., Terlevich et al. 1987; Kennicutt et al. 1987) also reported correlations based on UV and optical starbursts.

All the afore-described studies rely on optical images to identify bars and therefore suffer from the associated caveats outlined in Subsect. 3.6. Taken at face value, they suggest that relatively luminous starbursts have a higher frequency of large-scale stellar bars than normal galaxies. A natural explanation exists for the bar–starburst correlation. The most powerful starbursts

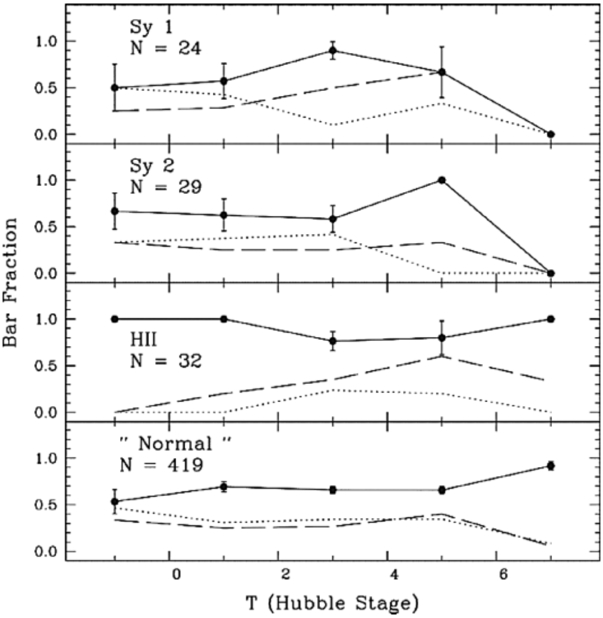

Fig. 9. Relation between large-scale bars and nuclear starburst/AGN: The fraction of barred galaxies as a function of Hubble type and nuclear types is shown for the extended 12 μm Galaxy Sample. Nuclear types (Hii, Sy 1, Sy 2, LINER) are taken from NED. Quiescent nuclei without any of these signatures are denoted as "normal". Optically-based RC3 bar classes are shown as dotted lines (unbarred), dashed lines (weakly barred SAB), and solid lines (weakly and strongly barred SAB+SB). The data bins are as in Fig. 3. Numbers under the panel label give the number of galaxies in each subsample. (Figure is from Hunt & Malkan 1999)

typically occur in the central few 100 pc of galaxies once large supercritical gas densities (often as large as several 1000 $M_\odot$ pc^{-2}; Jogee et al. 2004c) build up. A spontaneously or tidally induced large-scale bar is an ideal fueling mechanism for luminous circumnuclear starbursts because it efficiently drains angular momentum from gas on exactly the right spatial scales (several kpc to a few hundred pc; Fig. 2) relevant for building the pre-requisite large concentrations.

6.3 Correlations Between Large-Scale Bars and AGN

The study of Hunt & Malkan (1999) based on the E12GS sample and RC3 optical bar classes finds that there is no excess of bars in Seyferts. Two recent NIR-based studies (Mulchaey & Regan 1997; Knapen et al. 2000) have investigated the fraction f of large-scale bars in Seyferts and normal galaxies using matched control samples, high resolution NIR images, and ellipse fits to characterize bars. Mulchaey & Regan (1997) report a similar incidence of bars ($f \sim 70\%$) in Seyferts and normal galaxies while Knapen et al. (2000) find a higher fraction of bars in Seyferts (79% ± 8% vs. 59% ± 9%) at a significance level of 2σ. More recently, Laurikainen, Salo, & Buta (2004) classified bars in the Ohio State University sample (Eskridge et al. 2002) using Fourier decomposition of NIR images (Fourier bars), and report a higher fraction (72%) of such Fourier bars in Seyfert galaxies, LINERs, and HII/starburst galaxies, as compared to 55% in the inactive galaxies. It is not entirely clear at this time how the Fourier bars identified in this study compare to bars identified by other methods such as ellipse fits, and how they are impacted by other $m = 2$ modes (e.g., spirals). Furthermore, the nuclear types (HII, Seyfert, and LINERs) in the OSU sample are not homogeneously classified via spectroscopic observations. Thus, it is fair to conclude that at this time, the question of whether Seyferts have an excess of large-scale bars compared to inactive galaxies remains open.

I discuss below what one might expect with regards to correlations between Seyferts and large-scale bars, based on theoretical considerations, and outline some areas where future work is needed.

1. The first question one might ask is whether all barred galaxies are expected to show AGN activity. It is clear from Fig. 2 and preceding discussions that a large-scale bar efficiently drives gas only down to scales of a few 100 pc. At that stage, the specific angular momentum L of the gas is still 1000 times too high to be digestible by a BH. Thus, unless other mechanisms are present to reduce L further, the gas will not fuel the central BH. Furthermore, even if a barred system does go through an AGN phase, the lifetime of a bar is expected to be at least 1 Gyr, while a typical AGN duty cycle is 10–100 times shorter. *Thus, we do not expect all barred galaxies to show AGN activity.*
2. A different question is whether we would expect all local Seyfert host galaxies to have a bar. This question is tantamount to asking whether we need to transport gas from the outer regions (several kpc) to the inner few 100 pc of a galaxy in order to *directly or indirectly* fuel Seyferts. Let us first consider the required mass budget. The mass of gas required to fuel a typical Seyfert at a rate of 10^{-2} $M_\odot$ yr^{-1} over a nominal duty cycle of 10^8 years is *only* 10^6 $M_\odot$ *or* $10^{-3} - 10^{-2}$ *of the typical gas content* (10^8–10^9 $M_\odot$) found in the inner kpc of a present-day spiral. One might argue, therefore, that we would not expect a strong correlation between Seyfert activity and strong/moderate bars because even weak/inefficient large-scale fueling mechanisms are more than adequate to drive such tiny amounts of gas from kpc scales down to the inner few 100 pc. Examples of such weak fueling mechanisms would be oval features which are not conventionally classified as bars, weak non-axisymmetries easily induced in minor mergers/interactions, and dynamical friction slowly sinking a gas-rich satellite (see Subsect. 5.1). It has in fact been argued that the

strong excess of rings seen in Seyferts (see point 4) represent such weal oval perturbations. In summary, from the point of view of *large-scale gas transport,*one would not expect strong/moderate large-scale bars to be a pre-requisite for fueling Seyferts.
However, we should bear in mind that a correlation between Seyfert and strong/moderate large-scale bars might *indirectly* result due to requirements for gas transport on *small* scales (100s of pc) rather than large (kpc) scales. Even if only 0.1%–1% of the gas present on scales of a few 100 pc provides an adequate mass budget for AGN activity over many duty cycles, the fueling of the BH can only occur if there exist mechanisms on scales of a few 100 pc which can drain the angular momentum of this gas by more than 99.99%. If some of these mechanisms, such as dynamically decoupled secondary nuclear bars, are favored by the presence of a moderate/strong large-scale bar, a correlation between the latter and Seyferts might result. However, one may also counter-argue that with only 0.1%–1% of the circumnuclear gas present being involved in AGN fueling, strong nuclear fueling mechanisms such as nuclear bars may not be needed. Instead, localized low-energy processes such as SNe shocks and cloud-cloud collisions may be enough to significantly reduce the angular momentum on one ambient 10^6 $M_\odot$ cloud and drive it from 100s of pc down to tens of pc.

3. The uncertain question of whether bars can self-destroy and reform recurrently over a Hubble time (see Subsect. 6.1) adds another pertinent dimension to the interpretation of statistical correlations or lack thereof between bars and Seyferts. In some models of bar destruction (e.g., Hasan & Norman 1990; Friedli & Benz 1993; Norman, Sellwood, & Hasan 1996), a bar which dutifully brings a large gas concentration into the inner kpc might dissolve away once other mechanisms in the inner kpc start to relay the fuel to 100 pc scales and eventually to the BH. Some observational support for this picture comes from studies reporting that the bar strength is weaker in Seyferts than in inactive galaxies (Shlosman et al. 2000; Laurikainen et al. 2004).
4. A strong correlation is seen between Seyferts and the presence of large-scale rings in the host galaxies (Hunt & Malkan 1999): the frequency of outer rings and of (inner +outer) rings is higher by 3–4 times in Seyfert galaxies compared to normal galaxies. It has been argued that this correlation may result by chance because both Seyferts and rings (particularly outer rings) tend to be more common in early-type systems. Another possibility is that these oval rings represent the type of weak non-axisymmetric distortions discussed in point 2. A third possibility is that a large fraction of these rings might be remnants of bar dissolution. Both latter possibilities need to be explored further theoretically and observationally.
5. To date, all studies between AGN and large-scale bars have focused on local galaxies. Yet, AGN activity is known to increase with redshift, with

the optically bright QSO phases peaking at $z \sim 2.5$ (Subsect. 3.4). It is important to extend studies of bars and AGN to earlier epochs. The ongoing work on the impact and evolution of bars over the last 9 Gyr (out to $z \sim 1.3$) based on the GEMS *HST* survey and Chandra Deep Field South data (Jogee et al. 2004a,b) will help constrain how bars relate to AGN activity at these epochs.

7 Nuclear Bars

7.1 Nuclear Bars: Theory and Observations

A large-scale bar efficiently drives gas from the outer disk down to scales of a few hundred pc where the gas inflow stalls after crossing an ILR. Theory and simulations (Shlosman et al. 1989; Friedli & Martinet 1993; Heller & Shlosman 1994; Maciejewski & Sparke 2000) have suggested that a nuclear bar (so called "secondary" bar) nested within the large-scale bar (so called "primary" bar) can break the status quo and gravitationally torque digestible fuel closer to the galactic center. Several scenarios exists for the formation and evolution of nuclear bars. The nuclear bar can decouple from the primary bar such that its pattern speed (Ω_{n}) is higher (Shlosman et al. 1989; Friedli & Martinet 1993; Heller & Shlosman 1994) or lower (Heller, Shlosman & Englmaier 2001) than the primary pattern (Ω_{p}), depending on whether it forms via a self-gravitational or non-self-gravitational instability. The nuclear bar can also be in a coupled phase with ($\Omega_{\mathrm{n}} = \Omega_{\mathrm{p}}$), while in the case of a merger remnant it may even counter-rotate with respect to the primary bar. In simulations, the decoupled phase with $\Omega_{\mathrm{n}} > \Omega_{\mathrm{p}}$ is particularly effective in removing angular momentum from the gas and in helping to fuel the BH.

Most observational studies on nuclear bars have focused on the *morphological* properties of bars determined from optical and IR images (Wozniak et al. 1995; Friedli et al. 1996; Mulchaey & Regan 1997; Jogee, Kenney, & Smith 1999; Laine et al. 2002; Erwin & Sparke 2002; Erwin 2004). Studies using space-based and ground-based IR and optical images reveal that about 20%–30% of S0-Sc galaxies host double bars and 20%–40% of barred galaxies host a second bar (e.g., Laine et al. 2002; Erwin & Sparke 2002). Confirmed morphologically-identified double bars exist primarily in galaxies of Hubble type no later than Sbc. They have ellipticities of 0.27–0.6, semi-major axes of 200–1600 pc, and position angles which both lead and trail the primary bar by varying amounts. The latter fact suggests that a significant fraction of nuclear bars must be dynamically decoupled with respect to the primary bar.

NGC 2782 provides a nice observational case of a nuclear bar driving gas down to 100 pc scales. It hosts a nuclear stellar bar which is associated with $\sim 2.5 \times 10^9$ $\mathrm{M}_{\odot}$ of molecular gas and appears to be channeling gas into the central 100 pc where an M82-class powerful central starburst resides (Fig. 10;

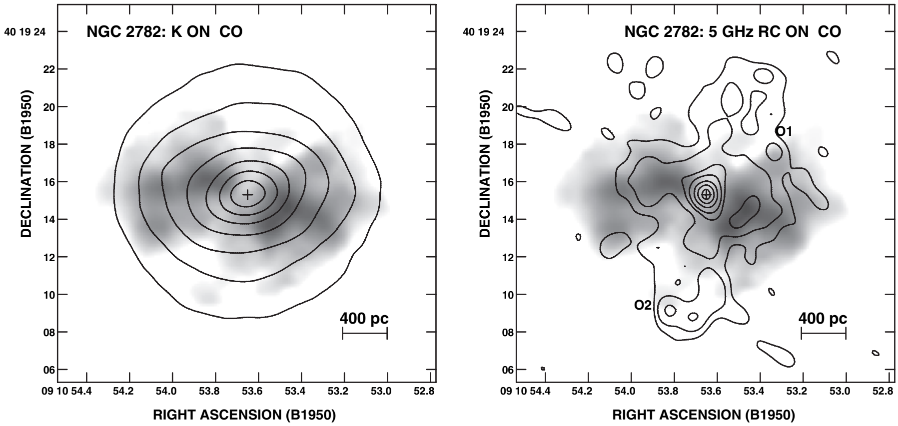

Fig. 10. A nuclear stellar bar feeding gas into a powerful starburst within the inner 100 pc radius of NGC 2782: *Left:* K-band contours are overlaid on the ($2.1'' \times 1.5''$) CO intensity map (greyscale). A nuclear stellar bar (identified via isophotal fitting of the K-band image) is present at a PA of ∼ 100 °and is itself nested within a large-scale oval/bar which is visible in a larger I-band image. The cold gas traced in CO has a bar-like distribution which leads the nuclear stellar bar, and its velocity field (not shown here) reveals weak bar-like streaming motions. *Right:* 5 GHz radio continuum (contours) are overlaid on the CO map (greyscale). The nuclear stellar bar appears to be feeding molecular gas into an intense starburst which peaks in RC within the inner 100 pc radius and has a SFR of 3–6 $M_\odot$ yr^{-1}. The starburst in turn appears to be driving an outflow associated with the northern and southern RC bubbles. (From Jogee, Kenney, & Smith 1999)

Jogee, Kenney, & Smith 1999). The nuclear stellar bar is identified via isophotal fits to K-band image showing a characteristic plateau in position angle (PA) as the ellipticity rises to a maximum. The molecular gas distribution is bar-like, leads the nuclear stellar bar, and shows weak bar-like streaming motions (Jogee et al. 1999). NGC 2782 may be witnessing the early decoupling phase of a nuclear stellar bar induced by gas inside the OILR of the large-scale bar, which itself appears to be dissolving.

Aside from NGC 2782, only a handful of other nuclear bar candidates have high resolution CO maps. These show a wide range of CO morphologies in double-barred galaxies. The double-barred galaxy NGC 5728 shows CO clumps in a very disordered configuration (Petitpas & Wilson 2003). In NGC 4314, the CO emission forms a clumpy ring at the end of the nuclear bar (Benedict et al. 1996; Jogee et al. 2004c). The differences in CO distributions in these nuclear bars may be linked to evolutionary differences (Jogee et al. 2004c) or differences in gas properties (Petitpas & Wilson 2003).

7.2 Correlations Between Nuclear Bars and AGN

Multiple investigations have searched for a nuclear bar–AGN correlation. Statistics from earlier studies which used offset dust lanes to identify nuclear bars (e.g., Regan & Mulchaey 1999; Martini & Pogge 1999) must be re-evaluated because simulations (Maciejewski et al. 2002; Shlosman & Heller 2002) show that the gas flow in nested nuclear bars differs fundamentally from that along large-scale bars and does not lead to large-scale *offset* shocks and dust lanes. More recent studies based on isophotal fits to NIR images report that the fraction of nuclear bars is similar (20%–30%) in Seyfert and non-Seyfert hosts (Laine et al. 2002; Erwin & Sparke 2002). There are several possible explanations for the lack of statistical correlations observed between nuclear bars and AGN activity.

1. Nuclear bars help to solve the angular momentum problem one step further than large-scale bars, but the gas still has to lose several orders of magnitude in L even at pc scales (see Fig. 2).
2. Not all morphologically-identified nuclear bars are expected to be equally effective in removing angular momentum from the gas. Theoretically, the most effective ones are those with $\Omega_{\rm n} > \Omega_{\rm p}$. However, to date limited observations exist on *kinematic properties* and pattern speeds of nuclear bars, and this is an area where much progress has yet to be made.
3. Once large mass concentrations build in the core of the nuclear bar, it can cause first the nuclear bar, then the large-scale primary bar, to dissolve. The lifetime of secondary nuclear bars is not precisely determined, but is expected to be short and $\sim$ a few bar rotation timescales.
4. As discussed in Subsect. 6.3, only 10^6 $\mathrm{M}_\odot$ or 0.1%–1% of the gas present on scales of a few 100 pc can adequately fuel a Seyfert over a nominal duty cycle of 10^8 years at a rate of 10^{-2} $\mathrm{M}_\odot$ yr^{-1}. Thus, in lieu of strong fueling mechanisms such as nuclear bars, localized processes such as SNe shocks and cloud-cloud collisions may be adequate for driving one ambient 10^6 $\mathrm{M}_\odot$ cloud from 100s of pc down to 10s of pc (see Subsect. 6.3 point 2).

8 Nuclear Spirals and AGN/Starburst Activity

There is mounting evidence from high resolution ground-based and *HST* observations of galaxies that nuclear dust spirals are common on scales of a few tens to hundreds of pc (e.g., Elmegreen et al. 1998; Carollo, Stiavelli, & Mack 1998; Laine et al. 1999; Regan & Mulchaey 1999; Martini & Pogge 1999 ; Jogee et al. 2002; Martini et al. 2003). A variety of nuclear spirals are present, including flocculent, chaotic, and two-armed grand-design spirals. Nuclear dust spirals are likely to trace shocks since their arm-to-interarm contrast (Martini & Pogge 1999; Elmegreen, Elmegreen & Eberwein 2002; Laine et al. 1999) implies a mass density enhancement ≥ 2. Since shocks dissipate orbital

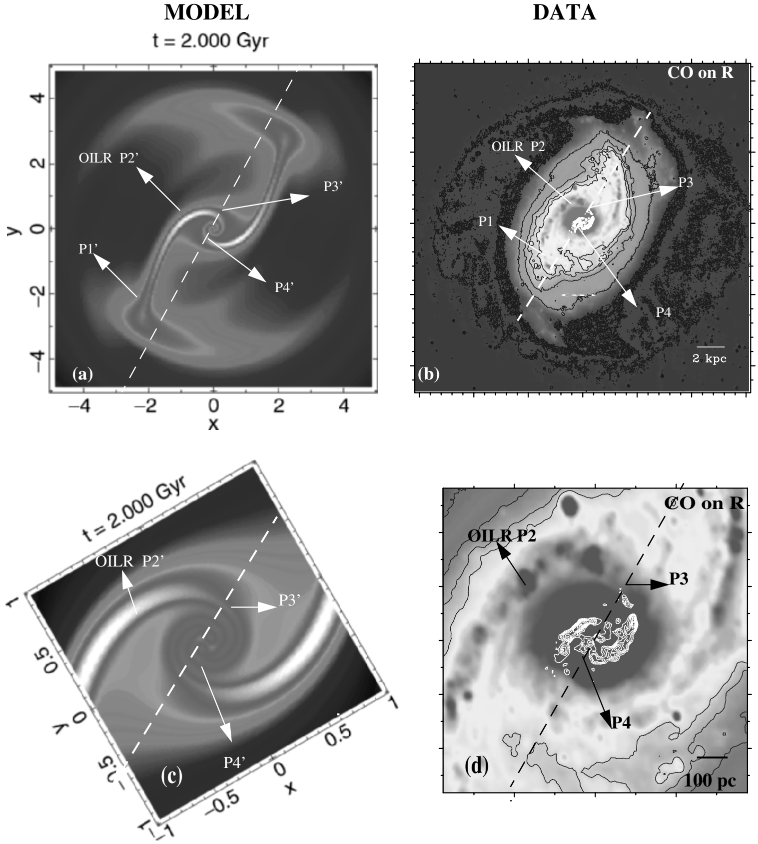

Fig. 11. Bar-driven gaseous spiral density waves (SDWs) forming a grand-design nuclear dust spiral well inside the OILR of NGC 5248: Comparison of the CO and *R*-band data (*right*) with hydrodynamical models (*left*; from Englmaier & Shlosman 2000) of a bar-driven gaseous SDW. Points P1' to P4' in the models correspond to points P1 to P4 in the data. The *top* panel shows the entire galaxy, and the *bottom* panel the central 40″ (3.0 kpc) only. In NGC 5248, continuous spirals in CO and dust extend inwards from 6 kpc, cross the OILR (P2) of the bar, continue inwards (P3, P4), and connect to nuclear dust spirals which extend from 300 pc to 70 pc. Comparison of the data with the hydrodynamical models suggest that the continuous connected spiral structure is generated by a gaseous SDW which is driven by the large-scale bar and penetrates very deep inside the OILR due to the low central mass concentration (From Jogee et al. 2002b)

energy and lead to the outward transfer of angular momentum, some authors have suggested that nuclear dust spirals may play some role in the fueling of AGN.

How do nuclear dust spirals form? Most of them are believed to be in non-self-gravitating central gaseous disks based on the lack of significant SF

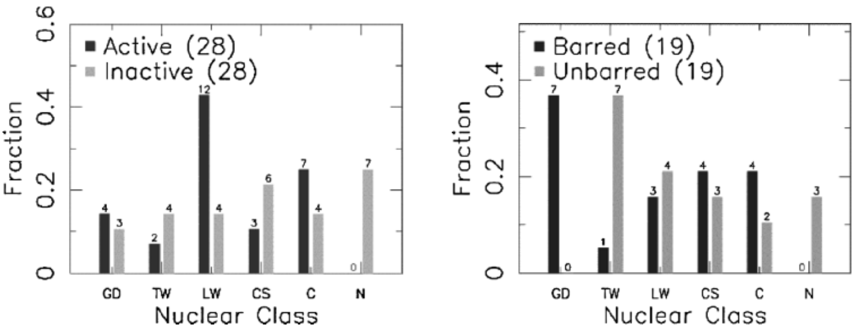

Fig. 12. Nuclear dust spirals in AGN hosts and barred galaxies – *Left:* Distribution of matched samples of 28 active and 28 inactive galaxies. The nuclear dust morphology is divided into six classes denoted as GD: grand-design nuclear dust spiral, TW: tightly wound nuclear spiral, LW: loosely wound nuclear spiral, CS: chaotic spiral, C: chaotic dust structure, N: no dust structure present. The total number of galaxies in each class is shown in parentheses and is used to normalize the histogram bars. Nuclear spirals occur with comparable frequency in active and inactive nuclei. *Right:* Same as above, but for the matched sample of 19 barred and 19 unbarred galaxies. Grand-design nuclear spirals are found only in barred galaxies. (Adapted from Martini et al. 2003)

observed in the spirals. This belief is corroborated in a few cases by rough estimates of the neutral gas mass density $\Sigma(\mathrm{H})$ and Toomre Q values associated with the nuclear dust spirals (Martini & Pogge 1999). In such a non-self-gravitating gas disk, nuclear spirals may form in several ways. **(1)** Hydrodynamical simulations (Englmaier & Shlosman 2000) suggest that bar-driven shocks which exist near the ILRs can trigger gaseous SDWs which propagate inwards across the ILRs, and lead to grand-design two-armed nuclear dust spirals. Observational support for the model comes from detailed comparisons of the simulations with the data in NGC 5248 (Jogee et al. 2002b; Fig. 11) which hosts a grand-design nuclear spiral. **(2)** Repeated compression from acoustic turbulence followed by shearing can lead to flocculent nuclear dust spirals (Elmegreen, Elmegreen, & Eberwein 2002; Montenegro et al. 1999). **(3)** Simulations including gas dynamical effects (rotation, shear, shocks), thermal cooling effects, and feedback from SF (e.g., Wada & Norman 2001) lead to a multi-phase ISM where chaotic and flocculent spirals recur.

Is there a correlation between nuclear spirals and AGN activity? A survey based on NIR *HST* images and color maps of 123 nearby galaxies by Martini & Pogge (2003) reports that nuclear dust spirals occur with comparable frequency in both active and inactive galaxies (Fig. 12). This suggests that the low level inflow rate on scales of 10s of pc which might be triggered by shocks in the nuclear spirals is not a universal mechanism for feeding even low level AGN activity. The study also reports that grand-design nuclear dust spirals are found only in galaxies with a large-scale bar, consistent with afore-mentioned idea that these features form via bar-driven gaseous SDWs (Fig. 11).

9 From Hundred pc to Sub-pc Scales

It is far from understood how gas is driven from scales of 100 pc down to sub-pc scales where few direct observational constraints exist. At 100 pc, matter still needs to reduce *its specific angular momentum L by a factor of more than 100* before it can reach sub-pc scales, and eventually the last stable radius of a massive BH (see Fig. 2). I discuss a few mechanisms without attempting to do an exhaustive review:

- *Dynamical friction and feedback from SF:* We already discussed dynamical friction in the context of minor mergers where satellite galaxies located at tens of kpc sink towards the disk of the parent galaxy (Subsect. 5.2) while inducing non-axisymmetric instabilities in this disk. Since dynamical friction operates on a timescale which is $\propto$ (R^2 v/M lnΛ), it becomes increasingly important for massive gas clumps located at small radii. For instance, Jogee et al. (1999) estimate that $t_{df} \sim 5 \times 10^7$ for a $M \sim 10^8$ $M_\odot$ gas clump at $R \sim$ 700 pc using the approximation below:

$$t_{\rm df} = 7 \times 10^7 \text{ yr} \left(\frac{M}{10^8 M_\odot}\right)^{-1} \left(\frac{V_{\rm c}}{300 \ km \ s^{-1}}\right) \left(\frac{R}{700 \text{ pc}}\right)^2 \qquad (6)$$

  Feedback effects from powerful circumnuclear starbursts of several $M_\odot$ yr^{-1} which are common in the central few hundred pc of galaxies (e.g., Jogee et al. 2004c) may also contribute towards removing angular momentum in localized clouds.
- *Tidal Disruption of Gas Clumps:* Numerical investigations by Bekki (2000) suggest that once clumps get down to tens of pc (for instance via dynamical friction), the tidal gravitational field of the MBH transforms the clump into a moderately thick gaseous disk or torus. A few percent of the gas mass (corresponding to a few $\times 10^5$ $M_\odot$) can be subsequently transferred from this gaseous disk to the central sub-parsec region around the MBH within a few $\times 10^6$ yr via viscous torques.
- *Runaway self-gravitational instabilities:* It has been suggested that in a gas-rich nuclear disk, self-gravitational instabilities on repeatedly small scales could lead to three or more bars which are nested within each other (e.g., Shlosman et al. 1990). However, while the presence of three nested bars or triaxial features have been reported in a handful of active galaxies (e.g., Friedli et al. 1996; Laine et al. 2002), the current limited spatial resolution of NIR surveys and molecular gas surveys does not yet allow systematic observational tests of this scenario.

- *Stellar mass loss and disruption of stellar clusters:* As pointed out by Ho et al. (1997b), nominal stellar mass loss rates of $\sim 10^{-5}$–10^{-6} $M_\odot$ yr^{-1} in the central regions of galaxies with luminosity densities of 10^3–10^4 $M_\odot$ pc^{-3} may provide enough fuel for low luminosity AGN. However, while the mass loss rates may be adequate, we yet have to identify mechanisms which can reduce the specific angular momentum of the gas down to 10^{24} cm^2 s^{-1} (see Fig. 2). In the same vein, the tidal disruption of stellar clusters (Emsellem & Combes 1997; Bekki 2000) passing by a central MBH in the accretion disk has been invoked as a source of fuel. However, it still remains to be investigated how exactly the angular momentum of such clusters would be drained and whether this is facilitated by clusters on radial orbits.
- *On pc scales:* On pc ands sub-pc scales viscous torques (e.g., Shlosman et al. 1989; Pringle 1996), warping induced in an accretion disk due to radiation pressure from a central source (Pringle 1996), and hydromagnetic outflows in AGN (Emmering, Blandford, & Shlosman 1992) have been invoked. In the latter model, it is postulated that a hydromagnetic wind containing dense molecular clouds is accelerated radiatively and centrifugally away from an accretion disk, removing angular momentum from the disk, and forming the broad emission lines seen in AGN (Emmering et al. 1992).

10 Summary and Future Perspectives

Faced with the fallible task of providing an "objective" summary of this review, I can only forewarn the reader with these perennial words:

> *To command the professors of astronomy to confute their own observations is to enjoin an impossibility, for it is to command them to not see what they do see, and not to understand what they do understand, and to find what they do not discover. – Galileo Galilei*

1. **Symbiotic evolution of BHs and bulges:** The mass of a central BH appears to be tightly correlated with the stellar velocity dispersion of the bulge of the host galaxy. SMBHs with a wide range of masses (10^6–10^{10} $M_\odot$) follow the same $M_{\rm bh}$–σ relation, although they reside in a wide array of host galaxies including quiescent early-type (E/S0 Sb–Sc) galaxies, local Seyferts, and luminous QSOs out to $z \sim 3$. Numerous variants of the $M_{\rm bh}$–σ relation have by now been proposed, including tight correlations between $M_{\rm bh}$ and quantities such as the bulge luminosity, the light concentration of galaxies, the Sersic index, and the mass of the DM halo. It thus appears that active and quiescent BHs bear a common relationship to the surrounding triaxial component of their host galaxies over a wide range of cosmic epochs and BH masses. This points towards a symbiotic evolution of BHs and the central triaxial features of their hosts.

2. **Census and growth epoch of BHs:** A census of the BH mass density at different epochs suggests that accretion with a standard radiative efficiency during the quasar era ($z = 0.2$–5) can readily account for the BH mass density (few $\times\ 10^5\ \mathrm{M}_\odot\ \mathrm{Mpc}^{-3}$) found in local ($z < 0.1$) early-type galaxies. Furthermore, only a small fraction of this local BH mass density is in the form of active Seyfert galaxies, and in the latter systems, the inferred mass accretion rates at the BH are typically 10^3 times lower than in QSOs. Taken together, the evidence suggests that there is no significant growth of BHs at the present epoch compared to the quasar era. Therefore, in the context of AGN fueling, we should bear in mind that the dominant fueling mechanisms for luminous QSOs out to $z \sim 2.5$ may be markedly different from those impacting local Seyferts. For instance, tidal interactions and mergers are likely to be more important in activating AGN activity at early epochs than in the present-day.
3. **The angular momentum problem:** One of the most important challenges in fueling AGN is arguably the angular momentum problem. The specific angular momentum (L) of matter located at a radius of a few kpc must be reduced by more than 10^5 before it is fit for consumption by a BH (Fig. 2). The angular momentum barrier is a more daunting challenge than the amount of gas *per se*. For instance, while there may be ample material (e.g., $10^6\ \mathrm{M}_\odot$ clouds) in the inner 200 pc radius to fuel typical Seyferts over nominal duty cycles (10^8 years), the challenge is to understand what fueling mechanisms can drain their angular momentum by 99.99% so that they are digestible by a BH.
4. **Fueling mechanisms on different scales:** There is no universal fueling mechanism which operates efficiently on all spatial scales, from several kpc all the way down to the last stable orbit of a BH. Instead, different fueling mechanisms such as gravitational torques (e.g., from large-scale bars and nuclear bars), dynamical friction (on massive circumnuclear gas clumps), hydrodynamical torques (shocks), and viscous torques may relay each other at different radii in terms of their effectiveness in draining angular momentum. According to simulations, large-scale bars (whether spontaneously or tidally induced) are the most efficient mechanisms for driving large gas inflows from several kpc down to the inner few hundred pc. This holds not only in the case of an isolated barred galaxy, but also for some classes of minor (1:10) mergers, most intermediate (1:3) mass ratio mergers, and the early phases of most major (1:1) mergers. During the late stages of a major merger, strongly-varying gravitational torques and strong shocks on crossing orbits can subsequently drive the circumnuclear gas further in at large rates ($\gg 1\ \mathrm{M}_\odot\ \mathrm{yr}^{-1}$).
5. **Correlations between AGN and interactions:** Statistically significant correlations between morphological signs of interactions and AGN are seen in systems with high mass accretion rates ($\dot{M} \geq 10\ \mathrm{M}_\odot\ \mathrm{yr}^{-1}$) such as very luminous or radio-loud QSOs. The presence of a correlation

only at very high mass accretion rates holds to reason because such accretion rates are primarily realized in nature during violent processes such as major/intermediate mass-ratio interactions. However, it must be noted that the reverse does not hold true: not *all* major interactions lead to extreme activity because their effectiveness in inducing large gas inflows depend on many merger parameters (speeds, energies, spin-orbital angular momenta alignments).

For moderate luminosity QSOs and typical Seyferts, no clear correlation between activity and interactions is seen. The lack of a correlation may be due to the fact that the low $\dot{M}$ required in Seyferts and lower luminosity QSOs can be provided not only by strong fueling mechanisms such as interactions which impact the bulk of the gas, but also by localized low-energetic processes which impact only 10^{-3}–10^{-2} of the circumnuclear gas (see Subsect. 5.4).

6. **Correlations between AGN and host galaxy properties:**
   - Local AGN (Sy 1 and Sy 2) and moderate luminosity ($M_{\mathrm{B}} \simeq -23$) AGN in the redshift range $z = 0.4$–1.1 tend to reside primarily in early-type (bulge-dominated) galaxies.
   - AGN early-type hosts show enhanced blue global rest-frame optical colors compared to early-type inactive galaxies. This holds in low redshift ($z < 0.2$) SDSS studies, as well as in intermediate redshifts ($z \sim 0.5$–1.1) samples. These colors are consistent with the presence of young stellar populations, suggesting that the onset of AGN activity is intimately linked to the recent onset of global SF in the hosts.
   - The frequency of large-scale stellar bars is significant higher in starburst galaxies than normal galaxies. The bar–starburst correlation is consistent with the idea that a bar efficiently drains angular momentum of gas on exactly the right spatial scales (several kpc to a few hundred pc; Fig. 2) relevant for building the pre-requisite large gas concentrations for circumnuclear starbursts. At this time, the question of whether Seyferts have an excess of large-scale bars compared to inactive galaxies remains open. Studies based on high resolution NIR images, different samples, and different bar identification methods (ellipse fits, Fourier methods) yield different conflicting results. The reader is referred to Subsect. 6.3 for a detailed discussion.
   - Seyferts appear to have *weaker* bar strengths compared to inactive galaxies. These results may lend observational support to the long-claimed (but little-tested) idea that a large build-up of mass concentration via gas inflow into the inner 100 pc can weaken or even destroy the large-scale bar. However, the question of whether bars are long-lived or whether they dissolve and reform recurrently over a Hubble time is also highly controversial at the moment (Subsect. 6.1).
   - The frequency of outer rings and of (inner + outer) rings appears to be higher by a factor of 3-4 in Seyfert galaxies compared to normal

galaxies. Various interpretations of this strong correlation have been proposed. The rings may be weak non-axisymmetric oval distortions or they be remnants of dissolved bars. Alternatively, the correlation may be the result of both Seyferts and rings existing preferentially in early-type systems.

- The frequency of both nuclear stellar bars (identified morphologically) and nuclear dust spirals is found to be similar in Seyferts and normal galaxies. This lack of correlation suggests that these features are not universal mechanisms for fueling an AGN or/and that their lifetime is short ($\leq$ few $\times 10^8$ years).

7. **Future perspectives:** I outline below a few of the many exciting developments in AGN research we might wish/expect in the upcoming decade. *(a)* Now that we have a relatively solid census of BH mass density from $z \sim 0.1$–5 (Sect. 4), we need to systematically map the molecular gas content and structural properties of AGN hosts as a function of redshift in order to investigate why AGN activity has declined sharply since $z \sim 2.5$. NIR integral field spectroscopy on 8m-class telescopes, inflowing 24 µm *Spitzer* data, and the advent of sub-millimeter facilities like the 50-m LMT (*circa* 2007) and ALMA (*circa* 2010) will provide key constraints. *(b)* To date, all studies between AGN/starbursts and large-scale bars have focused on local galaxies. Yet, both the cosmic SFR density and AGN activity increase out to t $z \sim 1$. The ongoing work on the impact of bars and interactions over the last 9 Gyr (out to $z \sim 1.3$) on central starbursts and AGN based on the GEMS *HST* survey and Chandra Deep Field South data (Jogee et al. 2004a,b) will help constrain how bars and external triggers relate to the activities and structural evolution of galaxies at these epochs. *(c)* There is a dire lack of high resolution interferometric observations of molecular gas for a *statistically significant sample* of AGN, even locally Three ongoing surveys are alleviating this problem. The Molecular gas in Active and Inactive Nuclei (MAIN; Jogee, Baker, Sakamoto, & Scoville 2001) high resolution, multi-line survey covers forty (Seyfert, LINER, and HII) nuclei. MAIN aims at constraining the drivers of activity levels in galactic nuclei, and complements the ongoing Nuclei of GAlaxies (NUGA) survey (Garcia-Burillo et al. 2003) and the multiple line survey of Seyferts (Kohno et al. 2001); *d)* The advent of large (30–100 m) diffraction-limited telescopes such as the Giant Magellan Telescope will help test/extend the $M_{\rm bh}$–σ relation for late-type spirals, IMBHs, and low surface brightness ellipticals.

Acknowledgements

For comments and discussions, I thank numerous colleagues, in particular, L. Ferrarese, R. van der Marel, F. Combes, S. Laine, the GEMS collaboration, I. Shlosman, N. Grogin, J. H. Knapen, & L. Hernquist. I acknowledge support from the National Aeronautics and Space Administration (NASA) under LTSA Grant NAG5-13063 issued through the Office of Space Science.

References

1. Adams, F.C., Graff, D.S., & Richstone, D.O. 2001, ApJL, 551, L31
2. Athanassoula, E. 1992, MNRAS, 259, 328
3. Bahcall, J.N., Kirhakos, S., Saxe, D.H., & Schneider, D.P. 1997, ApJ, 479, 642
4. Baker, A.J. 2000, Ph.D. thesis, California Institute of Technology
5. Balick, B., & Heckman, T.M. 1982, ARAA, 20, 431
6. Barnes, J.E., & Hernquist, L.E. 1991, ApJL, 370, L65
7. Barnes, J.E., & Hernquist, L. 1996, ApJ, 471, 115
8. Bekki, K. 2000, ApJ, 545, 753
9. Baumgardt, H., Makino, J., Hut, P., McMillan, S., & Portegies Zwart, S. 2003, ApJL, 589, L25
10. Benedict, G.F., Smith, B.J., & Kenney, J.D.P. 1996, AJ, 112, 1318
11. Binney, J., & Tremaine, S. 1987, Princeton, NJ, Princeton University Press, 1987
12. Blandford, R.D., & McKee, C.F. 1982, ApJ, 255, 419
13. Blandford, R.D., & Begelman, M.C. 1999, MNRAS, 303, L1
14. Böker, T., van der Marel, R.P., & Vacca, W.D. 1999, AJ, 118, 831
15. Bournaud, F., & Combes, F. 2002, A&A, 392, 83
16. Bromm, V., & Loeb, A. 2003, ApJ, 596, 34
17. Burkert, A., & Silk, J. 2001, ApJL, 554, L151
18. Buta, R., & Combes, F. 1996, Fundamentals of Cosmic Physics, 17, 95
19. Carollo, C.M., Stiavelli, M., & Mack, J. 1998, AJ, 116, 68
20. Cid Fernandes, R., Heckman, T., Schmitt, H., Delgado, R.M.G., & Storchi-Bergmann, T. 2001, ApJ, 558, 81
21. Chokshi, A., & Turner, E.L. 1992, MNRAS, 259, 421
22. Colbert, E.J.M., & Miller, M.C. 2005 (astro-ph/0402677)
23. Combes, F., & Gerin, M. 1985, A&A, 150, 327 Combes 1994;
24. Combes, F. 2003, ASP Conf. Ser. 290: Active Galactic Nuclei: From Central Engine to Host Galaxy, eds. S. Collin, F. Combes, & I. Shlosman (ASP), 411
25. Contopoulos, G., & Papayannopoulos, T. 1980, A&A, 92, 33
26. Cowie, L.L., & Barger, A.J., Chap. 9, "Supermassive Black Holes in the Distant Universe", ed. A.J. Barger, Kluwer Academic Publishers, in press
27. Di Matteo, T., Croft, R.A.C., Springel, V., & Hernquist, L. 2003, ApJ, 593, 56
28. Filippenko, A.V., & Ho, L.C. 2003, ApJL, 588, L13
29. de Vaucouleurs, G., de Vaucouleurs, A., Corwin, H.G., Buta, R.J., Paturel, G., & Fouque, P. 1991, Volume 1-3, XII, Springer-Verlag
30. Dahari, O. 1984, AJ, 89, 966
31. Disney, M.J., et al. 1995, Nature, 376, 150
32. Dubinski, J. 1994, ApJ, 431, 61
33. Ebisuzaki, T., et al. 2001, ApJL, 562, L19
34. El-Zant, A.A., Shlosman, I., Begelman, M.C., & Frank, J. 2003, ApJ, 590, 641
35. Elmegreen, B.G. 1994, ApJ, 427, 384
36. Elmegreen, B.G., et al. 1998, ApJL, 503, L119

37. Elmegreen, D.M., Elmegreen, B.G., & Eberwein, K.S. 2002, ApJ, 564, 234
38. Emmering, R.T., Blandford, R.D., & Shlosman, I. 1992, ApJ, 385, 460
39. Englmaier, P., & Shlosman, I. 2000, ApJ, 528, 677
40. Erwin, P., & Sparke, L.S. 2002, AJ, 124, 65
41. Erwin, P. 2004, A&A, 415, 941
42. Eskridge, P.B., et al. 2002, ApJS, 143, 73
43. Fabian, A.C., & Iwasawa, K. 1999, MNRAS, 303, L34
44. Ferrarese, L. 2002, ApJ, 578, 90
45. Ferrarese, L. 2002, Current high-energy emission around black holes, 3
46. Ferrarese, L., Pogge, R.W., Peterson, B.M., Merritt, D., Wandel, A., & Joseph, C.L. 2001, ApJL, 555, L79
47. Ferrarese, L., & Merritt, D. 2000, ApJL, 539, L9
48. Ferrarese, L., & Ford, H.C. 1999, ApJ, 515, 583
49. Ferrarese, L., Ford, H.C., & Jaffe, W. 1996, ApJ, 470, 444
50. Friedli, D., & Martinet, L. 1993, A&A, 277, 27
51. Friedli, D., & Benz, W. 1993, A&A, 268, 65
52. Friedli, D., Wozniak, H., Rieke, M., Martinet, L., & Bratschi, P. 1996, A&A Suppl., 118, 461
53. García-Burillo, S., et al. 2003, A&A, 407, 485
54. Gebhardt, K., et al. 2000, ApJL, 539, L13
55. Gebhardt, K., Rich, R.M., & Ho, L.C. 2002, ApJL, 578, L41
56. Geha, M., Guhathakurta, P., & van der Marel, R.P. 2002, AJ, 124, 3073
57. Genzel, R., Pichon, C., Eckart, A. et al. 2000, MNRAS, 317, 348
58. Gerssen, J., van der Marel, R.P., Gebhardt, K., Guhathakurta, P., Peterson, R.C., & Pryor, C. 2003, AJ, 125, 376
59. Ghez, A.M., et al. 2003, ApJL, 586, L127
60. Giavalisco, M., et al. 2004, ApJL, 600, L93
61. Graham, A.W., Erwin, P., Caon, N., & Trujillo, I. 2001, ApJL, 563, L11
62. Grogin, N. et al. 2004, ApJ, submitted
63. Haehnelt, M.G., & Kauffmann, G. 2000, MNRAS, 318, L35
64. Häring, N., & Rix, H. 2004, ApJL, 604, L89
65. Harms, R.J., et al. 1994, ApJL, 435, L35
66. Hasan, H., & Norman, C. 1990, ApJ, 361, 69
67. Hawarden, T.G., et al. 1986, MNRAS, 221, 41
68. Heckman, T.M., Armus, L., & Miley, G.K. 1990, ApJS, 74, 833
69. Heller, C.H., & Shlosman, I. 1994, ApJ, 424, 84
70. Heller, C., Shlosman, I., & Englmaier, P. 2001, ApJ, 553, 661
71. Hernquist, L. 1989, Nature, 340, 687
72. Hernquist, L., & Mihos, J.C. 1995, ApJ, 448, 41
73. Ho, L.C., Filippenko, A.V., & Sargent, W.L.W. 1997a, ApJS, 112, 315
74. Ho, L.C., Filippenko, A.V., & Sargent, W.L.W. 1997b, ApJ, 487, 591
75. Hummel, E., van der Hulst, J.M., Kennicutt, R.C., & Keel, W.C. 1990, A&A, 236, 333
76. Hunt, L.K., & Malkan, M.A. 1999, ApJ, 516, 660
77. Hutchings, J.B., Frenette, D., Hanisch, R., Mo, J., Dumont, P.J., Redding, D.C., & Neff, S.G. 2002, AJ, 123, 2936
78. Hutchings, J.B. 1987, ApJ, 320, 122
79. Jahnke, K. et al. 2004, ApJ, in press
80. Jogee, S., Kenney, J.D.P., & Smith, B.J. 1998, ApJL, 494, L185

81. Jogee, S., Kenney, J.D.P., & Smith, B.J. 1999, ApJ, 526, 665
82. Jogee, S. 1999, Ph.D. Thesis, Yale University
83. Jogee, S., Baker, A.J., Sakamoto, K., Scoville, N.Z., & Kenney, J.D.P. 2001, ASP Conf. Ser. 249: The Central Kiloparsec of Starbursts and AGN: The La Palma Connection, 612
84. Jogee, S., Knapen, J.H., Laine, S., Shlosman, I., Scoville, N.Z., & Englmaier, P. 2002a, ApJL, 570, L55
85. Jogee, S., Shlosman, I., Laine, S., Englmaier, P., Knapen, J.H., Scoville, N.Z., & Wilson, C.D. 2002b, ApJ, 575, 156
86. Jogee, S., et al. 2003, IAU Symposium, 216
87. Jogee, S., Rix, H.-W., Barazza, F., Heyer, I., Davies, J., Shlosman, I., Lubell, G., et al. 2004a, in Penetrating Bars through Masks of Cosmic Dust eds. D. Block, K. Freeman, R. Groess, I. Puerari, & E.K. Block (Dordrecht: Kluwer), in press.
88. Jogee, S., Rix, H.-W., Barazza, F., Heyer, I., Davies, J., Shlosman, I., Lubell, G., et al. 2004b, ApJL, submitted
89. Jogee, S., Scoville, N.Z., & Kenney, J. 2004c, ApJ, in press
90. Kaspi, S., Smith, P.S., Netzer, H., Maoz, D., Jannuzi, B.T., & Giveon, U. 2000, ApJ, 533, 631
91. Kauffmann, G., et al. 2003, MNRAS, 346, 1055
92. Keel, W.C., Kennicutt, R.C., Hummel, E., & van der Hulst, J.M. 1985, AJ, 90, 708
93. Kenney, J.D.P., Wilson, C.D., Scoville, N.Z., Devereux, N.A., & Young, J.S. 1992, ApJL, 395, L79
94. Kennicutt, R.C., Roettiger, K.A., Keel, W.C., van der Hulst, J.M., & Hummel, E. 1987, AJ, 93, 1011
95. Kirhakos, S., Bahcall, J.N., Schneider, D.P., & Kristian, J. 1999, ApJ, 520, 67
96. Knapen, J.H., Beckman, J.E., Heller, C.H., Shlosman, I., & de Jong, R.S. 1995, ApJ, 454, 623
97. Knapen, J.H., Shlosman, I., & Peletier, R.F. 2000, ApJ, 529, 93
98. Knapen, J.H. 2004, ASP Conf. Ser.: The Neutral ISM in Starburst Galaxies, eds. S. Aalto, S. Hüttemeister, & A. Pedlar, in press (astro-ph/0312172)
99. Kohno, K., et al. 2001, ASP Conf. Ser. 249: The Central Kiloparsec of Starbursts and AGN: The La Palma Connection, 672
100. Kormendy, J., & Richstone, D. 1995, ARAA, 33, 581
101. Laine, S., Shlosman, I., & Heller, C.H. 1998, MNRAS, 297, 1052
102. Laine, S., Knapen, J.H., Perez-Ramirez, D., Doyon, R., & Nadeau, D. 1999, MNRAS, 302, L33
103. Laine, S., Shlosman, I., Knapen, J.H., & Peletier, R.F. 2002, ApJ, 567, 97
104. Laurikainen, E., & Salo, H. 1995, A&A, 293, 683
105. Laurikainen, E., Salo, H., & Rautiainen, P. 2002, MNRAS, 331, 880
106. Laurikainen, E., Salo, H., & Buta, R. 2004, ApJ, 607, 103
107. Lehnert, M.D., Heckman, T.M., Chambers, K.C., & Miley, G.K. 1992, ApJ, 393, 68
108. Lynden-Bell, D. 1969, Nature, 223, 690
109. Maciejewski, W., & Sparke, L.S. 2000, MNRAS, 313, 745
110. Maciejewski, W., Teuben, P.J., Sparke, L.S., & Stone, J.M. 2002, MNRAS, 329, 502
111. Martin, P., & Roy, J. 1994, ApJ, 424, 599

112. Martini, P., & Pogge, R.W. 1999, AJ, 118, 2646
113. Martini, P., Regan, M.W., Mulchaey, J.S., & Pogge, R.W. 2003, ApJ, 589, 774
114. Martini, P. 2004, in Proc. IAU 222: The Interplay among Black Holes, Stars and ISM in Galactic Nuclei, eds. Th. Storchi Bergmann, L.C. Ho, & H.R. Schmitt, in press
115. Merritt, D., & Ferrarese, L. 2001, MNRAS, 320, L30
116. Mihos, J.C., & Hernquist, L. 1996, ApJ, 464, 641
117. Mihos, J.C., Walker, I.R., Hernquist, L., Mendes de Oliveira, C., & Bolte, M. 1995, ApJL, 447, L87
118. Miyoshi, M., et al. 1995, Nature, 373, 127
119. Mobasher, B., Jogee, S., Dahlen, T., de Mello, D., Lucas, R.A., Conselice, C.J., Grogin, N.A., & Livio, M. 2004, ApJL, 600, L143
120. Mulchaey, J.S., & Regan, M.W. 1997, ApJL, 482, L135
121. Naab, T., & Burkert, A. 2001, ASP Conf.Ser. 230: Galaxy Disks and Disk Galaxies, 451
122. Narayan, R., Igumenshchev, I.V., & Abramowicz, M.A. 2000, ApJ, 539, 798
123. Negroponte, J., & White, S.D.M. 1983, MNRAS, 205, 1009
124. Noguchi, M. 1988, A&A, 203, 259
125. Norman, C., & Scoville, N. 1988, ApJ, 332, 124
126. Norman, C.A., Sellwood, J.A., & Hasan, H. 1996, ApJ, 462, 114
127. Ostriker, J.P., & Tremaine, S.D. 1975, ApJL, 202, L113
128. Padovani, P., Burg, R., & Edelson, R.A. 1990, ApJ, 353, 438
129. Petitpas, G.R., & Wilson, C.D. 2003, ApJ, 587, 649
130. Peterson, B.M. 1993, PASP, 105, 247
131. Pringle, J.E. 1996, MNRAS, 281, 357
132. Quillen, A.C., Frogel, J.A., Kenney, J.D.P., Pogge, R.W., & Depoy, D.L. 1995, ApJ, 441, 549
133. Quinn, P.J., Hernquist, L., & Fullagar, D.P. 1993, ApJ, 403, 74
134. Regan, M.W., Vogel, S.N., & Teuben, P.J. 1997, ApJL, 482, L143
135. Regan, M.W., & Mulchaey, J.S. 1999, AJ, 117, 2676
136. Richstone, D., et al. 1998, Nature, 395, A14
137. Rix, H., et al. 2004, ApJS, 152, 163
138. Rees, M.J. 1984, ARAA, 22, 471
139. Roberts, M.S., & Haynes, M.P. 1994, ARAA, 32, 115
140. Sakamoto, K., Okumura, S.K., Ishizuki, S., & Scoville, N.Z. 1999, ApJ, 525, 691
141. Sanchez, S. et al. 2004,ApJ, in press
142. Sanders, D.B., & Mirabel, I.F. 1996, ARAA, 34, 749
143. Schmitt, H.R. 2001, AJ, 122, 2243
144. Schwarz, M.P. 1984, MNRAS, 209, 93
145. Shen, J., & Sellwood, J.A. 2004, ApJ, 604, 614
146. Sheth, K., Regan, M.W., Scoville, N.Z., & Strubbe, L.E. 2003, ApJL, 592, L13
147. Shields, G.A., Gebhardt, K., Salviander, S., Wills, B.J., Xie, B., Brotherton, M.S., Yuan, J., & Dietrich, M. 2003, ApJ, 583, 124
148. Shimada, M., Ohyama, Y., Nishiura, S., Murayama, T., & Taniguchi, Y. 2000, AJ, 119, 2664
149. Shlosman, I., Frank, J., & Begelman, M.C. 1989, Nature, 338, 45
150. Shlosman, I., & Begelman, M.C. 1989, ApJ, 341, 685
151. Shlosman, I., Begelman, M.C., & Frank, J. 1990, Nature, 345, 679

152. Shlosman, I., Peletier, R.F., & Knapen, J.H. 2000, ApJL, 535, L83
153. Shlosman, I., & Heller, C.H. 2002, ApJ, 565, 921
154. Shlosman, I. 2003, ASP Conf. Ser. 290: Active Galactic Nuclei: From Central Engine to Host Galaxy, eds. S. Collin, F. Combes, & I. Shlosman (Astronomical Society of the Pacific), 427
155. Schödel, R., Ott, T., Genzel, R., Eckart, A., Mouawad, N., & Alexander, T. 2003, ApJ, 596, 1015
156. Soltan, A. 1982, MNRAS, 200, 11
157. Storchi-Bergmann, T., et al. 2001, ApJ, 559, 147
158. Struck, C. 1997, ApJS, 113, 269
159. Terlevich, R., Melnick, J., & Moles, M. 1987, IAU Symp. 121: Observational Evidence of Activity in Galaxies, 121, 499
160. Tremaine, S., et al. 2002, ApJ, 574, 740
161. Veilleux, S., Kim, D.-C., & Sanders, D.B. 2001, QSO Hosts and Their Environments, 165
162. van der Marel, R.P., & van den Bosch, F.C. 1998, AJ, 116, 2220
163. van der Marel, R.P. 2003, in Carnegie Observatories Astrophysics Series, Vol. 1, Coevolution of Black Holes and Galaxies, ed. L.C. Ho (Cambridge University Press)
164. Wada, K., & Norman, C.A. 2001, ApJ, 547, 172
165. Walker, I.R., Mihos, J.C., & Hernquist, L. 1996, ApJ, 460, 121
166. Wandel, A., Peterson, B.M., & Malkan, M.A. 1999, ApJ, 526, 579
167. Wozniak, H., Friedli, D., Martinet, L., Martin, P., & Bratschi, P. 1995, A&A Suppl., 111, 115
168. Wyithe, J.S.B., & Loeb, A. 2003, ApJ, 595, 614
169. Yates, M.G., Miller, L., & Peacock, J.A. 1989, MNRAS, 240, 129
170. Yu, Q., & Tremaine, S. 2002, MNRAS, 335, 965

X-Ray Survey Results on Active Galaxy Physics and Evolution

W.N. Brandt[1], D.M. Alexander[2], F.E. Bauer[2], and C. Vignali[3]

[1] Department of Astronomy & Astrophysics, The Pennsylvania State University,
525 Davey Lab, University Park, PA 16802, USA
`niel@astro.psu.edu`
[2] Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK
`dma@ast.cam.ac.uk, feb@ast.cam.ac.uk`
[3] INAF – Osservatorio Astronomico di Bologna, Via Ranzani 1,
40127 Bologna, Italy
`cristian@anastasia.bo.astro.it`

Abstract. This "pedagogical" review describes the key *Chandra* and *XMM-Newton* extragalactic surveys to date and details some of their implications for active galaxy physics and evolution. We additionally highlight two topics of current widespread interest: (1) X-ray constraints on the active galactic nucleus content of luminous submillimeter galaxies, and (2) the demography and physics of high-redshift ($z > 4$) active galaxies as revealed by X-ray observations. Finally, we discuss prospects for future X-ray surveys with *Chandra*, *XMM-Newton*, and upcoming missions.

1 Introduction

The cosmic X-ray background (XRB) is largely due to accretion onto supermassive black holes integrated over cosmic time. Thus, extragalactic X-ray surveys offer the potential to contribute substantially to our understanding of the physics of Active Galactic Nuclei (AGN) as well as the evolution of the AGN population. Such surveys have dramatically advanced over the past four years, largely due to the flood of data from the *Chandra X-ray Observatory* (hereafter *Chandra*) [56] and the *X-ray Multi-Mirror Mission-Newton* (hereafter *XMM-Newton*) [34]. The superb X-ray mirrors and charge-coupled device (CCD) detectors on these observatories provide

- Sensitive imaging spectroscopy in the ≈ 0.5–10 keV band, with up to 50–250 times (depending upon the energy band considered) the sensitivity of previous X-ray missions. X-ray surveys have finally reached the depths needed to complement the most sensitive surveys in the radio, submillimeter, infrared, and optical bands.
- X-ray source positions with accuracies of ≈ 0.3–$3''$. These high-quality positions are essential for matching to (often faint) multiwavelength counterparts.

W.N. Brandt et al.: *X-Ray Survey Results on Active Galaxy Physics and Evolution*, Lect. Notes Phys. **693**, 185–209 (2006)
DOI 10.1007/3-540-34621-X_7

- Large source samples allowing reliable statistical inferences to be drawn about the overall extragalactic X-ray source population. In a fairly deep *Chandra* or *XMM-Newton* observation, $\gtrsim$100–200 sources can be detected.

The extragalactic survey capabilities of *Chandra* and *XMM-Newton* are complementary in several important respects. The sub-arcsecond imaging of *Chandra* provides the best possible source positions, and with long exposures *Chandra* can achieve the highest possible sensitivity at energies of $\approx$ 0.5–6 keV; unlike the case for *XMM-Newton*, even the deepest *Chandra* observations performed to date do not suffer from significant source confusion. *XMM-Newton*, in comparison, has a substantially larger photon collecting area than *Chandra*, allowing efficient X-ray spectroscopy. In addition, *XMM-Newton* has better high-energy response than *Chandra* and can carry out the deepest possible surveys from $\approx$ 7–10 keV. Even *XMM-Newton*, however, does not cover the peak of the X-ray background at 20–40 keV. Finally, the field of view for *XMM-Newton* is $\sim$2.5 times that of *Chandra.*

Chandra and *XMM-Newton* have resolved $\gtrsim$80–90% of the 0.5–10 keV XRB into discrete sources, extending earlier heroic efforts with missions including *ROSAT*, *ASCA*, and *BeppoSAX*. The main uncertainties in the precise resolved fraction are due to field-to-field cosmic variance (which leads to spatial variation in the XRB flux density) and instrumental cross-calibration limitations. With the recent advances, attention is now focused on (1) understanding the nature of the X-ray sources in detail and their implications for AGN physics, and (2) understanding the cosmological evolution of the sources and their role in galaxy evolution. In this review, we briefly describe the key *Chandra* and *XMM-Newton* extragalactic surveys to date (Sect. 2) and detail some of their implications for AGN physics and evolution (Sect. 3). In Sect. 3 we highlight two topics of current widespread interest: (1) X-ray constraints on the AGN content of luminous submillimeter galaxies, and (2) the demography and physics of high-redshift ($z > 4$) AGN as revealed by X-ray observations. We also discuss prospects for future X-ray surveys with *Chandra*, *XMM-Newton*, and upcoming missions (Sect. 4).

Throughout this paper, we adopt $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_{\mathrm{M}} = 0.3$, and $\Omega_\Lambda = 0.7$ (flat cosmology).

2 Chandra and XMM-Newton Extragalactic Surveys

To learn about AGN physics and evolution in a complete manner, both “deep” and “wider” X-ray surveys are required; the trade-off between the two, of course, is between sensitivity and solid-angle coverage on the sky. None of the *Chandra* and *XMM-Newton* surveys discussed in this paper is

Table 1. Some Deep X-ray Surveys with *Chandra* and *XMM-Newton*

Survey Name	Exposure	Representative Reference or Note
	Chandra	
Chandra Deep Field-North	1950 ks	D.M. Alexander et al., 2003, AJ, 126, 539
Chandra Deep Field-South	940 ks	R. Giacconi et al., 2002, ApJS, 139, 369
HRC Lockman Hole	300 ks	PI: Murray
Extended CDF-S	250 ks	PI: Brandt
Groth-Westphal	200 ks	PI: Nandra
Lynx	185 ks	D. Stern et al., 2002, AJ, 123, 2223
LALA Cetus	177 ks	PI: Malhotra
LALA Boötes	172 ks	J.X. Wang et al., 2004, AJ, 127, 213
SSA13	101 ks	A.J. Barger et al., 2001, AJ, 121, 662
3C295	100 ks	V. D'Elia et al., 2004, astro-ph/0403401
Abell 370	94 ks	A.J. Barger et al., 2001, AJ, 122, 2177
SSA22 "protocluster"	78 ks	L.L. Cowie et al., 2002, ApJ, 566, L5
ELAIS	75 ks	J.C. Manners et al., 2003, MNRAS, 343, 293
WHDF	75 ks	PI: Shanks
	XMM-Newton	
Lockman Hole	766 ks	G. Hasinger et al., 2001, A&A, 365, L45
Chandra Deep Field-South	317 ks	A. Streblyanska et al., 2004, astro-ph/0309089
13 hr Field	200 ks	M.J. Page et al., 2003, AN, 324, 101
Chandra Deep Field-North	180 ks	T. Miyaji et al., 2003, AN, 324, 24
Subaru Deep	100 ks	PI: Watson
ELAIS S1	100 ks	PI: Fiore
Groth-Westphal	80 ks	T. Miyaji et al., 2004, astro-ph/0402617

The Extended *Chandra* Deep Field-South is comprised of four fields (each 250 ks), the *XMM-Newton* ELAIS S1 survey is comprised of four fields (each 100 ks), and the *Chandra* ELAIS survey is comprised of two fields (each 75 ks). The *XMM-Newton* Subaru Deep survey also has seven flanking fields (each $\approx$ 50 ks). Only the first $\approx$ 100 ks of the *XMM-Newton* Lockman Hole data have been published at present.

truly wide-field, in that the widest still only cover $\lesssim 1\%$ of the sky.[4] Both deep and wider X-ray surveys are reviewed briefly below.

2.1 Deep X-Ray Surveys

Table 1 makes it clear that deep *Chandra* and *XMM-Newton* surveys are a major "industry." The 21 surveys listed there have a total exposure exceeding 70 days, and $\gtrsim 50$ scientists have invested substantial effort on the analysis and interpretation of these data. Comparable effort has also been expended

---

[4] For this reason, we denote these surveys as "wider" (relative to the deep *Chandra* and *XMM-Newton* surveys) rather than "wide-field."

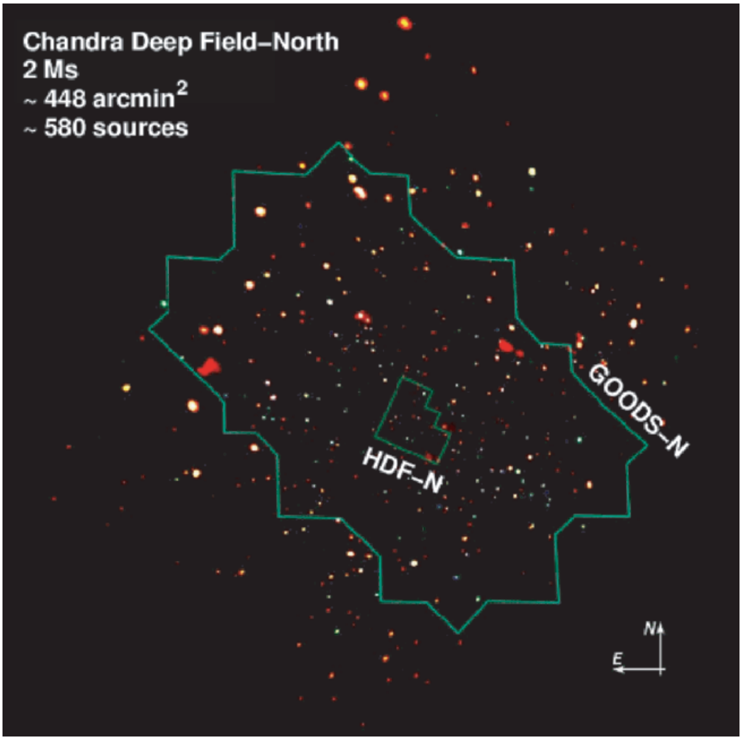

Fig. 1. Adaptively smoothed image of the 2 Ms CDF-N, constructed from data in the 0.5–2 keV (*red*), 2–4 keV (*green*), and 4–8 keV (*blue*) bands. Nearly 600 sources are detected in the ≈ 448 arcmin^2 field. The regions covered by the HDF-N and GOODS-N surveys are denoted. Adapted from [4]

on multiwavelength follow-up studies of these surveys; due to the small solid angles under investigation, superb multiwavelength coverage can be obtained relatively economically.

The two most sensitive surveys in Table 1, by a significant factor, are the 2 Ms *Chandra* Deep Field-North (CDF-N; see Fig. 1) [4] and 1 Ms *Chandra* Deep Field-South (CDF-S) [25]. Both are situated in intensively studied regions of sky with little Galactic foreground X-ray absorption. They reach 0.5–2 keV fluxes of ≈ $(2.5\text{–}5)\times10^{-17}$ erg cm^{-2} s^{-1}, corresponding to count rates of $\lesssim 1$ count every 2–4 days. At these flux levels, even moderate-luminosity AGN (similar to Seyfert galaxies in the local universe) can be detected to $z \gtrsim 10$. The CDF-N and CDF-S are clearly "pencil-beam" surveys, each covering ≈ 400 arcmin^2; for reference, this is ∼1/2 the solid angle of the full Moon and ∼75 times the solid angle of the original Hubble Deep Field-North (HDF-N; see Fig. 1). Public X-ray source catalogs are available for both the CDF-N and CDF-S (see the references in Table 1) [15]; these contain ≈ 580 and ≈ 370 sources, respectively.

Table 2. Some Wider X-ray Surveys with *Chandra* and *XMM-Newton*

Survey Name	Ω (deg^2)	Representative Reference or Note
		Chandra
ChaMP	14	D.W. Kim et al., 2004, ApJS, 150, 19
Clusters Serendipitous	1.1	P. Gandhi et al., 2004, MNRAS, 348, 529
CYDER	$\cdots$	F.J. Castander et al., 2003, AN, 324, 40
Lockman Hole NW	0.4	A.T. Steffen et al., 2003, ApJ, 596, L23
MUSYC	1	PI: van Dokkum
NOAO DWFS	9.3	PI: Jones
SEXSI	2.2	F.A. Harrison et al., 2003, ApJ, 596, 944
SWIRE Lockman	0.6	PI: Wilkes
1 hr Field	0.2	PI: McHardy
13 hr Field	0.2	I.M. McHardy et al., 2003, MNRAS, 342, 802
		XMM-Newton
AXIS	$\cdots$	X. Barcons et al., 2002, A&A, 382, 522
CFRS	0.6	T.J. Waskett et al., 2003, MNRAS, 341, 1217
HBS28	9.8	A. Caccianiga et al., 2004, A&A, 416, 901
HELLAS2XMM	2.9	A. Baldi et al., 2002, ApJ, 564, 190
LSS	64	M. Pierre et al., 2004, astro-ph/0305191
Survey Science Center	$\cdots$	M.G. Watson et al., 2001, A&A, 365, L51
VIMOS	2.3	PI: Hasinger
2dF	1.5	A. Georgakakis et al., 2003, MNRAS, 344, 161

The second column above lists estimated survey solid angles; survey sensitivities are not uniform but rather vary significantly across these solid angles. In some cases, survey solid angles are not well defined and thus are not listed. In these cases, the reader should consult the listed reference or note for further details.

The other deep X-ray surveys in Table 1 [26, 29, 41, 57], have generally been performed in regions of sky where (1) extensive coverage already exists at one-to-several wavelengths, and/or (2) some interesting astronomical object is present (e.g., 3C295, Abell 370, or the SSA22 "protocluster"). They are all sensitive enough to detect moderate-luminosity AGN to $z \sim 3$–5, and in total the surveys in Table 1 cover a solid angle of $\sim$3.5 deg^2 ($\sim$16 Moons).

2.2 Wider X-Ray Surveys

Wider *Chandra* and *XMM-Newton* surveys (see Table 2) are a comparably large and important "industry" to the deep surveys. These typically involve investigation of X-ray sources in $\sim$4–150 X-ray observations of moderate exposure (usually 20–60 ks, but sometimes as short as ≈ 5 ks); the observations are sometimes obtained from the public data archives. The wider surveys serve to bridge the observational "gap" between the deepest *Chandra* observations and the deepest observations made by previous X-ray missions

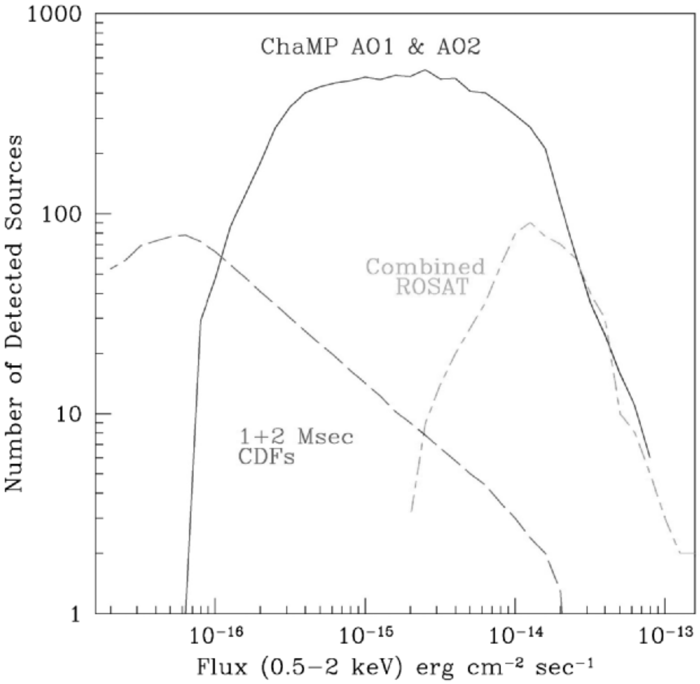

Fig. 2. Number of sources predicted from the ChaMP survey (for 137 ChaMP fields from *Chandra* Cycle 1 and Cycle 2) compared to numbers of sources from the *Chandra* Deep Fields and the *ROSAT* surveys analyzed by T. Miyaji, G. Hasinger, & M. Schmidt, 2000, A&A, 353, 25. An impressive ≈ 6000 ChaMP sources are expected in total, and these largely lie at intermediate 0.5–2 keV flux levels of $(4\text{–}60)\times 10^{-16}$ erg cm^{-2} s^{-1}. From [35]

(e.g., *ROSAT*; see Fig. 2), and they effectively target the intermediate 0.5–8 keV flux levels (10^{-15}–10^{-13} erg cm^{-2} s^{-1}) which contribute most significantly to the XRB.

The wider X-ray surveys provide a broad census of the X-ray source population, often generating enormous numbers of sources (1000–6000 or more; e.g., see Fig. 2). They thereby allow discovery of both intrinsically rare source types as well as low-redshift examples of sources found in the deep X-ray surveys. However, complete multiwavelength follow-up, e.g. [28], often must be compromised for reasons of observational economy; thus many of the wider surveys target specific source types of interest. Often targeted are sources with unusually hard X-ray spectra, sources with unusually large X-ray-to-optical flux ratios, or sources that appear likely to be at high redshift based upon optical imaging data.

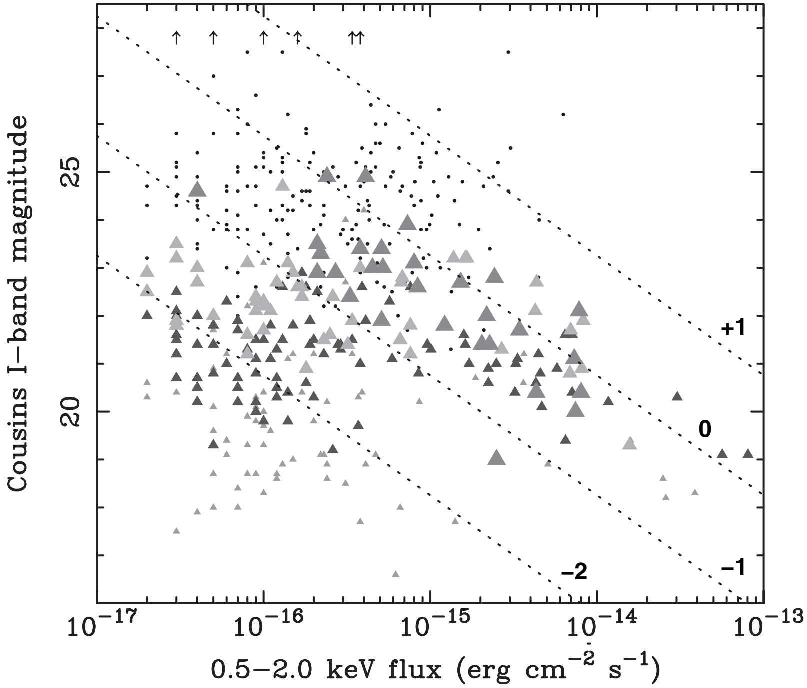

Fig. 3. *I*-band magnitude versus 0.5–2 keV flux for extragalactic X-ray sources in the CDF-N. Sources with redshifts of 0–0.5, 0.5–1, 1–2, and 2–6 are shown as *violet*, *blue*, *green*, and *red filled triangles*, respectively (symbol sizes also increase with redshift). *Small black dots* indicate sources without measured redshifts. The *slanted, dotted lines* indicate constant values of $\log(f_X/f_I)$; the respective $\log(f_X/f_I)$ values are labeled. Adapted from [4] and [8]

3 Some Implications for AGN Physics and Evolution

3.1 Properties of the X-Ray Sources

Basic Nature

A broad diversity of X-ray sources is found in the recent *Chandra* and *XMM-Newton* surveys. This is apparent in even basic flux-flux plots such as that shown in Fig. 3; at the faintest X-ray flux levels in the CDF-N, the extragalactic sources range in optical flux by a factor of $\gtrsim 10,000$.

Classification of the X-ray sources is challenging for several reasons. First of all, many of the sources are simply too faint for efficient optical spectroscopic identification with 8–10 m class telescopes (note the small dots in Fig. 3). Intensive optical identification programs on the deepest *Chandra* and *XMM-Newton* fields typically have ≈50–70% completeness at best. Furthermore, many of the X-ray sources have modest apparent optical luminosities,

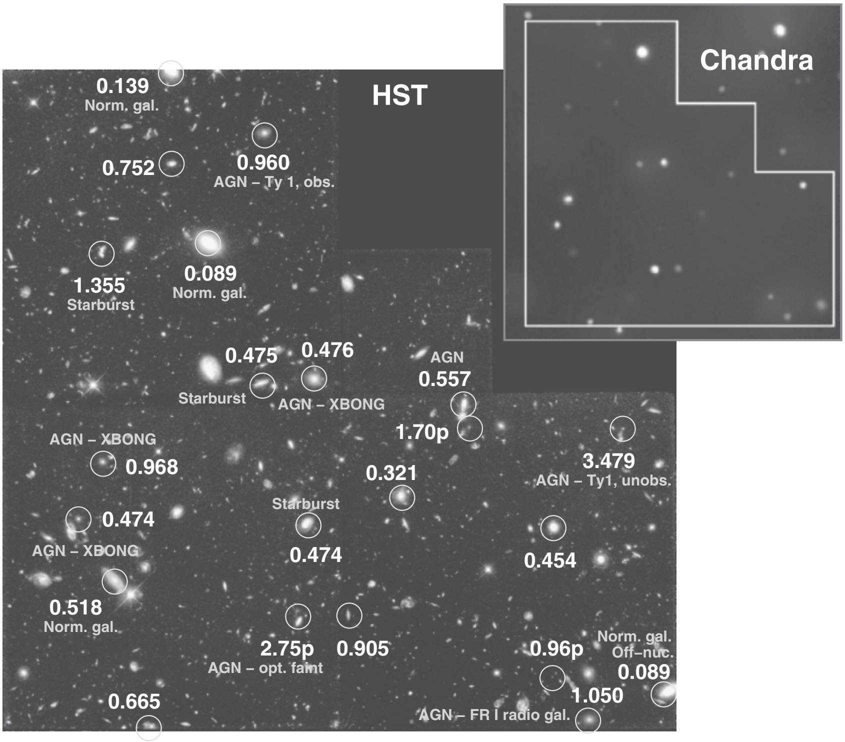

Fig. 4. *Chandra* and *HST* images of the HDF-N. The 22 *Chandra* sources are circled on the *HST* image; the circles are much larger than the *Chandra* source positional errors. The numbers are source redshifts; redshifts followed by a "p" are photometric. Basic source type information for many of the sources is also given

and thus their host galaxies make substantial diluting contributions to the flux measured in a spectroscopic aperture (e.g. Fig. 4). Finally, another challenge is an apparent "schism" between optical (type 1 vs. type 2) and X-ray (unobscured vs. obscured) schemes of classification; not all X-ray obscured AGN have type 2 optical spectra, and not all AGN with type 1 optical spectra are unobscured.

Considering X-ray, optical, and multiwavelength information, e.g. [1, 2, 10, 13, 14, 18, 24, 27, 31, 32, 48, 50], the primary extragalactic source types are found to be the following:

- *Unobscured AGN.* Blue, broad-line AGN are found that do not show signs of obscuration at either X-ray or optical/UV wavelengths. They are found over a broad range of redshift ($z \approx 0$–5), and they comprise a significant fraction of the brightest X-ray sources. At $z \gtrsim 1.5$ they also comprise a substantial fraction of all X-ray sources with spectroscopic identifications (certainly in part because these objects are the most straightforward to identify spectroscopically).

- *Obscured AGN with clear optical/UV AGN signatures.* Some objects showing X-ray evidence for obscuration have clear AGN signatures in their rest-frame optical/UV spectra. Notably, such AGN can have both type 1 and type 2 optical/UV classifications. Most of these objects have $z \lesssim 1.5$.
- *Optically faint X-ray sources.* These sources have $I \gtrsim 24$ and thus usually cannot be identified spectroscopically. Many, however, appear to be luminous, obscured AGN at $z \approx 1$–3 when their X-ray properties, optical photometric properties (including photometric redshifts), and multiwavelength properties are considered. Thus, these objects largely represent an extension of the previous class to higher redshifts and fainter optical magnitudes.
- *X-ray bright, optically normal galaxies (XBONGs).* XBONGs have X-ray luminosities ($\approx 10^{41}$–10^{43} erg s^{-1}) and X-ray-to-optical flux ratios suggesting some type of moderate-strength AGN activity. Some also have hard X-ray spectral shapes suggesting the presence of X-ray obscuration. Optical spectra give redshifts of $z \approx 0.05$–1, but AGN emission lines and non-stellar continua are not apparent. The nature of XBONGs remains somewhat mysterious. Some may just be Seyfert 2s where dilution by host-galaxy light hinders optical detection of the AGN, but others have high-quality follow up and appear to be truly remarkable objects. These "true" XBONGs may be (1) AGN with inner radiatively inefficient accretion flows, or (2) AGN that suffer from heavy obscuration covering a large solid angle ($\approx 4\pi$ sr), so that optical emission-line and ionizing photons cannot escape the nuclear region.
- *Starburst galaxies.* At the faintest X-ray flux levels in the deepest *Chandra* surveys, a significant fraction of the detected sources appear to be $z \approx 0$–1.3 dusty starburst galaxies. They are members of the strongly evolving starburst population responsible for creating much of the infrared background. The observed X-ray flux appears to be the integrated emission from many X-ray binaries and supernova remnants.
- *"Normal" galaxies.* Apparently normal galaxies are also detected in the deepest *Chandra* surveys out to $z \approx 0.2$. The observed X-ray emission is again probably largely from X-ray binaries and supernova remnants; these objects and the starburst galaxies above are probably not distinct but rather constitute a single population of galaxies with star formation of varying intensity. Low-luminosity AGN are likely present in some cases as well. Some normal galaxies sport luminous X-ray sources clearly offset from their nuclei. At even fainter X-ray flux levels, normal and starburst galaxies should be the dominant class of extragalactic X-ray sources [47, 60].

Most of the AGN found in X-ray surveys are "radio quiet" in the sense that the ratio (R) of their rest-frame 5 GHz and 4400 Å flux densities are $R < 10$.

Figure 4 shows some of the source classifications in the HDF-N, which is at the center of the CDF-N (see Fig. 1) and thus has the most sensitive X-ray coverage available. Note, for example, that three of the brightest X-ray sources are XBONGs. These were not recognized as AGN prior to the *Chandra* observations, despite the many intensive studies of the HDF-N.

Luminosity and Redshift Distributions

The combined results from deep and wider X-ray surveys show that the sources comprising most of the XRB have X-ray luminosities comparable to those of local Seyfert galaxies, such as NGC 3783, NGC 4051, and NGC 5548 (e.g., see Fig. 5). While a few remarkable obscured quasars have been found, these appear fairly rare and only make a small contribution to the XRB. Indeed, it appears that the fraction of obscured AGN drops with luminosity from ≈ 60–70% at Seyfert luminosities to ≈ 30% at quasar luminosities.

Most spectroscopically identified AGN in the deep X-ray surveys have $z \lesssim 2$, although a significant minority have $z \approx 2$–5 [33, 37]. This is partly due to spectroscopic incompleteness bias, as is apparent by comparing the filled circles, open circles, and stars in Fig. 5. However, as will be described further in Subsect. 3.2, there is a real enhancement of AGN at $z \lesssim 1$ relative to expectations from pre-*Chandra* AGN-synthesis models of the XRB. An impressive ∼60% of the 2–8 keV XRB arises at $z < 1$.

AGN Sky Density

Most (≈ 70–100%) of the extragalactic X-ray sources found in both the deep and wider X-ray surveys with *Chandra* and *XMM-Newton* are AGN of some type. Starburst and normal galaxies make increasing fractional contributions at the faintest X-ray flux levels, but even in the CDF-N they represent $\lesssim$ 20–30% of all sources (and create $\lesssim 5\%$ of the XRB). The observed AGN sky density in the deepest X-ray surveys is ≈ 6500 deg^{-2}, about an order of magnitude higher than that found at any other wavelength. This exceptional effectiveness at finding AGN arises because X-ray selection (1) has reduced absorption bias and minimal dilution by host-galaxy starlight, and (2) allows concentration of intensive optical spectroscopic follow-up upon high-probability AGN with faint optical counterparts (i.e., it is possible to probe further down the luminosity function).

Completeness of X-Ray AGN Selection

Are there significant numbers of luminous AGN that are not found even in the deepest X-ray surveys? This could be the case if there is a large population of AGN that are X-ray weak due either to absorption or an intrinsic inability to produce X-rays. This question can be partially addressed by looking for AGN

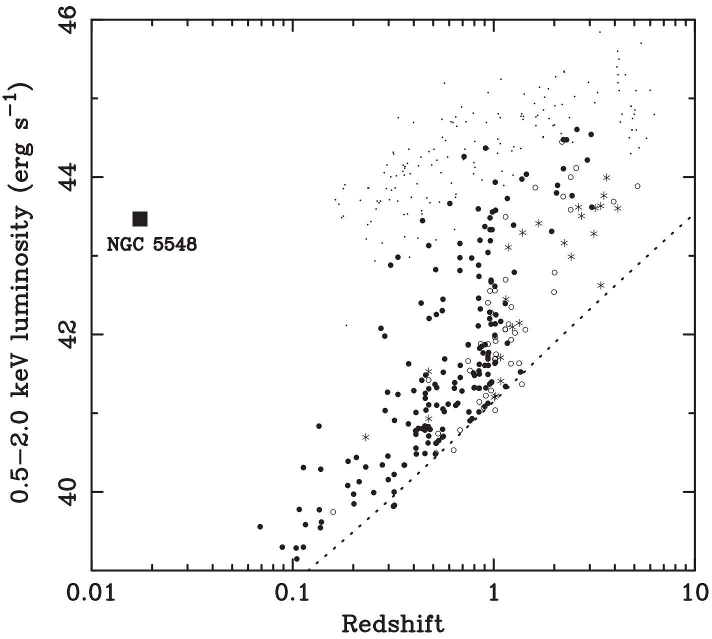

Fig. 5. Luminosity in the 0.5–2 keV band (computed from the 0.5–2 keV flux assuming a power-law spectrum with a photon index of $\Gamma = 2$) versus redshift for extragalactic sources in the CDF-N with spectroscopic redshifts. Sources with $I = 16$–22, $I = 22$–23, and $I > 23$ are indicated with filled circles, open circles, and stars, respectively. The dotted curve shows the approximate sensitivity limit near the center of the CDF-N. Also shown are the well-studied Seyfert 1 galaxy NGC 5548 (filled square) and Sloan Digital Sky Survey (SDSS) quasars from the SDSS Early Data Release with X-ray coverage in archival *ROSAT* data (small dots; the relevant solid angle covered by pointed *ROSAT* observations is ≈ 15 deg^2). Note that NGC 5548 could have been detected to $z \sim 10$ in the CDF-N. Note also that the CDF-N and SDSS populations are nearly disjoint, as a consequence of the different solid angle coverages (a factor of ~120) and depths. Adapted from [4, 8] and [53]

found at other wavelengths that are not detected in X-rays. In the CDF-N, one of the most intensively studied regions of sky at all wavelengths, there are only 1–2 such AGN known. The most conspicuous is 123725.7+621128, a radio-bright (≈ 6 mJy at 1.4 GHz) wide angle tail source that is likely at $z \approx 1$–2 (although the redshift of this source remains uncertain). This is one of the brightest radio sources in the CDF-N but has been notoriously difficult to detect in X-rays. Manual analysis of the 2 Ms *Chandra* data at the AGN position indicates a likely, but still not totally secure, detection (at a false-positive probability threshold of 3×10^{-5} using the standard *Chandra* wavelet source detection algorithm). The 0.5–2 keV luminosity is $\lesssim 5 \times 10^{41}$ erg s^{-1}. The only other known AGN in the CDF-N without an X-ray detection is

123720.0+621222, a narrow-line AGN at $z = 2.45$; its 0.5–2 keV luminosity is $\lesssim 2 \times 10^{42}$ erg s^{-1}.

Despite the spectacular success of X-ray surveys at finding AGN, appropriate humility is required when assessing the AGN selection completeness of even the deepest X-ray surveys. This is made clear by consideration of "Compton-thick" AGN, which comprise a sizable fraction (≈40%) of AGN in the local universe, e.g. [38, 39, 40] and [43]. Such AGN are absorbed by intrinsic column densities of $N_{\mathrm{H}} \gg 1.5 \times 10^{24}$ cm^{-2}, within which direct line-of-sight X-rays are effectively destroyed via the combination of Compton scattering and photoelectric absorption. Such AGN are often only visible via weaker, indirect X-rays that are "reflected" by neutral material or "scattered" by ionized material.[5] Many of the local Compton-thick AGN (e.g., NGC 1068, NGC 6240, Mrk 231), if placed at $z \gtrsim 0.5$–1.5, would remain undetected in even the deepest *Chandra* surveys. Thus, it appears plausible that ≈ 40% of AGN at such redshifts may have been missed (the number, of course, could be higher or lower if the fraction of Compton-thick AGN evolves significantly with redshift). Deeper observations with *Chandra* (≈ 10 Ms; see Subsect. 4.1) may be able to detect the indirect X-rays from any missed Compton-thick AGN, and observations with *Spitzer* may be able to detect "waste heat" from such objects at infrared wavelengths.

Another way to address AGN selection completeness in X-ray surveys is to consider "book-keeping" arguments: can the observed sources explain the observed 20–40 keV XRB intensity, and can all the observed accretion account for the local density of supermassive black holes? The answer is plausibly "yes" in both cases, but with some uncertainty. In the first case, one must make a significant spectral extrapolation from 5–10 keV and worry about mission-to-mission cross-calibration uncertainties. In the second, significant uncertainties remain in bolometric correction factors, accretion efficiencies, and the local density of supermassive black holes. The current book-keeping arguments cannot rule out the possibility that a significant fraction of the AGN population (e.g., Compton-thick AGN) is still missed in X-ray surveys. Indeed, some book-keepers find better agreement with the local black-hole mass function after making a substantial correction for missed accretion in Compton-thick AGN.

3.2 Recent X-Ray Results on AGN Evolution

Optical studies of AGN evolution have typically focused on luminous quasars. These have been known to evolve strongly with redshift since ~1968 [45],

---

[5] In some "translucent" cases, where the column density is only a few $\times 10^{24}$ cm^{-2} (i.e., a few Thomson depths), direct "transmission" X-rays from a Compton-thick AGN may become visible above rest-frame energies of ~10 keV. For comparison, the column density through your chest is $\sim 1 \times 10^{24}$ cm^{-2}; if you stood along the line-of-sight to an AGN, you could almost render it Compton thick!

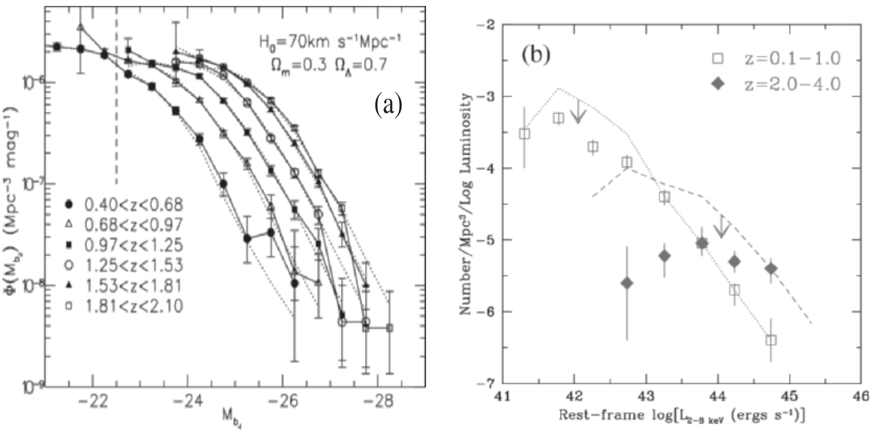

Fig. 6. (**a**) Optical luminosity functions in 6 redshift "shells" spanning $z =$ 0.40–2.10 for ∼16,800 luminous AGN from the 2dF and 6dF surveys. Note the clear positive evolution with increasing redshift at high luminosity (i.e., the comoving number density of luminous AGN increases with redshift from $z = 0.40$–2.10). From [22]. (**b**) X-ray (2–8 keV) luminosity functions in two redshift "shells" (as labeled) for moderate-to-high luminosity AGN from the CDF-N, Abell 370, SSA13, and SSA22 *Chandra* surveys (see Table 1) as well as several earlier X-ray surveys. The dotted and dashed curves show the maximum possible luminosity functions after allowing for incompleteness of the follow-up spectroscopy. Note the apparent negative evolution with increasing redshift at moderate luminosity. Adapted from [20]

having a comoving space density at $z \approx 2$ that is $\gtrsim$100 times higher than at $z \approx 0$. Figure 6a shows optical luminosity functions in 6 redshift "shells" spanning $z = 0.40$–2.10 for ∼16,800 luminous AGN from the 2dF and 6dF surveys. Clear positive evolution with redshift is observed, and pure luminosity evolution (PLE) models provide an acceptable fit to these data. New optical AGN surveys, such as COMBO-17, have recently discovered significant numbers of moderate-luminosity AGN (with $M_{\rm B} > -23$) at $z \approx 1$–4, allowing investigation of their evolution. As for luminous quasars, the AGN found in these surveys also appear to peak in comoving space density at $z \approx 2$. Both PLE and pure density evolution (PDE) models can acceptably fit the COMBO-17 data alone. Although a systematic combination of the COMBO-17 data with a large sample of higher-luminosity AGN has yet to be published, there are hints that the redshift at which the comoving space density peaks is smaller at lower luminosities, e.g. [19].

As noted in Subsect. 3.1, the deepest X-ray surveys efficiently select AGN even fainter than those found by COMBO-17 out to high redshift (e.g., see Fig. 5). X-ray AGN samples show a clear dependence of AGN evolution upon luminosity, with strong positive evolution only being seen at high luminosities (see Fig. 6b). Lower luminosity AGN appear to be about as common at

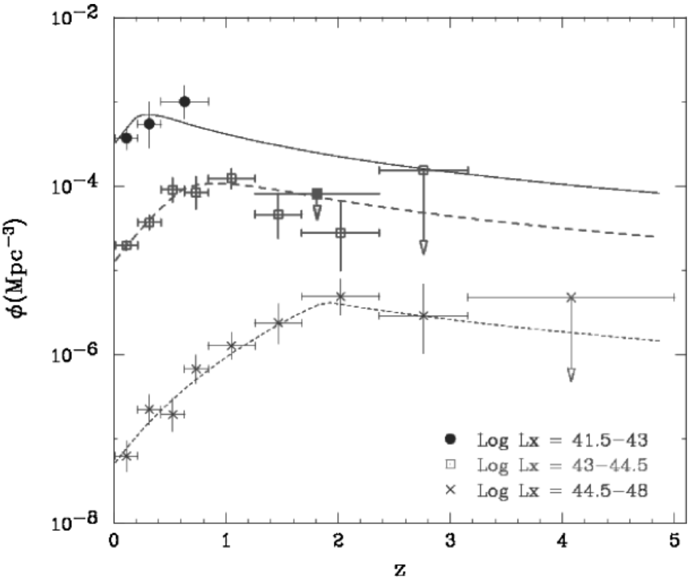

Fig. 7. The comoving spatial density of AGN in three X-ray luminosity ranges as a function of redshift, derived using data from several X-ray surveys. From [51]

$z \approx 0$–1 as they ever were, consistent with the trend hinted at by COMBO-17 [58]. These results are robust to incompleteness of the spectroscopic follow up, although clearly they are still dependent upon the completeness of AGN X-ray selection (see Subsect. 3.1). It appears that while the SMBH in rare, luminous AGN could grow efficiently at high redshift, the SMBH in most AGN had to wait longer to grow, e.g. [23, 42].

Figure 7 shows estimates of the comoving spatial density of AGN in three X-ray luminosity ranges as a function of redshift. These have been constructed utilizing a combination of *Chandra*, *ASCA*, and *HEAO1* surveys at photon energies above 2 keV (with 247 AGN in total). The data are best fit with luminosity-dependent density evolution (LDDE) out to some cutoff redshift (z_c), where z_c increases with luminosity; as a result, the ratio of the peak spatial density to that at the present day is higher for more luminous AGN. At a basic level, LDDE also seems more physically plausible than PLE or PDE; simple PLE models tend to overpredict the number of $\gtrsim 10^{10}$ $\mathrm{M}_\odot$ black holes in the local universe, while simple PDE models tend to overpredict the local space density of quasars.

3.3 X-Ray Emitting AGN in Luminous Submillimeter Galaxies

The deepest *Chandra* and *XMM-Newton* surveys have finally provided the necessary X-ray sensitivity to complement the most sensitive surveys at submillimeter and infrared wavelengths. One notable instance where obtaining the highest possible X-ray sensitivity has been essential is in studies of the AGN content of distant submillimeter galaxies detected with the SCUBA instrument on the James Clerk Maxwell Telescope. Most of these galaxies are thought to contain intense starbursts with star-formation rates of $\gtrsim 100\ \mathrm{M}_\odot\ \mathrm{yr}^{-1}$, yet they are not notable in optical galaxy surveys due to dust obscuration of the corresponding starlight. The SCUBA galaxy population is thought to be mostly at $z \approx 1.5$–3, e.g. [17], and such galaxies were $\sim$1000 times more common at $z \sim 2$ than in the local universe. The obscured starlight in submillimeter galaxies is re-radiated in the rest-frame infrared (observed-frame submillimeter), e.g. [11].

What fraction of submillimeter galaxies contains actively accreting supermassive black holes? Sensitive X-ray studies play an important role in addressing this question, since they allow effective searching for AGN in the majority of submillimeter galaxies that are optically faint (and thus challenging to study in detail with optical spectroscopy). Early comparisons between $\approx$20–150 ks *Chandra* surveys and submillimeter surveys yielded little ($\lesssim 10\%$) source overlap. However, the latest analysis of the 2 Ms CDF-N data [7, 30], reveals that seven of the 13 ($\approx$54%) bright submillimeter galaxies (with 850 µm flux densities of >5 mJy) in the CDF-N central region mapped with SCUBA have X-ray counterparts (see Fig. 8) [6]; these counterparts have

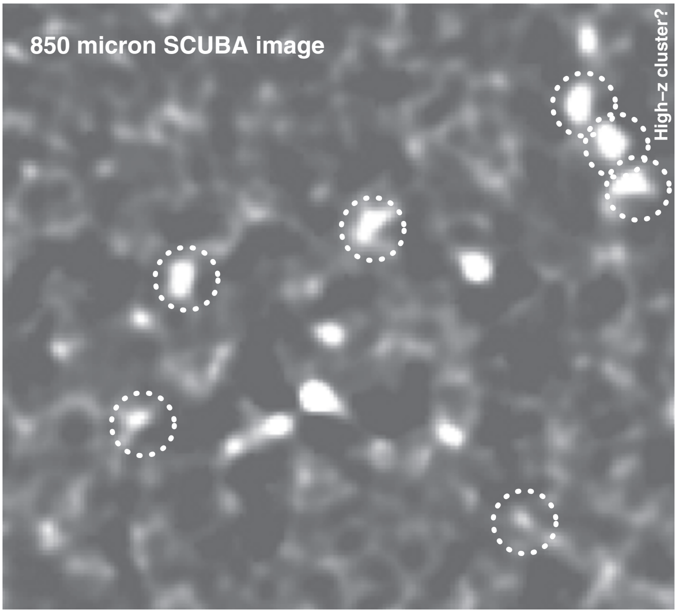

Fig. 8. Map at 850 µm of the central region of the CDF-N; the map is $\approx 6'$ on a side. Sources at 850 µm with X-ray detections are enclosed by *dotted circles*. The three clustered 850 µm/X-ray sources near the upper-right corner are also coincident with an extended X-ray source, perhaps a high-redshift cluster. Adapted from [3] and [12]

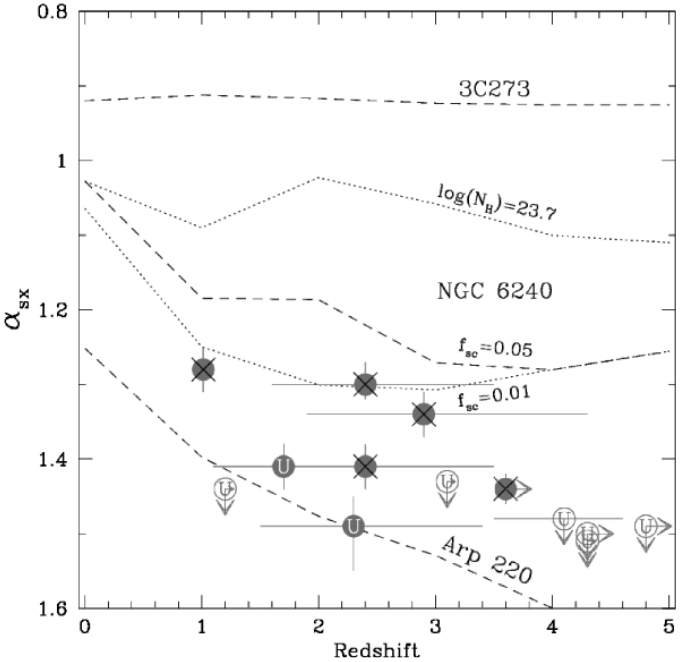

Fig. 9. Submillimeter-to-X-ray spectral index ($\alpha_{\rm sx}$) versus redshift. Submillimeter sources in the central part of the CDF-N with (without) X-ray detections are shown as *solid* (*open*) *circles*. The five circles with overlaid crosses are likely AGN according to their X-ray properties, while those with overlaid "U" are of unknown X-ray type. *Dashed curves* show $\alpha_{\rm sx}$ values for 3C 273, NGC 6240, and Arp 220 adopting their observed amounts of X-ray absorption; alternative *dotted curves* show an AGN like NGC 6240 but with less internal absorption ($N_{\rm H} = 5 \times 10^{23}$ cm^{-2}) and a smaller scattered flux fraction ($f_{\rm sc} = 0.01$). Adapted from [3]

≈15–200 counts in the full *Chandra* bandpass. Five of the seven X-ray detected submillimeter galaxies likely host obscured AGN based upon their observed X-ray luminosities, X-ray spectral shapes, and X-ray-to-submillimeter flux ratios (see Fig. 9). The remaining two have X-ray emission properties consistent with those expected from star formation activity, although it is possible that they host weak AGN as well. If the latter two sources are indeed powered mainly by star formation, they would be the most X-ray luminous ($\approx 4 \times 10^{42}$ erg s^{-1}) starburst galaxies known.

Do the X-ray emitting AGN found in many submillimeter galaxies make a significant contribution to these galaxies' total energy output? Answering this question requires assessment of the amount of X-ray absorption present since, for a given observed X-ray flux, a Compton-thick AGN can be much more luminous than a Compton-thin AGN (see Subsect. 3.1). Basic X-ray spectral fitting suggests that three of the five submillimeter galaxies hosting

AGN in the CDF-N central region have Compton-thin absorption, while only one is likely to have Compton-thick absorption (the final object has poor X-ray spectral constraints). Armed with this knowledge, consideration of the observed X-ray-to-submillimeter flux ratios (see Fig. 9) suggests that $\lesssim 10\%$ of the total energy output from these submillimeter galaxies is ultimately due to an AGN. Star-formation is apparently the dominant power source for the infrared/submillimeter emission, even when an AGN is also present.

The results above are currently being extended, utilizing redshifts from ongoing deep optical spectroscopy. Thus far, these extended results confirm the main conclusions above, e.g. [5].

3.4 High-Redshift ($z > 4$) AGN Demography and Physics

As is apparent from Figs. 5 and 10, deep X-ray surveys can detect $z > 4$ AGN that are $\gtrsim 10$–30 times less luminous than those found in wide-field optical AGN surveys such as the SDSS. At least in the local universe, such moderate-luminosity AGN are much more numerous and thus more representative than the rare, highly luminous quasars. Furthermore, unlike the rest-frame ultraviolet light sampled at $z > 4$ in ground-based AGN surveys, X-ray surveys suffer from progressively *less* absorption bias as higher redshifts are surveyed. At $z > 4$, hard $\approx$2–40 keV rest-frame X-rays are accessed; these can penetrate large column densities up to several $\times 10^{24}$ cm^{-2}, e.g. [16].

Spectroscopic follow-up of moderate-luminosity X-ray detected AGN at $z > 4$ is challenging, since such objects are expected to have z magnitudes of 23–26 (provided they have not "dropped out" of the z bandpass entirely). Nevertheless, significant constraints on the sky density of such objects have been set via large-telescope spectroscopy and Lyman-break selection. In the latter case, objects can be selected that either have appropriate optical/near-infrared colors to be at $z > 4$ or alternatively have no optical/near-infrared detections. The "bottom line" from these demographic studies in the CDF-N and CDF-S is that there are $\lesssim 12$ AGN at $z > 4$ detectable in a 1–2 Ms *Chandra* field, and that only $\approx$4 of these have a z magnitude of $<$25 (this limit on the sky density is still $\sim$260 times the sky density of $z > 4$ quasars from the SDSS), e.g. [9]. These sky-density constraints are sufficient to rule out some pre-*Chandra* predictions by about an order of magnitude, and the combined X-ray and SDSS results, e.g. [21, 36, 59], indicate that the AGN contribution to reionization at $z \approx 6$ is small.

Once high-redshift AGN have been identified, via either X-ray or optical surveys, broad-band spectral energy distribution analyses and X-ray spectral fitting can provide information on their accretion processes and environments. The currently available data, albeit limited, suggest that $z > 4$ AGN are accreting and growing in roughly the same way as AGN in the local universe; there is no evidence that their inner X-ray emitting regions have been affected by, for example, accretion-disk instabilities or radiation-trapping effects. Figure 10 plots X-ray versus optical flux for $z > 4$ AGN from both X-ray and

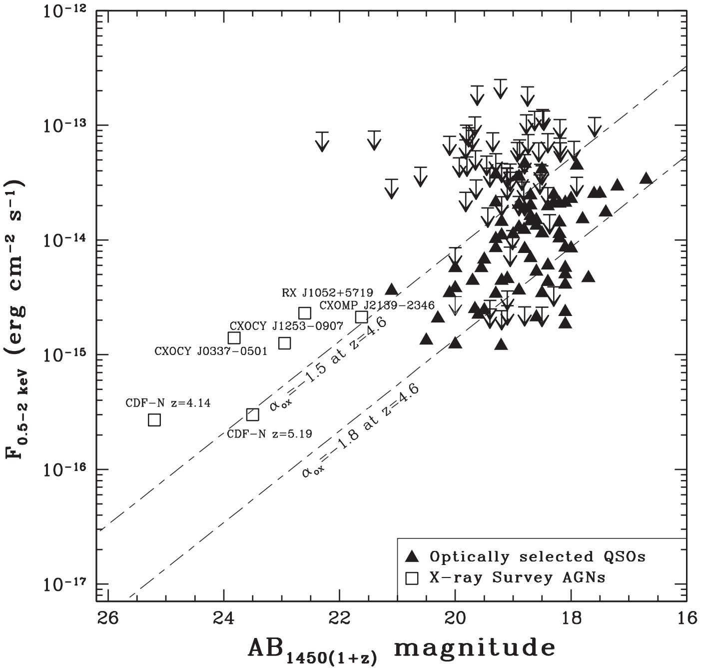

Fig. 10. Observed-frame, Galactic absorption-corrected 0.5–2 keV flux versus $\mathrm{AB}_{1450(1+z)}$ magnitude for $z \geq 4$ AGN found both in optical and X-ray surveys (the X-ray *upper limits* shown are all for AGN from optical surveys). The *slanted, dashed lines* show the $\alpha_{\mathrm{ox}} = -1.5$ and $\alpha_{\mathrm{ox}} = -1.8$ loci at $z = 4.6$. Adapted from [54]

optical surveys. The X-ray-to-optical spectral indices, α_{ox}, for these objects are consistent with those of AGN in the local universe, once luminosity effects and selection biases are taken into account. These biases and effects likely explain, for example, why the moderate-luminosity, X-ray selected AGN in Fig. 10 have notably higher X-ray-to-optical flux ratios than the luminous, optically selected quasars.

Two recent X-ray spectral fitting results on $z > 4$ AGN are shown in Fig. 11. Figure 11a shows the X-ray spectrum of the highest redshift AGN discovered thus far in the CDF-N, a low-luminosity quasar at $z = 5.186$. It was only possible to obtain a respectable-quality X-ray spectrum for such an object due to the 2 Ms CDF-N exposure. Spectral fitting yields a power-law photon index of $\Gamma = 1.8 \pm 0.3$, consistent with observations of similar objects at low redshift, and there is no evidence for intrinsic X-ray absorption. Figure 11b shows a "stacked" spectrum of 46 luminous radio-quiet quasars at $z = 4.0$–6.3 (their median redshift is $z = 4.43$); this spectrum has 750 counts in total. Joint fitting of the 46 individual spectra, using the Cash statistic, yields a power-law photon index ($\Gamma = 1.9 \pm 0.1$) that is again consistent

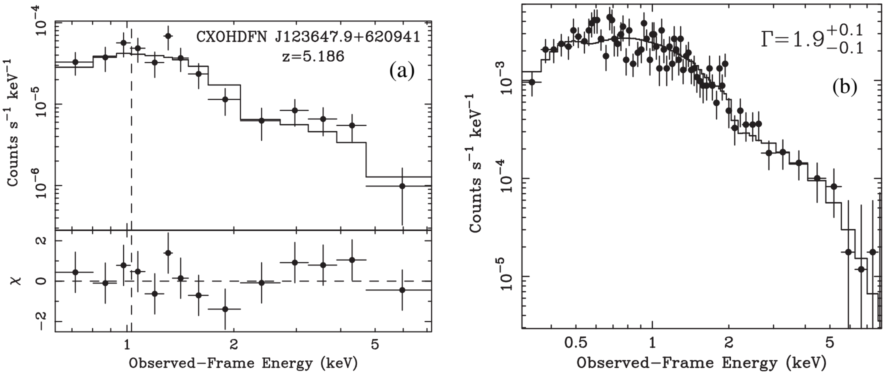

Fig. 11. Observed-frame X-ray spectra for (**a**) the $z = 5.186$ CDF-N AGN CXOHDFN J123647.9+620941 and (**b**) 46 radio-quiet quasars at $z = 4.0$–6.3 that have been stacked together. The best-fitting power-law models with Galactic absorption are also shown; see the text for fitting results. In (a) the *lower panel* shows the fit residuals in units of sigma, and the *vertical dashed line* indicates the energy of the (undetected) 6.4 keV iron Kα line. Adapted from [52] and [55]

with observations at low redshift. A fairly tight limit on any intrinsic X-ray absorption of $N_{\rm H} \lesssim 9 \times 10^{20}$ cm^{-2} is also set. The overall picture emerging, then, is that while the AGN population shows enormous changes in number density over cosmic time, individual AGN X-ray emission regions appear to be remarkably stable entities.

4 Some Future Prospects

4.1 Future Prospects for Chandra and XMM-Newton

Future prospects for learning more about AGN physics and evolution via X-ray surveys appear wonderful! Follow-up studies for most of the ≈40 surveys listed in Table 1 and Table 2 are ongoing, and many exciting results are thus guaranteed even if no more X-ray data are taken. Fortunately, however, both *Chandra* and *XMM-Newton* continue to generate torrents of superb new data that can provide even more impressive advances.

Where can the capabilities of *Chandra* and *XMM-Newton* be best applied in future observations? Figure 12 presents one useful way of thinking about this issue, via a plot of 0.5–2 keV flux limit versus solid angle for selected X-ray surveys. Key parts of this diagram remain to be explored. For example, very little solid angle has been surveyed at 0.5–2 keV flux levels of (2–20) $\times 10^{-17}$ erg cm^{-2} s^{-1}, and thus our understanding of the X-ray universe at these flux levels suffers from limited source statistics and likely cosmic variance. These flux levels are below the *XMM-Newton* confusion limit, and thus multiple 0.25–2 Ms *Chandra* observations are required. Specific science

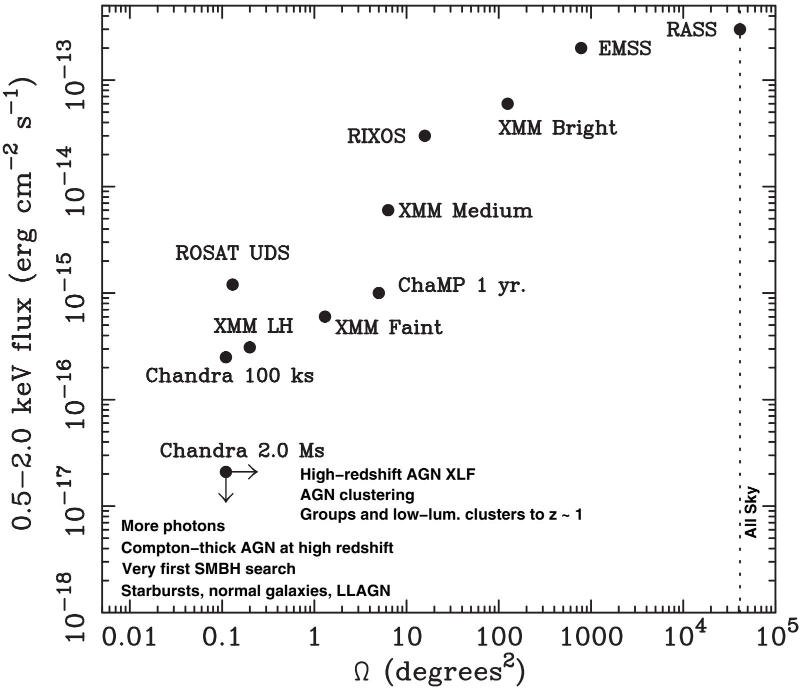

Fig. 12. A selection of extragalactic X-ray surveys in the 0.5–2 keV flux limit versus solid angle, *Ω*, plane. Shown are the *ROSAT* All-Sky Survey (RASS), the *Einstein* Extended Medium-Sensitivity Survey (EMSS), the *ROSAT* International X-ray/Optical Survey (RIXOS), the *XMM-Newton* Serendipitous Surveys (*XMM* Bright, *XMM* Medium, *XMM* Faint), the *Chandra* Multiwavelength Project (ChaMP), the *ROSAT* Ultra Deep Survey (*ROSAT* UDS), the ≈100 ks *XMM-Newton* survey of the Lockman Hole (*XMM* LH), *Chandra* 100 ks surveys, and *Chandra* 2 Ms surveys (i.e., the CDF-N). Although each of the surveys shown clearly has a range of flux limits across its *solid angle*, we have generally shown the most sensitive flux limit. The vertical *dot-dashed line* shows the *solid angle* of the whole sky. Some key science goals achievable by extending deep *Chandra* surveys both wider and deeper are also listed

goals that can be advanced with this approach include (1) pinning down the X-ray luminosity function of moderate-luminosity AGN at $z \approx 2$–6, (2) tracing AGN clustering out to high redshift; this is ideally done with contiguous, deep coverage, and (3) measuring the evolution and properties of groups and low-luminosity clusters out to $z \approx 1$. Figure 13a depicts the ongoing Extended *Chandra* Deep Field-South survey, which has been guided by the philosophy above. It will cover a contiguous ∼1/4 deg^2 area at a 0.5–2 keV flux level of $(1\text{–}2) \times 10^{-16}$ erg cm^{-2} s^{-1}, and it should generate ≈400 new AGN (in addition to the ≈300 already known in the CDF-S). Almost all of these will have superb *HST* imaging and multiwavelength coverage.

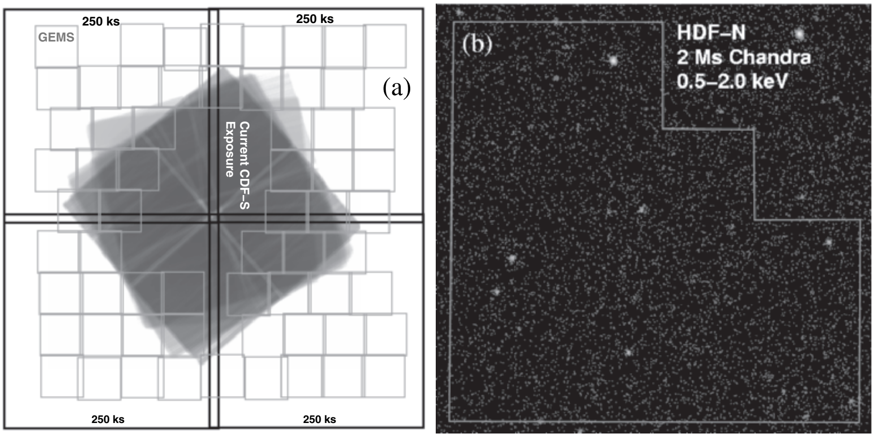

Fig. 13. (**a**) Schematic illustration of the Extended *Chandra* Deep Field-South survey. The underlying grayscale image shows the current CDF-S exposure map. The four *large black squares* show the coverage of the upcoming four 250 ks *Chandra* observations. The 63 *small gray squares* show the coverage of *HST* ACS observations made by the GEMS project (the GOODS survey provides *HST* ACS coverage for the central region not covered by GEMS). (**b**) *Chandra* 0.5–2 keV image of the central part of the 2 Ms CDF-N centered on the HDF-N (shown in outline). Note that most (≈94%) pixels are *black*, indicating no background. *Chandra* is essentially in the photon-limited regime with a 2 Ms exposure, and it can remain in this regime even with an ≈10 Ms exposure (for 0.5–2 keV sources near the field center)

An equally important guiding philosophy is to observe one field with *Chandra* as sensitively as possible (see Fig. 12). Reaching 0.5–2 keV flux levels of $\approx 5 \times 10^{-18}$ erg cm^{-2} s^{-1} is entirely feasible; *Chandra* could remain nearly photon limited near the field center (see Fig. 13b), and source confusion is unlikely even for source densities exceeding 100,000 deg^{-2}. The total required exposure time on a field is ≈10 Ms. Specific science goals include (1) determining if there is a significant population of Compton-thick AGN at $z \approx 0.5$–4 that has been missed to date (see Subsect. 3.1), (2) tightening constraints on moderate-luminosity AGN at $z \approx 4$–10, (3) detecting hundreds of normal and starburst galaxies out to high redshift (these should outnumber the AGN), and using their X-ray emission as an independent, extinction-free measure of star-formation rate, and (4) obtaining significant numbers of X-ray photons on the faint X-ray source populations currently known, so that X-ray spectral and variability analyses can be applied effectively to determine their nature. Such a sensitive X-ray observation will not be possible again for 10–20 years (see Fig. 14)! Performing such an observation now can provide information on the sources that will be the primary targets of future missions such as *XEUS* and *Generation-X*; it will thereby bolster the science cases for these missions and aid their optimal design.

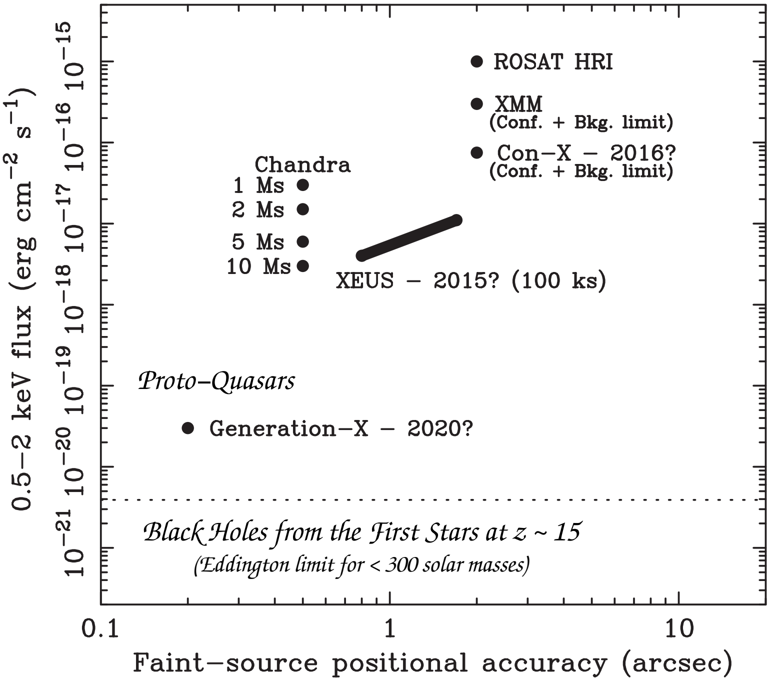

Fig. 14. Flux limit from 0.5–2 keV versus faint-source positional accuracy for some past, present, and future X-ray missions (the locations in the diagram and launch dates for future missions are approximate). With a 5–10 Ms exposure, *Chandra* can achieve sensitivities comparable to those discussed for *XEUS*. Furthermore, *Chandra* positions are likely to be the best available for $\gtrsim 15$ yr. Also shown are the expected X-ray fluxes from (1) the black holes made by the deaths of the first stars at $z \sim 15$, and (2) proto-quasars containing black holes of mass $\sim 10^3$–10^4 $M_\odot$ at $z \sim 10$–15

4.2 Upcoming and Planned X-Ray Missions

In the future, both large ($\gtrsim US\,\$1$ billion, or $\gtrsim 600$ billion Chilean pesos; see Fig. 14) and small-to-medium class ($\approx US\,\$120$–180 million) X-ray missions should substantially advance the AGN X-ray survey work described above. *Constellation-X*, for example, should enable high-quality X-ray spectroscopy for some of the remarkable brighter sources found in X-ray surveys. *XEUS* should be able to generate hundreds of fields that are as sensitive as the deepest *Chandra* surveys, while also providing superior photon statistics to those available presently. Fitting of high-quality *XEUS* spectra should allow direct redshift determination in many cases. Ultimately, *Generation-X* will reach flux limits ~100 times better than those of *Chandra* and *XEUS* (see Fig. 14). This improved sensitivity should allow detection and study of ~ 1000 $M_\odot$ “proto-quasars” at $z \approx 10$–15, enabling investigation of how the stellar-mass black holes made by the deaths of the first stars grew to make the first AGN.

Future small-to-medium class X-ray missions, at least one to be launched soon, will sensitively survey large areas of sky at high X-ray energies; some will access the poorly explored ≈10–200 keV band covering the peak of the XRB. After its 2004–2005 launch, for example, *Swift* will serendipitously conduct the most sensitive ≈10–150 keV survey to date with its Burst Alert Telescope. A large fraction of the sky should be covered over the lifetime of *Swift*, and ≈200–400 AGN should be detected. In the 2007–2010 timeframe, proposed missions such as the *Dark Universe Observatory (DUO)* and the *Nuclear Spectroscopic Telescope Array (NuSTAR)* will also hopefully conduct sensitive surveys in the 0.3–8 keV and 6–80 keV bands, respectively. *DUO* would detect ∼160,000 AGN in its surveys of the North Galactic Cap (the SDSS area) and South Galactic Pole, while *NuSTAR* would carry the first highly sensitive, focusing telescope for >10 keV X-rays. Other planned small-to-medium class missions include Japan's *Monitor of All-sky X-ray Image (MAXI)* and *New X-ray Telescope (NeXT)* as well as Europe's *LOBSTER* and *ROSITA*. The *Black Hole Finder Probe*, defined as part of NASA's Beyond Einstein program, should ultimately obtain an all-sky census of accreting black holes using a wide-field imaging telescope in the ≈10–600 keV band.

Acknowledgements

We gratefully acknowledge support from NSF CAREER award AST-9983783, CXC grant GO2-3187A, the Royal Society (DMA), the PPARC (FEB), and Italian Space Agency contract ASI/I/R/057/02 (CV). We thank all of our collaborators.

Some Recent Key References

Clearly the relevant literature is extensive, and the reader is urged to consult not only the references below but also the works cited in these references.

1. D.M. Alexander, W.N. Brandt, A.E. Hornschemeier, et al., 2001, AJ, 122, 2156
2. D.M. Alexander, H. Aussel, F.E. Bauer, et al., 2002, ApJ, 568, L85
3. D.M. Alexander, F.E. Bauer, W.N. Brandt, et al., 2003, AJ, 125, 383
4. D.M. Alexander, F.E. Bauer, W.N. Brandt, et al., 2003, AJ, 126, 539
5. D.M. Alexander, F.E. Bauer, S.C. Chapman, et al., 2004, in Multiwavelength Mapping of Galaxy Formation and Evolution, ed. R. Bender, A. Renzini (Springer-Verlag, Berlin), in press (astro-ph/0401129)
6. O. Almaini, S.E. Scott, J.S. Dunlop, et al., 2003, MNRAS, 338, 303
7. A.J. Barger, L.L. Cowie, A.T. Steffen, et al., 2001, ApJ, 560, L23
8. A.J. Barger, L.L. Cowie, P. Capak, et al., 2003, AJ, 126, 632

9. A.J. Barger, L.L. Cowie, P. Capak, et al., 2003, ApJ, 584, L61
10. F.E. Bauer, D.M. Alexander, W.N. Brandt, et al., 2002, AJ, 124, 2351
11. A.W. Blain, I. Smail, R.J. Ivison, et al., 2002, Phys. Rep., 369, 111
12. C. Borys, S. Chapman, M. Halpern, et al., 2003, MNRAS, 344, 385
13. W.N. Brandt, A. Laor, B.J. Wills, 2000, ApJ, 528, 637
14. W.N. Brandt, A.E. Hornschemeier, D.M. Alexander, et al., 2001, AJ, 122, 1
15. W.N. Brandt, D.M. Alexander, A.E. Hornschemeier, et al., 2001, AJ, 122, 2810
16. W.N. Brandt, C. Vignali, D.P. Schneider, et al., 2004, Adv. Space Res., 34, 2478
17. S.C. Chapman, A.W. Blain, R.J. Ivison, et al., 2003, Nature, 422, 695
18. A. Comastri, M. Mignoli, P. Ciliegi, et al., 2002, ApJ, 571, 771
19. L.L. Cowie, A.J. Barger, M.W. Bautz, et al., 2003, ApJ, 584, L57
20. L.L. Cowie, G.P. Garmire, M.W. Bautz, et al., 2002, ApJ, 556, L5
21. S. Cristiani, D.M. Alexander, F.E. Bauer, et al., 2004, ApJ, 600, L119
22. S.M. Croom, R.J. Smith, B.J. Boyle, et al., 2004, MNRAS, 349, 1397
23. F. Fiore, M. Brusa, F. Cocchia, et al., 2003, A&A, 409, 79
24. P. Gandhi, C.S. Crawford, A.C. Fabian, et al., 2004, MNRAS, 348, 529
25. R. Giacconi, A. Zirm, J.X. Wang, et al., 2002, ApJS, 139, 369
26. M. Giavalisco, H.C. Ferguson, A.M. Koekemoer, et al., 2004, ApJ, 600, L93
27. R. Gilli, 2004, Adv. Space Res., 34, 2470
28. P.J. Green, J.D. Silverman, R.A. Cameron, et al., 2004, ApJS, 150, 43
29. G. Hasinger, B. Altieri, M. Arnaud, et al., 2001, A&A, 365, L45
30. A.E. Hornschemeier, W.N. Brandt, G.P. Garmire, et al., 2000, ApJ, 541, 49
31. A.E. Hornschemeier, F.E. Bauer, D.M. Alexander, et al., 2003, AJ, 126, 575
32. A.E. Hornschemeier, D.M. Alexander, F.E. Bauer, et al., 2004, ApJ, 600, L147
33. M.P. Hunt, C.C. Steidel, K.L. Adelberger, et al., 2004, ApJ, 605, 625
34. F. Jansen, D. Lumb, B. Altieri, et al., 2001, A&A, 365, L1
35. D.W. Kim, R.A. Cameron, J.J. Drake, et al., 2004, ApJS, 150, 19
36. A.M. Koekemoer, D.M. Alexander, F.E. Bauer, et al., 2004, ApJ, 600, L123
37. A. Marconi, G. Risaliti, R. Gilli, et al., 2004, MNRAS, 351, 169
38. G. Matt, 2002, Phil. Trans. R. Soc. Lond. A, 360, 2045
39. G. Matt, A.C. Fabian, M. Guainazzi, et al., 2000, MNRAS, 318, 173
40. E.C. Moran, A.V. Filippenko, R. Chornock, 2002, ApJ, 579, L71 (Subsect. 3.1)
41. R.F. Mushotzky, L.L. Cowie, A.J. Barger, et al., 2000, Nature, 404, 459
42. B.M. Peterson, 1997, An Introduction to Active Galactic Nuclei (Cambridge University Press, Cambridge), pp. 183–193
43. G. Risaliti, R. Maiolino, M. Salvati, 1999, ApJ, 522, 157
44. H.-W. Rix, M. Barden, S.V.W. Beckwith, et al., 2004, ApJS, 152, 163
45. M. Schmidt, 1968, ApJ, 151, 393
46. D.P. Schneider, X. Fan, P.B. Hall, et al., 2003, AJ, 126, 2579
47. P. Severgnini, A. Caccianiga, V. Braito, et al., 2003, A&A, 406, 483
48. I.A.G. Snellen, P.N. Best, 2001, MNRAS, 328, 897
49. A.T. Steffen, A.J. Barger, L.L. Cowie, et al., 2003, ApJ, 596, L23
50. G.P. Szokoly, J. Bergeron, G. Hasinger, et al., 2004, ApJS, 155, 271
51. Y. Ueda, M. Akiyama, K. Ohta, et al., 2003, ApJ, 598, 886
52. C. Vignali, F.E. Bauer, D.M. Alexander, et al., 2002, ApJ, 580, L105
53. C. Vignali, W.N. Brandt, D.P. Schneider, 2003, AJ, 125, 433
54. C. Vignali, W.N. Brandt, D.P. Schneider, et al., 2003, AJ, 125, 2876
55. C. Vignali, W.N. Brandt, D.P. Schneider, et al., 2005, AJ, 129, 2519

56. M.C. Weisskopf, H.D. Tananbaum, L.P. Van Speybroeck, et al., 2000, Proc. SPIE, 4012, 2
57. R.E. Williams, B. Blacker, M. Dickinson, et al., 1996, AJ, 112, 1335
58. C. Wolf, L. Wisotzki, A. Borch, et al., 2003, A&A, 408, 499
59. D.G. York, J. Adelman, J.E. Anderson, et al., 2000, AJ, 120, 1579
60. F. Yuan & R. Narayan, 2004, ApJ, 612, 724

The Links Between AGN and Galaxy Formation

O. Almaini

Department of Physics and Astronomy, University of Nottingham, Nottingham, NG7 2RD, UK
`omar.almaini@nottingham.ac.uk`

In this chapter I review the growing evidence that Active Galactic Nuclei (AGN) may play a critical role in the process of galaxy formation. I begin by summarising the observational evidence for a tight link between the formation of supermassive black holes and their hosts. I then review our current understanding for how galaxies form within the paradigm of hierarchical structure formation. Some of the areas in which these models fail, particularly the need to rapidly quench star formation in the most massive galaxies, might be explained by massive energy input from AGN. This input may take place directly (through AGN driven winds) or by the subtle effects of gentle mechanical heating in a galaxy cluster.

1 Supermassive Black Holes in Local Galaxies

1.1 The Black Hole Paradigm

The evidence for a supermassive black hole as the power source for AGN activity is now overwhelming [38]. For over 30 years the best evidence came from the rapid variability observed in the continuum emission from AGN, which strongly argued for a very compact central power source. The most rapid variations are observed in the X-ray waveband, where large amplitude changes can occur on timescales of minutes, suggesting a central engine roughly the same size as our solar system. Given the enormous power output from AGN, often several hundred times the entire bolometric output of their host galaxies, an incredibly compact and efficient energy generating mechanism must be responsible. A gravitational power source provides a natural explanation, based on liberating the potential energy of matter falling onto a supermassive black hole.

The radio jets observed in many AGN provide additional evidence for the presence of supermassive black holes. Many are collimated and aligned to tremendous precision over many millions of light years, strongly suggesting a gyroscopically-stable power source (such as a rotating black hole). The

O. Almaini: *The Links Between AGN and Galaxy Formation*, Lect. Notes Phys. **693**, 211–228 (2006)
DOI 10.1007/3-540-34621-X_8

observed relativistic velocities of many knots in these jets are also difficult to explain without a gravitationally-powered central engine.

By the 1970s, the accumulation of this evidence led to general acceptance of the current paradigm in which AGN are powered by material accreting onto supermassive black holes. The only viable alternative was to suggest a central power-house driven by a succession of nuclear supernova explosions. Such models can explain many of the characteristic features of the AGN phenomenon, but they have great difficulty in two important areas. One is in explaining the observed jet activity in AGN. The other is in explaining the rapid variability, particularly variations on timescales of minutes in the X-ray waveband.

Further evidence emerged in the 1990s, when X-ray observations with the *ASCA* satellite revealed very broad, highly-skewed iron emission lines in a number of local AGN. This was arguably the first direct evidence for emission from the immediate vicinity of a supermassive black hole, with relativistic effects providing a natural explanation for the skewed shape and broad red wing on these iron lines. More recent observations with the Chandra and XMM-Newton observatories have shown that the interpretation of these X-ray spectra is not entirely straightforward, but in at a least a few key AGN the evidence for a broad, relativistically-skewed X-ray iron line remains very strong.

Arguably the strongest evidence now comes from observations of our own Milky Way. The radio source Sgr A^* has long been thought to represent the location of a supermassive black hole. High-resolution infrared speckle imaging in the vicinity of Sgr A^* has revealed stars orbiting in clear Keplerian orbits around the nucleus, with extreme velocities of up to 1350 km s^{-1}. The acceleration of these stars in particular provides very strong constraints on the central enclosed mass, which is found to be $2.6 \pm 0.2 \times 10^6$ $M_\odot$ within the central 0.02 pc [21, 20]. A central black hole provides the only plausible explanation.

1.2 The Ubiquity of Black Holes

Active galaxies were once considered rather exotic, unusual phenomena, but observations of the luminosity function of AGN at high redshift (see contribution by Brandt) revealed that quasar activity was much more common in the past. This led to a considerable debate: did all galaxies once go through a relatively short-lived quasar phase (for 107–108 years)? Searches for supermassive black holes in local galaxies have now essentially answered this question.

Over the last 15 years it has become clear that almost all galaxies contain a supermassive black hole at their nucleus. The evidence has come from a mixture of ground based studies of stellar and gas dynamics in the inner regions of local galaxies. By the early 1990s ground-based dynamical searches found strong evidence for central dark objects in approximately 10 local galaxies,

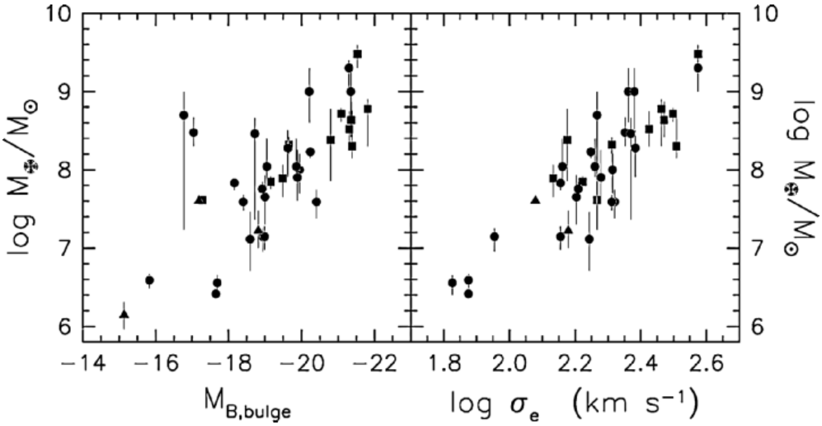

Fig. 1. The observed correlation between black hole mass with the absolute magnitude of the bulge component of the host galaxy (*left*) and the luminosity-weighted bulge velocity dispersion (*right*). *Filled circles* indicate measurements from stellar dynamics, *squares* are based on ionised gas dynamics and *triangles* from maser disk dynamics. Taken from [29]. (Reproduced by permission of the AAS)

based on a mixture of stellar and gas dynamics. By far the tightest constraint came from a study of H_2O maser emission in the LINER galaxy NGC4258. Radio observations (including high-resolution milli-arcsec observations with the Very Long Baseline Array) reveal a clear Keplerian rotation curve, consistent with gas rotating in a disk around a central mass of $M_\bullet = 4.1 \times 107\mathrm{M}_\odot$, strongly constrained to reside within the central 0.005 arcsec (= 0.18pc).

Ground based studies of local galaxies were soon followed up by high resolution observations with the Hubble Space Telescope. Several dozen massive galaxies have been observed so far, all of which show evidence for a massive central dark object. Furthermore, one can now compare the present day black hole mass density with the integrated accretion from AGN over the history of the Universe (estimated from either the X-ray background or the AGN luminosity function). The results agree remarkably well (to within a factor of two), suggesting that most of the build-up of black hole mass has actually taken place through luminous accretion. Every galaxy must therefore have been through at least one quasar epoch at some stage in its lifetime.

As the number of galaxies with measured black holes began to grow, by the late 1990s two intriguing correlations emerged (see Fig. 1) [29]. The first was the discovery of a correlation between the mass of the black hole and the luminosity of the bulge component of the host galaxy [32]. The second was an apparently much tighter relationship between the black hole mass ($M_\bullet$) and the luminosity-weighted velocity dispersion (σ_*) of the host bulge [17, 19]:

$$M_\bullet = 1.3 \times 108 \; M_\odot \left(\frac{\sigma_*}{200 \text{ km s}^{-1}} \right)^{3.65} \tag{1}$$

The expectation was for only a crude relationship between these quantities, related to the availability of fuel feeding both the AGN and star-forming processes, but it would be fair to say that the tightness of the $M_\bullet - \sigma_*$ relationship was an unexpected discovery. Remarkably, this is even tighter than the observed relationship between the stellar velocity dispersion and the bulge luminosity.

This relation is particularly surprising given the vast difference in scale between the black hole (light-minutes in diameter) and the host galaxy bulge (thousands of light-years in diameter). Furthermore, the galaxies used in these studies are an inhomogeneous collection of objects, likely to have very different star-formation and merging histories. Under these circumstances the observed relationship is strongly suggestive of a much more fundamental link between black hole growth and galaxy formation. The fact that the black hole mass scales with only the mass of the bulge component (rather than the total galaxy mass, or the mass of the dark-matter halo) presents a rather important clue.

But does the black hole influence the formation of its host, or vice versa? This remains unclear, but a growing body of evidence is emerging to suggest that AGN can have a major influence on their environment. In this review I explore the current paradigm of hierarchical galaxy formation with the central black hole in mind. One of the major problems is in producing the most massive galaxies early enough in the Universe and then abruptly terminating any subsequent star formation. AGN may offer the missing ingredient.

2 Models for Galaxy Formation

2.1 The Hierarchical Paradigm

Our current understanding for the growth of structure in the Universe is a hierarchical model. In this scenario, small objects form first and then merge and collapse to form larger structures. This hierarchical picture emerges as a direct consequence of the currently-favoured *Cold Dark Matter* (CDM) cosmology (recently modified by the addition of a "dark energy" component). At the present time the best-fitting cosmological model has the following properties: (i) ordinary matter making up approximately one third of the critical density of the Universe ($\Omega_m \simeq 0.3$); (ii) a "dark energy" component ($\Omega_X \simeq 0.7$, which may be a cosmological constant); (iii) baryons making up only a very small fraction of the mass of the Universe ($\Omega_b \simeq 0.04$); (iv) a present-day Hubble constant $H_0 \simeq 65$ km s^{-1} Mpc^{-1}; (v) present-day mass fluctuations on a scale of 8 Mpc of the order $\sigma_8 \simeq 0.9$.

A key feature of this model is that the baryonic matter is dominated by a non-interacting component with random velocities too small to affect structure formation ("cold dark matter"). In such a Universe, structure forms as overdensities grow slowly over time by gradual gravitational contraction.

The resulting dark matter haloes then virialize and eventually merge, leading to the build-up of structure in a hierarchical fashion.

The hierarchical CDM model for our Universe appears be very strongly confirmed by a wide range of observational evidence. These range from observations of the spectrum of fluctuations in the Cosmic Microwave Background, the cosmological distance scale, the space density and baryonic fraction in galaxy clusters, the clustering of galaxies on large scales (e.g. from the 2DF and SLOAN surveys), the observation of distant type Ia supernovae, the age of the oldest stars and the abundance of light elements. In fact, the CDM + Dark Energy model is now in such good agreement with observations that many cosmologists now place this picture on equal footing with the Big Bang model itself as the fundamental description of the constituents of our Universe.

2.2 Semi-Analytic Models and the Messy Process of Galaxy Formation

While the basic hierarchical paradigm for the evolution of dark matter haloes is widely accepted, our understanding of the rather more complex process of galaxy formation is still far from complete. The N-body simulations conducted so far have largely concentrated on the behaviour of the dark matter haloes. Instead, the approach taken is a "semi-analytical" compromise, which generally involves taking the dark matter haloes from N-body ΛCDM simulations and then adding basic prescriptions for the gas physics, radiative cooling, complex processes of star formation, the output initial mass function, supernovae feedback, dust, enrichment etc., attempting to reproduce the space density, luminosities, colours and other properties of the observed galaxy population [9]. Given the wide range of poorly understood phenomena, it is perhaps somewhat surprising that semi-analytic models succeed as well as they do!

The current crop of models have great success in reproducing the local optical and infrared galaxy luminosity functions over a wide range in luminosity. They can also reproduce the general form of the Tully-Fisher relationship between the luminosity and rotational velocity in spiral galaxies. Going back in cosmological time, the integrated global history of star formation (e.g. the general integrated form of the "Madau" diagram) is also successfully reproduced, as is the general form for the build up of stellar mass (for optically-selected galaxies) from $z \simeq 5$ to the present day.

There are important "classic" problems with semi-analytic models, however. The first is the so-called "dwarf" problem, namely that the models predict far too many dwarf galaxies than are observed in the local neighbourhood [25]. This is also known as the "sub-structure" problem. Recent work suggests that this may be at least partially overcome by the effects of a photoionising background, which heats up gas in the smallest haloes and suppresses gas cooling, but the issue is far from resolved. Another is the "cuspy core" problem, which arises because of the "cold" nature of cold dark matter. The cores of galaxies should have much more sharply-peaked profiles than are actually observed. The transfer of angular momentum from the baryonic matter to the central dark matter has been suggested as a possible solution, perhaps through an interaction with a central bar, but this also remains a serious issue.

The third problem is in reproducing the sharp turn over in the galaxy luminosity function above L^*. What is the physical mechanism that appears to restrict the formation of the most massive galaxies? Why are there no galaxies in the Universe with velocity dispersions above $\sim$300 km s^{-1}?

Related to this, recent pioneering work by Kauffmann and collaborators [27, 28] has revealed a striking bimodality in local galaxy properties. Divided rather sharply at $M^* \simeq 3 \times 10^{10}$ M$_\odot$, corresponding to a virial velocity of $\simeq$ 100 km s^{-1}, this is comparable to the L^* of the Schechter luminosity function. Below this value local galaxies tend to be disky galaxies which have undergone recent bursts of star formation. Above M^*, local galaxies are dominated by old, red spheroidal systems. Semi-analytic models have great difficulty in producing such a bimodal distribution. In particular, a mechanism is required for switching off star-formation in the most massive galaxies at early times.

In recent years, related problems have emerged which relate to the ability of semi-analytic models to produce massive and/or old galaxies at high redshift. In particular, they fail rather badly in producing two important new classes of high-redshift galaxy: the old, red spheroids at $z = 1 - 2$ and the highly luminous, dust enshrouded galaxies recently detected in sub-mm surveys.

3 Observations of the High-Redshift Universe

3.1 Massive Galaxies at High Redshift

Recent deep infrared surveys have revealed a large population of distant galaxies with very red colours [11, 30, 33]. These "extremely red objects" (EROs) are classified on the basis of their optical to infrared colours (e.g. $(R - K)_{\mathrm{Vega}} > 5$). A large fraction of these galaxies appear to be old, passively evolving spheroidal systems at redshifts $0.8 < z < 1.8$, which appear very red as the 4000Å break (indicative of an old stellar population) falls between the two observed passbands. They are also very strongly clustered, with clustering strengths comparable to present-day elliptical galaxies [40], suggesting that these are comparably massive systems [36]. Semi-analytic models fail to produce the observed space density of these galaxies by almost an order of magnitude (Fig. 2). Very recently, the study of EROs has been extended to the $z > 2$ epoch by a combination of ultradeep imaging and

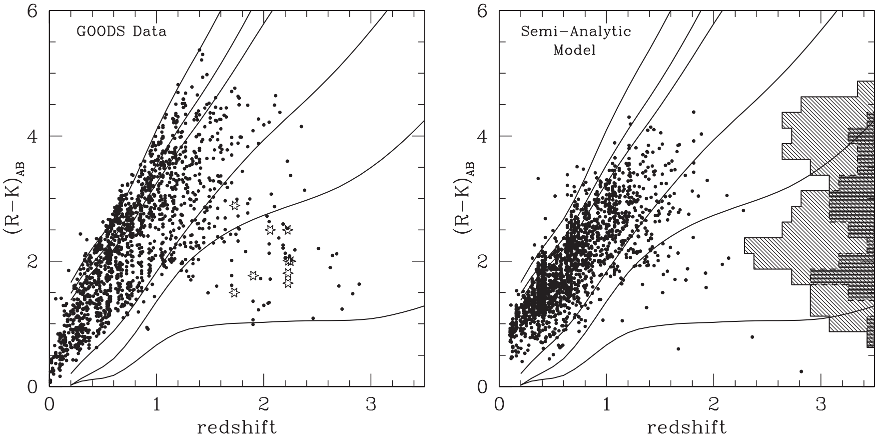

Fig. 2. Observed $(R-K)_{AB}$ colour vs redshift for the *GOODS* CDFS [*left*] and the semi-analytic mock catalog [*right*]. Tracks are shown for single age, stellar metallicity populations with ages of 13.5, 5.8, 3.2, 1, 0.5, and 0.1 Gyr (unreddened), from *top* to *bottom*. The color distribution of galaxies with $K < 22$ in the redshift range $1.7 < z < 2.5$ is shown in the right panel as histograms (diagonal hatched: *GOODS*; horizontal fill: mock catalog). Taken from [43] (Reproduced by permission of the AAS)

$J - K$ colour selection [18]. The discovery of old passive systems at $z > 2$ may ultimately place even tighter constraints on galaxy formation models.

On a similar timescale, surveys in the sub-mm waveband have revealed a population of highly luminous, dust enshrouded galaxies [4, 13, 24]. Originally discovered by the SCUBA array at the James Clerk Maxwell Telescope [41, 42], these sources represent enormous bursts of star formation of $\sim 1000\,M_{\odot}/\mathrm{yr}^{-1}$ [31], although progress beyond their initial discovery has been limited. This is due to an unfortunate combination of their optical faintness and the relatively large 10–15 arcsec SCUBA beam, which makes unambiguous identification extremely difficult [35]. This deadlock has recently been partially overcome with deep radio observations at the VLA, which have yielded precise locations for significant samples of sub-mm galaxies for the first time. Follow-up spectroscopy by Chapman and collaborators [7] has produced the first reliable $N(z)$ estimate for the population of SCUBA galaxies, showing a median redshift of $z = 2.4$ with only a small fraction (a few per cent) of galaxies at $z < 1$. The sub-mm sources therefore appear to be high-z analogues of the Ultra-Luminous Infrared Galaxies (ULIRGs) which have been observed in the local Universe, but unlike ULIRGs (which are very rare and unusual events) the sub-mm galaxies may actually dominate the star-formation at $z > 2$. Semi-analytic models have great difficulty producing so many luminous objects at these redshifts [43] without resorting to extreme model parameters (Lacey, private communication). Furthermore,

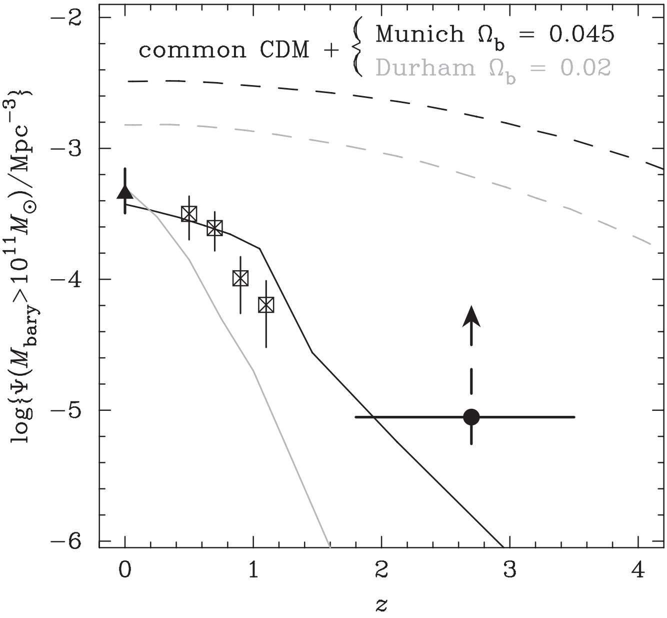

Fig. 3. Comoving number densities of galaxies with baryonic masses $\geq 10^{11}\ M_\odot$ as a function of redshift. The *triangle* and *open squares* show densities of massive stellar systems at $z = 0$ [10] and $z \sim 1$ [12]; the *filled circle* shows the density for massive sub-mm galaxies at $z \sim 2.7$, with a factor of 7 correction for burst lifetime. *Solid curves* show the predictions of semi-analytic modeling by the Munich group [26] and the Durham group (an updated version of [5]); *dashed curves* show the corresponding number densities of haloes with *available* baryonic masses $\geq 10^{11}\ M_\odot$. Taken from [44]. (Reproduced by permission of the AAS)

recent observations of sub-mm galaxies using the Spectrometer for Infrared Faint Field Imaging (SPIFFI) at the VLT have produced the first reliable mass estimates for these galaxies [44]. The spatially-resolved gas kinematics implies very large baryonic masses of $\sim 10^{11} \mathrm{M}_\odot$. The inferred space density of such massive systems at $z > 2$ is substantially higher than current semi-analytic models predict, although well within the number of dark matter haloes with enough *available* baryonic mass at these epochs (Fig. 3).

It seems likely that the two populations are related, and that the passively-evolving ERO galaxies are the remnants of the explosive bursts of star-formation observed in the sub-mm at $z > 2$. What seems to be required to reconcile the theoretical models is much more rapid star-formation in the most massive galaxies at high-redshift (possibly fueled by more gas-rich merging?) and then some mechanism to rapidly quench it. Feedback from the central AGN is one attractive possibility, which could also be related to the mechanism required to produce a sharp turn-over in the galaxy luminosity function above L^* [6].

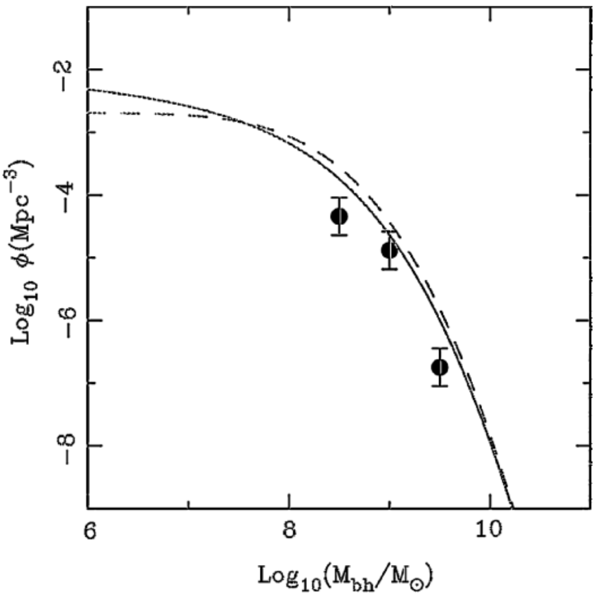

Fig. 4. Mass function of black holes at $z \simeq 2$, based on virial black hole estimates from the SLOAN quasar survey. The *solid* and *dashed lines* represent two estimates of the $z = 0$ black hole mass function. Adapted from [34]

3.2 Massive Black Holes at High Redshift

Related to the high-redshift galaxy populations described above, it is also now emerging that the most massive black holes were also in place very early on in the Universe. The first determination of the black hole mass function at high redshift suggests that most black holes above a mass of $10^{8-9} M_{\odot}$ were already in place by $z \simeq 2$ [34], mirroring the early formation for their host galaxies (see Fig. 4).

Recent determinations of the AGN X-ray luminosity function (see review by Brandt) show that these observations are consistent with a global "cosmic-down-sizing" effect, in which the most massive black holes peak in their activity at substantially earlier epochs than their smaller, less luminous counterparts. Current semi-analytic models have great difficulty in reproducing the rapid decline in activity for the most luminous quasars, which may be related to the difficulties outlined above in quenching star-formation in the most massive bulges.

3.3 Can We Observe Joint Black-Hole/Bulge Formation?

The local $M_{\bullet}/\sigma_{*}$ relation points to a tightly-linked formation process, so one might expect the most luminous newly-forming spheroids (e.g. the sub-mm

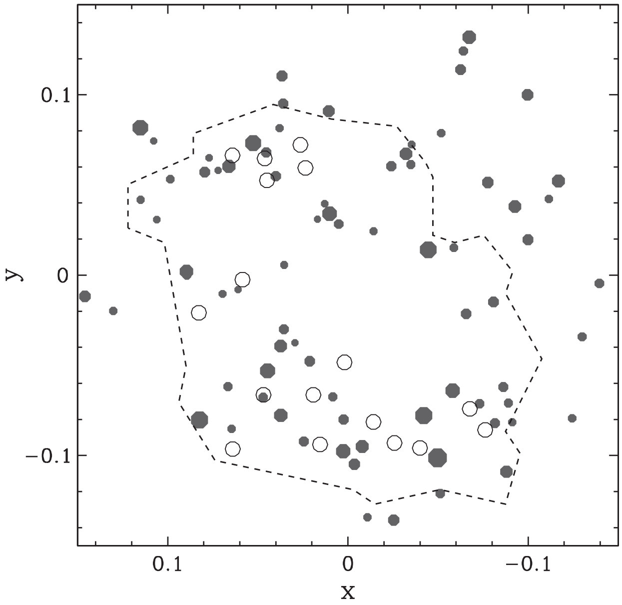

Fig. 5. Illustrating the distribution of Chandra sources (filled points) and SCUBA sources (*open circles*) in the ELAIS N2 field. Only one of the 17 SCUBA sources is detected by Chandra, strongly ruling out the possibility that the remainder contain luminous quasars unless they are heavily obscured by Compton-thick material. Taken from [3]

sources) to contain powerful accreting black holes. Somewhat surprisingly, however, X-ray observations of the most luminous sub-mm sources [3, 14, 23] suggest that the majority are *not* hosting powerful quasars (Fig. 5). Are the central black holes dormant, still growing, or heavily obscured by Compton-thick material? This remains unclear, and further observational work will be required to study the black-hole activity in these sub-mm sources. However, the deepest (2Ms) Chandra observations of the Hubble Deep Field SCUBA sources [1] have recently detected a large fraction as moderate-luminosity (possibly absorbed) AGN, which may favour the "still-growing" hypothesis.

4 The Impact of AGN on their Environments

4.1 Outflows and Winds

In this section I review some of the mechanisms by which AGN are observed to transfer energy to their immediate environment. The potential importance of the black hole can be readily illustrated when one considers the total energy generated by a quasar over its lifetime. If we assume growth by accretion to a total mass M with the usual efficiency μ ($\sim$0.1), folding in the typical black-hole/bulge mass ratio of $\sim$0.1%, the energy released by the quasar is therefore $10^{-4}M_{bulge}c2$. If we compare this to the binding energy of a galaxy bulge, $\frac{1}{2}M_{bulge}\sigma 2 \simeq 10^{-6}M_{bulge}\sigma^2_{300}c2$, we find that for a typical massive bulge (with a velocity dispersion of 300 km s^{-1}) only 1% of the energy generated by accretion would be sufficient to completely destroy the host. The key issue is to determine what fraction of the quasar energy is transferred to the host galaxy (e.g. by winds) compared to that which escapes as radiation.

There is now growing evidence that large-scale winds and bulk outflows are very common, if not ubiquitous, in active galaxies. These are most easily observed in the population of Broad Absorption Line (BAL) quasars, which represent at least 10% of the quasar population. These show deep, broad absorption lines from highly ionised UV resonance transitions. These blue-shifted absorption features are readily explained by massive outflowing winds with velocities of 0.1–0.3c intersecting our line of sight.

More recently, sensitive X-ray observations with the Chandra and XMM-Newton telescopes have routinely detected X-ray outflows in BAL quasars [39]. Since X-rays are generated in the innermost region of the central engine, these observations can potentially offer new insights into the wind phenomena at smaller radii. Among these is the high-redshift (z = 3.91) quasar APM 08279+5255, which is boosted in flux by a factor of $\sim$100 by gravitational lensing (Fig. 6). Chartas and collaborators [8] have observed this system with

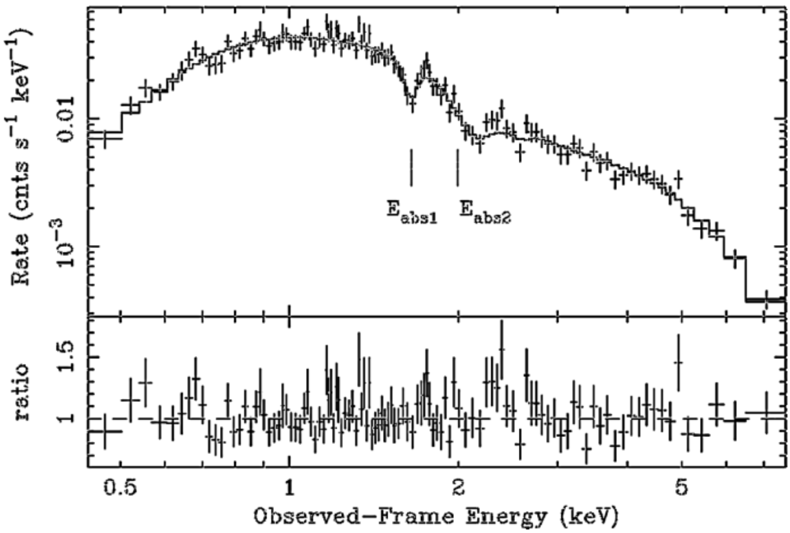

Fig. 6. Chandra X-ray spectrum of the z = 3.91 gravitationally-lensed quasar APM 08279+5255. The two absorption features appear to be relativistically blue-shifted Fe XXV Kα, suggested high column-density outflows at $\sim$0.2 and 0.4c. Taken from [8] (Reproduced by permission of the AAS)

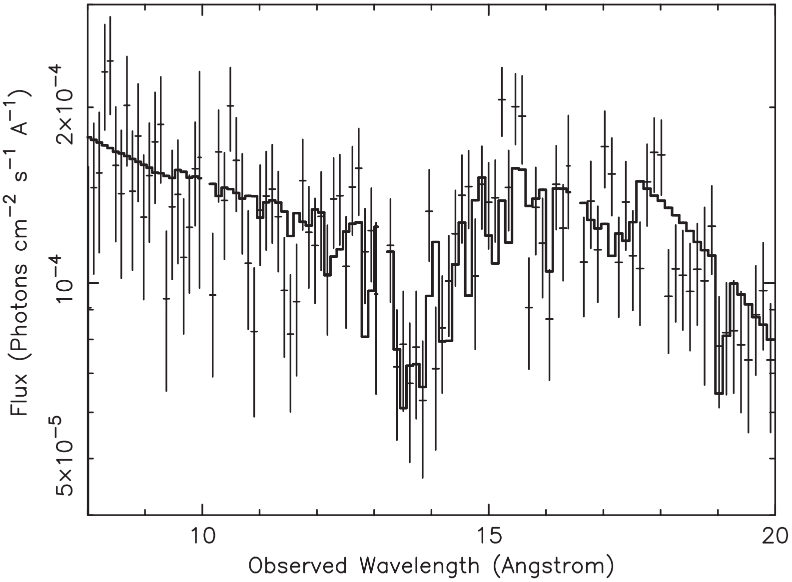

Fig. 7. XMM-Newton RGS-2 spectrum of the $z = 0.184$ quasar PDS 456. A strong absorption trough is present at 12-15 Å, observed frame, suggesting a high outflow velocity of ∼50000 km s^{-1} from L-shell iron transitions in the ionization range Fe XVII – Fe XXIV. Taken from [39]

Chandra and discovered relativistic, broad absorption lines corresponding to wind speeds of ∼0.2 and ∼0.4c.

Another is the local quasar PDS456, a particularly luminous local system with a very high accretion rate. As such it may be a close analogue of growing black holes at high-redshift, which ought to be accreting at close to their maximum rate as they accrete in the gas-rich environment of a newly-forming galaxy. PDS456 was observed by [39] with XMM-Newton, revealing relativistic broad X-ray outflows consistent with an enormous mass outflow rate of $\sim 10 M_{\odot} yr^{-1}$ (Fig. 7). The implied momentum outflow is comparable to that carried in the Eddington-limited radiation field.

Future X-ray mission may enable us to determine how common such outflows are in quasars in the crucial $z > 2$ epoch.

4.2 The Impact of AGN in Galaxy Clusters

There is strong evidence that the formation and evolution of galaxy *clusters* may also be heavily influenced by AGN activity. Very simple arguments show that the cooling times (due to Bremsstrahlung and line radiation) in the central regions of a massive galaxy cluster ought to be comparable to the age of the cluster. As the gas cools, the resulting drop in central temperature and pressure should lead to a massive infall of material (a "cooling-flow") of up to

1000 $M_\odot yr^{-1}$. Gas cooling below ∼1 keV produces distinctive strong emission lines from cool species, but Chandra and XMM-Newton observations in recent years have failed to observe these emission lines at the predicted strengths (e.g. [16, 37]. It therefore appears that gas is not cooling as expected, and hence the "cooling-flow" picture appears incomplete. This is potentially of great importance, since our whole understanding of the process of gas cooling may therefore be fundamentally flawed.

The most obvious solution to this problem is to invoke additional heat input, for which a very likely source is the central AGN. Since most galaxy clusters contain very massive central cluster galaxies, the associated black holes are potentially an enormous source of energy. Further circumstantial evidence for AGN heat input comes from the excess entropy observed in many galaxy clusters. In addition, the relationship between the X-ray luminosity and temperature is much steeper than predicted by hierarchical merging models, strongly suggesting the need for additional heat input.

There have been recent arguments that the energy input from the AGN must be gradual and gentle, since the excess entropy appears to be spread over large radii. Furthermore, direct X-ray observations show few examples of shocks in the vicinity of the central AGN. Instead, many researchers now believe that the heat could be transferred gradually by buoyant bubbles and sound waves. It has been known since the time of ROSAT that the plasma in radio lobes can displace the hot thermal gas, producing holes in the X-ray emission. Recent Chandra observations [15] have shown that such cavities are very common (Fig. 8). These bubbles of buoyant gas should rise through the intra-cluster medium and inject substantial thermodynamic pdV work as they move through the intra-cluster medium. This should transfer heat through sound waves, direct evidence for which may have recently been detected in the Perseus and Virgo clusters.

Recent theoretical modeling and numerical simulations appear to confirm that this mechanism could be sufficient to counteract the cooling-flow, and account for the excess of entropy observed at large radii in present-day clusters. Whatever the exact mechanism, it now seems very likely that AGN have a major influence on the cooling rate in largest dark matter haloes. Black holes may therefore play a very important role in the formation and evolution of some of the largest bound structures in the Universe. This could also be related to the long-sought mechanism responsible for the sharp turn over at the high-mass end of the *galaxy* luminosity function, suppressing the formation of the most massive galaxies by restricting their gas cooling.

5 Models for the Black-Hole/Bulge Relation

The discovery of the tight black-hole/bulge mass relation led to a flurry of theoretical papers. It is beyond the scope of this review to do justice to them all, so below I review only the two main categories of models which have

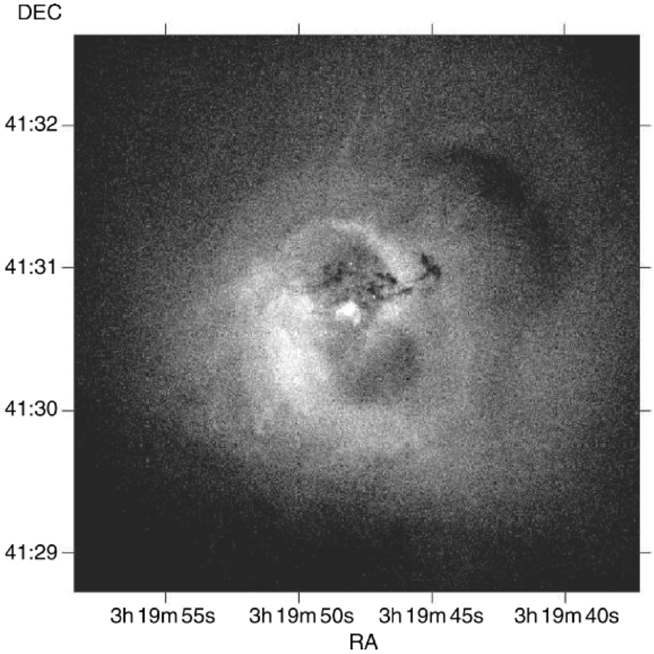

Fig. 8. Chandra X-ray image of the central region of the Perseus cluster in the 0.3 to 1.5 keV band. Pixels are 0.49 arcsec in dimension. North is to the top of this image. Cavities evacuated by the central radio lobes are clearly visible. Taken from [15]

gained most acceptance in the literature. It should be made clear that these are by no means the only plausible models. In particular there has been very interesting recent work on a model which relies on radiation drag from bulge stars to extract angular momentum and to fuel the central black hole, while another is based on the likelihood of stellar capture by the accretion disk. Details of these and other (more exotic) models can be readily obtained with a literature search.

The first major class of models is based on the available supply of (low angular-momentum) gas to feed the central black hole. Relating this to the fueling and growth of the galaxy bulge is achieved within the context of hierarchical merging in semi-analytic models. These methods do succeed in explaining the observed $M_\bullet - \sigma_*$ relations, but they require a certain amount of fine tuning, particularly regarding the black hole fueling rate as a function of redshift. Furthermore, as described in Subsect. 1.2, in any realistic scenario it is very hard to explain the *tightness* of this relation given the huge difference in scales between the black hole and the bulge, and the wide range of environments and galaxy merger histories. Current semi-analytic models also have fundamental problems in producing massive galaxies early enough, as described in Sect. 3.

Another class of models make use of black hole winds to regulate the gas fueling. Originally proposed by Silk & Rees (1998), these models have since been refined and modified by Blandford, Fabian, Haehnelt, King and other authors. I will briefly outline the model by King which is the latest incarnation of this class. It also has the additional nice feature that it effectively contains no free parameters. The critical new ingredient is a prescription for the mass outflow rate for a black hole in the final stages of growth. King shows that such outflows are a natural consequence of super-Eddington accretion. The Eddington limit restricts the amount of material which enters the black hole, so it is assumed that any excess material is expelled in the form of a wind. It can be shown that the expected outflow rate is given by:

$$\dot{M} v_{\mathrm{out}} \simeq \frac{L_{\mathrm{Edd}}}{c} \tag{2}$$

Intriguingly, this is precisely the outflow momentum observed in a number of local X-ray BALs (see Subsect. 4.1). These outflows are then assumed to drive a wind bubble out through the centre of the galaxy, with an analogous treatment to the theory of stellar wind bubbles. In solving the equation of motion for the outflowing shell, the swept up gas as a function of radius can be related to the bulge velocity dispersion (since $M_{\mathrm{gas}}(r) \propto M_{\mathrm{total}}(r) \propto \sigma_*^2 R$) and this leads naturally to a predicted shell velocity:

$$v2_{\mathrm{shell}} \propto L_{\mathrm{Edd}}/\sigma_*^2 \tag{3}$$

Thus as the black hole grows, its Eddington luminosity grows in proportion and therefore the outflow wind velocity increases. Eventually the wind velocity reaches the escape velocity of the bulge ($v_{esc} = \sigma$) and the gas is expelled. Any subsequent black hole (or bulge) growth is then terminated. At this point, inspection of the equation above reveals that the final black hole mass will relate to the bulge velocity dispersion by $M_\bullet \propto \sigma_*^4$ (since $M_\bullet \propto L_{\mathrm{Edd}}$) which is very close to the observed relation. A full treatment also gives a constant of proportionality which is strikingly similar to the observed value.

While far from the last word on the matter, these elegant arguments at least serve to illustrate the feasibility of producing the black-hole/bulge relation through AGN-driven outflows. Observations of the predicted wind bubble around high accretion-rate quasars would provide a strong confirmation of this picture.

6 Summary and Future Prospects

Over the last 10 years it has become clear that essentially every local galaxy contains a supermassive black hole. It would appear that most of this black-hole growth took place through luminous accretion processes, implying that every galaxy has been through at least one quasar phase. There is also growing evidence that this quasar activity may have had a profound influence on the process of galaxy formation. The surprisingly tight relationship between the black hole mass and the velocity dispersion of the host galaxy bulge strongly suggests a feedback mechanism, linking the growth of the black hole (on a scale of light-minutes) to the build up of the galaxy bulge (on scales of kpc).

On a similar timescale, it has also become clear that our models for galaxy formation are having great difficulty in several key areas. The observed bimodality in local galaxy properties is one. Why are the most massive galaxies uniformly old, red ellipticals, while below the critical M^* most galaxies are star-forming disks? Another (related) problem is in explaining the abundance of old, massive galaxies in the Universe at redshifts $z = 1 - 3$. What terminates the star-formation in these galaxies so abruptly? Why indeed is there a clear upper limit to the mass a galaxy can attain? Current semi-analytic models have great difficulty in reproducing these properties. Some mechanism is required for switching off both the star-formation and the fueling of quasars in the most massive galaxies.

Feedback from a central AGN may provide a solution to these problems. The energy produced by a supermassive black hole is certainly sufficient, so the key issue is to determine how much energy is transferred to the host galaxy by either mechanical winds, radiation pressure or some related process. There are growing indications that a large number of AGN are producing relativistic bulk outflows. If such winds are found to be ubiquitous, particularly for high accretion-rate AGN, then this would be a profound discovery. Recent theoretical models based on such AGN winds provide a rather elegant explanation for the observed black-hole/bulge mass relation while simultaneously providing a mechanism for expelling gas from a newly-forming proto-spheroid.

So far, however, the evidence for major AGN feedback is rather circumstantial, or based on plausibility arguments. It is far from clear that AGN are really having a major influence on their hosts, and the precise mechanism for regulating the growth of a massive galaxy (and its central black hole) still remains unclear. The next few years ought to be an exciting time as we attempt to disentangle the processes of quasar and galaxy formation. Among the key questions that need to be addressed:

- We need to determine the formation epoch for the most massive galaxies, and how rapidly their mass is assembled. Future major infrared surveys (such at the UKIDSS Ultra-Deep survey; see `www.ukidss.org`) may be the key to measuring these very fundamental properties.
- We need a detailed prescription for AGN feedback, ideally worked into the semi-analytic models for galaxy formation.
- We need to understand how AGN influence gas cooling in the most massive haloes. Can they explain the apparent upper limit in the stellar mass of a galaxy?

- We need a detailed investigation of relativistic outflows in AGN. How common are these outflows, and how do their properties and outflow rates depend on accretion rate? Future high-throughput X-ray missions such as XEUS and Constellation-X will be crucial for extending this work to the epoch of peak quasar activity.

The next 5–10 years will be a very exciting time. It will be a strange twist of nature if relatively tiny black holes do indeed play a key role in the growth of the largest galaxies in the Universe.

Key References

Here we note some key references which are relevant to this review. This is far from an exhaustive list. The most up-to-date reference is often deliberately chosen in preference to the original discovery paper, to enable the reader to quickly enter the literature in this fast-moving field.

References

1. Alexander, D.M. et al., 2003, AJ, 125, 383
2. Almaini, O., Lawrence, A., Boyle, B.J., 1999, MNRAS, 305, L59
3. Almaini, O. et al., 2003, 338, 303
4. Barger, A. et al., 1998, Nat, 394, 248
5. Baugh, C.M., Benson, A.J., Cole, S., Frenk, C.S., Lacey, C., 2003, in “The masses of galaxies at low and high redshift”, ed. R. Bender & A. Renzini, (Berlin: Springer Verlag), 91
6. Benson, A., 2003, ApJ, 599, 38
7. Chapman, S.C., Blain, A.W., Ivison, R.J., Smail, I., 2003, Nature, 422, 695
8. Chartas, G., Brandt, W.N., Gallagher, S.C., Garmier, G.P., 2002, ApJ, 579, 169
9. Cole, S. et al., 1994, MNRAS, 271, 781
10. Cole, S. et al., 2001, MNRAS, 326, 255
11. Daddi, E. et al., 2000, A&A, 361, 535
12. Drory, N. et al., 2001, ApJ, 561, L111
13. Eales, S. et al., 1999, ApJ, 515, 518
14. Fabian, A.C. et al., 2000, MNRAS, 315, L8
15. Fabian, A.C., Sanders, J.S., Allen, S.W., Crawford, C.S., Iwasawa, K., Johnstone, R.M., Schmidt, R.W., Taylor, G.B., 2003, MNRAS, 344, L43
16. Fabian, A.C., 2003, RMxAC, 17, 303
17. Ferrarese, L. & Merrit D., 2000, ApJL, 539, 9
18. Franx, M. et al., 2003, ApJ, 587, L79
19. Gebhardt, K. et al., 2000, ApJL, 539, 13
20. Genzel, R., Pichon, C., Eckart, A., Gerhard, O.E., Ott, T., 2000, MNRAS, 317, 348
21. Ghez, A.M., Morris, M., Becklin, E.E., Tanner, A., Kremenek, T., 2000, Nature, 407, 349

22. Granato, G.L. et al., 2001, MNRAS, 324, 757
23. Hornschemeier, A.E. et al., 2001, ApJ, 554, 742
24. Hughes, D.H. et al., 1998, Nat, 394, 241
25. Kauffmann, G. et al., 1993, MNRAS, 264, 201
26. Kauffman, G., Coldberg, J.M., Diaferio, A., White, S.D.M., 1999, MNRAS, 303, 1988
27. Kauffmann, G. et al., 2003, MNRAS, 341, 33
28. Kauffmann, G. et al., 2003, MNRAS, 341, 54
29. Kormendy, J., Gebhardt, K., 2001, AIPC, 586, 363
30. Labbé, I. et al., 2003, AJ, 125, 1107
31. Lilly, S.J. et al., 1999, ApJ, 518, 641
32. Magorrian, J. et al., 1998, AJ, 115, 2285
33. Moustakas, L.A. et al., 2004, ApJ, 600, L131
34. McLure, R.J., Dunlop, J.S., 2004, MNRAS, 352, 1390
35. Neri, R. et al., 2003, ApJ, 597, L113
36. Percival, W.J., Scott, D., Peacock, J.A., Dunlop, J.S., 2003, MNRAS, 338, L31
37. Peterson, J.R., Kahn, S.M., Paerels, F.B.S., Kaastra, J.S., Tamura, T., Bleeker, J.A.M., Ferrigno, C., Jernigan, J.G., 2003, ApJ, 590, 207
38. Rees, M.J., 1984, ARA&A, 22, 471
39. Reeves, J.N., O'Brien, P.T., Ward, M.J., 2003, ApJ, 593, 65
40. Roche, N.D., Almaini, O., Dunlop, J., Ivison, R.J., Willott, C.J., 2002, MNRAS, 337, 1282
41. Smail, I., Ivison, R.J., Blain, A.W., 1997, ApJ, 490, L5.
42. Smail, I., Ivison, R.J., Blain, A.W., Kneib, J.P., 2002, MNRAS, 331, 495
43. Somerville, R.S. et al., 2004, ApJ, 600, L135
44. Tecza, M. et al., 2004, ApJ, 605, L198

Appendix: Acronyms Used in this Book

AGN	Active Galactic Nuclei
ALMA	Atacama Large Millimeter Array
ATCA	Australian Telescope Compact Array
BAL-QSO	Broad Absorption QSO
BAL-Quasar	Broad Absorption Quasar
BH	Black Hole
BLR	Broad Line Region
CCD	Charge Coupled Device
CDF-N(S)	Chandra Deep Field North (South)
CDM	Cold Dark Matter
CMB	Cosmic Microwave Background
CSO	Caltech Submillimeter Observatory
CSS	Compact Steep Spectrum
EELR	Extended Emission Line Region
ENLR	Extended Narrow Line Region
ERO	Extremely Red Object
FRI(II)	Faranov and Riley I (II)
GPS	GHz Peaked Spectrum
HDF-N(S)	Hubble Deep Field North (South)
HST	Hubble Space Telescope
IGM	Inter Galactic Medium
IILR	Inner ILR
ILR	Inner Lindbland Resonance
IMBH	Intermediate Mass Black Hole
IPCS	Image Photon Counting System
ISM	Inter Stellar Medium
LDDE	Luminosity-Dependent Density Evolution
LINER	Low Ionization Nuclear Emission-line Region
LOC	Locally Optimally-emitting Cloud
LOS	Line Of Sight
MDO	Massive Dark Object
MERLIN	Multi Element Radio Linked Interferometer Network
MIR	Mid-Infrared
NFW	Navarro-Frenck-White
NIR	Near-Infrared

NLR	Narrow Line Region
NLS1	Narrow Line Seyfert 1
OILR	Outer ILR
PDE	Pure Density Evolution
PLE	Pure Luminosity Evolution
Quasar	Quasi Stellar Radio Source
QSO	Quasi Stellar Object
SCUBA	Submillimetre Common-User Bolometer Array
SED	Spectral Energy Distribution
SFR	Star Formation Rate
SMBH	Super Massive Black Hole
SPH	Smooth-Particle Hydrodynamics
SPIFFI	SPectrometer for Infrared Faint Field Imaging
SSC	Synchrotron Self-Compton
UKIDSS	UKIRT Infrared Deep Sky Survey
UKIRT	United Kingdom IR Telescope
ULIRG	Ultra Luminous IR Galaxy
VLA	Very Large Array
VLBA	Very Long Baseline Array
VLBI	Very Long Baseline Interferometry
VLTI	Very Large Telescope Interferometer
XBONG	X-ray Bright, Optically Normal Galaxy
XRB	X-ray Background
XSPEC	X-ray Spectral fitting package

Index

Absorption
  broad absorption lines 30, 31, 221
  free-free 109
Acceleration of particles 62
Accretion
  efficiency 3, 147
  flow 143
  spherical, Bondi 3, 101
Accretion disk 4–10, 35
  ADAF 11
  corona 10
  geometry 4, 8, 10
  spectrum 9
  structure 4, 8
  temperature 6, 7
  viscosity 7
Active galactic nuclei (AGN)
  classification 1, 52, 216
  density of 201
  environmental impact 222
  evolution 196
  unified model 25, 62, 101, 102, 111, 114
Angular momentum
  release 174
  specific 149, 174
Auger ionization 14

Baldwin relation 80
Bars
  large-scale 151, 160
  nuclear 169
BL Lac objects 63, 67
Black holes
  evidence for 77, 185, 211, 212
  formation 219, 223
  mass 89, 144, 213
  mass density 151
  mass-sigma relation 146, 213, 219, 224
  properties 1, 212
Blazars 70
Bloated stars 19, 33, 92
Broad line region (BLR) 25, 77
  model 91
  size 86

Cen A 54, 65–67
Charge exchange 14
Clumpiness 115
CO observations 104, 117
Collisional
  excitation 16
  ionization 14, 15
Confinement 19, 25, 26, 64
Continuum emission 17
Cooling flow 222
Cygnus A 64

Dielectronic recombination 14
Doppler factor 47
Dust
  in BLR/NLR 27
  in torus 102, 104, 109, 114, 115
  silicate feature 114
  sublimation 27, 103, 115
  temperature 103, 115
Dynamical friction 150, 174

Eddington
  accretion rate 3
  luminosity 2, 81
  time 3
Emission lines
  coronal 27
  profiles BLR/NLR 26, 92–95

Equipartition 57, 59

Faraday rotation 60, 67
Faranov-Riley classes (FRI, FRII) 52, 54
Fluorescence (line and continuum) 18, 19
Force multiplier 23
Free-free absorption 109
Fueling, growth 2, 101, 143, 148, 151, 160, 169

Galaxy formation 214, 217
Gaunt factor 44

H2 emission 104
HI absorption 106, 108
Highly ionized gas (HIG) 30
Host galaxy 154, 160
Hot intercloud medium (HIM) 19, 26
Hot spots 53
Hydrostatic equilibrium 48

Identifications 191, 217
Inflow 151, 160
Infrared emission 113
Intergalactic medium (IGM) 48, 64, 223
Inverse Compton scattering 42, 47, 61, 69
Ionization
  cone 102
  parameter 17, 78
Isothermal gas 49

Jets
  acceleration 62
  composition 59
  confinement 64
  high power 52
  interaction with external medium 48
  low power 54
  magnetic fields 56, 63, 67
  speed 52, 56, 61

K-shell lines 19, 102

Lindblad resonances 160
Line cooling 15
Lorentz factor 40, 42, 46, 69
Luminosity function 197, 212, 216, 219

M87 56
Mach number 52
Masers 110
Mechanical heating 16
Mergers 149, 156, 159
Molecular gas 104, 106

Narrow line region (NLR) 26, 78, 102
NGC 1068 104, 113, 118

Outflows, winds 19, 22, 36, 94, 95, 175, 220, 222, 225

Photoionization 13, 78
Polarization 67, 69

Radiation
  pressure 2, 22, 24
  self-Compton (SSC) 43, 47, 58
  synchrotron 40, 56, 70
  thermal 44
Radiative transfer 16, 114
Radio galaxies, radio sources 52, 70
Rankine-Hugoniot conditions 54
Recombination 13, 14
Relativistic beaming 47, 58
Reverberation mapping 27, 81

Sag A* 212
Scattering 10
SCUBA sources 199, 217
Shocks 52, 150
Spectral energy distribution (SED) 11, 113, 117
Stability curve 20
Starburst 16, 33, 152, 156, 160, 165, 171, 193
Superluminal motion 46
Surveys 185, 198

Temperature
  dust 103, 115
  electron 78
Thermal
  balance 15

  radiation 44
Tidal/gravitational torque 5, 150, 151
Torus 33, 115, 116

UV absorber 25

Variability
  of BLR 81
  time delay 82
  velocity-delay map 83
Viscosity
  stress 7
  torque 150, 175

Warm absorber 30

X-ray
  absorption 30, 102, 106
  background (XRB) 194, 196
  Compton thick absorption 196
  emission 198, 200
  space missions 185, 206
  variability 113, 117

Lecture Notes in Physics

For information about earlier volumes
please contact your bookseller or Springer
LNP Online archive: springerlink.com

Vol.648: S. G. Karshenboim, E. Peik (Eds.), Astrophysics, Clocks and Fundamental Constants

Vol.649: M. Paris, J. Rehacek (Eds.), Quantum State Estimation

Vol.650: E. Ben-Naim, H. Frauenfelder, Z. Toroczkai (Eds.), Complex Networks

Vol.651: J. S. Al-Khalili, E. Roeckl (Eds.), The Euroschool Lectures of Physics with Exotic Beams, Vol.I

Vol.652: J. Arias, M. Lozano (Eds.), Exotic Nuclear Physics

Vol.653: E. Papantonoupoulos (Ed.), The Physics of the Early Universe

Vol.654: G. Cassinelli, A. Levrero, E. de Vito, P. J. Lahti (Eds.), Theory and Application to the Galileo Group

Vol.655: M. Shillor, M. Sofonea, J. J. Telega, Models and Analysis of Quasistatic Contact

Vol.656: K. Scherer, H. Fichtner, B. Heber, U. Mall (Eds.), Space Weather

Vol.657: J. Gemmer, M. Michel, G. Mahler (Eds.), Quantum Thermodynamics

Vol.658: K. Busch, A. Powell, C. Röthig, G. Schön, J. Weissmüller (Eds.), Functional Nanostructures

Vol.659: E. Bick, F. D. Steffen (Eds.), Topology and Geometry in Physics

Vol.660: A. N. Gorban, I. V. Karlin, Invariant Manifolds for Physical and Chemical Kinetics

Vol.661: N. Akhmediev, A. Ankiewicz (Eds.) Dissipative Solitons

Vol.662: U. Carow-Watamura, Y. Maeda, S. Watamura (Eds.), Quantum Field Theory and Noncommutative Geometry

Vol.663: A. Kalloniatis, D. Leinweber, A. Williams (Eds.), Lattice Hadron Physics

Vol.664: R. Wielebinski, R. Beck (Eds.), Cosmic Magnetic Fields

Vol.665: V. Martinez (Ed.), Data Analysis in Cosmology

Vol.666: D. Britz, Digital Simulation in Electrochemistry

Vol.667: W. D. Heiss (Ed.), Quantum Dots: a Doorway to Nanoscale Physics

Vol.668: H. Ocampo, S. Paycha, A. Vargas (Eds.), Geometric and Topological Methods for Quantum Field Theory

Vol.669: G. Amelino-Camelia, J. Kowalski-Glikman (Eds.), Planck Scale Effects in Astrophysics and Cosmology

Vol.670: A. Dinklage, G. Marx, T. Klinger, L. Schweikhard (Eds.), Plasma Physics

Vol.671: J.-R. Chazottes, B. Fernandez (Eds.), Dynamics of Coupled Map Lattices and of Related Spatially Extended Systems

Vol.672: R. Kh. Zeytounian, Topics in Hyposonic Flow Theory

Vol.673: C. Bona, C. Palenzula-Luque, Elements of Numerical Relativity

Vol.674: A. G. Hunt, Percolation Theory for Flow in Porous Media

Vol.675: M. Kröger, Models for Polymeric and Anisotropic Liquids

Vol.676: I. Galanakis, P. H. Dederichs (Eds.), Half-metallic Alloys

Vol.677: A. Loiseau, P. Launois, P. Petit, S. Roche, J.-P. Salvetat (Eds.), Understanding Carbon Nanotubes

Vol.678: M. Donath, W. Nolting (Eds.), Local-Moment Ferromagnets

Vol.679: A. Das, B. K. Chakrabarti (Eds.), Quantum Annealing and Related Optimization Methods

Vol.680: G. Cuniberti, G. Fagas, K. Richter (Eds.), Introducing Molecular Electronics

Vol.681: A. Llor, Statistical Hydrodynamic Models for Developed Mixing Instability Flows

Vol.682: J. Souchay (Ed.), Dynamics of Extended Celestial Bodies and Rings

Vol.683: R. Dvorak, F. Freistetter, J. Kurths (Eds.), Chaos and Stability in Planetary Systems

Vol.684: J. Dolinšek, M. Vilfan, S. Žumer (Eds.), Novel NMR and EPR Techniques

Vol.685: C. Klein, O. Richter, Ernst Equation and Riemann Surfaces

Vol.686: A. D. Yaghjian, Relativistic Dynamics of a Charged Sphere

Vol.687: J. W. LaBelle, R. A. Treumann (Eds.), Geospace Electromagnetic Waves and Radiation

Vol.688: M. C. Miguel, J. M. Rubi (Eds.), Jamming, Yielding, and Irreversible Deformation in Condensed Matter

Vol.689: W. Pötz, J. Fabian, U. Hohenester (Eds.), Quantum Coherence

Vol.690: J. Asch, A. Joye (Eds.), Mathematical Physics of Quantum Mechanics

Vol.691: S. S. Abdullaev, Construction of Mappings for Hamiltonian Systems and Their Applications

Vol.692: J. Frauendiener, D. J. W. Giulini, V. Perlick (Eds.), Analytical and Numerical Approaches to Mathematical Relativity

Vol.693: D. Alloin, R. Johnson, P. Lira (Eds.), Physics of Active Galactic Nuclei at all Scales