

Cambridge astrophysics series

X-ray emission from clusters of galaxies

For Jane

X-ray Emission from Clusters of Galaxies

Craig L. Sarazin

Department of Astronomy, University of Virginia

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

Contents

Preface

1 Introduction

2 Optical observations

 2.1 Catalogs

 2.2 Redshifts

 2.3 Richness—the number of galaxies in a cluster

 2.4 Luminosity function of galaxies

 2.5 Morphological classification of clusters

 2.6 Velocity distribution of galaxies

 2.7 Spatial distribution of galaxies

 2.8 Cluster masses—the missing mass problem

 2.9 Dynamics of galaxies in clusters

 2.9.1 Two-body relaxation

 2.9.2 Violent relaxation

 2.9.3 Ellipsoidal clusters

 2.9.4 Dynamical friction

 2.10 Galactic content of clusters

 2.10.1 cD galaxies

 2.10.2 Proportion of spiral, S0, and elliptical galaxies

 2.11 Extensions of clustering

 2.11.1 Poor clusters

 2.11.2 Superclusters and voids

3 Radio observations

 3.1 General radio properties

 3.2 Correlations between X-ray and radio emission

 3.3 Head-tail and other distorted radio structures

 3.4 Cluster radio haloes

 3.5 Cosmic microwave diminution (Sunyaev–Zel'dovich effect)

 3.6 Faraday rotation

 3.7 21 cm line observations of clusters

4 X-ray observations

 4.1 Detections and identifications

 4.2 X-ray luminosities and luminosity functions

 4.3 X-ray spectra

 4.3.1 Continuum features in the spectrum

 4.3.2 Line features—the 7 keV iron line

 4.3.3 Lower energy lines

 4.4 The spatial distribution of X-ray emission

 4.4.1 X-ray centers, sizes, and masses

 4.4.2 X-ray images of clusters and the morphology of the intracluster gas

 4.5 Individual clusters

 4.5.1 Coma

 4.5.2 Perseus

 4.5.3 M87/Virgo

 4.5.4 A1367

 4.6 X-ray-optical correlations

 4.7 Poor clusters

 4.8 High redshift clusters and X-ray cluster evolution

5 Theoretical progress

 5.1 Emission mechanisms

 5.1.1 Inverse Compton emission

 5.1.2 Individual stellar X-ray sources

5.1.3	Thermal bremsstrahlung from intracluster gas
5.2	Ionization and X-ray emission from hot, diffuse plasma
5.2.1	Ionization equilibrium
5.2.2	X-ray emission
5.2.3	Resulting spectra
5.3	Heating and cooling of the intracluster gas
5.3.1	Cooling
5.3.2	Infall and compressional heating
5.3.3	Heating by ejection from galaxies
5.3.4	Heating by galaxy motions
5.3.5	Heating by relativistic electrons
5.4	Transport processes
5.4.1	Mean free paths and equilibration time-scales
5.4.2	Thermal conduction
5.4.3	Effects of the magnetic field
5.4.4	Viscosity
5.4.5	Diffusion and settling of heavy ions
5.4.6	Convection and mixing
5.5	Distribution of the intracluster gas–hydrostatic models
5.5.1	Isothermal distributions
5.5.2	Adiabatic and polytropic distributions
5.5.3	More complicated distributions
5.5.4	Empirical gas distributions derived by surface brightness deconvolution
5.5.5	Total masses and mass distributions in clusters—the hydrostatic method
5.5.6	Chemically inhomogeneous equilibrium models
5.6	Wind models for the intracluster gas
5.7	Cooling flows and accretion by cDs
5.7.1	Cooling flows
5.7.2	Accretion by central galaxies
5.7.3	Thermal instability and optical filamentation
5.7.4	Accretion-driven star formation
5.7.5	Cooling flow models with star formation
5.7.6	Evolution of cooling flows and active galaxies
5.8	X-ray emission from individual galaxies
5.8.1	Massive haloes around M87 and other central galaxies
5.8.2	Other models for M87 and other central galaxies
5.8.3	X-ray emission from noncentral cluster galaxies
5.9	Stripping of gas from galaxies in clusters
5.10	The origin and evolution of the intracluster medium
5.10.1	Infall models
5.10.2	Ejection from galaxies
6	Prospects for the future and AXAF
	References
	Index

Preface

As this book was being completed, observational X-ray astronomy was in a relatively quiet period between the demise of the *Einstein* X-ray observatory, and the launch of the next generation X-ray observatory, AXAF. This seemed like a good time to summarize what we have learned about the X-ray emission of clusters of galaxies. This book is mainly devoted to a review of the observational properties of X-ray clusters of galaxies and of the theoretical understanding we currently have of the physical state, dynamics, and origin of the hot intracluster gas. The book also contains less complete reviews of the optical and radio properties of clusters of galaxies, and of their dynamics.

Much of the material in this book first appeared in a review article in *Reviews of Modern Physics* (Sarazin, 1986). I should like to thank Ed Salpeter for suggesting this review, and for many helpful suggestions.

Andy Fabian, Richard Mushotzky, Paul Nulsen, Simon White and an anonymous referee very kindly provided detailed comments on an early draft of the manuscript and caught many significant errors. I really would like to thank them for all their help. I am also indebted to John Bahcall, James Binney, Pat Henry, Hernan Quintana, Yoel Rephaeli, Herb Rood, Hy Spinrad, Mel Ulmer, and Ray White for helpful comments, suggestions, and unpublished results. I should particularly like to thank John Bahcall for his encouragement. Figures for the original review paper and for this book were very kindly provided by Neta Bahcall, Jack Burns, Claude Canizares, Christine Jones, Andy Fabian, Bill Forman, Paul Gorenstein, Mark Henriksen, Roger Lynds, George Miley, Richard Mushotzky, Chris O'Dea, Herb Rood, Steve Strom, and Simon White. I should like particularly to thank Christine Jones and Bill Forman for producing special figures from the *Einstein* X-ray data.

Much of the original review was written while I was a visitor at the Institute for Advanced Study in Princeton, N.J., and I would like to thank the Institute and particularly John Bahcall for their hospitality. Karen Jobes at the Institute for Advanced Study helped considerably with the work processing. Parts of the review was written during visits to the Aspen Center for Physics and the National Radio Astronomy Observatory, and I would also like to thank them.

The manuscript for the book was completed while I was a Visiting Fellow at the Joint Institute for Laboratory Astrophysics of the University of Colorado and the National Bureau of Standards. I should like to thank the JILA Fellows for their hospitality. During this sabbatical leave I was also supported in part by a Sesquicentennial Fellowship from the University of Virginia.

This research was supported in part by NSF Grant AST 81-20260 and NASA Astrophysics Theory Center Grant NAGW-764.

This work would have been impossible without the support and patient understanding of my wife Jane.

1 Introduction

Regular clusters of galaxies are the largest organized structures in the universe. They typically contain hundreds of galaxies, spread over a region of whose size is roughly 10^{25} cm. Their total masses exceed 10^{48} gm. They were first studied in detail by Wolf (1906), although the tendency for galaxies to cluster on the sky had been noted long before this. A great advance in the systematic study of the properties of clusters occurred when Abell compiled an extensive, statistically complete catalog of rich clusters of galaxies (Abell, 1958). For the last quarter century, this catalog has been the most important resource in the study of galaxy clusters. Optical photographs of several of the best studied clusters of galaxies are shown in Figure 1.¹ The Virgo cluster (Figure 1a) is the nearest rich cluster to our own galaxy; the Coma cluster (Figure 1b) is the nearest very regular cluster.

In 1966, X-ray emission was detected from the region around the galaxy M87 in the center of the Virgo cluster (Byram *et al.*, 1966; Bradt *et al.*, 1967; Figure 1a). In fact, M87 was the first object outside of our galaxy to be identified as a source of astronomical X-ray emission. Five years later, X-ray sources were also detected in the directions of the Coma (Figure 1b) and Perseus (Figure 1c) clusters (Fritz *et al.*, 1971; Gursky *et al.*, 1971a,b; Meekins *et al.*, 1971). Since these are three of the nearest rich clusters, it was suggested that clusters of galaxies might generally be X-ray sources (Cavaliere *et al.*, 1971). The launch of the *Uhuru* X-ray astronomy satellite permitted a survey of the entire sky for X-ray emission (Giacconi *et al.*, 1972) and established that this was indeed the case. These early *Uhuru* observations indicated that many clusters were bright X-ray sources, with luminosities typically in the range of 10^{43-45} erg/s. The X-ray sources associated with clusters were found to be spatially extended; their sizes were comparable to the size of the galaxy distribution in the clusters (Kellogg *et al.*, 1972; Forman *et al.*, 1972). Unlike other bright X-ray sources but consistent with their spatial extents, cluster X-ray sources did not vary temporally in their brightness (Elvis, 1976). Although several emission mechanisms were proposed, the X-ray spectra of clusters were most consistent with thermal bremsstrahlung from hot gas.

This interpretation requires that the space between galaxies in clusters be filled with very hot ($\approx 10^8$ K), low density ($\approx 10^{-3}$ atoms/cm³) gas. Remarkably, the total mass in this intracluster medium is comparable to the total mass in all the stars in all the galaxies in the cluster. As to the origin of this gas, it was widely assumed that it had simply fallen into the clusters from the great volumes of space between them, where it had been stored since the formation of the universe (Gunn and Gott, 1972).

In 1976, X-ray line emission from iron was detected from the Perseus cluster of galaxies (Mitchell *et al.*, 1976), and shortly thereafter from Coma and Virgo as well (Serlemitsos *et al.*, 1977). The emission mechanism for this line is thermal, and its detection confirmed the thermal interpretation of cluster X-ray sources. However, the only known sources for significant quantities of iron or any other heavy element in astronomy are nuclear reactions in stars, and no significant population of stars has been observed which do not reside in galaxies. Since the abundance of iron in the intracluster gas was observed to be similar to its abundance in stars, a substantial portion of this gas must have been ejected from stars in galaxies in the cluster (Bahcall and Sarazin, 1977). This is despite the fact that the total mass of intracluster gas is on the same order as the total mass of stars presently observed in the clusters. Obviously, these X-ray observations suggest that galaxies in clusters have had more interesting histories than might otherwise have been assumed.

In this paper, the X-ray observations of clusters of galaxies and the theories for the intracluster gas will be reviewed. Because clusters are still largely defined by their optical properties, I shall first review the optical observations of clusters (Chapter 2). I shall particularly emphasize information on their dynamical state, and the possibility that the galaxy population has been affected by the intracluster gas. Radio observations of clusters also provide information on the intracluster gas, which is summarized in Chapter 3. For example, certain distortions seen in radio sources in clusters are most naturally explained as arising from interactions with this gas. Moreover, extensive searches have been made for ‘shadows’ in the cosmic microwave radiation due to electron scattering by intracluster gas. Then the X-ray observations will be reviewed (Chapter 4),

¹Unless otherwise indicated, all the figures in this book showing optical, X-ray, or radio brightness on the sky have north at the top and east at the left. When coordinates are given, the east–west coordinate is right ascension (in hours, minutes, and seconds) and the north–south coordinate is declination (in degrees, minutes, and seconds).

including the recent results of X-ray imaging and spectroscopy from the *Einstein* X-ray satellite. In Chapter 5 theories for the X-ray emission mechanism, the physical state, the distribution, the origin, and the history of the intracluster medium will be reviewed. Finally, I shall comment briefly on the prospects for further observations of X-ray clusters, particularly with the AXAF satellite, in Chapter 6.

Review articles on clusters of galaxies emphasizing their optical properties include Abell (1965, 1975), van den Bergh (1977b), Bahcall (1977a), Rood (1981), White (1982), and particularly Dressler (1984). Superclusters of galaxies are reviewed by Oort (1983). Some recent reviews which include the X-ray properties of clusters are Gursky and Schwartz (1977), Binney (1980), Cavaliere (1980), Cowie (1981), Canizares (1981), and Holt and McCray (1982). Fabian *et al.* (1984b) give an excellent review of cooling flows in X-ray clusters (see Section 5.7). Up-to-date reviews of the spectroscopic properties of X-ray clusters are given by Mushotzky (1984, 1985). Forman and Jones (1982) give a comprehensive review of the X-ray images of clusters. This book is based in large part on my review paper on X-ray clusters (Sarazin, 1986a).

2 Optical observations

2.1 Catalogs

The two most extensive and often cited catalogs of rich clusters of galaxies are those of Abell (1958) and Zwicky and his collaborators (Zwicky *et al.*, 1961–1968). As is conventional, in this book Abell clusters will be denoted by giving A and then the number of the cluster in Abell’s list. Both of these catalogs were constructed by identifying clusters as enhancements in the surface number density of galaxies on the National Geographic Society–Palomar Observatory Sky Survey (Minkowski and Abell, 1963) and thus are confined to northern areas of sky (declination greater than -20° for Abell and -3° for Zwicky). Abell surveyed only clusters away from the plane of our galaxy.

As clustering exists on a very wide range of angular and intensity scales (Peebles, 1974), it is not possible to give a unique and unambiguous definition of a ‘rich cluster’. Thus the membership of a catalog of clusters is determined by the criteria used to define a rich cluster. These criteria must specify the required surface number density enhancement for inclusion and the linear or angular scale of the enhancement. The scale is necessary in order to exclude small groupings of galaxies; for example, a close pair of galaxies can represent a very large enhancement above the background galaxy density on a small angular scale. Alternatively, specifying the surface density and scale is equivalent to specifying the number of galaxies (the ‘richness’ of the cluster) and the scale size. Because the number of galaxies observed increases as their brightness diminishes (Section 2.4), one must also specify the range of magnitudes of the galaxies included in determining the cluster richness. Finally, because galaxies grow fainter with increasing distance, the catalog can only be statistically complete out to a limiting distance or redshift, and only clusters within this distance range should be included in a statistically complete sample.

Abell's criteria were basically (1) that the cluster contain at least 50 galaxies in the magnitude range m_3 to $m_3 + 2$, where m_3 is the magnitude of the third brightest galaxy; (2) that these galaxies be contained within a circle of radius $R_A = 1.7/z$ minutes of arc or $3 h_{50}^{-1}$ megaparsecs,² where z is the estimated redshift of the cluster (Section 2.2); (3) that the estimated cluster redshift be in the range $0.02 \leq z \leq 0.20$. R_A is called the Abell radius of the cluster. The Abell catalog contains 2712 clusters, of which 1682 satisfy all these criteria. The other 1030 were discovered during the search and were included to provide a more extensive finding list for clusters. The Abell catalog gives estimates of the cluster center positions (see also Sastry and Rood, 1971), distance, and richness of the clusters, as well as the magnitude of the tenth brightest galaxy m_{10} .

For the Zwicky catalog, the criteria were as follows: (1) the boundary (scale size) of the cluster was determined by the contour (or isopleth) where the galaxy surface density fell to twice the local background density; (2) this isopleth has to contain at least 50 galaxies in the magnitude range m_1 to $m_1 + 3$, where m_1 is the magnitude of the first-brightest galaxy. No distance limits were specified, although in practice, very nearby clusters such as Virgo (Figure 1a), were not included because they extended over several Sky Survey plates. Obviously, the Zwicky catalog criteria are much less strict than Abell's, and the Zwicky catalog thus contains many more clusters that are less rich. For each cluster, the Zwicky catalog gives a classification (Section 2.5), and estimates of the coordinates of the center, the diameter, the richness, and the redshift. Finding charts showing the cluster isopleths and positions of brighter galaxies and stars are also presented.

A number of smaller catalogs have been compiled, consisting of clusters in the southern sky or clusters at higher redshifts ($z \gtrsim 0.2$). Early southern cluster catalogs or lists include those of de Vaucouleurs (1956), Klemola (1969), Snow (1970), Sersic (1974), and Rose (1976). Until recently, the search for southern clusters was severely handicapped by the lack of deep survey plates. The first deep optical survey of the south, the European Southern Observatory Quick Blue Survey (ESO-B), was completed in 1978 (West, 1974; Holmberg *et al.*, 1974). A catalog of southern clusters from the first portion of this survey was prepared by Duus and Newell (1977) and a list of southern clusters near X-ray sources was given by Lugger (1978). More recently, portions of the ESO/SRC-J survey of very deep blue plates have been used to compile southern cluster catalogs by Braid and MacGillivray (1978) and White and Quintana (1985). Before his untimely death, Abell was preparing a southern continuation of the Abell catalog in collaboration with Corwin. The red portion of the ESO/SRC survey is currently being done, and these plates are being used to detect high redshift clusters (West and Frandsen, 1981).

The discovery of higher redshift clusters ($z \gtrsim 0.2$) is of great importance to cosmological studies; lists of such clusters include Humason and Sandage (1957), Gunn and Oke (1975), Sandage, Kristian, and Westphal (1976), Spinrad *et al.* (1985), and Vidal (1980). In addition, the lists of southern clusters from the ESO/SRC surveys discussed in the last paragraph contain many high redshift clusters.

2.2 Redshifts

The clusters in the Abell (1958) catalog were assigned to distance groups, based on the redshift estimated from the magnitude of the tenth brightest galaxy in the cluster. Leir and van den Bergh (1977) have given improved estimates of redshifts for 1889 rich Abell clusters, using the magnitudes of the first and tenth brightest galaxies, and an estimate of the cluster radius. Their distance scale is calibrated using measured redshifts for 101 clusters. Photometric distance estimators derived from a larger sample of redshifts have been given by Sarazin *et al.* (1982). Similarly, clusters in the Zwicky catalog were placed in distance groups, based on the magnitudes and sizes of the brightest cluster galaxies.

Spectroscopic redshifts are now available for about 500 Abell clusters. Sarazin (1986a, Table I) gives an extensive list of the redshifts for Abell clusters, taken primarily from Sarazin *et al.* (1982), Noonan (1981), Struble and Rood (1982, 1985), and Hoessel, Gunn, and Thuan (1980).

Of course, many redshifts are known for non-Abell clusters as well. The compilation of Noonan (1981) includes many such redshifts.

2.3 Richness—the number of galaxies in a cluster

The richness of a cluster is a measure of the number of galaxies associated with that cluster. Because of the presence of background galaxies, it is not possible to state with absolute confidence that any given galaxy belongs to a given cluster. Thus one cannot give an exact tally of the number of galaxies in a cluster. Richness is a statistical measure of the population of a cluster, based on some operational definition of cluster

²For convenience, in this book the Hubble constant H_o will be parameterized as $H_o = 50h_{50}$ km/sec/Mpc.

membership. The richness will be more useful if it is defined in such a way as to be reasonably independent of the distance to and morphology of a cluster.

Zwicky *et al.* (1961-1968) define the richness of their clusters as the total number of galaxies visible on the red Sky Survey plates within the cluster isopleth (Section 2.1); the number of background galaxies expected is subtracted from the richness. These richnesses are clearly very dependent on a cluster redshift (Abell, 1962; Scott, 1962). First, a wider magnitude range is counted for nearby clusters, because the magnitude of the first brightest galaxy is further from the plate limit. Second, a larger area of the cluster is counted for nearby clusters, because the point at which the surface density is twice that of the background will be farther out in the cluster.

Abell (1958) has divided his clusters into richness groups using criteria that are nearly independent of distance (see Section 2.1); that is, the magnitude range and area considered do not vary with redshift. Just (1959) has found a slight richness-distance correlation in Abell's catalog; however, the effect is small and is probably explained by a slight incompleteness (10%) of the catalog for distant clusters (Paal, 1964). When more accurate determinations have been made, it has been found that the Abell richness generally correlated well with the number of galaxies, but that the Abell richness may be significantly in error in some individual cases (Mottmann and Abell, 1977; Dressler, 1980a). Thus the Abell richnesses are very useful for statistical studies, but must be used with caution in studies of individual clusters. Note also that it is generally preferable to use the actual Abell counts rather than just the richness group (Abell, 1982).

2.4 Luminosity function of galaxies

The luminosity function of galaxies in a cluster gives the number distribution of the luminosities of the galaxies. The integrated luminosity function $N(L)$ is the number of galaxies with luminosities greater than L , while the differential luminosity function $n(L)dL$ is the number of galaxies with luminosities in the range L to $L + dL$. Obviously, $n(L) \equiv -dN(L)/dL$. Luminosity functions are often defined in terms of galaxy magnitudes $m \propto -2.5 \log_{10}(L)$; $N(\leq m)$ is the number of galaxies in a cluster brighter than magnitude m . Observational studies of the luminosity functions of clusters include Zwicky (1957), Kiang (1961), van den Bergh (1961a), Abell (1962, 1975, 1977), Rood (1969), Gudehus (1973), Bautz and Abell (1973), Austin and Peach (1974b), Oemler (1974), Krupp (1974), Austin *et al.* (1975), Godwin and Peach (1977), Mottmann and Abell (1977), Dressler (1978b), Bucknell *et al.* (1979), Carter and Godwin (1979), Thompson and Gregory (1980), and Kraan-Korteweg (1981). Figure 2 gives the observed integrated luminosity function for a composite of 13 rich clusters as derived by Schechter (1976) from Oemler's (1974) data.

Three types of functions are commonly used for fitting the luminosity function. Zwicky (1957) proposed the form

$$N(\leq m) = K(10^{0.2(m-m_1)} - 1), \quad (2.1)$$

where K is a constant and m_1 is the magnitude of the first brightest galaxy. In general, the Zwicky function fits the faint end of the luminosity function adequately, but does not fall off rapidly enough for brighter galaxies (see Figure 2). Clearly, equation (2.1) implies that $N(L) = K[(L_1/L)^{1/2} - 1]$, where L_1 is the luminosity of the first brightest galaxy, and K is the expected number of galaxies in the range $1/4 L_1 \leq L \leq L_1$.

Abell (1975) has suggested that the luminosity function $N(L)$ be fit by two intersecting power laws, $N(L) = N^*(L/L^*)^{-\alpha}$, where $\alpha \approx 5/8$ for $L < L^*$, and $\alpha \approx 15/8$ for $L > L^*$. L^* is the luminosity at which the two power laws cross, and N^* is the expected number of galaxies with $L \geq L^*$. Of course, this form is intended only as a simple and practical fit to the data; the real luminosity function certainly has a continuous derivative $n(L)$, unlike Abell's function. The magnitude luminosity function corresponding to Abell's form is often written as

$$\log_{10} N(\leq m) = \begin{cases} K_1 + s_1 m & m \leq m^* \\ K_2 + s_2 m & m > m^* \end{cases} \quad (2.2)$$

where K_1 and K_2 are constants. The slopes are approximately $s_1 \approx 0.75$ and $s_2 \approx 0.25$, and the power laws cross at $m = m^*$, so that $K_1 + s_1 m^* = K_2 + s_2 m^*$. As shown in Figure 2, the Abell form fits both the bright and faint ends of the luminosity function adequately.

Schechter (1976) proposed an analytic approximation for the differential luminosity function

$$n(L)dL = N^*(L/L^*)^{-\alpha} \exp(-L/L^*)d(L/L^*) \quad (2.3)$$

where L^* is a characteristic luminosity, $N^*\Gamma(1 - \alpha, 1)$ is the number of galaxies with $L > L^*$, and $\Gamma(a, x)$ is the incomplete gamma function; Schechter derives a value for the faint end slope of $\alpha = 5/4$. The integrated luminosity function corresponding to equation (2.3) is $N(L) = N^*\Gamma(1 - \alpha, L/L^*)$.

The advantages of the Schechter function are that it is analytic and continuous, unlike Abell's form, and that it is a real statistical distribution function, unlike Zwicky's form, which requires that the magnitude of the first brightest galaxy (which ought to be a random variable) be specified. An expression similar to the Schechter form was predicted by a simple analytical model for galaxy formation (Press and Schechter, 1974). The Schechter function is steeper than Abell's function at the bright end because it contains an exponential. The Schechter differential luminosity function decreases monotonically with luminosity, while the Abell function has a local maximum at $L = L^*$ (Abell, 1975); there is some weak evidence for such a peak, especially near the centers of rich clusters (Rood and Abell, 1973).

The Schechter function fits the observed distribution in many clusters reasonably well from the faint to the bright end (Figure 2), as long as the very brightest galaxies, the cD galaxies, are excluded (Schechter, 1976). These can have luminosities as large as $10L^*$, and thus are extremely improbable if equation (2.3) holds exactly. However, cD galaxies have a number of distinctive morphological characteristics which suggest that they were formed by special processes which occur primarily in the centers of rich clusters (see Section 2.10.1). In any case, they apparently can be excluded by morphological (as opposed to luminosity) criteria, and equation (2.3) can be taken to be the non cD luminosity function.

The parameter N^* in the Abell or Schechter functions is a useful measure of the richness of the cluster. If these luminosity functions are adequate approximations, then fitting the luminosity function to determine N^* ought to give a more accurate measure of richness than counting galaxies in magnitude ranges. Note that while the total number of galaxies predicted by Abell's, Zwicky's, or Schechter's functions diverges at the faint end, the total luminosity is finite; for example, the total cluster optical luminosity is $L_{opt} = N^*\Gamma(2 - \alpha)L^*$ for the Schechter function, and N^* thus measures the total cluster luminosity.

Because of the break in the luminosity function near L^* , this parameter represents a characteristic luminosity of cluster galaxies. Moreover, L^* is nearly the same for many clusters; it corresponds to an absolute visual magnitude $M_V^* \approx -21.2 + 5\log_{10} h_{50}$ for Abell's form (Austin *et al.*, 1975), and $M_V^* \approx -21.9 + 5\log_{10} h_{50}$ for Schechter's form (Schechter, 1976). By comparing these values with the apparent magnitude m^* derived from the observed luminosity function of a cluster, one can derive a distance estimate for the cluster (Bautz and Abell, 1973; Schechter and Press, 1976). Unfortunately, this requires that magnitudes be available for a large number of galaxies.

The magnitude of the brightest galaxy in a cluster has often been used as a distance indicator in cosmological studies (see, for example, Hoessel, Gunn, and Thuan, 1980). It is obviously easier to determine observationally than the luminosity function. Moreover, the luminosities of the brightest cluster galaxies show a very small variation from cluster to cluster (Sandage, 1976) and depend only weakly on richness. There has been a controversy concerning whether the luminosities of these brightest galaxies are determined statistically by the cluster luminosity function, or whether the brightest galaxies are produced by some special physical processes operating in clusters (see Dressler, 1984, for a review). Under the statistical hypothesis, all of the galaxy luminosities are random variables drawn from the luminosity function (Geller and Peebles, 1976). Then, the observed number of galaxies in any luminosity range will have a Poisson distribution about the expectation given by the luminosity function, and the luminosity L_1 of the brightest galaxy will thus be distributed as $\exp[-N(L_1)]n(L_1)dL_1$. As $N(L_1)$ is proportional to richness, the statistical model predicts that L_1 increases with cluster richness, which is not really observed (Sandage, 1976). However, Schechter and Peebles (1976) have argued that the near constancy of L_1 results from a selection effect (that is, the observed sample is biased), and that the statistical hypothesis may still be valid.

Alternatively, the brightest cluster galaxies may be affected by special physical processes, such as the tidal interactions or mergers of galaxies (Peach, 1969; Ostriker and Tremaine, 1975; Hausman and Ostriker, 1978; Richstone, 1975). Evidence that this may indeed be the case is given in Section 2.10.1.

While the luminosity functions of many clusters are reasonably well represented by the Schechter or Abell form with a universal value of M^* , significant departures exist in a number of clusters (Oemler, 1974; Mottmann and Abell, 1977; Dressler, 1978b). These departures include variations in the value of M^* , variations in the slope of the faint end of the luminosity function (α in Schechter's form, equation (2.3)), and variations in the steepness of the bright end of the luminosity function (Dressler, 1978b). These variations

are, in many cases, correlated with cluster morphology (Section 2.5). The variations in M^* and α probably reflect variations in the conditions in the cluster at its formation, while the variations in the bright end slope may result from evolutionary changes, such as the tidal interaction or merging of massive galaxies (Richstone, 1975; Hausman and Ostriker, 1978). In particular, the clusters with the steepest luminosity functions at the bright end often contain cD galaxies (Dressler, 1978b); this may indicate that the brighter galaxies were either eliminated by mergers to form the cD or diminished in brightness through tidal stripping (Section 2.10.1).

Turner and Gott (1976a) have shown that the luminosity function of galaxies in small groups is well represented by equation (2.3). In fact, Bahcall (1979a) has suggested that the luminosity function of all galaxian systems – from single galaxies (in or out of clusters) to the groups and clusters themselves – can be fit in a single function similar to the Schechter form (equation 2.3).

2.5 Morphological classification of clusters

A number of different cluster properties have been used to construct morphological classification systems for clusters. Somewhat surprisingly, these different systems are highly correlated, and it appears that clusters can be represented very crudely as a one-dimensional sequence, running from *regular* to *irregular* clusters (Abell, 1965; 1975). There is considerable evidence that the regular clusters are dynamically more evolved and relaxed than the irregular clusters. The various morphological classification schemes are described below, and the way in which they fit into the one-dimensional sequence is summarized in Table 1, which is adapted from Abell (1975) and Bahcall (1977a).

Zwicky *et al.* (1961–1968) classified clusters as *compact*, *medium compact*, or *open*. A *compact* cluster has a single pronounced concentration of galaxies, with more than ten galaxies appearing in contact as seen on the plate. A *medium compact* cluster has either a single concentration with ten galaxies separated by roughly their own diameters, or several concentrations. An *open* cluster lacks any pronounced concentration of galaxies.

Bautz and Morgan (1970) give a classification system based on the degree to which the cluster is dominated by its brightest galaxies. Bautz–Morgan *Type I* clusters are dominated by a single, central cD galaxy; cD galaxies have the most luminous and extensive optical emission found in galaxies (see Section 2.10.1). In *Type II* clusters, the brightest galaxies are intermediate between cD and normal giant ellipticals, while in *Type III*, there are no dominating cluster galaxies. *Type I–II* and *Type II–III* are intermediates. Leir and van den Bergh (1977) have classified 1889 rich Abell clusters on the Bautz–Morgan system, and some of the newer southern catalogs (e.g., White and Quintana, 1985) give Bautz–Morgan types for their clusters.

The original Rood–Sastry (1971) classification system is based on the nature and distribution of the ten brightest cluster galaxies. Basically, the six Rood–Sastry (RS) classes are defined as follows:

- cD: the cluster is dominated by a central cD galaxy (example: A2199).
- B: binary—the cluster is dominated by a pair of luminous galaxies (example: A1656 (Coma)).
- L: line—at least three of the brightest galaxies appear to be in a straight line (example: A426 (Perseus)).
- C: core—four or more of the ten brightest galaxies form a cluster core, with comparable galaxy separations (example: A2065 (Corona Borealis)).
- F: flat—the brightest galaxies form a flattened distribution on the sky (example: A2151 (Hercules)).
- I: irregular—the distribution of brightest galaxies is irregular, with no obvious center or core (example: A400).

Rood and Sastry (1971) give classifications for low redshift Abell clusters on this system. They show that these classifications form a bifurcated sequence, which can be represented by a ‘tuning-fork’ diagram (Figure 3a). This sequence is correlated with the sequence of *regular* to *irregular* clusters in the sense that clusters on the left of the diagram (*cD* and *B*) are regular and those to the right (*F* and *I*) are irregular. Rich clusters are more or less equally distributed among the three arms of the diagram.

Recently, Struble and Rood (1982, 1985) have proposed a revised version of the RS classification system. The definitions have been revised slightly, and a number of subclasses of the main RS classes have been proposed. More significantly, Struble and Rood have rearranged the tuning fork diagram into a ‘split linear’ diagram (Figure 3b), based on systematic trends in the galaxy distribution and content of clusters. This new scheme was devised in part from a comparison to numerical N-body simulations of the collapse of clusters

(White, 1976c; Carnevali *et al.*, 1981; Farouki *et al.*, 1983; also see Figure 5). Struble and Rood propose that this sequence represents an evolutionary sequence of clusters from irregular *I* to *cD* clusters.

Morgan (1961) and Oemler (1974) have constructed classification systems based on the galactic content of clusters (that is, the fraction of cluster galaxies which are spirals (Sps), disk galaxies without spiral structure (SOs), or elliptical (Es)). Morgan (1961) classified clusters as *type i* if they contained large numbers of spirals and as *type ii* if they contained few spirals. Oemler (1974) has refined this system, defining three classes of clusters: *spiral-rich* clusters, in which spirals (Sp) are the most common galaxies; *spiral-poor* clusters, in which spirals are less common and SOs are the most common galaxies; and *cD* clusters, which are dominated by a central cD galaxy and in which the great majority of galaxies are ellipticals or SOs.

These systems of classification are empirically found to be highly correlated, and can roughly be mapped into a one-dimensional sequence running from *regular* clusters to *irregular* clusters (Abell 1965, 1975). As shown in Table 1, *regular* clusters are highly symmetric in shape, and have a core with a high concentration of galaxies toward the center. Subclustering is weak or absent in *regular* clusters. In contrast, *irregular* clusters have little symmetry or central concentration, and often show significant subclustering. This suggests that the *regular* clusters are, in some sense, dynamically relaxed systems, while the *irregular* clusters are dynamically less-evolved and have preserved roughly their distribution of formation. Additional evidence that regular clusters are dynamically relaxed is provided in Sections 2.6, 2.7, and 2.10.1, and the nature of the dynamical processes that might produce this relaxation is discussed in Section 2.9.

Regular clusters tend to be *compact* (Zwicky type), Bautz–Morgan *Type I* to *II*, Rood–Sastry types *cD* or *B*, Morgan *type ii*, Oemler *spiral-poor* or *cD* clusters. These last four correlations indicate a connection between the dynamical state and galactic content of clusters. There is no one-to-one correlation between the morphology of a cluster and its richness; *regular* clusters are always rich, while *irregular* clusters may be either rich or sparse. However, *regular* clusters tend to have higher central galaxy densities than *irregular* clusters, because they are at least as rich and more compact.

2.6 Velocity distribution of galaxies

The existence of the morphological sequence of clusters from irregular to regular clusters (Section 2.5) suggests that the regular clusters may have undergone some sort of dynamical relaxation. The nature of this relaxed distribution is examined in the next two sections through the distribution of cluster galaxy velocities and positions.

The redshifts in Section 2.2 are determined from the mean radial velocity of galaxies in a cluster; in fact, the radial velocities of individual galaxies are distributed around this mean. It has been conventional to characterize this distribution by the dispersion σ_r of radial velocities about the mean

$$\sigma_r = \langle (v_r - \langle v_r \rangle)^2 \rangle^{1/2} \quad (2.4)$$

where v_r is the radial velocity, which is the component of the galaxy velocity along the line-of-sight. The dispersion completely characterizes the radial velocity distribution function if it is Gaussian:

$$p(v_r)dv_r = \frac{1}{\sigma_r\sqrt{2\pi}} \exp(-(v_r - \langle v_r \rangle)^2/2\sigma_r^2)dv_r \quad (2.5)$$

Here, $p(v_r)dv_r$ is the probability that an individual cluster galaxy has a radial velocity in the range v_r to $v_r + dv_r$. While the Gaussian distribution has usually been adopted simply for convenience, statistical tests reveal that it is a consistent fit to the observed total distribution function in many clusters, at least if high velocities ($|v_r - \langle v_r \rangle| > 3\sigma_r$) are excluded (Yahil and Vidal, 1977). However, the velocity dispersion in a given cluster generally decreases with distance from the cluster center; in Coma and Perseus the decline is about a factor of two from the center to the outer edge (Rood *et al.*, 1972; Kent and Gunn, 1982; Kent and Sargent, 1983). Moreover, the velocity dispersion can differ in different clumps of an irregular cluster showing subclustering (Geller and Beers, 1982; Bothun *et al.*, 1983).

A Gaussian distribution for a single component of the velocity obtains for a system of non-identical particles in thermodynamic equilibrium, in which case we identify the velocity dispersion with $\sigma_r \equiv (kT/m)^{1/2}$, where T is the galaxy ‘temperature’ and m the galaxy mass. While the Gaussian velocity distribution found in clusters suggests that they are at least partially relaxed systems, they are not fully relaxed to thermodynamic equilibrium. In thermodynamic equilibrium, all components of the cluster would have equal temperatures;

what is observed is that it is the velocity dispersion (not temperature) which is nearly independent of galaxy mass and position (Rood *et al.*, 1972; Kent and Gunn, 1982; Kent and Sargent, 1983).

Some variation of the velocity dispersion with galaxy mass or cluster position is observed (Rood *et al.*, 1972; Kent and Gunn, 1982; Kent and Sargent, 1983). In the Coma and Perseus clusters, the most luminous galaxies have a somewhat smaller velocity dispersion than the less luminous galaxies, and the velocity dispersion decreases with increasing projected distance from the cluster center. In fact, the latter effect must necessarily occur if clusters are finite, bound systems. Then, the velocities of bound galaxies at any point in the cluster cannot exceed the escape velocity at that position. As the escape velocity decreases with increasing distance from the cluster center, the velocity dispersion must also decrease with projected distance from the cluster center.

The observed mean galaxy velocity $\langle v_r \rangle$ will depend on the projected position in the cluster if the cluster is rotating. The projected shapes of many clusters are substantially flattened. Of course, this is true of the L (line) and F (flat) clusters (Section 2.5); however, many regular (cD and B) clusters are also significantly flattened (Abell, 1965; Dressler, 1981). If these clusters are actually oblate in shape due to significant rotational support, the variation in $\langle v_r \rangle$ across the cluster would be expected to be comparable to σ_r . In fact, such large velocity gradients are not observed (Rood *et al.*, 1972; Gregory and Tifft 1976; Schipper and King, 1978; Dressler, 1981); apparently, the flattening of clusters is not due to rotation.

Useful compilations of velocity dispersions for clusters have been given by Hintzen and Scott (1979), Danese *et al.* (1980), and particularly Noonan (1981).

2.7 Spatial distribution of galaxies

The most regular clusters show a smooth galaxy distribution with a concentrated core (Figure 4; Table 1). In general, models to describe the galaxy distribution in these clusters will possess at least five parameters, which can be taken to be the position of the cluster center on the sky, the central projected density of galaxies per unit area of the sky σ_o , and two distance scales r_c and R_h . The core radius r_c is a measure of the size of the central core, and is usually defined so that the projected galaxy density at a distance r_c from the cluster center is one half of the central density σ_o . The halo radius R_h measures the maximum radial extent of the cluster. Of course, the observed value of the central density σ_o must depend on the range of galaxy magnitudes observed, and the values of other parameters may also depend on galaxy magnitude. If the cluster is elongated, at least two more parameters are necessary; these can be taken to be the orientation of the semimajor axis of the cluster on the sky, and the ratio of semimajor to semiminor axes. However, spherically symmetric galaxy distributions will be discussed first, and r_c and R_h will be assumed to be independent of galaxy mass or luminosity. Then, one can write

$$\begin{aligned} n(r) &= n_o f(r/r_c, r/R_h) \\ \sigma(b) &= \sigma_o F(b/r_c, b/R_h), \end{aligned} \quad (2.6)$$

where $n(r)$ is the spatial volume density of galaxies at a distance r from the cluster center, n_o is the central ($r = 0$) density, $\sigma(b)$ is the projected surface density at a projected radius b , and f and F are two dimensionless functions. Obviously,

$$\sigma(b) = 2 \int_b^{R_h} \frac{n(r)rdr}{(r^2 - b^2)^{1/2}}. \quad (2.7)$$

A number of models have been proposed to fit the distribution of galaxies. Among the simplest are the isothermal models, which assume a Gaussian radial velocity distribution for galaxies (equation 2.5). If one further assumes that the velocity distribution is isotropic and independent of position, that the galaxy distribution is stationary, and that galaxy positions are uncorrelated, then one can write the galaxy phase space density $f(\mathbf{r}, \mathbf{v})$ as

$$f(\mathbf{r}, \mathbf{v}) d^3r d^3v = n(r) (2\pi\sigma_r^2)^{-3/2} \exp\left[-\frac{1}{2}\left(\frac{v}{\sigma_r}\right)^2\right] d^3v d^3r. \quad (2.8)$$

Now, the time-scale for two-body gravitational interactions in a cluster is much longer than the time for a galaxy to cross the cluster (see Section 2.9). Thus the galaxies can be considered to be a collisionless gas, and the phase space density f is conserved along particle trajectories (Liouville's theorem); f is then a function

of only the integrals of the motion ('Jeans' theorem'). If the velocity distribution is isotropic, f does not depend on the orbital angular momentum of the galaxies, and can only depend on the energy per unit mass $\epsilon = \frac{1}{2}v^2 + \phi(r)$, where $\phi(r)$ is the gravitational potential of the cluster. If we measure $\phi(r)$ relative to the cluster center, the galaxy spatial distribution is thus $n(r) = n_o \exp[-\phi(r)/\sigma_r^2]$. If the mass in the cluster is distributed in the same way as the galaxies, then Poisson's equation for the gravitational potential becomes

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\phi}{dr} \right) = 4\pi G n_o m \exp(-\phi/\sigma_r^2), \quad (2.9)$$

where m is the mass per galaxy (Chandrasekhar, 1942). It is conventional to make the change of variables $\psi \equiv \phi/\sigma_r^2$, $\xi \equiv r/\beta$, $\beta \equiv \sigma_r/(4\pi G n_o m)^{1/2}$. Then the equation for ψ is

$$\frac{1}{\xi^2} \frac{d}{d\xi} \left(\xi^2 \frac{d\psi}{d\xi} \right) = e^{-\psi}, \quad (2.10)$$

subject to the boundary conditions (assuming the density is regular at the cluster center) of $\psi = 0$, $d\psi/d\xi = 0$ at $\xi = 0$. Equation (2.10) is identical to the equation for an isothermal gas sphere in hydrostatic equilibrium (Chandrasekhar, 1939).

The galaxy density distribution in this isothermal sphere model is then $n(\xi) = n_o \exp[-\psi(\xi)]$. The projected galaxy density can be calculated from equation (2.7) and written as $\sigma(b) = \sigma_o F_{isot}(b/\beta)$. For convenience, the core radius r_c is defined as $r_c \equiv 3\beta$, because $F_{isot}(3) \approx 0.502$, which is obviously close to one-half. The central volume density and projected density are related by $\sigma_o \approx 6.06n_o\beta = 2.02n_o r_c$. Unfortunately, neither $\psi(\xi)$ nor F_{isot} can be represented by simple analytic functions. Relatively inaccurate tables of these functions are given in Zwicky (1957, p. 139), and more accurate economized analytic approximations to ψ and F_{isot} have been given by Flannery and Krook (1978), and Sarazin (1980), respectively.

At large radii $\xi \gg 1$, $n(\xi) \approx 2n_o/\xi^2$, and the total number of galaxies and total mass diverge in proportion to r . Thus the isothermal sphere cannot accurately represent the outer regions of a finite cluster. A number of methods to truncate the isothermal distribution have been used. If one is primarily concerned with representing the galaxy distribution near the core, and if one is not concerned with determining dynamically consistent velocity and spatial distributions, one can simply truncate the isothermal distribution at some radius. Zwicky (1957, p. 140) and Bahcall (1973a) have truncated the isothermal distribution with a uniform surface density cutoff C :

$$\sigma(b) = \frac{\sigma_o [F_{isot}(b/\beta) - C/6.06]}{[1 - C/6.06]}. \quad (2.11)$$

The cluster galaxy density then falls to zero at a radius R_h given by $F_{isot}(R_h/\beta) = C/6.06$. Bahcall (1973a) also defines a modified central surface density parameter $\alpha \equiv \sigma_o/(6.06 - C)$; then, for small $C \ll 2\pi$, the central volume density is just $n_o \approx \sigma_o [1 - (C/2\pi)^2]/(6.06\beta) \approx \alpha/\beta$. The solid curve for the surface density in Figure 4 is given by equation 2.11.

King (1966) has developed self-consistent truncated density distributions for clusters. The phase-space density he assumes is

$$f(\mathbf{r}, \mathbf{v}) d^3r d^3v \propto \exp \left[\frac{\phi(0) - \phi(r)}{\sigma_{r\infty}^2} \right] \times \left\{ \exp \left(-\frac{v^2}{2\sigma_{r\infty}^2} \right) - \exp \left(\frac{\phi(r)}{\sigma_{r\infty}^2} \right) \right\} d^3r d^3v \quad (2.12)$$

where $\sigma_{r\infty}$ is the radial velocity dispersion in an untruncated cluster. The velocity distribution is thus truncated at the escape velocity v_e : $f(\mathbf{r}, |\mathbf{v}| \geq v_e) = 0$ where $v_e^2 = -2\phi(r)$, and the potential goes to zero at infinity $\phi(\infty) = 0$. King showed that equation (2.12) gave an approximate solution to the Fokker–Planck equation for a finite cluster subject to two-body gravitational encounters. As shown in Section 2.9, galaxy clusters are nearly collisionless; however, it is possible that equation (2.12) is a reasonable approximation for the truncated phase-space density. The phase-space density f in equation (2.12) is a function of only the energy per unit mass ϵ and the parameter $\sigma_{r\infty}^2$, and thus satisfies Jeans' theorem. King integrates equation

(2.12) over all velocities to give the density $n(r)$ as a function of the potential $\phi(r)$, and then solves Poisson's equation to give a self-consistent potential. The density $n(r)$ in these models falls continuously to zero at a finite radius R_h . The models can be scaled in distance and central density as with the unbounded isothermal models described earlier. The only characteristic parameter is $\sigma_r(0)/\sigma_{r\infty}$ or equivalently R_H/r_c (where r_c is again the core radius). King prefers to use the potential difference between cluster center and edge, $W_o \equiv [\phi(R_h) - \phi(0)]/\sigma_{r\infty}^2$. These models predict that the velocity dispersion declines in the outer portions of the cluster, as is observed in Coma (Rood *et al.*, 1972).

Unfortunately, none of these bounded or unbounded isothermal models can be represented exactly in terms of simple analytic functions. However, King (1962) showed that the following analytic functions were a reasonable approximation to the inner portions of an isothermal function:

$$\begin{aligned} n(r) &= n_o [1 + (r/r_c)^2]^{-3/2}, \\ \sigma(b) &= \sigma_o [1 + (b/r_c)^2]^{-1}, \end{aligned} \quad (2.13)$$

where $\sigma_o = 2n_o r_c$. At large radii $r \gg r_c$, $n(r) \approx n_o (r_c/r)^3$ in the analytic King model; thus the cluster mass and galaxy number diverge as $\ln(r/r_c)$. Although this is a slower divergence than the unbounded isothermal model, this analytic King model also must be truncated at some finite radius R_h .

Another analytic model is that of de Vaucouleurs (1948a), which was proposed to fit the distribution of surface brightness in elliptical galaxies. However, this distribution also fits many regular clusters (de Vaucouleurs, 1948b). The projected density is

$$\sigma(b) = \sigma_o \exp[-7.67(b/r_e)^{1/4}] \quad (2.14)$$

where r_e is an effective radius such that one half of the galaxies lie at projected radii $b \leq r_e$. Accurate tables of the three-dimensional density and potential for this model have been given by Young (1976). The de Vaucouleurs form has several advantages over the isothermal function. It has only one distance scale, the effective radius r_e . It also converges to a finite total number of galaxies and cluster mass without a cutoff radius. Numerical simulations of the collapse of clusters seem to lead to distributions similar to this form (see Section 2.9.2). Unfortunately, the de Vaucouleurs form has not been widely used to fit galaxy distributions in clusters, and there have been few attempts to determine objectively whether it or the isothermal models give better fits to the actual distributions.

One major difference between the isothermal functions and the de Vaucouleurs law is that the latter has a density cusp at the cluster center; in fact, as can be seen from Figure 4, many clusters show these cusps, which much be removed in order to fit isothermal sphere models to the galaxy distributions. Beers and Tonry (1986) show that the galaxy distribution in clusters is very sensitive to the position chosen for the cluster center, and that many clusters have central number density spikes if the cluster center is assumed to correspond to the position of a cD galaxy (Section 2.10.1) or the maximum of the X-ray surface brightness (Section 4.4.1). They find that the surface density near this cusp varies roughly as $\sigma(b) \propto b^{-1}$, which is consistent with a singular isothermal sphere (one with $r_c \rightarrow 0$), or with an anisotropic galaxy velocity distribution, with an excess of radial orbits. The presence of these cusps is important to understanding the occurrence of multiple nuclei and companions about cD galaxies (Section 2.10.1).

Another useful fitting form is the Hubble (1930) profile, which is sometimes used to fit the light distribution in elliptical galaxies. It is

$$\sigma(b) = \sigma_o \left(1 + \frac{b}{r_c}\right)^{-2}, \quad (2.15)$$

which has the same asymptotic distribution as equation (2.13).

These models (equations 2.11–2.15) have been used to fit the projected distribution of galaxies. In most cases, the galaxy distributions have been fit to the truncated isothermal model (equation 2.11). Figure 4 shows the surface number density distributions in 15 regular clusters, from Bahcall (1975), along with the fitting function (equation 2.11). A compilation of the values of the core radii which have been derived for clusters is given in Table III of Sarazin (1986a), which includes values from Abell (1977), Austin and Peach (1974a), Bahcall (1973a, 1974a, 1975), Bahcall and Sargent (1977), Birkenshaw (1979), Bruzual and Spinrad (1978a,b), des Forets *et al.* (1984), Dressler (1978c), Havlen and Quintana (1978), Johnston *et al.* (1981),

Koo (1981), Materne *et al.* (1982), Quintana (1979), Sarazin (1980), Sarazin and Quintana (1987), and Zwicky (1957).

Bahcall (1975) has suggested that the core radii of regular clusters are all very similar, with an average value

$$r_c = (0.25 \pm 0.04) h_{50}^{-1} \text{ Mpc.} \quad (2.16)$$

Sarazin and Quintana (1987) find that this may be true for the most compact, regular clusters. However, they also find that the core radius and galaxy distribution depend on the morphology of the cluster (Section 2.5).

The statistical uncertainty in the determination of the core radius or central density of a cluster tends to be rather large ($\gtrsim 30\%$), because even in a rich cluster only a small fraction of the galaxies are within the core. However, the errors in σ_o and r_c are highly anticorrelated, so that the product $(\sigma_o r_c)$ is relatively well-determined. The reason for this is that the number of galaxies within a projected radius b such that $r_c < b \ll R_h$ is roughly $N(b) \approx 2.17(\sigma_o r_c)b$ for an isothermal model. Thus the uncertainty in the product $(\sigma_o r_c)$ tends to be determined by Poisson statistics on the total number of cluster galaxies observed, and not just by the smaller number in the core (Sarazin and Quintana, 1987). Bahcall (1977b, 1981) has defined a related quantity \bar{N}_o as the number of ‘bright galaxies’ with projected positions within 0.5 Mpc of the cluster center. Here, bright galaxies are those no more than two magnitudes fainter than the third brightest galaxy ($m \leq m_3 + 2$). Of course, the magnitude of the third brightest galaxy m_3 itself depends on richness; \bar{N}_o is the number corrected for richness assuming a universal luminosity function (Section 2.4). From the discussion above, it is clear that $\bar{N}_o \propto (\sigma_o r_c)$, if σ_o is taken at a standard luminosity level (for example, L^* (Section 2.4)), and if $r_c \lesssim 0.5 \text{ Mpc} \ll R_h$.

Because they are better determined statistically than the core radius r_c or center surface density σ_o , $(\sigma_o r_c)$ and \bar{N}_o are often more useful as richness parameters when searching for correlations of integral properties of clusters in the optical, radio, and X-ray region. However, when comparing detailed spatial distributions the core radius is needed.

For example, from the arguments given above $\sigma_o r_c \propto \bar{N}_o \propto n_o r_c^2$, if σ_o and n_o are taken at a standard luminosity level. From equation (2.9), $r_c^2 = 9\sigma_r^2/4\pi G n_o m$, where m is the average galaxy mass $m \equiv \rho_o/n_o$, and ρ_o is the central mass density. As n_o is defined at a fixed luminosity level, this gives $(\sigma_o r_c) \propto \bar{N}_o \propto (M/L)^{-1}\sigma_r^2$, where (M/L) is the mass-to-light ratio of the cluster (Section 2.8). Bahcall (1981) finds the empirical correlation $\bar{N}_o \approx 21(\sigma_r/10^3 \text{ km/s})^{2.2}$, which suggests that cluster mass-to-light ratios decrease with σ_r . This relationship may be useful for providing quick estimates of the velocity dispersions of clusters.

Several other size scales can be determined for clusters. The halo radius R_h gives the outermost limit of the cluster. Unfortunately, this is very poorly determined, because it depends critically on the assumed background. Moreover, clusters often have very extended haloes or are embedded in extended regions of enhanced density (superclusters). For Coma, the main isothermal distribution of galaxies extends to roughly $4h_{50}^{-1}$ Mpc; there is then a low-density halo extending to $\approx 10h_{50}^{-1}$ Mpc, which blends into the Coma supercluster which extends to a radius of about $35h_{50}^{-1}$ Mpc (Rood *et al.*, 1972; Rood, 1975; Chincarini and Rood, 1976; Abell, 1977; Gregory and Thompson, 1978; Shectman, 1982). However, studies of the galaxy covariance function (Peebles, 1974) suggest that there are no preferred scales for galaxy clustering and that the outer regions of clusters and superclusters represent a continuous distribution of clustering.

Other size scales for clusters have been measured that are intermediate between the core and halo size; they include the harmonic mean galaxy separation (Hickson, 1977), which is related to the gravitational radius R_G of a cluster (Section 2.8), the de Vaucouleurs effective radius r_e defined by equation (2.14), the mean projected distance from the cluster center (Noonan, 1974; Capelato *et al.*, 1980), and the Leir and van den Bergh radius (1977).

While the distribution functions for galaxies discussed above are spherically symmetric, most clusters appear to be at least slightly elongated, and some are highly elongated (Sastry, 1968; Rood and Sastry, 1972; Rood *et al.*, 1972; Bahcall, 1974a; MacGillivray *et al.*, 1976; Thompson and Gregory, 1978; Carter and Metcalfe, 1980; Dressler, 1981; Binggeli, 1982). Carter and Metcalfe (1980) and Binggeli (1982) give ellipticities and position angles for samples of Abell clusters. Their results suggest that clusters have average intrinsic ellipticities of $\approx 0.5 - 0.7$; thus clusters are actually much more elongated on average than elliptical galaxies.

Carter and Metcalfe (1980) and Binggeli (1982) find that the position angles for the long axes of clusters are significantly aligned with the axis of the first-brightest cluster galaxy. In Sections 2.9.3 and 2.10.1, it is shown that such alignments may result if the first-brightest galaxies are produced by the merger of smaller galaxies through dynamical friction. They might also be produced during the collapse of the cluster.

Thompson (1976) has suggested that the axes of many of the elliptical galaxies in clusters may be aligned with the cluster axis; Adams *et al.* (1980) find a similar effect in two linear (L; see Section 2.5) clusters. Helou and Salpeter (1982) and Salpeter and Dickey (1985) do not find such alignments for the axes of spiral galaxies in the Virgo or Hercules clusters.

Binggeli (1982) finds that the long axes of Abell clusters tend to point at one another, even when the clusters are separated by as much as $\approx 30h_{50}^{-1}$ Mpc. The alignments of nearby clusters were found to show evidence of a correlation even up to distances a factor of three larger.

In the previous discussion, it has been assumed that the galaxy density decreases monotonically with distance from the cluster center. However, Oemler (1974) found a plateau or local minimum in the projected distribution of galaxies in many clusters at a radius of about $0.4R_G$ (where R_G is the gravitational radius). These features would imply the existence of significant oscillations in the three-dimensional galaxy density, although the process of deprojecting to counts is rather unstable (Press, 1976). Although these features are not statistically very significant in any one case, they do appear in a large number of clusters (Omer *et al.*, 1965; Bahcall, 1971; Austin and Peach, 1974a).

2.8 Cluster masses—the missing mass problem

The masses of clusters of galaxies can be determined if it is assumed that they are bound, self-gravitating systems. Cluster masses were first derived by Zwicky (1933) and Smith (1936). They found that the masses greatly exceed those which would be expected by summing the masses of all the cluster galaxies. Reviews of cluster mass determinations include Faber and Gallagher (1979) and Rood (1981).

If clusters were not bound systems, they would disperse rather quickly (in a crossing time $t_{cr} \approx 10^9$ years). Because many clusters appear regular and relaxed, and because their galactic content is quite different from that of the field, it is very unlikely that the regular compact clusters are flying apart. Assuming they are held together by gravity, one limit on the mass of clusters comes from the binding condition,

$$E = T + W < 0. \quad (2.17)$$

Here E is the total energy, T the kinetic energy, and W the gravitational potential energy:

$$\begin{aligned} T &= \frac{1}{2} \sum_i m_i v_i^2, \\ W &= -\frac{1}{2} \sum_{i \neq j} \frac{Gm_i m_j}{r_{ij}}, \end{aligned} \quad (2.18)$$

where the sums are over all galaxies, m_i and v_i are the galaxy mass and velocity, and r_{ij} is the separation of galaxies i and j .

A stronger mass limit results if we assume that the cluster distribution is stationary. The equations of motion of galaxies can be integrated to give

$$\frac{1}{2} \frac{d^2 I}{dt^2} = 2T + W, \quad (2.19)$$

where $I = \sum m_i r_i^2$ is the moment of inertia measured from the center of mass. For a stationary configuration, the left-hand-side of equation (2.19) is zero, and

$$W = -2T, \quad E = -T. \quad (2.20)$$

This is the virial theorem. The total cluster mass is $M_{tot} = \sum m_i$; we can define a mass weighted velocity dispersion $\langle v^2 \rangle \equiv \sum m_i v_i^2 / M_{tot}$, and gravitational radius

$$R_G \equiv 2M_{tot}^2 \left(\sum_{i \neq j} \frac{m_i m_j}{r_{ij}} \right)^{-1}. \quad (2.21)$$

Then, the virial theorem gives

$$M_{tot} = \frac{R_G \langle v^2 \rangle}{G} \quad (2.22)$$

Now, $\langle v^2 \rangle$ and R_G can be evaluated from the radial velocity distribution and the projected spatial distribution of a fair sample of galaxies if one assumes that the positions of galaxies and the orientation of their velocity vectors are uncorrelated. Then, $\langle v^2 \rangle = 3\sigma_r^2$, where σ_r is the (mass-weighted) radial velocity dispersion, and $R_G = (\frac{\pi}{2}) b_G$ (Limber and Mathews, 1960), where

$$b_G = 2M_{tot}^2 \left(\sum \frac{m_i m_j}{b_{ij}} \right)^{-1} \quad (2.23)$$

and b_{ij} is the projected separation of galaxies i and j . Alternatively, R_G can be calculated from a fit to the galaxy distribution (as in Section 2.7) or from strip counts of galaxies (Schwarzschild, 1954). Combining these we have

$$M_{tot} = \frac{3R_G \sigma_r^2}{G} = 7.0 \times 10^{14} M_\odot \left(\frac{\sigma_r}{1000 \text{ km/s}} \right)^2 \left(\frac{R_G}{\text{Mpc}} \right). \quad (2.24)$$

Since $\sigma_r \approx 10^3 \text{ km/s}$ and $R_G \approx 1 \text{ Mpc}$ for an average rich cluster, one finds $M_{tot} \approx 10^{15} M_\odot$ typically.

The virial theorem requires that one sum over all the particles in a cluster; it can only be directly applied to galaxies alone if they contain all of the mass in the cluster. One can also derive the mass of a cluster by treating the galaxies as test particles in dynamical equilibrium in the gravitational potential of the cluster. Such calculations generally give results for the total mass which are consistent with the virial theorem determinations; usually, one must assume that the galaxies and the total mass have the same spatial distribution. The most accurate of these determinations involve comparing numerical or analytical models for the galaxy distribution in a cluster with the observed spatial and velocity distributions (White, 1976c; Kent and Gunn, 1982; Kent and Sargent, 1983). The best data exist for the Coma and Perseus clusters; the masses for both these clusters are about $3 \times 10^{15} M_\odot$ within a projected radius of 3° from the cluster center (Kent and Gunn, 1982; Kent and Sargent, 1983).

These analyses give surprisingly large masses for clusters, particularly when compared to the total luminosity of a cluster $L_{tot} \approx 10^{13} L_\odot$. The conventional method of quantifying this comparison is to calculate the mass-to-light ratio of a system in solar units $(M_{tot}/L_{tot})/(M_\odot/L_\odot)$. Obviously, any system composed solely of stars like our Sun would have a mass-to-light ratio of unity. Mass-to-light ratios have been derived for a number of clusters using the virial theorem and the measured magnitudes of the galaxies at visual (V) or blue (B) magnitudes. Typically (see Faber and Gallagher, 1979, and Rood, 1981, and references in these reviews) one finds

$$(M/L_V)_{tot} \approx 250 h_{50} M_\odot / L_\odot, \quad (M/L_B)_{tot} \approx 325 h_{50} M_\odot / L_\odot. \quad (2.25)$$

The mass-to-light ratios found for the luminous portions of galaxies range from $(M/L_B)_{gal} \approx (1-12) h_{50} M_\odot / L_\odot$, with the large values corresponding to the E and S0 galaxies which predominate in compact regular clusters. *Thus only about 10% of the mass in clusters can be accounted for by material within the luminous portions of galaxies.* This is the so-called ‘missing mass’ problem, although this is something of a misnomer. It is not the mass which is missing; what is missing is any other evidence for this invisible matter. The application of the virial theorem to clusters and the discovery that most of their mass was nonluminous was first made by Zwicky (1933) and Smith (1936).

How secure are these mass determinations for clusters? As long as a fair sample of galaxies are observed, the virial theorem masses are insensitive to velocity anisotropies, although this can be a problem if only galaxies near the cluster center are used. The dynamical model estimates for the masses are also not affected much by velocity anisotropies (Kent and Gunn, 1982). The virial theorem applies even if most galaxies are bound in binaries or subclusters, as long as the velocities and positions are measured for a fair sample of galaxies and R_G is computed directly from the galaxy positions. However, galaxies are not generally bound in binaries in regular clusters (Abell, Neyman, and Scott, 1964), and in any case typical binary or group velocities are too small to contribute significantly to σ_r (however, see Noonan, 1975). If the clusters are bound but not relaxed, equation 2.17 would give a mass lower by only a factor of two than

equation 2.20. The velocity dispersion and R_G of a cluster might be contaminated by field galaxies. This background contamination problem can be reduced by determining the mass-to-light ratio for the core of a regular cluster, using the core parameters of an isothermal model fit (Section 2.7). Again, one typically finds $(M/L_V)_{\text{core}} \approx 300h_{50} M_\odot/L_\odot$ (Rood *et al.*, 1972; Bahcall, 1977a; Dressler, 1978b).

In principle, the missing mass could be associated with a large number of low luminosity, large M/L galaxies (see Noonan, 1976; Section 2.4). Such galaxies could affect the mass-to-light ratio if they made a significant contribution to the total light of a cluster which was not included in the usual determinations of the total luminosity because the galaxies were too faint to be detected individually. However, these galaxies would then produce a significant amount of diffuse light between galaxies in a cluster. Observations of the diffuse light in the Coma cluster show that its luminosity is at most comparable to that in bright galaxies (de Vaucouleurs and de Vaucouleurs, 1970; Melnick *et al.*, 1977; Thuan and Kormendy, 1977). Thus the required mass-to-light ratios cannot be greatly affected by low mass galaxies. Abell (1977) included a large number of low mass galaxies in his determination of the mass-to-light ratio in Coma, and found a value about a factor of two smaller than that in equation (2.25).

There is now considerable direct and indirect evidence for invisible matter associated directly with individual galaxies as well as clusters (see Faber and Gallagher, 1979, and Rood, 1981, for extensive reviews). It has been suggested that the invisible matter forms an extended halo around non-cluster galaxies, extending significantly further than the luminous material and having at least several times more mass (Einasto *et al.*, 1974; Ostriker *et al.*, 1974).

The nature of this missing mass component in clusters and galaxies remains one of the most important unsolved problems in astrophysics. Careful searches have been made for diffuse radiation in clusters, which might be produced if the missing mass were gaseous or stellar. These have included searches for radio free-free emission (Davidson and Welch, 1974), optical and UV line emission (Crane and Tyson, 1975; Bohlin *et al.*, 1973), optical continuum emission (de Vaucouleurs and de Vaucouleurs, 1970; Melnick *et al.*, 1977; Thuan and Kormendy, 1977), and 21 cm radio emission from neutral hydrogen (Shostak *et al.*, 1980). The observations of X-ray emission from clusters, which are the primary subject of this book, also give limits on the mass of hot gas (Section 5.4.1). These studies and other similar searches have shown that the missing mass is not diffuse gas at any temperature, not dust grains of the type found in the interstellar medium, and not luminous stars.

In general, most suggestions as to the identity of the missing mass fall into three categories:

- (1) substellar mass condensations, such as stars with mass $< 0.1M_\odot$ (Ostriker, Peebles, and Yahil, 1974), planetary size bodies, or comets (Tinsley and Cameron, 1974);
- (2) invisible remnants of massive stars (black holes, neutron stars, or cool white dwarfs);
- (3) stable, weakly interacting elementary particles, such as massive neutrinos, magnetic monopoles, axions, photinos, etc. (Blumenthal *et al.*, 1984).

Where is this missing mass located? A number of arguments suggest that the total mass distribution is similar to the galaxy distribution (Rood *et al.*, 1972). The fact that the galaxy distributions in compact, regular clusters are reasonably represented by isothermal spheres and that core and total mass-to-light ratios are equal within the errors supports this view. On the other hand, Smith *et al.* (1979b) and Smith (1980) have argued that the galaxy and mass distributions can be very dissimilar if the galaxies collapsed after the cluster potential was established. However, in this case violent relaxation (Section 2.9.2) is not effective, clusters could not have compact cores, and the mass-to-light ratio in the inner regions would not be the same as that for the entire cluster.

Is the missing mass bound to individual galaxies in clusters? Probably not. The rate of two-body relaxation (dynamical friction) of galaxies in clusters is proportional to galaxy masses (Section 2.9.1). If the missing mass were distributed among galaxies in proportion to their luminosity, very significant mass segregation would be expected (Rood, 1965; White, 1977b), which is not seen (Section 2.7). Moreover, if the missing mass were bound to galaxies in haloes having velocity dispersions similar to those in their visible portions, then the haloes would have to extend roughly 0.5 Mpc from the galaxy center (White, 1985). This is much larger than the typical separations of galaxies in the cores of clusters. Thus the missing mass probably forms a continuum, occupying the entire volume of the cluster, but with an overall density distribution fairly similar to that of the galaxies.

Although the missing mass is probably not currently bound to individual galaxies, it may initially have

formed massive haloes around individual galaxies, which were stripped by tidal interactions during and after the formation of the cluster (Richstone, 1976, and Section 2.10.1). As mentioned above, there is considerable evidence that more isolated galaxies are immersed in extensive massive haloes (see reviews by Faber and Gallagher, 1979, and Rood, 1981). However, the time scales for tidal interactions and dynamical friction are similar; it is not clear, therefore, that tidal interactions will strip the massive haloes before galaxies merge. The problem of the competition between tidal forces and dynamical friction is the subject of a number of recent investigations (Richstone and Malumuth, 1983; Malumuth and Richstone, 1984; Merritt, 1983, 1984a,b, 1985; Miller, 1983), which unfortunately reach contradictory conclusions (Section 2.10.1). This problem is particularly serious for the massive binary galaxies which dominate B clusters (Section 2.5).

Alternatively, the missing mass in clusters may never have been associated with galaxies. For example, the missing mass may be more extensive than the luminous material; the luminous matter, may only be the ‘tip of the iceberg’ of the real mass distribution. Rood *et al.* (1970) found that M/L increased with the scale size of the system (see also Rood, 1981), which supports this hypothesis; however, Turner and Sargent (1974) have argued that this was an artifact of calculating M from equation (2.22).

In many ways, a much better way to determine the masses of clusters of galaxies and the distribution of the missing mass is to apply the hydrostatic equation to the distribution of hot intracluster gas. This method, which is discussed in Section 5.5.5, avoids any assumption about the shapes of the orbits of galaxies, since the velocities of the particles in the gas are isotropic.

2.9 Dynamics of galaxies in clusters

In Sections 2.5, 2.6, and 2.7, evidence was presented indicating that the velocity and spatial distribution of galaxies in regular, compact clusters are in a relaxed, quasistationary state. In this section, the nature of the relaxation processes of galaxies in clusters will be briefly discussed.

If clusters result from the gravitational growth of initially small perturbations, we expect the initial state of material in the protocluster to be irregular. The relaxation of the cluster thus involves the spatial motion of the galaxies, and a lower limit to the relaxation time is the crossing time

$$t_{cr}(r) \equiv \frac{r}{v_r} \approx 10^9 \text{ yr} \left(\frac{r}{\text{Mpc}} \right) \left(\frac{\sigma_r}{10^3 \text{ km/s}} \right)^{-1}, \quad (2.26)$$

where the radial velocity dispersion σ_r has been substituted for the radial velocity v_r . An upper limit on the age of the cluster is the Hubble time (age of the universe)

$$t_H \approx (1.3 - 2.0) h_{50}^{-1} \times 10^{10} \text{ yr}, \quad (2.27)$$

where the numerical coefficient depends on the cosmological model. Thus the outer parts of a cluster or surrounding supercluster for which $r \gtrsim 10 \text{ Mpc}$ cannot possibly have relaxed and are expected to be irregular, as is observed (Sections 2.7 and 2.11.2).

2.9.1 Two-body relaxation

The phase-space distribution of galaxies in the central parts of spherical regular clusters can be represented as $f(\mathbf{r}, \mathbf{v}) \propto \exp(-\epsilon/\sigma_r^2)$, where $\epsilon \equiv \frac{1}{2}v^2 + \phi(r)$ is the energy per unit mass in the cluster (Section 2.7). This is very similar to a Maxwell–Boltzmann distribution, except that it is the energy per unit mass (velocity dispersion) which is determined, and not the energy (temperature). We first consider the possibility that clusters have relaxed thermodynamically.

Thermodynamic equilibrium could result from elastic collisions between galaxies, and would be expected if the time scale for energy transfer in such collisions were shorter than the age of the cluster or the time-scale for the loss of kinetic energy through dissipative processes. The most important elastic two-body collisions in a cluster are gravitational. The resulting relaxation times have been calculated analytically by Chandrasekhar (1942) and Spitzer and Harm (1958), and numerically through N-body simulations, including a realistic luminosity function for galaxies by White (1976c, 1977b) and Farouki and Salpeter (1982). A useful characteristic time-scale (0.94 times the Spitzer ‘reference time’) is

$$t_{rel} = \frac{3\sigma_r^3}{4\sqrt{\pi}G^2 m_m n_f \ln \Lambda} \quad (2.28)$$

for the relaxation of a galaxy of mass m in a background field of galaxies of mass m_f and number density n_f . Λ is the ratio of maximum to minimum impact parameters of collisions which contribute to the relaxation; the maximum impact parameter is on the order of half the gravitation radius of the cluster R_G (equation 2.21), while the minimum impact parameter is roughly the larger of the galaxy radius r_g or the ‘turning radius’ $Gm_f/3\sigma_r^2$ (White, 1976b). Thus

$$\Lambda \approx \min \left[\left(\frac{3R_G\sigma_r^2}{2Gm_f} \right), \left(\frac{R_G}{2r_g} \right) \right]. \quad (2.29)$$

Usually, the second value applies. Thus, for $r_g \approx 20$ kpc, $R_G \approx 1$ Mpc, $\ln \Lambda \approx 3$.

In order to give a lower limit to the relaxation time, we assume for the moment that all the mass in a cluster (including the missing mass (Section 2.8)) is bound to individual galaxies; we later show this is unlikely to be the case. Then, we define an average density $\langle \rho \rangle \equiv 3M_{tot}/(4\pi R_G^3)$, where M_{tot} is the total mass of the cluster. If we assume a Schechter luminosity function (equation 2.3), and a fixed galaxy mass-to-light ratio, we find

$$t_{rel}(m, r) \gtrsim 0.24 t_{cr}(R_G) N^* \left[\left(\frac{m}{m^*} \right) \left(\frac{\rho(r)}{\langle \rho \rangle} \right) \ln \Lambda \right]^{-1}, \quad (2.30)$$

where m^* is the characteristic galaxy mass (corresponding to L^*), N^* is the characteristic galaxy number (richness), t_{cr} is the cluster crossing time (equation 2.26), and $\rho(r)$ is the total cluster density at r . Equation (2.30) is a lower limit because it assumes all the cluster mass is bound to individual galaxies. Typically, $R_G \approx 1$ Mpc, $\sigma_r \approx 1000$ km/s, $N^* \approx 1000$, and $\ln \Lambda \approx 3$, giving $t_{cr}(R_G) \approx 10^9$ yr and $t_{rel}(m^*, R_G) \gtrsim 3 \times 10^{11}$ yr. This is much longer than a Hubble time (equation 2.27); it is therefore unlikely that the apparently relaxed state of regular clusters results from two-body collisions. However, two-body relaxation processes can affect the more massive galaxies ($m \gg m^*$) near the cluster center ($\rho(r) \gg \langle \rho \rangle$), as is discussed below.

2.9.2 Violent relaxation

The fact that clusters exhibit a nearly constant velocity dispersion (rather than kinetic temperature) suggests that the relaxation is produced by collective gravitational effects. Lynden-Bell (1967) showed that collective relaxation effects can result in a very rapid quasi-relaxation (‘violent relaxation’). The existence of these effects involves a somewhat subtle point; collective relaxation develops through collisionless interactions, and thus the detailed (‘fine-grain’) phase-space distribution of galaxies is conserved. However, if the relaxation is sufficiently violent (for example, the energy of particles is changed by a significant fraction), then initially adjoining units of phase-space will be widely separated in the final state (‘phase mixing’). Thus, if one averages the fine-grain phase-space density over any observable volume to give a ‘coarse-grain’ distribution, this coarse-grain distribution can be an equilibrium distribution and independent of the details of the initial state.

If clusters were formed by the growth and collapse of initial perturbations, then during the collapse the gravitational potential ϕ fluctuated violently. This would cause a change in the energy per unit mass $\epsilon \equiv \frac{1}{2}v^2 + \phi(r)$ of

$$\frac{D\epsilon}{Dt} \approx \frac{\partial \phi}{\partial t}. \quad (2.31)$$

For example, if the cluster collapsed from a stationary state ($v = 0$) to a virialized final state (Section 2.8), then $\Delta\epsilon \approx \Delta\phi \approx \epsilon$. The time scale for collapse t_{coll} is roughly a dynamical time scale or crossing time

$$t_{coll} \approx (G\langle \rho \rangle)^{-1/2} \approx (R_G^3/GM_{tot})^{1/2} \approx t_{cr}(R_G). \quad (2.32)$$

As the energy of a galaxy changes by $\approx 100\%$ during a collapse time, violent relaxation can be completed during the collapse of the cluster, after which time the potential is constant, and the galaxy distribution is in stationary virial equilibrium.

Since the equation of motion of a particle in the cluster’s mean gravitational field is independent of mass (equation 2.31), the equilibrium is independent of mass. Lynden-Bell (1967) derived an equilibrium state by assuming that the system relaxed to the most probable coarse-grained phase-space state subject to conservation of energy, particle number, and fine-grain phase density. This equilibrium state was found

to be a Fermi–Dirac distribution that reduces to the Maxwell–Boltzmann distribution for the appropriate number densities in clusters (see also Shu, 1978),

$$f(\mathbf{r}, \mathbf{v}) \propto \exp(-\epsilon/\sigma_r^2). \quad (2.33)$$

This phase space density produces a Gaussian velocity distribution and isothermal spatial distribution of galaxies, which are roughly consistent with the observed distributions in the inner parts of clusters (Sections 2.6 and 2.7). The distribution is also independent of mass, roughly as observed. If equation (2.33) held for all radii, the cluster mass would be infinite (Section 2.7). However, galaxies at large radii or with large energies never reach equilibrium because their crossing times (equation 2.26) are longer than the Hubble or collapse times (equations 2.27 and 2.32). Because the relaxation occurs only while the cluster is collapsing, it is not clear that this equilibrium state can ever be achieved in any real collapsing system.

The state of equilibrium of a collisionless, gravitationally collapsing system can be studied directly through numerical N-body experiments. These experiments do show that isolated collapsing or merging systems relax rapidly (in a few crossing times), and they agree roughly on the nature of the equilibrium state. In general, they *do not* find that the systems become isothermal spheres; instead they find spatial distributions that are reasonably represented by the de Vaucouleurs (1948a,b; equation 2.14) or Hubble (1930; equation 2.15) forms (White, 1979; Villumsen, 1982; van Albada, 1982). These distributions are more centrally condensed and fall off more rapidly at large radii than the isothermal sphere. However, it is possible that the subsequent collapse of surrounding material can increase the density at large distances; Gunn (1977) has shown that for some initial conditions this process can lead to nearly isothermal mass distributions.

The idea that the distribution of galaxies in clusters is determined by violent relaxation during the formation of the cluster provides a simple explanation for the one-dimensional morphological sequence of clusters running from irregular to regular (Table 1; Gunn and Gott, 1972; Jones *et al.*, 1979; Dressler, 1984). The regular clusters are those old enough to have collapsed and relaxed, while the irregular ones have not. As the collapse time is $t_{coll} \approx (R_h^3/GM_{tot})^{1/2} \approx (G\rho_i)^{-1/2}$, where ρ_i is the initial density, higher density protoclusters will collapse more rapidly. Since the age of clusters is limited by the Hubble time t_H , *regular clusters will be produced by higher density protoclusters, and irregular clusters by lower density protoclusters*. Thus we expect regular clusters to have higher densities than irregular clusters, as is observed. Moreover, violent relaxation and phase mixing will eliminate subclustering and produce a centrally condensed, symmetric distribution, as observed in regular clusters.

White (1976c) has followed the collapse of a model cluster numerically, with an N-body code; the results are shown in Figure 5. He finds that the cluster first forms irregular subcondensations around massive galaxies (like an I cluster). These continually merge until the galaxy distribution is elongated and has two large clumps (like an F cluster), which merge to form a smooth, regular cluster with a prominent core (a C cluster) (see also Henry *et al.*, 1981; Forman and Jones, 1982). Recently, Cavaliere *et al.* (1983) have produced many similar models for cluster collapse.

The other aspects of the morphological sequence have to do with galactic content—the fractions of spiral, S0, and elliptical galaxies, or domination by supergiant (B, D, and cD) galaxies. In Section 2.10 evidence is presented which suggests that the galactic content of clusters is also determined by the density of the cluster, although the mechanism is still controversial.

2.9.3 Ellipsoidal clusters

The distribution in equation (2.33) is isotropic in velocity space and spherically symmetric in real space. However, many of the most regular clusters have observed galaxy distributions that are not symmetric on the sky (Section 2.7). In principle, this asymmetry could be due to rotation; however, the velocity fields in clusters show no evidence for dynamically significant rotation (Section 2.6).

Numerical N-body simulations of the formation of clusters show that if the initial distribution of galaxies is aspherical, the final distribution after violent relaxation will be aspherical (Figure 5d; Aarseth and Binney, 1978). The most general final configurations have compact cores and are regular, but the surfaces of constant density are basically triaxial ellipsoids rather than spheres. The triaxial shape is maintained by an ellipsoidal Gaussian velocity distribution $p(v) \propto \exp[-(2\Sigma_{ij})^{-1}v_i v_j]$, where Σ is the velocity dispersion tensor. The principal axes of Σ and the spatial figure of the galaxy are parallel and nearly proportional to one another (Binney, 1977), as expected from the tensor virial theorem (Chandrasekhar, 1968). Thus the velocity dispersion is largest parallel to the longest diameter of the cluster.

2.9.4 Dynamical friction

Once the collapse of a cluster is completed, violent relaxation is ineffective, and further relaxation occurs through two-body interactions. As shown above, these effects are probably not important for a typical galaxy at an average position in the cluster, but they can be significant for a more massive galaxy ($m \gg m^*$) in the cluster core ($\rho \gg \langle \rho \rangle$; see equation 2.30). Chandrasekhar (1942) showed that a massive object of mass m moving at velocity \mathbf{v} through a homogeneous, isotropic, Maxwellian distribution of lighter, collisionless particles suffers a drag force given by

$$\frac{d\mathbf{v}}{dt} = -\mathbf{v} \left\{ \frac{4\pi G^2 m}{v^3} \ln(\Lambda) \rho_{tot} [\text{erf}(x) - x \times \text{erf}'(x)] \right\} \quad (2.34)$$

where $x \equiv v/(\sqrt{2}\sigma_r)$, σ_r is the radial velocity dispersion of the lighter particles, ρ_{tot} is their total mass density, erf is the error function, and Λ is given by equation (2.29). Note that this ‘dynamical friction’ force is independent of the mass of the lighter particles, and thus depends only on the total density ρ_{tot} of such particles (regardless of whether they are galaxies or collisionless missing mass particles).

The relaxation time in the cluster core can be evaluated by noting that, for an isothermal cluster distribution (Section 2.7), the central density ρ_o , velocity dispersion σ_r , and core radius r_c are related by

$$4\pi G \rho_o r_c^2 = 9\sigma_r^2. \quad (2.35)$$

Let $m = (m/m^*)m^* = (m/m^*)(M/L_V)_{gal}L_V^*$, where L_V^* and m^* are the characteristic visual luminosity and corresponding mass of galaxies in the Schechter luminosity function (Schechter, 1976, and Section 2.4), and M/L_V is the visual mass-to-light ratio. Schechter finds $L_V^* \approx 4.9 \times 10^{10} L_\odot$. Substituting this value, $\ln \Lambda \approx 3$, and the ρ_o from equation (2.35) into equation (2.28) gives for the dynamical friction time-scale

$$t_{rel}(m, r = 0) \approx 6 \times 10^9 \text{ yr} \left(\frac{\sigma_r}{1000 \text{ km/s}} \right) \times \left(\frac{r_c}{0.25 \text{ Mpc}} \right)^2 \left(\frac{m}{m^*} \right)^{-1} \left[\frac{(M/L_V)_{gal}}{10M_\odot/L_\odot} \right]^{-1}. \quad (2.36)$$

Dynamical friction slows down the more massive galaxies near the center of a spherical cluster; they spiral in toward the cluster center (Lecar, 1975). The kinetic energy removed from the massive galaxies is transferred to the lighter particles (galaxies or missing mass components), which then expand. Thus dynamical friction produces mass segregation in a cluster; more massive galaxies are found preferentially at smaller radii. Moreover, at a fixed radius the velocity dispersion of the more massive galaxies will be lower, as they have been slowed down.

If the cluster is aspherical because of an anisotropic velocity dispersion, the dynamical friction force will not be parallel to the velocity (Binney, 1977). Because the force depends inversely on velocity, it is strongest along the shortest axis of the cluster. Thus dynamical friction increases the anisotropy of the most massive galaxies in a cluster. This can explain the formation of L clusters such as Perseus, in which the brightest galaxies form a narrow chain along the long axis of the cluster (Binney, 1977).

The massive galaxies that spiral into the cluster center will eventually merge to form a single supergiant galaxy if they are not tidally disrupted first. In Section 2.10.1 evidence is given which suggests that this is the mechanism by which cD galaxies form; however, there are also strong arguments that galaxies lose mass due to tidal effects so rapidly that dynamical friction is not important (Merritt, 1983, 1984a, 1985). The merger of the most massive and therefore brightest galaxies in the cluster core might cause the luminosity function to fall off more rapidly at high luminosities, as may have been observed in some cD clusters (Dressler, 1978b), and might produce a deficiency of brighter galaxies (other than the cD) in the cluster core (White, 1976a).

The fact that the dynamical friction time scale depends inversely on galaxy mass provides a method to measure the mass bound to individual galaxies in a cluster, as opposed to the total virial mass of the cluster. If the mass-to-light ratio derived for the cluster as a whole (typically $(M/L_V) \approx 250M_\odot/L_\odot$; equation (2.25)) is substituted in equation (2.36), $t_{rel} \approx 2 \times 10^8 \text{ yr} (m/m^*)^{-1}$, which is much smaller than the probable age of a cluster of about a Hubble time $t_H \approx 10^{10} \text{ yr}$ (for typical parameters for a compact, regular cluster). Thus,

if the missing mass in clusters were bound to individual galaxies, the massive galaxies $m \gtrsim m^*$ would be highly relaxed (at least in the cluster core). All of these galaxies would have segregated into a very small core and possibly merged to form a single galaxy. Observations of the degree of mass segregation, luminosity function, and luminosity of the brightest galaxies in well studied clusters are all inconsistent with this much relaxation. From the observations one can limit the mass-to-light ratios of massive galaxies in cluster cores to $(M/L_V) \lesssim 30M_\odot/L_\odot$. This indicates that the missing mass cannot be bound to individual, massive galaxies, but rather must form a continuum. This important result was first given by Rood (1965), and has been verified by numerical integration of equation (2.34) and N-body models by White (1976a, 1977b) and Merritt (1983).

2.10 Galactic content of clusters

In the discussion of the morphological classification of clusters (Section 2.5), it was noted that the regular, compact clusters have a galactic content that differs markedly from that of the field. First, these clusters are often dominated by a single very luminous (cD) galaxy, or by a pair of very bright galaxies. Second, elliptical and S0 galaxies predominate over spiral galaxies in regular, compact clusters, where the opposite is true in the field. In this section possible origins for these differences are described and are related to the overall picture of cluster evolution given in the preceding section.

2.10.1 cD galaxies

cD galaxies were defined by Mathews, Morgan, and Schmidt (1964) as galaxies with a nucleus of a very luminous elliptical galaxy embedded in an extended amorphous halo of low surface brightness. They are usually found at the center of regular, compact clusters of galaxies (Morgan and Lesh, 1965; Bautz and Morgan, 1970), and about 20% of all rich clusters contain cD galaxies. However, some galaxies that appear to be cDs have been found in poor clusters and groups (Morgan, Kayser, and White, 1975; Albert, White, and Morgan, 1977).

cD galaxies are extremely luminous; if one excludes nuclear sources (Seyfert galaxies, N galaxies, and quasars), they are, as a class, the most luminous galaxies known. Sandage (1976) and Hoessel (1980) find average absolute magnitudes $\langle M_V \rangle_{cD} \approx -23.7 + 5 \log h_{50}$ and $-22.7 + 5 \log h_{50}$ for apertures of $43/h_{50}$ kpc and $19/h_{50}$ kpc, respectively. Since the galaxies extend well beyond this, total luminosities are at least a magnitude brighter. Moreover, the magnitudes of cDs show a rather small dispersion (≈ 0.3 mag) and are only weakly dependent on cluster richness (Sandage, 1976). cD galaxies are usually considerably brighter (often by a full magnitude) than other galaxies in the same cluster.

The question naturally arises as to whether cD galaxy luminosities simply represent the high end of the normal galaxy luminosity function, or whether cDs are ‘special’. The distribution of brightest galaxy luminosities L_1 in a cluster would be $p(L_1)dL_1 = \exp(-N(L_1))dN(L_1)$ if the brightest galaxies were drawn at random from the integrated luminosity function $N(L)$ (Section 2.4). This distribution would produce a large dispersion in L_1 and a strong dependence on cluster richness (neither of which are observed), unless the luminosity function were much steeper than is observed at the bright end (Sandage, 1976). However, cD luminosities themselves are too high to be drawn from the galaxy luminosity function (Schechter, 1976), unless it were less steep than is observed. While selection effects may account for the small dispersion of brightest elliptical and S0 galaxies in groups and clusters (Schechter and Peebles, 1976), it does not appear that this can explain the high luminosities of cD galaxies. A number of statistical tests (Sandage, 1976; Tremaine and Richstone, 1977; Dressler, 1978a) indicate that the magnitudes of cD galaxies cannot be drawn statistically from a general galaxy luminosity function.

cD galaxies are more extended than the other giant elliptical galaxies in two ways (see Figure 6a and 6b where the surface brightness of the cD in A1413 is compared to the giant elliptical NGC1278 in the Perseus cluster). First, the core regions of cDs are apparently larger. Hoessel (1980) finds that cD galaxies have core radii that average $\langle r_c(gal) \rangle \approx 4/h_{50}$ kpc, while typical giant elliptical galaxies have $\langle r_c(gal) \rangle \approx 0.4/h_{50}$ kpc (here, the surface brightness of the galaxy at projected radius b assumed to vary as $\{1 + [b/r_c(gal)]^2\}^{-1}$). Unfortunately, these core radii represent fairly small angles and their determination may be ambiguous (Dressler, 1984). If one fits galaxy surface brightnesses to a de Vaucouleurs (1948a) form (equation 2.14), the effective radii r_e of cD galaxies are roughly a factor of two larger than the effective radii of normal giant elliptical galaxies. Alternatively, one can define the integrated luminosity slope

$$\alpha \equiv \frac{d \ln L(b)}{d \ln b}, \quad (2.37)$$

where $L(b)$ is the luminosity observed within the projected radius b . Hoessel (1980) finds $\langle \alpha \rangle_{cD} \approx 0.59$ for $b = 19/h_{50}$ kpc, which is about twice the value found for typical giant ellipticals.

In cD galaxies in rich clusters, the giant elliptical-like core of the galaxy is embedded in a very extended low surface brightness halo (Oemler, 1973, 1976; Carter, 1977; Dressler, 1979). In Figures 6a,b the surface brightness of the very extended cD in the cluster A1413 is compared to that of a typical giant elliptical galaxy NGC1278 in the Perseus cluster (Oemler, 1976). The solid line shows a de Vaucouleurs fit (equation 2.14) to each galaxy. This profile fits the inner parts of either galaxy reasonably well, but the cD in A1413 has a halo of low surface brightness extending to beyond 1 Mpc from the galactic center. The surface brightness in cD haloes generally falls off as roughly the 3/2 power of projected distance from the galaxy center.

Masses of cD galaxies have been estimated by a measurement of the stellar velocity dispersion in the outer parts of the cD in A2029 (Dressler, 1979) and by measurements of the velocities of smaller companion galaxies which are assumed to be bound to the cD (Wolf and Bahcall, 1972; Jenner, 1974). Typically, one finds $M \approx 10^{13} M_\odot$, but it is difficult to separate the galaxy and cluster mass distributions in the outermost parts of the cD (Dressler, 1979). For the same central velocity dispersion, cD galaxies are about 60% brighter than other giant ellipticals (Malumuth and Kirshner, 1981).

cD galaxies are usually found very near the centers of compact clusters (Morgan and Lesh, 1965; Leir and van den Bergh, 1977; White, 1978a). They also have velocities very near the mean velocity of galaxies in the cluster (Quintana and Lawrie, 1982), and may in fact give a better estimate of the cluster mean velocity than the average value for a few bright galaxies. These results suggest that cDs are usually sitting at rest at the bottom of the cluster gravitational potential well.

cD galaxies often have double or multiple nuclei (Minkowski, 1961; Morgan and Lesh, 1965; Hoessel, 1980; Schneider and Gunn, 1982); that is, there are several peaks in the surface brightness within the central part of the cD. Many cDs appear to be in binary or multiple galaxy systems (Leir and van den Bergh, 1977; Rood and Leir, 1979; Struble and Rood, 1981); when two cD nuclei appear to be surrounded by a common halo, these are referred to as ‘dumbbell’ galaxies. One important problem with these multiple nuclei is whether they are physically associated with and bound to the cD, or whether they are chance projections. In many cases, the multiple nuclei have rather large velocities relative to the cD, and cannot be bound to it (Jenner, 1974; Tonry, 1984, 1985a,b; Hoessel *et al.*, 1985). These multiple nuclei must then be chance projections. Such projections would be unlikely if galaxies in clusters have isotropic velocity dispersions and constant surface number density cores. However, they are much more likely if galaxies have radial orbits, and the cluster density has a cusp near the center (Tonry, 1985a; Merritt, 1984b, 1985), as appears to be the case in many clusters (Beers and Tonry, 1986; Section 2.7).

The special structural and kinematic properties of cD galaxies suggest that they have been formed or modified by dynamical processes in clusters. Gallagher and Ostriker (1972) and Richstone (1975, 1976) have suggested that cDs consist of the debris from galaxy collisions. In a rich cluster, the outer envelopes of galaxies will be stripped by tidal effects during these collisions. The rate of mass loss due to tidal collisions in a cluster was derived by Richstone (1976); if his rate is integrated over a Schechter luminosity function (Section 2.4) and an isothermal galaxy density function with core radius r_c (Section 2.7), one finds

$$\left(\frac{dM}{dt} \right)_{tidal} \approx 3 \times 10^4 M_\odot \text{ yr}^{-1} \left(\frac{\sigma_r}{10^3 \text{ km/s}} \right)^4 \left(\frac{\sigma_*}{200 \text{ km/s}} \right)^{-1} \left(\frac{r_t}{r_c} \right), \quad (2.38)$$

where σ_r is the cluster line-of-sight velocity dispersion (Section 2.6), and r_t and σ_{gal} are the tidal (outermost) radius and line-of-sight stellar velocity dispersion of a typical galaxy ($L = L^*$; see Section 2.4). Equation (2.38) assumes that all the cluster mass is in galaxies; if this is not the case, then the rate of mass loss from galaxies is reduced by the square of the fraction of mass in galaxies.

Knoblock (1978a,b) and Da Costa and Knoblock (1979) have argued that Richstone’s expression overestimates the rate of mass loss, because the tidal stripping of a galaxy halo is limited by the rate of diffusion of stars into the halo. However, they take as their basic model a galaxy in which the halo mass is very small; this does not agree with the determinations of the mass distributions in galaxies (see, for example, Faber and Gallagher, 1979). For realistic mass models, it appears that the claimed discrepancy is a factor of roughly (σ_*/σ_r) .

If most of the mass in a cluster were initially in extended ($r_t \approx r_c \approx 300$ kpc) haloes of dark material bound to individual galaxies, equation (2.38) shows that these dark haloes could have been stripped by tidal interactions within the age of a cluster, as was suggested in Section 2.8.

The outer portions of the luminous material would also be stripped (Strom and Strom, 1978a, 1979); if we assume a tidal radius of $r_t \approx 0.1r_c \approx 30$ kpc, we would expect $\gtrsim 10^{12}L_\odot$ of luminous material to be stripped from galaxies in a compact cluster. The stripped material would settle to the center of the cluster gravitational potential; because of the dependence of the collision rate on the square of galaxy density, the stripped material would probably be somewhat more centrally condensed than the galaxy distribution in the cluster. This tidal debris might thus be observed as the extended halo about a central cD galaxy. The stripped material would be centered on the cluster center, be at rest (on average) relative to the cluster center-of-mass, and have a mass and spatial distribution similar to that observed for cD haloes. This model for the formation of cD haloes (Gallagher and Ostriker, 1972; Richstone, 1975, 1976) predicts that they have high velocity dispersions, as has been observed for the cD in A2029 by Dressler (1979). This tidal debris model may explain the observed properties of cD haloes; however, it cannot naturally explain the giant elliptical galaxy nucleus of the cD.

Ostriker and Tremaine (1975), Gunn and Tinsley (1976), and White (1976a, 1977a) have suggested that cD galaxies are produced by the merger of massive galaxies within the core of a cluster. As discussed in Section 2.9, dynamical friction causes the orbits of massive cluster galaxies to decay. As such galaxies reach the cluster center, they merge to form a single supergiant galaxy, which swallows any galaxies that subsequently pass through the cluster center. The merger hypothesis (called ‘galactic cannibalism’ by Ostriker and Hausman, 1977) provides an attractive explanation for the formation of cD galaxies in a cluster.

First of all, since a cD galaxy would be produced by the merger of many of the more luminous galaxies in a cluster, the high luminosities of cDs (in excess of that expected from the galaxy luminosity function) is naturally explained. Once a massive galaxy reaches the cluster center, it will swallow other galaxies. Initially, its luminosity will increase at the rate (Ostriker and Tremaine, 1975)

$$\frac{d(L_{cD}/L^*)}{d(t/10^{10} \text{ yr})} \approx 13h_{50}^{-1} \left(\frac{r_t}{100 \text{ kpc}} \right)^3 \left(\frac{r_c}{250 \text{ kpc}} \right)^{-4} \times \left(\frac{\sigma_r}{10^3 \text{ km/s}} \right) \left(\frac{M/L_V}{30M_\odot/L_\odot} \right), \quad (2.39)$$

where r_t is the tidal radius of the central galaxy, r_c and σ_r are the core radius and velocity dispersion of the cluster, and M/L_V is the visual mass-to-light ratio of all galaxies in the cluster. (Very recently, Merritt (1985) has shown that equation 2.39 greatly overestimates the actual merger rate at early times in a cluster.) Once most of the massive galaxies within the core have been accreted, the luminosity grows as $t^{1/2}$ rather than as t (Ostriker and Tremaine, 1975):

$$\left(\frac{L_{cD}}{L^*} \right) \approx 8 \left(\frac{\sigma_r}{10^3 \text{ km/s}} \right)^{3/2} \left(\frac{M/L_V}{30M_\odot/L_\odot} \right)^{1/2} \left(\frac{t}{10^{10} \text{ yr}} \right)^{1/2}. \quad (2.40)$$

Thus luminosities of $L_{cD} \approx 10L^*$ can be produced during the lifetime of a compact cluster.

The merger product should be larger than the initial galaxy, because the kinetic energy of the merging galaxies ‘heats’ and inflates the final product (Ostriker and Hausman, 1977; Hausman and Ostriker, 1978). Hausman and Ostriker (1978) argue that the galaxy core radius $r_c(\text{gal})$ will increase by roughly a factor of ten, and the structure parameter α by a factor of at least two, in rough agreement with the observations (Hoessel, 1980).

With the large increase in the luminosity of the first-brightest galaxy through mergers (equation 2.40), it might seem surprising that these galaxies show a rather small dispersion in absolute magnitude (Sandage, 1976). The observed magnitudes are for fixed measuring apertures, which are generally much smaller than the halo size of the cD; depending on the amount of swelling the cD undergoes, the observed magnitude may either increase or decrease (Gunn and Tinsley, 1976). In the simulations of Hausman and Ostriker (1978), the apparent magnitude within an aperture of 16 kpc remains roughly constant once the massive galaxies in the core of the cluster have been swallowed. This may explain the small dispersion in observed cD absolute magnitudes.

The first-brightest galaxy grows in luminosity by swallowing other massive galaxies; this increases the contrast between the first-brightest and the other bright galaxies (Gunn and Tinsley, 1976; Hausman and

Ostriker, 1978), as is observed (Sandage, 1976; Tremaine and Richstone, 1977; Dressler, 1978a; McGlynn and Ostriker, 1980).

Of course, the process of dynamical friction and merging leaves the merger product nearly at rest relative to the average cluster galaxy and nearly at the center of the cluster, as is observed for cDs.

If a cluster is elongated, then the merger product at its center will tend to be elongated in the same direction. In general, cD galaxies are more highly flattened than other ellipticals (Leir and van den Bergh, 1977; Dressler, 1978c), and the axes of the cDs align with those of their clusters, as predicted (Sastry, 1968; Rood and Sastry, 1972; Dressler, 1978c, 1981; Carter and Metcalfe, 1980; Binggeli, 1982).

If cD galaxies are produced by mergers of massive galaxies, one expects mergers to take place roughly every 10^9 years (Ostriker and Hausman, 1977). The actual merger will occur over roughly an orbital period in the cD galaxy $\approx 3 \times 10^8$ years (White, 1978b), and for this period the nucleus of the merging galaxy will be visible within the cD envelope. Thus one would expect about 1/4 of cD galaxies to have multiple nuclei, roughly as is observed (Hoessel, 1980). However, a major problem with this interpretation of multiple nuclei in cDs is that these nuclei, in many cases, have rather large velocities relative to the cD (Jenner, 1974; Tonry, 1984, 1985). This suggests that they are not bound to the cD, but might only be cluster galaxies seen in projection against the cD.

Rood and Leir (1979) find that about 1/4 of all cD galaxies are binaries (actually dumbbells); they argue that this fraction is much larger than is expected due to dynamical friction and merging. The reason is that in these systems the two galaxies in the binary differ by less than a magnitude in apparent luminosity. However, the primary component of the system would have swallowed roughly five other galaxies in the merger picture, and therefore should be much brighter than its companion. Since one magnitude corresponds to a factor of 2.5 in brightness, and magnitudes at a fixed aperture do not increase directly with total luminosity because of swelling, it is not clear how serious a discrepancy this is (see also Tremaine, 1981).

The dynamical friction and merger theory may explain most of the observed properties of cD galaxies except for their very extended haloes. The tidal debris theory explains the extended haloes, but not the properties of the inner elliptical galaxy. It seems natural to suggest that cD galaxies result from the action of both types of processes.

However, Merritt (1983, 1984a,b, 1985) has recently pointed out some very serious problems with this simple picture of cD formation. He argues that the tidal effect of the global cluster potential (as opposed to that due to individual galaxy interactions) dominates the evolution of cluster galaxies. Tidal stripping of galaxies would then lower their masses and prevent any mergers. If this argument is correct and if mergers formed cDs, then the mergers must have occurred before the cluster collapsed, perhaps in smaller subclusters or groups (Carnevali *et al.*, 1981; van den Bergh, 1983a; Dressler, 1984).

Theories of the origin of cD galaxies can be tested through observations of the apparent cD galaxies in 23 poor clusters which were discovered by Morgan, Kayser, and White (1975) (hereafter MKW) and Albert, White, and Morgan (1977) (hereafter AWM). These clusters range in Abell richness (Section 2.3) from < 10 to ≈ 50 galaxies; none are richer than Abell richness class zero (Bahcall, 1980). These clusters have velocity dispersions σ_r (Stauffer and Spinrad, 1978) and core radii r_c (Bahcall, 1980) which are probably about 1/2 of those of rich, compact clusters. Since the dynamical friction rate (when summed over all galaxies) varies as σ_r/r_c (equation 2.30) and the merger rate varies as σ_r/r_c^4 (equation 2.39), one expects mergers to form remnants in these poor clusters even larger than those in rich clusters. On the other hand, the tidal stripping rate (equation 2.38) varies as σ_r^4/r_c , and is thus much lower in poor clusters. If mergers form the extended elliptical galaxy body of a cD and tidal debris makes up the very extended halo, the cDs in poor clusters should lack such haloes. Surface photometry of the cDs in poor clusters (Oemler, 1976; Stauffer and Spinrad, 1980; Thuan and Romanishin, 1981) indicates that this is indeed the case. In Figure 6c, the surface photometry of the cD in the poor cluster AWM4 as observed by Thuan and Romanishin (1981) is shown.

While the processes of merging and tidal stripping, perhaps in preexisting subclusters and groups, seem to provide a possible explanation of many of the unusual properties of cD galaxies, the core of a rich compact cluster is a very active physical environment in which many other processes may be important. For example, recent X-ray and optical observations suggest that central, dominant galaxies are accreting vast quantities of gas, as much as $400 M_\odot/\text{yr}$ in the case of NGC1275 in the Perseus cluster (Sections 4.3.3 and 5.7). The accreting galaxies include a number of cDs. There are several arguments which suggest that the accreted gas is being converted into low mass stars (Cowie and Binney, 1977; Fabian *et al.*, 1982a; Sarazin and O'Connell,

1983); if so, accretion can significantly increase the core luminosities of cDs (Fabian *et al.*, 1982a; Sarazin and O'Connell, 1983).

Finally, I would like to comment on a semantic issue that has arisen concerning cDs and other similar galaxies. It should be clear from the discussion above that cDs represent the extreme result of a number of dynamical processes which must be occurring continuously in clusters and groups. There are a number of related types of galaxies which share many but not all of the properties of cDs; they may be extended, and be the brightest galaxy in a cluster (or nearly so), and be located at the spatial and velocity center of the cluster. In Section 4.4.2 we shall see that such galaxies may affect the X-ray morphology of clusters, even if they do not satisfy the technical definition of a cD galaxy; examples are M87 in the Virgo cluster and NGC1275 in the Perseus cluster. It has sometimes been argued that the definition of cD galaxies should be extended in some way to include all such objects. It seems to me that this is not a good idea; it is more useful to retain the degree of specificity in the original definition of cD galaxies, based on their optical properties as defined by Morgan and his collaborators (Mathews, Morgan, and Schmidt, 1964). I suggest that a new term, 'central dominant galaxy' (abbreviated cd?) be used for galaxies that are the brightest cluster member or within 0.5 magnitudes of the brightest, and that appear to be at rest at the cluster center.

2.10.2 Proportion of spiral, S0, and elliptical galaxies

Elliptical and S0 galaxies are more common than spiral galaxies in the inner portions of regular compact clusters, while the opposite is true in irregular clusters and in the field (Table 1). Many explanations have been proposed for the origin of this systematic variation in galactic content. In general, these theories fall into two broad classes. In the first class, the proportion of galaxy types is set by the conditions when the galaxies form, and once the galaxies form they do not alter their morphology. Thus, in regions that are or will become regular clusters, more ellipticals are formed. In the second group of theories, galaxies may form with the same distribution of morphological types everywhere, but physical processes that depend on environment cause galaxies to alter their morphology. That is, in compact regular clusters, spirals become S0s or ellipticals, and S0s become ellipticals. Often an analogy is made between the generation of variations in the character of galaxy populations and human populations; in the first case, galaxy morphology is determined at conception (the 'heredity' hypothesis), while in the second case it is influenced primarily by the 'environment' in later life.

From the point of view of the environment theories, the primary difference between spiral galaxies and S0 or elliptical galaxies is that spirals contain much more gas. As a result, spirals have active star formation, and the gas allows shocks, which delineate the spiral structure. The basic idea behind most of the environment theories, first suggested by Spitzer and Baade (1951), is that spiral galaxies would become S0 galaxies if their gas were removed. A number of mechanisms have been proposed to remove the gas from spiral galaxies in clusters; these are reviewed in detail in Section 5.9.

Spitzer and Baade (1951) suggested that collisions between spiral galaxies in the cores of compact clusters remove the interstellar medium from the disks of these galaxies. Subsequent increases in the estimates of the distance scale to clusters have had the effect of seriously reducing the efficiency of this process (Section 5.9). It certainly could strip some of the spirals in a cluster. However, if the cluster has a significant amount of intracluster gas, there are several other processes which are more efficient.

Gunn and Gott (1972) suggested that S0 galaxies are formed when spiral galaxies lose their interstellar medium through ram pressure ablation on intracluster gas. This process has now been studied fairly extensively, and is reviewed in Section 5.9. In general, these studies indicate that a galaxy will be stripped almost completely during a single passage through the core of a cluster if the intracluster gas density exceeds roughly 3×10^{-4} atoms/cm³.

Another mechanism for removing gas from galaxies, which can operate even when the galaxies are moving slowly through the intracluster gas, is evaporation (Cowie and Songaila, 1977; Sections 5.9 and 5.4.2). Heat is conducted into the cooler galactic gas from the hotter intracluster gas, and if the rate of heat conduction exceeds the cooling rate, the galactic gas will heat up and flow out of the galaxy (equation 5.118). The evaporation rate can be significantly reduced if the conductivity saturates (Section 5.4.2), or because of the magnetic field (Section 5.4.3). If the conductivity is not suppressed by these effects, this mechanism can play an important role in stripping gas from galaxies.

The X-ray observations that are the primary subject of this book have shown that many clusters have intracluster gas densities high enough to make ram pressure stripping effective (Section 5.9; equation 5.115).

In addition, there is a strong inverse correlation between the X-ray luminosity and the fraction of spiral galaxies in a cluster (Bahcall, 1977c; Melnick and Sargent, 1977; Tytler and Vidal, 1979; equation 4.9). The spiral fraction also decreases as the velocity dispersion of the cluster increases, as required for ram pressure ablation, and the spirals that are observed in X-ray clusters are on average at large projected distances from the cluster center (Gregory, 1975; Melnick and Sargent, 1977), where the intracluster gas density is presumably much lower than in the cluster core. Although not directly connected with stripping from spiral galaxies, the X-ray observations of M86 (an elliptical galaxy in the Virgo cluster) suggest that it is currently undergoing ram pressure ablation (Forman *et al.*, 1979; Fabian *et al.*, 1980).

Ram pressure ablation or evaporation removes the gas from cluster galaxies, and this prevents ongoing star formation. Thus one would expect that even when spiral galaxies do occur in a cluster, they will have less gas than field spirals and less active star formation. Studies of the 21 cm radio line of hydrogen from spiral galaxies in clusters indicate that they have considerably less gas than field spirals, and that the amount of gas present increases with distance from the cluster center (Section 3.7). Spirals in clusters also show weaker optical line emission than those in the field, which also suggests they have less gas (Gisler, 1978); but see Stauffer 1983); the same is apparently true of cluster ellipticals and S0s (Davies and Lewis, 1973; Gisler, 1978). Moreover, many of the spirals in clusters have poorly defined spiral arms; they are classed as ‘anemic’ spirals by van den Bergh (1976). These anemic cluster spirals are intermediate in appearance and in color between field spirals and S0s, which suggests that they have less active star formation than the field spirals. Cluster spirals may also be smaller than field spirals (Peterson *et al.*, 1979). Finally, Gallagher (1978) and Kotanyi *et al.* (1983) present possible examples of spiral galaxies currently undergoing stripping.

Of course, the most direct test of the hypothesis that spiral galaxies evolve to become elliptical galaxies is the observation of spiral galaxies in high redshift clusters which may be undergoing this transformation. Unfortunately, the image sizes of galaxies in high redshift clusters (with ground based telescopes) are so small that the galaxies cannot be directly classified. However, Butcher and Oemler (1978a, 1984a,b) showed that a number of moderately high redshift ($z \approx 0.4$) clusters apparently contained a high proportion of blue galaxies. The blue galaxies lie at larger projected distances from the cluster center than the redder galaxies. When redshift effects were removed, these blue galaxies had colors indistinguishable from those of nearby spiral galaxies. No such population of blue galaxies occurs in nearby compact clusters (Butcher and Oemler, 1978b). These blue galaxies in high redshift clusters probably contain substantial quantities of gas and may be undergoing star-formation; they may indeed be spiral galaxies. If so, we would have fairly direct evidence that galactic populations evolve in rich clusters.

There are, however, a number of problems associated with this interpretation of the ‘Butcher–Oemler effect’. First, Mathieu and Spinrad (1981) and van den Bergh (1983b) have argued, based on the photometry and positions of the blue galaxies in the Butcher–Oemler cluster about 3C295, that the cluster is actually relatively poor, and that the blue galaxies are primarily background and foreground field galaxies; they claim that the actual cluster members have a more normal color distribution. Redshifts and spectra now exist for a reasonably large sample of galaxies in the Butcher–Oemler clusters (Dressler and Gunn, 1982; Dressler, 1984; Dressler *et al.*, 1985); they show that many of the blue galaxies are cluster members, although many are foreground galaxies. The blue galaxies that are cluster members do have colors which might suggest that they are spirals. However, many of them have spectra and surface brightnesses that are very different from those of present day normal spiral galaxies; they appear to consist of Seyfert-like active galaxies (Henry *et al.*, 1983; Dressler *et al.*, 1985) and galaxies that have recently undergone or are undergoing very large bursts of star formation (Butcher and Oemler, 1984b). There is some evidence from deep optical images that the blue galaxies are indeed disk galaxies (Thompson, 1986). The present rather uncertain situation concerning the Butcher–Oemler effect has been reviewed recently by Dressler (1984). Second, the effect may not be universal; high redshift clusters are known which appear not to contain such a blue population (see, for example, Koo, 1981). Third, X-ray observations show that the Butcher–Oemler clusters contain significant quantities of intracluster gas (Section 4.8). The ram pressure ablation time-scales are thus rather short ($\lesssim 10^9$ years). It seems unlikely that the clusters formed such a short time before they were observed; how then have the spirals survived? It is possible, given the fact that the blue galaxies lie in the outer parts of the clusters, that they have not passed through the cluster core yet. Another possibility, suggested by Gisler (1979), is that the rate of formation of interstellar gas in galaxies was higher at these redshifts. This could make it considerably more difficult to strip the galaxies through ram pressure ablation. (However, also see

Livio *et al.* (1980).) Finally, the redshifts of the clusters observed by Butcher and Oemler are not that large; why is it that similar clusters are not found at the present time?

Norman and Silk (1979) and Sarazin (1979) have suggested that nearly all the gas in a cluster was initially in the form of gaseous haloes around galaxies. These haloes would be stripped rather slowly by collisions until sufficient gas built up in the cluster core for ram pressure ablation to become effective. In this way, the stripping of spiral galaxies could be delayed for a significant fraction of a Hubble time, in agreement with the Butcher–Oemler effect. Larson *et al.* (1980) argued that the gaseous disks of galaxies are supplied by the continual infall of gas from the gaseous haloes proposed by Norman and Silk (1979) and Sarazin (1979). In this scenario, a galaxy will be transformed from a spiral into an S0 following the removal of its gaseous halo; star formation will then exhaust the interstellar medium in the disk in a few billion years. The Butcher–Oemler clusters might be in the midst of this transformation.

The stripping mechanisms described above (ram pressure ablation, evaporation, or the Spitzer–Baade effect) remove the gas from galaxies, but probably would not seriously affect the distribution of the stars because the mass fraction of gas in typical spirals is not terribly large (Farouki and Shapiro, 1980). Unfortunately, simply removing the gas from a spiral would leave behind a thin stellar disk; thus a stripped spiral galaxy might resemble an S0 galaxy, but never an elliptical galaxy. As the fraction of elliptical galaxies is higher in the centers of compact clusters than in the field (Table 1), the difference in galactic population between these two environments cannot result solely from the transformation of spirals into S0s. Moreover, S0s appear to have thicker disks than spirals (Burstein, 1979b). If galactic populations have evolved since formation, the stellar distribution in galaxies must have been modified during this evolution.

Of course, the process of gas removal could directly affect the stellar distribution if, during the removal of the gas, a significant portion were converted into stars which remained bound to the galaxy. This might produce the ‘thick disk’ component seen in S0s by Burstein (1979b).

Tidal gravitational effects during galaxy collisions can alter the stellar and mass distributions in galaxies. The possibility that massive haloes around galaxies might be removed by tidal collisions in clusters has already been discussed in Sections 2.8 and 2.10.1. Such tidal encounters might also puff up the disks of spiral galaxies and transform them into S0 or elliptical galaxies (Richstone, 1976). Unfortunately, detailed numerical simulations suggest that tidal interactions are not capable of transforming disk galaxies into ellipticals unless they do so before the cluster collapses (Da Costa and Knobloch, 1979; Farouki and Shapiro, 1981).

Another possibility is that elliptical galaxies are formed by the merger of spiral galaxies in clusters (Toomre and Toomre, 1972; White, 1979). There are a number of serious problems with this hypothesis, which have been summarized by Ostriker (1980) and Tremaine (1981). Most of these objections disappear if the mergers occur in subclusters before the formation of the cluster (see White, 1982).

Evidence given in support of the “heredity hypothesis” (the theory that galaxy morphology is determined at the time of galaxy formation) is primarily in the form of evidence against the “environment hypothesis”. Specifically, most of these arguments attack the extreme suggestion that the differences in galaxy morphology are determined solely by some environmental mechanism which removes the gaseous content of galaxies.

First, the total fraction of disk galaxies ($\text{Sp+S0}/(\text{Sp+S0+E})$) is not fixed, but varies from regular clusters to the field (Faber and Gallagher, 1976; Table 1). If the only change in galaxy morphology were the conversion of spirals into S0s, this ratio would be a constant. Second, some S0 and E galaxies are sometimes found in low density regions (the field), where ram pressure ablation and other environmental influences should be very weak (Sandage and Visvanathan, 1978; Dressler, 1980b), and these field S0s and Es have the same color distribution as the cluster S0s and Es. Third, Dressler (1980b) has found that distribution of galaxy morphologies correlates most strongly with the local galaxy density, and not as strongly with the global environment (which probably determines the density of intracluster gas). Fourth, the properties of S0 galaxies (colors, bulge-to-disk ratios, gaseous content, etc.) are generally intermediate between Sp and E galaxies (Sandage *et al.*, 1970; Faber and Gallagher, 1976). This would not necessarily be true if S0s were stripped Sp, but would be explained under the “heredity hypothesis” if the nature of galaxy formation were characterized by a single parameter, which is often taken to be the density of the region in which galaxy formation occurs (see below). Fifth, S0 galaxies have larger ratios of bulge-to-disk luminosity than spirals (Burstein, 1979a), and also absolutely larger and more luminous bulges (Dressler, 1980b). This would not be expected if S0s were simply spirals with their gas removed. Finally, S0 galaxies appear to have a thick, boxy

component to the disk, which is not present in spirals (Burstein, 1979b). Of course, one could argue that these thick disks actually arise during the process of stripping the gaseous disks from spirals; for example, the stripping process might induce star formation in the gas while it is being stripped, and produce a thick disk of stars supported by large velocity components perpendicular to the disk.

If galaxy morphology is determined by the conditions at the time of galaxy formation, what is the mechanism and to what conditions is it responsive? Most theories of galaxy formation assume that galaxies form by the collapse of initially gaseous matter (Eggen *et al.*, 1962; Gott, 1977). The stellar bulge components of galaxies all have distributions similar to the de Vaucouleurs profile (equation 2.14), which suggests that they have relaxed violently (Section 2.9) during the collapse. As discussed in Section 2.9, this implies that the collapse takes place on nearly a free-fall time (equation 2.32). If, prior to this collapse, most of the gas in the galaxy were converted to stars, these stars would act as a collisionless, dissipationless fluid, and only violent relaxation would occur. This would produce an ellipsoidal distribution of stars—that is, an elliptical galaxy. On the other hand, if star formation were ineffective, and most of the collapsing material remained gaseous, it would dissipate through radiation much of its energy, while maintaining its net angular momentum, and collapse to form a disk. With this hypothesis one can understand why the galaxies that contain significant quantities of gas are the disk-dominated spirals, and why the galaxies that lack gas are ellipticals and S0s.

Moreover, if galaxies initially have little gas compared to their stellar content, it is easier for them to remain free of gas. First, the stripping processes discussed above (ram pressure ablation, evaporation, etc.) are more effective if the density of interstellar gas is low. Second, even in the absence of such external mechanisms for removing gas, a galaxy with a high ratio of the density of stars to gas can clean itself of interstellar medium through the formation of a galactic wind (Mathews and Baker, 1971). The interstellar gas in a galaxy is heated by the stars, through supernovae, stellar winds, ionizing radiation, and the motion of mass-losing stars at high velocity through the ambient gas. If the gas density is sufficiently high compared to the stellar density, the gas will be able to cool efficiently and the energy input from stars will be radiated away. Conversely, if the gas density is low compared to the stellar density, the gas will heat up until thermal velocities in the gas exceed the escape velocity from the galaxy, and the gas leaves the galaxy in a transonic wind. Thus, if a galaxy starts with a small proportion of gas to stars, it can keep itself free of gas. Since the standard hypothesis is that E and S0 galaxies start with little gas, one can understand how they have managed to stay relatively gas free.

If it is the efficiency of star formation during the collapse of a protogalaxy which determines its morphology, what determines the efficiency of star formation? Why does it depend on the location of the protogalaxy? Two attempts to answer these questions have received particular attention. First, Sandage, Freeman, and Stokes (1970) argued that the efficiency of star formation in a protogalaxy is determined by the specific angular momentum content of the gas. If the angular momentum content of the gas is high, the collapse of protostars will be halted or delayed by centrifugal forces. Thus one would expect protogalaxies with a high angular momentum content to form spirals, and those with low angular momentum to form ellipticals. This hypothesis is in agreement with the observation that the specific angular momentum of disks significantly exceeds that of ellipsoidal components of galaxies.

Alternatively, Gott and Thuan (1976) have argued that the efficiency of star formation during the collapse of a galaxy is set by the density in the protogalaxy. Star formation requires that gas cool. Cooling processes generally involve two-body collisions. Therefore it is reasonable to assume that the cooling rate increases with density (see equation 5.23 for example). In protogalaxies with a sufficiently high initial density, the gas will be largely converted to stars during the collapse. If the initial density is sufficiently low, star formation is not effective and the gas collapses to a disk.

In Section 2.9 evidence was presented which indicated that the sequence of cluster morphology, from the field to irregular clusters to regular clusters, was a dynamical sequence resulting from increasing initial density. Thus regular, compact clusters formed from regions of high density, and irregular clusters from lower density regions. If density is also the factor that determines galaxy morphology, the relationship of galaxy morphology and cluster morphology can be understood.

In summary, it remains controversial whether galaxy morphology is determined primarily by conditions at the time of galaxy formation, or whether galaxy morphology evolves in response to the environment after formation. It seems unlikely that *all* galaxies have identical forms at birth, and that *all* the variation in galaxy morphology is due to environment. It seems reasonable that environment has played *some* role in

determining galaxy morphology; surely, somewhere at least one spiral galaxy has blundered into a core of a rich, compact cluster and been stripped. Thus I believe both mechanisms have probably significantly affected galaxy morphology. Note that two other possibilities further obscure the distinction between these two hypotheses. First, galaxies may have formed before clusters; then, the galaxies might have evolved in an environment different from that observed today (i.e., Roos and Norman, 1979). Second, the formation of disks in galaxies might be a slow and ongoing process (Larson *et al.*, 1980); then, there is no real distinction between the heredity and environment hypotheses, at least to the origin of disks.

Ultimately, the Hubble Space Telescope (Hall, 1982) will permit structural studies of galaxies in and out of clusters at large redshifts. These studies will show whether morphological evolution has occurred in galaxies, at least over the last half of the age of the universe.

2.11 Extensions of clustering

Rich clusters of galaxies represent only a portion of a spectrum of clustering (Peebles, 1974), which ranges from individual galaxies and binary galaxies to enormous regions of enhanced density ('superclusters') or reduced density ('voids').

2.11.1 Poor clusters

Lists of poor clusters and groups have been given by Sandage and Tamman (1975), de Vaucouleurs (1975), Turner and Gott (1976b), Hickson (1982), Beers *et al.* (1982), and Huchra and Geller (1982). Of particular interest are the poor clusters containing possible cD galaxies which have been catalogued by Morgan, Kayser, and White (1975, MKW) and Albert, White, and Morgan (1977, AWM). In Section 2.10.1 observations of the cDs in these poor clusters were used to constrain models for the formation of cD galaxies. Bahcall (1980) has studied the optical properties (richness, galaxy distribution, and galactic content) and finds that they represent a smooth continuation to lower richness of the properties of the Abell clusters. Recent optical studies of these clusters include Beers *et al.* (1984) and Malumuth and Kriss (1986).

2.11.2 Superclusters and voids

Clustering does not stop at the well-defined rich clusters, but extends to a much larger scale (Peebles, 1974). These superclusters appear in the distribution of galaxies, or in the distribution of clusters of galaxies. Recent reviews of superclustering include those of Rood (1981) and Oort (1983). Our own galaxy appears to lie within the Local Supercluster, a flattened system dominated by the Virgo cluster (de Vaucouleurs, 1953).

The recognition of distinct superclusters and of voids (nearly empty regions between superclusters) has largely become possible as larger samples of redshifts have become available for galaxies and clusters. With redshifts, one can study the three-dimensional distribution of galaxies, rather than just their angular distribution on the sky, and the confusing effects of projection can be reduced. Such studies have shown that Coma (A1656) and A1367 form part of a large ($\gtrsim 30$ Mpc) Coma supercluster (Rood *et al.*, 1972; Chincarini and Rood, 1976; Tifft and Gregory, 1976; Gregory and Thompson, 1978), and that the Perseus cluster is embedded in the Perseus supercluster (Gregory *et al.*, 1981). Other possible superclusters identified as groupings of clusters have been cataloged by Abell (1961), Rood (1976), Murray *et al.* (1978), and Thuan (1980).

Similar large scale clustering is observed directly in the distribution of galaxies in redshift surveys in small regions of the sky by Kirshner, Oemler, and Schechter (1978) and by the more extensive survey of Davis *et al.* (1982). Many of the superclusters appear to be highly elongated.

Large voids have also been found to lie between the superclusters. Such voids appear in front of the Coma, Perseus, and Hercules superclusters (Rood, 1981). An extremely large void (100–200 Mpc in size) in the galaxy distribution in the direction of Bootes has been discovered by Kirshner *et al.* (1981). A similarly large void in the distribution of Abell clusters was discovered recently by Bahcall and Soneira (1982).

3 Radio observations

3.1 General radio properties

The association between radio sources and clusters of galaxies was first made by Mills (1960) and van den Bergh (1961b). Radio emission from clusters has been reviewed recently by Robertson (1985). A brief review of the radio properties of clusters relevant to their X-ray emission will be given in this section. First, the general radio luminosities and spectra will be summarized. Second, possible correlations between the radio and X-ray emissions of clusters will be discussed. Third, two classes of radio source morphology (head-tail radio sources and cluster halo sources) which are unique to the cluster environment will be described. Fourth, possible observations of the effect of the intracluster medium on radio emission due to the cosmic blackbody background or background source will be reviewed. Finally, the use of the 21 cm hyperfine line to detect neutral hydrogen gas in clusters will be briefly discussed.

Figure 7 shows a contour plot of the radio emission in the Perseus cluster, superimposed on an optical photograph. The very strong radio emission from the central galaxy NGC1275 (Section 4.5.2) and the two head-tail radio galaxies NGC1265 and IC310 are shown.

The radio emission from clusters of galaxies (as well as most other extragalactic objects) is synchrotron emission due to the interaction of a nonthermal population of relativistic electrons (with a power-law energy distribution) with a magnetic field (Robertson, 1985). Such nonthermal synchrotron emission generally has a spectrum in which the intensity I_ν (erg/cm²-Hz-s) is well represented as a power-law over a wide range of frequencies ν ,

$$I_\nu \propto \nu^{-\alpha_r}, \quad (3.1)$$

where α_r is the radio spectral index. Typical extragalactic radio sources have $\alpha_r \approx 0.8$. Most of the radio emission from clusters is due to discrete sources, which can be associated with individual galaxies within the cluster. The properties of the nonthermal radio emission from radio galaxies have been reviewed recently by Miley (1980).

Radio emission from Abell clusters at a frequency of 1400 MHz has been surveyed by Owen (1975), Jaffe and Perola (1975), and Owen *et al.* (1982). At lower frequencies there are a number of older surveys (e.g., Fomalont and Rogstad, 1966), as well as an extensive list based on the 4C Cambridge survey (Slingo, 1974a,b; Riley, 1975; McHardy, 1978b). These observations suggested that only sources detected in the inner portions (within 1/3 Abell radii) of the Abell clusters were likely to belong to the cluster, the rest being background objects. Observations at higher frequencies have been made by Andernach *et al.* (1980, 1981), Haslam *et al.* (1978), and Waldhausen *et al.* (1979). The more distant Abell clusters were searched for radio emission by Fanti *et al.* (1983), while the richest Abell clusters were observed by Birkinshaw (1978). Jaffe (1982) surveyed a sample of high redshift clusters, and found evidence of evolution in the radio luminosity of clusters for the range of redshifts $0.25 < z < 0.95$.

In discussing the radio luminosity functions of clusters, it is important to distinguish the luminosity function of galaxies in clusters from the luminosity function of the cluster as a whole. The radio emission from clusters is mainly due to sources associated with individual radio galaxies. About 20% of the nearby strong radio galaxies are located in rich clusters of galaxies (McHardy, 1979). This appears to be mainly due to the fact that strong radio emission is primarily associated with giant elliptical galaxies, which occur preferentially in clusters. A galaxy of a given morphology (elliptical, for example) and optical luminosity apparently has the same radio luminosity function whether inside or outside of a cluster (Jaffe and Perola, 1976; Auriemma *et al.*, 1977; Guindon, 1979; McHardy, 1979). The radio luminosity function of the whole cluster can be fit as the result of superposing the luminosity function of an average of ≈ 5 radio galaxies per cluster (Owen, 1975). The cluster radio luminosity function does not appear to depend strongly on richness for the Abell clusters (Riley, 1975; Owen, 1975; McHardy, 1979).

There does appear to be a correlation between cluster radio emission and cluster morphology. Owen found that the more evolved RS types (cD,B,C, and L; see Section 2.5) have stronger radio emission. McHardy (1979) found that more evolved BM types (I,I-II; see Section 2.5) tend to contain stronger radio galaxies, at least at low frequencies.

Powerful radio sources are found most often near the cluster center (McHardy, 1979). They are usually associated with optically dominant galaxies, which often have multiple nuclei and are often cD galaxies (Guthrie, 1974; McHardy 1974, 1979; Riley, 1975).

Cluster radio sources generally have steeper radio spectra (values of $\alpha_r \gtrsim 1$) than radio sources in the field (Costain *et al.*, 1972; Baldwin and Scott, 1973; Slingo, 1974a,b; McHardy, 1979). The steepness of the spectrum (the value of α_r) increases with cluster richness and decreases as the BM type increases (Roland *et al.*, 1976; McHardy, 1979). The steepest radio spectra ($\alpha_r \approx 1.3$) in clusters are associated with radio sources in optically dominant galaxies (Riley, 1975).

De Young (1972) claimed that cluster radio galaxies were generally smaller than those in the field; this claim was not supported by larger surveys (Hooley, 1974; McHardy; 1979). Owen and Rudnick (1976a) found that cluster radio sources generally had extended emission; unresolved sources seen towards clusters are usually background objects. Guindon (1979) found that, while the average size of double or triple radio sources in clusters was the same as those in the field, clusters did lack the very largest sources.

Differences in the morphology of cluster radio sources and field sources are discussed in Sections 3.3 and 3.4 below.

Recently, the radio emission properties of poor clusters have been studied. In general, the sources in poor clusters also have steep spectra and show some of the same morphological distortions found in rich cluster sources (Burns and Owen, 1977). As in rich clusters, the strongest radio emission is associated with optically dominant galaxies near the cluster center, and these sources have especially steep spectra (White and Burns, 1980; Burns *et al.*, 1981b; Hanisch and White, 1981). There is no apparent difference between the radio emission of poor and rich cD clusters beyond the direct effect of richness (Burns *et al.*, 1980).

3.2 Correlations between X-ray and radio emission

Based on the small sample of X-ray clusters known at that time, Owen (1974) argued that there was a strong correlation between the radio and X-ray luminosity of clusters. Rowan-Robinson and Fabian (1975) did not find this correlation. Bahcall (1974b, 1977a) found that the X-ray emission of a cluster was increased if a strong radio source was located near the center of the cluster. McHardy (1978a) found that strong radio sources were more likely to be located in luminous X-ray clusters than in other clusters.

As discussed above, clusters generally contain steep-spectrum radio sources. It appears that radio sources in X-ray clusters have even steeper spectra, and that α_r correlates with X-ray luminosity (Erickson *et al.*, 1978; McHardy, 1978a; Cane *et al.*, 1981; Dagkesamansky *et al.*, 1982). This suggests that there might be a strong correlation between low frequency radio flux and X-ray luminosity (Erickson *et al.*, 1978; Cane *et al.*, 1981). Such a correlation is not found in larger samples of X-ray clusters (Mitchell *et al.*, 1979; Ulmer *et al.*, 1981). X-ray selected cluster samples from the *Einstein* observatory do not show a strong X-ray–radio correlation (Feigelson *et al.*, 1982; Johnson, 1981).

A possible correlation between X-ray emission and radio emission in poor clusters has also been found (Burns *et al.*, 1981c).

There appears to be a strong correlation between radio emission by central dominant galaxies in clusters and the presence of cooling flows, as evidenced by soft X-ray line emission or central spikes in the X-ray surface brightness (Burns *et al.*, 1981a; Valentijn and Bijleveld, 1983; Jones and Forman, 1984; Section 5.7.2).

To summarize, at present there may be a weak correlation between X-ray and radio luminosities, and there appears to be stronger correlation between the radio spectral index α_r and the X-ray luminosity L_x . Because of the many interrelationships between cluster properties, it is difficult to decide whether these possible correlations are primary or reflect other correlations (see Section 4.6). For example, the relationship between L_x and α_r may be a result of the fact that L_x correlates with cluster optical morphology (Section 4.6), as does α_r . Moreover, it is difficult to establish a clear causal basis for these correlations.

Costain *et al.* (1972) and Owen (1974) argued that the correlation between X-ray and radio emission implied that the same population of relativistic electrons produced both radio emission and X-ray emission. The radio emission is synchrotron emission; in this model the X-ray emission would be inverse Compton scattering of cosmic radiation photons by the same relativistic electrons. This ‘inverse Compton’ (IC) theory is described in more detail in Section 5.1.1. The IC model does require that cluster radio sources have steep spectra, as observed. However, the evidence against the IC model is now overwhelmingly strong (Sections 4.3 and Sec 5.1). Thermal emission by diffuse gas provides the main X-ray emission from clusters.

Another direct connection between the radio emitting electrons and the X-ray emitting thermal gas would be established if the thermal gas were heated by the relativistic electrons and/or any associated ‘cosmic ray’ nuclei. Such heating would occur through Coulomb interactions (Lea and Holman, 1978) and

might be enhanced by plasma interactions (Scott *et al.*, 1980; see Section 5.3.5 for more details). The heating is strongest for the lower energy electrons which produce very low frequency radio emission, and thus the heating requires that the radio sources in clusters have steep spectra, as observed. However, it is not clear that any ongoing heating of the thermal gas either is needed or does occur in the majority of X-ray clusters. First of all, there exist a reasonable number of strong X-ray clusters that do not have strong steep-spectrum radio sources. Second, the thermal energy per unit mass in the hot gas is roughly the same as the kinetic energy per unit mass in the galaxies. Thus the gas could have been heated initially by thermalizing its kinetic energy when it either was ejected from galaxies or fell into the cluster. In typical clusters, the cooling time in most of the gas is longer than the probable age of the cluster (the Hubble time), and no further heating of the gas would necessarily be required (Section 5.3).

A connection between the X-ray and radio luminosity of clusters might be produced if the radio emission were powered by accretion of gas which was initially part of the hot intracluster medium. In Sections 4.3 and 5.7 evidence is presented indicating that intracluster gas is cooling and being accreted by central dominant galaxies in many X-ray clusters. As mentioned previously, there is a correlation between radio emission by central dominant galaxies in clusters and the presence of these cooling flows (Burns *et al.*, 1981a; Valentijn and Bijleveld, 1983; Jones and Forman, 1984). The further accretion of this cooling gas onto a central massive object in the galaxy might produce the radio emission.

It is likely that the most important connection between the X-ray emitting gas in clusters and the relativistic electrons that produce the radio emission is dynamical. The hot gas provides pressure forces that can control the dynamics of the plasma of relativistic electrons. The current evidence suggests that radio emission from galaxies occurs when streams or blobs of relativistic nonthermal plasma are ejected from the nucleus of the galaxy (Miley, 1980). If the pressure and density of any surrounding medium is sufficiently large, the bulk motion and expansion of the radio emitting plasma will be retarded. This could explain the absence of very large radio galaxy sources associated with clusters. Moreover, the intracluster gas may confine the radio emitting plasma and prevent its adiabatic expansion. Expansion and synchrotron emission provide two competing energy loss mechanisms for the relativistic plasma. If expansion occurs, it weakens the radio emission but generally will not affect its spectrum. If the expansion is retarded by the pressure of a confining medium, synchrotron losses become important. These losses are most important for the highest energy electrons, and thus they cause the spectrum of the radio source to steepen. This mechanism probably provides the most plausible explanation of the steep spectrum of cluster radio sources.

Additional evidence for the dynamical effect of the hot intracluster gas on radio galaxies comes from distortions in the structure of the radio source produced if the radio galaxy is moving relative to the intracluster medium, as discussed below.

3.3 Head-tail and other distorted radio structures

There are two general types of radio source structures that occur predominantly in clusters of galaxies. In this section, distortions in the structures of single radio galaxies are discussed. In Section 3.4 cluster sized radio halo sources are reviewed.

A large proportion of relatively isolated radio galaxies have a fairly simple and symmetrical radio structure; for a review of the properties of radio galaxies see Miley (1980). Many of these galaxies have a compact radio source associated with the nucleus of the galaxy. There is also extended radio emission, generally in the form of double radio ‘lobes’, which lie on either side of the galaxy. These lobes are often of comparable brightness and projected distance from the nucleus, and most importantly, lie on a line through the nucleus of the galaxy. Recently observations have detected ‘jets’ of radio emission originating at the nucleus and extending out to the radio lobes; in some cases only one jet on one side of the galaxy has been found. The conventional theoretical scenario for the origin and energetics of these radio galaxies contains the following elements (Miley, 1980). First, the ultimate energy source for the radio emission (and all other nonthermal galaxy emission) is thought to be a very compact object in the nucleus of the galaxy. Energy is carried from this nonthermal engine out to the radio lobes by twin ‘beams’ of plasma; this plasma probably contains a mixture of thermal gas and relativistic nonthermal particles, which cause the beams to be observable in some cases as radio jets. The beams may be more or less continuous, or may consist of blobs of plasma (‘plasmoids’). The beams probably move outwards until they encounter a sufficient quantity of intergalactic (or intragalactic) gas, at which point their bulk kinetic energy of motion is converted into thermal energy and into the disordered relativistic motion of particles of nonthermal plasma.

This nonthermal plasma produces the emission from the radio lobes.

Radio galaxies in clusters show more complex radio structures, which generally tend to lack the symmetrical, aligned double structure of standard radio galaxies. These range from double lobed radio sources in which the lobes are not aligned with the galaxy nucleus ('bent-doubles' or 'wide-angle-tails') to sources in which all the radio emission lies in a tail on one side of the galaxy, and the galaxy itself forms the head of the tail ('head-tail' or 'narrow-angle-tail' radio galaxies).

The first head-tail (HT) radio galaxies discovered were NGC1265 and IC310 in the Perseus cluster (Ryle and Windram, 1968; see Figure 7), followed by the discovery of head-tail radio galaxies in Coma (Willson, 1970) and the 3C129 cluster (MacDonald *et al.*, 1968). Figure 8 shows a radio map of NGC1265, which is the archetypical head-tail radio galaxy. Some lists of head-tail radio galaxies and other distorted radio sources are those of Rudnick and Owen (1976a,b), Simon (1978, 1979), and Valentijn (1979a), and other observations of head-tail radio galaxies in rich clusters are given in Hill and Longair (1971), Vallee and Wilson (1976), Miley and Harris (1977), Gisler and Miley (1979), Burns and Ulmer (1980), Hintzen and Scott (1980), Bridle and Vallee (1981), Gavazzi *et al.* (1981), Vallee *et al.* (1981), and Dickey and Salpeter (1984).

NGC1265 and IC310, the first two head-tail radio galaxies discovered, are both in the Perseus cluster and have tails that lie on the line from the galaxy to the powerful radio galaxy NGC1275 at the cluster center (Figure 7). Ryle and Windram (1968) suggested that the radio emission from these galaxies was activated by a wind of relativistic particles from NGC1275, which also determined the directions of the tails. However, subsequent head-tail radio galaxies have not been found to show any alignment with the direction to powerful radio galaxies or the cluster center. The accepted explanation of head-tail radio galaxies, originally due to Miley *et al.* (1972), is that they are conventional radio galaxies moving at a high velocity through a static intracluster gas. The radio emitting beams or plasmoids are decelerated by the ram pressure of the intracluster gas and form a wake behind the galaxy. The high velocity of the galaxy is a result of the gravitational potential of the cluster (that is, the velocity v is comparable to the cluster velocity dispersion). The ram pressure acting on the radio blobs or beams is then

$$P_r = \rho_g v^2, \quad (3.2)$$

where ρ_g is the intracluster gas density (Section 5.9).

Miley (1973) found that the spectral index α_r and the fractional polarization of the radio emission increased with distance away from the galaxy along the tail. Synchrotron emission energy losses will steepen a radio spectrum, so the spectral variations are consistent with injection of particles at the galaxy. The polarization indicates that the magnetic field is highly ordered and directed along the direction of the tail, probably by the sweeping of the radio emitting plasma behind the galaxy. There are some indications that head-tail radio galaxies are more rapidly moving than typical cluster galaxies (Guindon and Bridle, 1978), but the effect is not very large (Ulrich, 1978), and in any case, only one component of the velocity can be measured. Head-tail radio galaxies are never cD galaxies and are seldom among the most luminous galaxies in a cluster, either at radio or at optical wavelengths (Rudnick and Owen, 1976a,b; Simon, 1978; McHardy, 1979; Valentijn, 1979c). Since cD galaxies are nearly at rest in the cluster potential (Section 2.10.1), they would not be expected to form head-tail radio galaxies.

Of course, the intracluster gas that produces the head-tail radio galaxies also produces X-ray emission; in general, the densities of intracluster gas ρ_g derived from X-ray observations are consistent with those needed to give a ram pressure sufficient to produce the observed radio structures (see, for example, Simon, 1979).

The first detailed theoretical work on head-tail radio galaxies was done by Jaffe and Perola (1973). They suggested two models; in the first, blobs of plasma were ejected in opposite directions (taken to be perpendicular to the direction of motion of the galaxy). They gave several arguments against this model. First, the adiabatic expansion of the blobs would produce large losses in their energy; if these losses exceeded losses due to synchrotron emission, the spectrum would not vary as observed. Second, they argued that the magnetic field was too well ordered to be an initially disordered field. Thus they proposed a second model, in which the radio galaxy possessed an extensive magnetosphere, which was swept behind the galaxy. The magnetosphere provided an ordered magnetic field, and they argued that it could confine the adiabatic expansion of the radio emitting blobs. This second argument was shown to be incorrect by Cowie and McKee (1975) and Pacholczyk and Scott (1976). Cowie and McKee showed that the large adiabatic losses

found by Jaffe and Perola were due in part to the assumption of high Mach number flow, which is not correct for the temperatures of the intracluster gas derived from X-ray observations (Section 4.3). As a result, subsequent theoretical work has largely been devoted to models for the interaction of free plasmoids (the first Jaffe–Perola model) or twin beams with intracluster gas (Cowie and McKee, 1975; Pacholczyk and Scott, 1976; Begelman *et al.*, 1979; Christiansen *et al.*, 1981).

One important problem with this model is that in many sources the age of the tail (projected length divided by the estimated galaxy velocity) is much longer than the lifetimes of the emitting electrons against synchrotron losses (Wilson and Vallee, 1977). Pacholczyk and Scott (1976) and Christiansen *et al.* (1981) have argued that these particles are reaccelerated by turbulence in the tails.

Recent high resolution radio observations have detected well defined radio jets leading out from the nucleus of the head-tail radio galaxy to the start of the tails (Owen *et al.*, 1978, 1979; Burns and Owen, 1980). Figure 9 shows these jets in NGC1265. Begelman *et al.* (1979) presented a twin-jet theory for the formation of HTs. Because the observed jets are narrow and follow well defined curved paths, it is unlikely that they consist of independent plasmoids, because these plasmoids would most likely have different masses and surface areas and would be bent by different amounts by the ram pressure force. The transition from a collimated jet into the swept back tail is fairly abrupt. Jones and Owen (1979) argued that the jets in the inner parts of the galaxies are protected from the ram pressure of the intracluster gas. They suggested that the head-tail radio galaxies contain regions of gas that are bound to the galaxy and too dense to be stripped by ram pressure (Section 5.9). The jets propagate through and are confined by this gas until they reach the intracluster gas and are swept back. However, the jets are curved in this inner region (Figure 9), which is difficult to understand if they propagate through static intragalactic gas (although the curvature could be due to buoyancy force if the intragalactic gas is compressed by ram pressure from the intracluster gas). In any case, Begelman *et al.* (1979) are able to fit the observed shape of the jets assuming ram pressure from intracluster gas and no intragalactic gas.

In some cases the tails in head-tail radio galaxies are observed to be curved or bent. Jaffe and Perola (1973), Miley and Harris (1977), and Vallee *et al.* (1979, 1981) suggested that the curvature reflects nonlinear galaxy motion, due to the orbit of the head-tail radio galaxies in the cluster potential, or due to a binary orbit. Jaffe and Perola (1973) also suggested that there might be an intracluster gas wind with shear. Cowie and McKee (1975) suggested that the bending might be due to buoyancy forces; the bendings in a number of head-tail radio galaxies are consistent with this (Gisler and Miley, 1979).

Another common morphological type of cluster radio source is wide-angle-tails' or WATs. These resemble classical double radio sources in which the two radio lobes and/or jets are not oppositely aligned, but lie at an angle. Figure 10 shows the radio image of the WAT 1919+479, which is in a Zwicky cluster. Some lists of these sources include those of Owen and Rudnick (1976b), Simon (1978), Guindon and Bridle (1978), Valentijn and Perola (1978), Valentijn (1979b,c), and van Breugel (1980). WATs are generally associated with optically dominant cluster galaxies, often cD galaxies (Owen and Rudnick, 1976b; Simon, 1978; Valentijn, 1979c), and tend to be more luminous radio sources than head-tail radio galaxies. They tend to be slower moving galaxies than those containing head-tail radio galaxies (Guindon and Bridle, 1978). A particularly fascinating case is the radio source 3C75 in the cluster A400, which has two intertwined pairs of jets which merge into a WAT (Owen *et al.*, 1985).

A major theoretical question about WATs is whether they are the result of the same physical process (ram pressure by intracluster gas) as head-tail radio galaxies. Owen and Rudnick (1976b) suggested that ram pressure was the mechanism behind WATs; since the WATs occur in slow moving (cD) galaxies, the ram pressure is less, and the radio structure is bent at a smaller angle. Another possible reason that WATs might be less bent is that they are associated with stronger radio sources. The plasmoids or beams associated with these sources may have more momentum than those in weaker radio sources, and may be harder to deflect. Valentijn and Perola (1978) and Valentijn (1979b) suggested that WATs were similar to HTs, except that the ejection angle of the beams or plasmoids was not nearly perpendicular to the velocity of the galaxy. This would not, by itself, explain why WATs occur in galaxies that are more prominent optically and in radio emission. Valentijn (1979b) also invokes the smaller galaxy velocity and more rigid beams discussed above, and suggests that these brighter galaxies may have more intragalactic gas, which shields the beams from the intracluster gas.

Burns (1981) and Burns *et al.* (1982) suggest that the WATs are bent mainly by buoyancy forces and

not ram pressure forces. Of course, if the galaxy associated with the WAT were a cD galaxy located at the center of the cluster potential and if the intracluster gas density were spherically symmetric about this center, no bending due to buoyancy could occur. Burns *et al.* (1982) argue that the cD associated with a WAT may not be exactly at the cluster center, and that accretion by the slowly moving cD may produce an aspherical density distribution in the intracluster gas.

Sparke (1983) argues that WATs and other distorted radio sources associated with cDs are indications that the clusters are in the process of collapsing. She notes that many of the WATs are associated with irregular clusters with clumpy X-ray emission and low X-ray temperatures (the irregular X-ray clusters of Table 3 below).

The association between these distorted radio morphologies and clusters of galaxies containing intracluster gas has been used to detect previously unobserved clusters and cluster X-ray sources. Burns and Owen (1979) found a number of distorted radio sources associated with poor Zwicky clusters, and suggested they might be X-ray sources. This conjecture was confirmed by Holman and McKee (1981). Fomalont and Bridle (1978) discovered a number of WATs in groups of galaxies.

Hintzen and Scott (1978) suggested that quasars with distorted radio structure were likely to be members of clusters of galaxies, and that radio observations of quasars could be used to detect high redshift clusters of galaxies and cluster X-ray sources. In general, optical studies have not found that quasars are associated with rich clusters, and any method that selected quasars in clusters or any type of high redshift clusters would be very useful. This method has in fact been used to detect clusters around several quasars (Hintzen *et al.*, 1981; Harris *et al.*, 1983a) and to provide a list of other candidates (Hintzen *et al.*, 1983).

3.4 Cluster radio haloes

The second type of radio morphology distinctly associated with clusters of galaxies is the cluster radio halo. They are diffuse, extended radio sources whose sizes are generally considerably larger than the cluster galaxy core radius, and smaller than the overall cluster size (for example, an Abell radius). The best studied example is Coma C, the halo source in the Coma cluster, which was first shown to be diffuse by Willson (1970), and which was studied further by Jaffe *et al.* (1976), Jaffe (1977), Valentijn (1978), Hanisch *et al.* (1979), Hanisch (1980), and Hanisch and Erickson (1980). Other clusters in which halo sources have probably been detected include A401 (Harris *et al.*, 1980; Roland *et al.*, 1981; but see Hanisch and Erickson, 1980), A1367 (Gavazzi, 1978; Hanisch, 1980; but this halo is considerably weaker and smaller than the others), A2255 (Harris *et al.*, 1980), A2256 (Bridle and Fomalont, 1976; Bridle *et al.*, 1979; but this is a very messy radio cluster), and A2319 (Grindlay *et al.*, 1977; Harris and Miley, 1978; Andernach *et al.*, 1980). While Ryle and Windram (1968) reported a rather large radio halo in the Perseus cluster, it does not appear to be a real source (Gisler and Miley, 1979; Jaffe and Rudnick, 1979; Birkinshaw, 1980; Hanisch and Erickson; 1980), although a smaller halo source has been reported recently (Noordam and de Bruyn, 1982).

Radio haloes do not appear to be very common, as a large number of surveys of clusters have failed to find them (Jaffe and Rudnick, 1979; Cane *et al.*, 1981; Andernach *et al.*, 1981; Hanisch, 1982a).

The cluster radio halos have very steep power-law radio spectra, with $\alpha_r \approx 1.2$. The power-law spectrum and some indications of polarization suggest that the emission process is synchrotron emission by nonthermal relativistic electrons, as in radio galaxies. The halo in Coma has a diameter (FWHM) of $\approx 1 \text{ Mpc}/h_{50}$, which is typical. In Coma, the spectrum of the halo is relatively uniform spatially (Jaffe, 1977). Although the sample of known radio haloes is small, they appear to be associated with clusters of intermediate optical morphology (BM II; RS B, C, L) (Hanisch, 1982b). These clusters are relaxed, but do not have a dominant cD galaxy (an exception may be A2319). The haloes are generally associated with clusters having a regular nXD X-ray morphology (Vestrand, 1982); Vestrand notes that many of these clusters have particularly luminous and extended X-ray emission and may have unusually high X-ray temperatures (see also Forman and Jones, 1982). In making these comparisons, the unusually weak and small halo associated with A1367 has not been included.

There is currently no consensus as to the origin of these haloes. Jaffe (1977) discussed observational and theoretical constraints on the origin of the nonthermal electrons producing the emission in the Coma cluster, and proposed that the electrons originate at strong radio sources in the cluster and diffuse out to form the halo. The observed spectral index α_r is about 0.5 larger than the spectral indices of strong cluster radio sources; such an increase occurs if there is a steady-state between the input of relativistic nonthermal electrons and synchrotron losses. Moreover, the number of electrons produced in strong radio sources is

sufficient to explain the halo radio emission if the magnetic field in the cluster is $B_c \approx 1\mu\text{G}$, which is consistent with limits on the hard X-ray emission from clusters (Section 4.3.1). However, the halo radio emission is less strongly peaked at the cluster center than the distribution of galaxies, particularly of strong radio galaxies (Jaffe, 1977). Thus the nonthermal electrons must be transported out from the cluster core. In order that synchrotron losses should not affect the spectrum of the electrons and cause the halo radio spectrum to steepen dramatically with radius (which is not observed; Jaffe, 1977), the particles must be transported at a velocity which is $\gtrsim 2000$ km/s. Convective fluid motions of this order would be supersonic and would involve a very high rate of energy dissipation. Thus Jaffe argued that the relativistic electrons must diffuse out into the cluster.

As discussed extensively by Jaffe, a diffusion velocity $\gtrsim 2000$ km/s would greatly exceed the Alfvén velocity

$$v_A \equiv \left(\frac{B_c^2}{4\pi\rho_g} \right)^{1/2} \quad (3.3)$$

in the intracluster plasma. (Here, ρ_g is the density of intracluster gas.) For typical values of the gas density from X-ray observations and the required magnetic field discussed above, $v_A \lesssim 100$ km/s. Particles that diffuse through a plasma faster than the Alfvén velocity excite plasma waves, which rapidly slow down the diffusion of the particles, and thus Jaffe argued that the Alfvén velocity acts as an upper limit on the diffusion speed of the relativistic electrons in radio halos. This velocity is much too small to allow the particles to diffuse without losses. A possible solution to this Alfvén speed limit problem, suggested by Holman *et al.* (1979), is that the plasma waves generated by electrons diffusing at speeds greater than v_A may be damped by ions in the background thermal plasma. This would allow diffusion at speeds up to the speed of these background ions, essentially the sound speed in the intracluster gas.

Another solution to the Alfvén speed problem was suggested by Dennison (1980b). He noted that the flux of relativistic, nonthermal particles at the Earth (cosmic rays) is dominated by protons, and suggested that this might also be true in radio sources. The protons would diffuse away from cluster radio galaxies at the Alfvén speed, but suffer no significant synchrotron losses because of their rigidity. In the cluster they would collide with thermal protons and produce secondary electrons by a number of processes. These relativistic, nonthermal secondaries would then produce the observed radio haloes in this model.

Harris and Miley (1978) suggested that the radio haloes are remnants of previous head-tail radio galaxies, whose spectra have steepened due to synchrotron losses. One problem with this idea is that HTs are typically not very luminous, and the clusters with radio haloes therefore would be required to have had a large number of bright radio galaxies in the past. However, radio haloes are rare, so this may not be a serious objection.

Jaffe (1977) considered the possibility that the nonthermal electrons are accelerated to relativistic energies within the cluster by turbulence in the intracluster gas. Roland *et al.* (1981) suggested that the turbulence was generated by the wakes of galaxies moving through the intracluster medium. Based on the small available sample, they suggested that the luminosity of radio haloes increases with the cluster X-ray luminosity L_x (a measure of the amount of gas in the cluster) and the velocity dispersion of galaxies σ_r , with $L_{halo} \propto L_x \sigma_r^2$. There are several problems with this hypothesis; first, unless the acceleration of relativistic electrons is very efficient, the rate of dissipation of the turbulent energy is unacceptably large (Jaffe, 1977). Second, no galactic wake has been detected as a radio source. They ought to appear as tailed galaxies without heads (no radio source in the nucleus of the galaxy).

The cluster A401 has been observed to possess a radio halo. With A399, this cluster forms a possible merging double system (Ulmer and Crudace, 1981; Section 4.4). Harris *et al.* (1980) suggested that radio haloes form during the coalescence of subclusters, possibly by the acceleration of relativistic particles in shocks which form in the intracluster gas. However, there are a reasonable number of other double clusters that do not show radio haloes.

Observations of radio haloes are important to the understanding of X-ray cluster emission because the nonthermal radio-emitting electrons and X-ray emitting thermal plasma coexist and may interact. Initially, it had been suggested that the X-ray emission might be inverse Compton emission from the nonthermal electrons (Section 5.1.1). However, the frequency of occurrence of X-ray emission and rarity of radio haloes is one of the many arguments against this theory. On the other hand, the nonthermal electrons may heat the thermal plasma and contribute to the X-ray emission indirectly (Lea and Holman, 1978; Rephaeli, 1979; Scott *et al.*, 1980; see also Sections 3.2 and 5.3.5). Vestrand (1982) has pointed out that radio halo clusters

have extended X-ray emission and may have higher X-ray temperatures than nonhalo clusters; he attributes this difference to the heating of intracluster gas by nonthermal electrons.

3.5 Cosmic microwave diminution (Sunyaev–Zel'dovich effect)

X-ray observations indicate that clusters of galaxies contain significant amounts of diffuse, hot gas. In this section and the next, the effect of this gas on sources of radio emission lying behind the cluster will be discussed. The free electrons in this intracluster plasma will have an optical depth for scattering low frequency photons given by

$$\tau_T = \int \sigma_T n_e dl, \quad (3.4)$$

where $\sigma_T = (8\pi/3)[e^2/(m_e c^2)]^2$ is the Thompson electron scattering cross section, n_e is the electron density, and l is the path length along any line of sight through the gas. For a typical X-ray cluster, $n_e \approx 10^{-3} \text{ cm}^{-3}$ and $l \approx 1 \text{ Mpc}$, and thus $\tau_T \approx 10^{-(2-3)}$. A fraction τ_T of the photons from any radio source behind a cluster will be scattered as the radiation passes through the cluster.

One ‘source’ of radio emission which lies ‘behind’ everything is the cosmic radiation which is a relic of the ‘big bang’ formation of the universe (Sunyaev and Zel'dovich, 1980a). This radiation has a spectrum that is nearly a blackbody, with a temperature of $T_r \approx 2.7 \text{ K}$. Because this radiation is nearly isotropic, simply scattering the radiation would not have an observable effect. However, because the electrons in the intracluster gas are hotter than the cosmic radiation photons, they heat the cosmic radiation photons and change the spectrum of the cosmic radiation observed in the direction of a cluster of galaxies. This effect was first suggested by Zel'dovich and Sunyaev (1969) (Sunyaev and Zel'dovich, 1972). Reviews of the theory and current observational status of this Sunyaev–Zel'dovich effect have been given recently by Sunyaev and Zel'dovich (1980a, 1981).

During an average scattering, a photon with frequency ν has its frequency changed by an amount $\Delta\nu/\nu = 4kT_g/m_e c^2$, where T_g is the electron temperature of the intracluster gas. In calculating the effect this has on the radiation spectrum, it is conventional to measure the intensity in terms of a ‘brightness temperature’ T_r ; this is defined as the temperature of a blackbody having the same intensity. Then, the change in the brightness temperature of the cosmic radiation due to passage through the intracluster gas is given by

$$\frac{\Delta T_r}{T_r} = \frac{\Delta I_\nu}{I_\nu} \frac{d \ln T_r}{d \ln I_\nu} = \tau_T \frac{kT_g}{m_e c^2} [x \coth(x/2) - 4], \quad (3.5)$$

where I_ν is the radiation intensity and $x \equiv h\nu/kT_r$. This expression is actually derived in the diffusion limit and is valid only for sufficiently small $x \lesssim 10$ (Sunyaev, 1981). The change in brightness temperature or intensity is negative for low frequencies $x < 3.83$ and positive for higher frequencies. This change occurs at a wavelength $\lambda_o = 0.14 \text{ cm}(2.7 \text{ K}/T_r)$. It is somewhat paradoxical that heating the background radiation lowers its brightness temperature at low frequencies. This is because Compton scattering conserves the number of photons, and shifting lower frequency photons to higher energies lowers the intensity at low frequencies.

Nearly all of the measurements that have been made of this effect have been at relatively low frequencies; taking the limit $x \rightarrow 0$ in equation (3.5) gives

$$\frac{\Delta T_r}{T_r} = - \int \frac{2kT_g}{m_e c^2} d\tau_T. \quad (3.6)$$

This reduction in the cosmic radiation in the direction of clusters of galaxies in the microwave region is often referred to as the ‘microwave diminution’. Calculations of ΔT_r and its variation with projected position for a large set of models for the intracluster gas have been given by Sarazin and Bahcall (1977), and specific predictions of the size of the effect for the Coma, Perseus, and Virgo/M87 clusters are given in Bahcall and Sarazin (1977). Models for the Coma cluster have been given by Gould and Rephaeli (1978) (but note that their basic model for the intracluster gas is not physically consistent) and Stimpel and Binney (1979) (these models included the effect of cluster ellipticity).

Unfortunately, it has proved to be very difficult to make reliable measurements of this very small microwave diminution effect. Pariiskii (1973) claimed to have detected a microwave diminution from the Coma cluster. Gull and Northover (1976) claimed to have detected both Coma and A2218, and found very small diminutions towards four other clusters. The claimed detections of Coma were at a level much too

high to be consistent with models for the X-ray emission from this well-studied cluster (Bahcall and Sarazin, 1977; Gould and Rephaeli, 1978), and subsequent observations have not confirmed the microwave diminution in Coma. Rudnick (1978) gave upper limits for five clusters, including Coma; his limits were consistent with models for the X-ray emission, but inconsistent with the previously claimed detections of Coma. Lake and Partridge (1977) claimed three detections of very rich clusters, but later withdrew the claim, saying that the measurements were undermined by systematic problems with the telescope. In a later survey of 16 clusters (Lake and Partridge, 1980), they detected only A576 at a level of $\Delta T_r = -1.3 \pm 0.3$ mK. (Note that $1\text{ mK} \equiv 10^{-3}\text{ K}$). Birkinshaw *et al.* (1978, 1981b) surveyed 10 clusters, detecting microwave diminutions in A576 ($\Delta T_r = -1.12 \pm 0.17$ mK), A2218 ($\Delta T_r = -1.05 \pm 0.21$ mK), and possibly A665 and A2319. A2218 was *not* detected by Lake and Partridge (1980); in fact, their measurements have the opposite *sign*. Perrenod and Lada (1979) made measurement at higher frequencies ($\lambda = 9\text{ mm} \gg \lambda_o$) in order to reduce the effects of contamination by radio galaxies and beam smearing. They detected A2218 at the same level as Birkinshaw *et al.*, and also had a marginal detection of A665 at $\Delta T_r = -1.3 \pm 0.6$ mK (A665 is the richest cluster in the Abell catalog). Lasenby and Davies (1983) did not detect either A576 or A2218.

The apparent microwave diminutions from A576 and A2218 require very large masses of gas (comparable to the virial masses) at very high temperatures $T_g \gtrsim 3 \times 10^8$ K. X-ray observations of A576 are completely inconsistent with this much gas at these temperatures (Pravdo *et al.*, 1979; White and Silk, 1980), and thus the measured microwave reductions must be due to some other effect. While earlier low spatial resolution studies of A2218 suggested that it was too weak an X-ray source to produce the claimed microwave diminution (Ulmer *et al.*, 1981), a detailed high spatial resolution study of the X-ray emission from A2218 with the *Einstein* observatory (Boynton *et al.*, 1982) indicates that the required amounts of gas are present in this cluster at the required temperatures $T_g = 10 - 30$ keV.

From equation (3.6), the microwave diminution effect is independent of distance as long as the cluster can be resolved. In fact, Birkinshaw *et al.* (1981a) have measured a diminution of $\Delta T_r = -1.4 \pm 0.3$ mK from the distant ($z = 0.541$) cluster 0016+16. Optically, this is a rich cluster (Koo, 1981), although the field is somewhat confused by a foreground cluster at $z = 0.30$. There is a strong X-ray source towards 0016+16, which would imply a very high X-ray luminosity if it is associated with the more distant cluster (White *et al.*, 1981b). In some ways, microwave diminution observations of distant clusters are more straightforward than observations of nearby clusters, because the reference positions are further outside the cluster core.

Recently, a new set of observations of the microwave diminution were published by Birkinshaw *et al.* (1984) (see also Birkinshaw and Gull, 1984). These observations used the Owens Valley Radio Observatory and are apparently less subject to systematic effects than earlier observations. They confirm the detections of 0016+16 ($\Delta T_r = -1.40 \pm 0.17$ mK), A665 ($\Delta T_r = -0.69 \pm 0.10$ mK), and A2218 ($\Delta T_r = -0.70 \pm 0.10$ mK). Several of these detections have also been confirmed by Uson and Wilkinson (1985). Since there are now several confirming observations of the microwave diminution in these three clusters, it may be that the effect has finally been observed unambiguously. However, in view of the disagreements between different observers in the past, the withdrawal of previously claimed detections, and the inconsistency of some of the radio results with X-ray measurements of the amount of gas present, I do not feel completely confident that the current microwave diminution results are conclusive. It is clear that the major sources of errors in the measurements are not statistical but systematic. These include very low level systematic problems with the response of the radio telescopes used (Lake and Partridge, 1980).

One major source of problems is the possible presence of radio sources in the cluster. If these are concentrated at the cluster core, they will increase the radio brightness of the cluster and mask the microwave diminution. All of the observations are corrected for the presence of strong radio sources, and the observers generally avoid observing clusters, such as Perseus, which contain very strong radio sources. There is still the danger that a larger number of harder to detect, weaker radio sources will make a significant contribution to the cluster radio brightness. Birkinshaw (1978) surveyed six clusters, including A576 and A2218, for weak radio source emission, and concluded that it was unlikely to affect the microwave diminution measurements. Schallwich and Wielebinski (1978) detected a weak radio source in the direction of A2218, and corrected the microwave diminution measurement of Birkinshaw *et al.* (1978) for this cluster. Unfortunately, this correction would destroy the agreement of this measurement with the shorter wavelength measurement of Perrenod and Lada (1979), because the radio source and microwave diminution have different spectral variations. Tarter (1978) has suggested that if clusters contained a small amount of ionized gas at a cooler temperature than

the X-ray emitting gas, the free-free radio emission from this gas could mask the microwave diminution. All of these radio source problems would generally mask the microwave diminution and might explain why some clusters that are predicted to have very strong diminutions, such as A2319, are in fact observed to have *positive* ΔT_r .

What about A576, in which a strong microwave diminution was initially observed, although very little gas is observed in X-rays? The microwave diminution measurements are generally relative measurements in which one compares the cosmic microwave brightness in the direction of the cluster core with the brightness at one or more positions away from the cluster core. A negative ΔT_r at the cluster core cannot be distinguished from a positive ΔT_r in these reference positions, which are generally not far outside the cluster core. Thus the observation of a negative ΔT_r in A576 may indicate that there is excess radio emission in the outer parts of the cluster. Cavallo and Mandolesi (1982) have suggested that this radio emission is produced by the stripping of gas from spiral galaxies in the outer parts of the cluster.

The microwave and X-ray observations of a cluster can be used to derive a distance to the cluster which is independent of the redshift (Cavaliere *et al.*, 1977, 1979; Gunn, 1978; Silk and White, 1978). From equation (3.6), ΔT_r depends on the electron density n_e , the gas temperature T_g , and the size of the emitting region. The X-ray flux f_x from the cluster depends on all of these, but also decreases with the inverse square of the distance to the cluster. Thus the distance can be determined by comparing the X-ray flux and the microwave diminution:

$$D_A \propto \left(\frac{\Delta T_r}{T_r} \right)^2 f_x^{-1} [T_g(0)]^{-3/2} \theta_c (1+z)^{-4}, \quad (3.7)$$

where D_A is the angular diameter distance (Weinberg, 1972), z is the redshift, θ_c is the angular radius of the cluster core (Section 2.7), $T_g(0)$ is the gas temperature at the cluster center, and the coefficient of proportionality depends on the distribution of gas in the cluster and the X-ray detector response (Cavaliere *et al.*, 1979). In fact, any assumptions about the gas distribution can be avoided by mapping the variation of both the X-ray surface brightness I_x and the microwave diminution as a function of the angle away from the cluster center θ . Silk and White (1978) find

$$D_A = \frac{F[T_g(0)]}{\pi(1+z)^4} \frac{\left\{ \int_0^\infty [\Delta T_r(0) - \Delta T_r(\theta)] \frac{d\theta}{\theta^2} \right\}^2}{\int_0^\infty [I_x(0) - I_x(\theta)] \frac{d\theta}{\theta^2}} \quad (3.8)$$

where F is a known function of gas temperature, which for a hot solar abundance plasma contains only atomic constants. This determination of the distance to the cluster is independent of the distribution of the gas as long as it is spherically symmetric. Applying this method to nearby clusters and comparing the distances with the redshift could allow the determination of the Hubble constant H_o . Mapping high redshift clusters ($z \approx 1$) could give the cosmological deceleration parameter q_o ; together, these two parameters determine the structure, dynamics, and age of the universe (Weinberg, 1972), yet remain very poorly determined after a half century of observational cosmology research.

Unfortunately, the difficulty of obtaining reliable microwave diminution measurements has made it impossible to apply this method at the present time (Birkinshaw, 1979; Boynton *et al.*, 1982). In general, cluster microwave diminutions have not been mapped with sufficient accuracy to allow the distance to be determined from equation (3.8). Even the optical data on clusters are not accurate enough to allow an accurate distance determination. In addition, the cause of false detections, such as A576, must be determined so that they can be weeded out of cluster samples. For example, if the detection of A576 were taken seriously, it would imply a distance to this cluster at least an order of magnitude more than its redshift distance (White and Silk, 1980).

Gould and Rephaeli (1978) suggested that it might be easier to detect the Sunyaev-Zel'dovich effect unambiguously at high frequencies ($\lambda < \lambda_o$) at which ΔT_r is positive. Some observations have been attempted at $\lambda = 1 - 3$ mm (Meyer *et al.*, 1983), but no cluster diminutions were detected, and this wavelength range straddles λ_o . Observations at shorter wavelengths must be made from satellites.

Sunyaev and Zel'dovich (1981) point out that although their effect is often thought of as a small change in the cosmic microwave background, at $\lambda < \lambda_o$ the effect may also be considered as an enormous source of submillimeter luminosity for the cluster. Because the surface brightness of the submillimeter emission is proportional to $\tau_T \propto n_e l$, where $l \approx r$ is the path length through the cluster and r is the radius of the gas,

and the surface area is proportional to r^2 , the luminosity is proportional to $n_e r^3$ or the mass M_g of the gas in the cluster. The submillimeter luminosity at frequencies above the critical frequency ($\lambda < \lambda_o$) is given by

$$L^+ = 8.7 \times 10^{43} \left(\frac{T_g}{10^8 \text{ K}} \right) \left(\frac{T_r}{2.7 \text{ K}} \right)^4 \left(\frac{M_g}{10^{14} M_\odot} \right) (1+z)^4 \text{ ergs/s} \quad (3.9)$$

where the factor $[T_r(1+z)]^4$ gives the cosmic radiation density at the redshift of the cluster. Thus clusters could be among the strongest sources of submillimeter radiation in the universe. Other strong submillimeter sources, such as quasars, would have different spectra, be more compact, and probably be variable.

The variation in the cosmic microwave intensity and polarization toward a cluster can also be used to determine the velocity of the cluster relative to the average of all material in that region of the universe (Sunyaev and Zel'dovich, 1980b). This velocity, measured relative to the local comoving cosmological reference frame, is as near as one can come to an absolute measure of motion in a relativistically invariant universe. In a sense, the cosmic background radiation acts as an 'ether'. Just as the thermal motion of electrons in a cluster changes the wavelength of the cosmic background radiation during scattering, their bulk motion has a similar effect. As long as $\tau_T \ll 1$, the variation in the brightness temperature due to cluster motion is independent of frequency and is given by

$$\frac{\Delta T_r}{T_r} = -\frac{v_r}{c} \tau_T, \quad (3.10)$$

where v_r is the radial component of the velocity. The tangential component of velocity (the component in the plane of the sky) can be detected if the polarization of the microwave background in the direction of clusters can be measured to very high accuracy. At low frequencies ($x \ll 1$), the polarization (which is in the direction of motion) is

$$p = \frac{1}{10} \beta_t^2 \tau_T + \frac{1}{40} \beta_t \tau_T^2 \quad (3.11)$$

to lowest order in τ_T and $\beta_t \equiv v_t/c$, where v_t is the tangential velocity.

The Sunyaev–Zel'dovich effect can also be used to determine the spectrum and angular distribution of the cosmic background radiation itself (Sunyaev and Zel'dovich, 1972; Fabbri *et al.*, 1978; Rephaeli, 1980, 1981; Sunyaev, 1981; Zel'dovich and Sunyaev, 1981). In principle, the different causes of variations in ΔT_r towards clusters (thermal motions of electrons, bulk motions, and variations in the cosmic background radiation itself) can be separated because they have different spectral variations (Sunyaev and Zel'dovich, 1981).

To summarize, the Sunyaev–Zel'dovich effect has a tremendous potential for providing information about the properties of hot gas in clusters and the nature of the universe as a whole. Unambiguous detections of this effect in clusters have proved elusive, but this situation may be improving.

3.6 Faraday rotation

A plasma containing a magnetic field is birefringent; the speed of propagation of an electromagnetic wave depends on its circular polarization (Spitzer, 1978). While natural sources of circularly polarized radiation are rare, synchrotron emission in an ordered magnetic field produces linearly polarized radiation, and many radio galaxies and quasars produce radio emission that is somewhat linearly polarized. The plane of polarization of linearly polarized radiation is rotated during passage through a birefringent medium. For a magnetized plasma, the angle of rotation ϕ is

$$\phi = R_m \lambda_m^2, \quad (3.12)$$

where it is conventional to give the wavelength λ_m of the radiation in meters. The rotation measure R_m is given by

$$R_m = \frac{e^3}{2\pi m_e^2 c^4} \int n_e B_{\parallel} dl = 8.12 \times 10^5 \int n_e B_{\parallel} dl_{pc} \text{ cm}^3 \text{ m}^{-2} \text{ G}^{-1} \text{ pc}^{-1} \quad (3.13)$$

where l is the path length through the medium, B_{\parallel} is the component of the magnetic field parallel to the direction of propagation of the radiation, and n_e is the electron density. On the far right of equation (3.13), l is given in pc, n_e in cm^{-3} , and B_{\parallel} in gauss (G). The wavelength dependence of the rotation is the feature that allows the observational separation of the initial polarization angle and the amount of rotation.

The rotation measures to radio sources lying within or behind clusters can be used to constrain the magnitude of the intracluster magnetic field and its geometry (Dennison, 1980a; Jaffe, 1980; Lawler and Dennison, 1982). One problem with these determinations is that the rotation measure due to other plasma along the line of sight to the radio source must be removed. If this can be done, and the electron density and path length through the gas are determined from X-ray observations of the cluster, the average value of B_{\parallel} can be determined. Since this average of a single component of B must be less than the average magnitude of B , this gives a lower limit to the intracluster magnetic field. The Faraday rotation of the halo radio source in M87 in the Virgo cluster has been used to give a lower limit $B \gtrsim 2\mu\text{G}$ ($\mu\text{G} \equiv 10^{-6}$ gauss) on the magnetic field in the gaseous halo around M87 (Andernach *et al.*, 1979; Dennison, 1980a). This limit implies that very little of the X-ray emission from M87 can be nonthermal.

Jaffe (1980) noted that the rotation measures to radio sources within or behind clusters are generally small $R_m \lesssim 100\text{m}^{-2}$, which for a cluster with $n_e \approx 3 \times 10^{-3}\text{ cm}^{-3}$ and a path length of 500 kpc gives $B_{\parallel} \lesssim 0.1\mu\text{G}$. On the other hand, the halo radio emission observed in some clusters (Section 3.4) implies that the intracluster magnetic field strengths are $B \gtrsim 1\mu\text{G}$. Since it is unlikely that the fields are preferentially ordered perpendicular to our line-of-sight to each cluster, the difference in these two limits must be due to the cancellation of components of B_{\parallel} along the path through the cluster. Jaffe argues that this implies that the field is tangled. If this tangled field can be thought of as consisting of cells of ordered field of size l_B randomly oriented along the path length l through the cluster, then statistically $\langle B_{\parallel} \rangle \approx B(l_B/l)^{1/2}$. Noting that only a portion of the rotation measure can be associated with the intracluster gas, Jaffe argued that the coherence length l_B of the intracluster magnetic field must be $l_B \lesssim 10$ kpc. Since tangled fields in a static medium can straighten and decay rapidly, he suggests that the fields are tangled by the turbulent wakes produced behind galaxies as they move through the intracluster gas.

With a larger sample of radio sources and clusters, Lawler and Dennison (1982) claimed that the sources seen through rich clusters had slightly larger rotation measures, although the two distributions only differ at the 80% confidence level. Attributing the difference to intracluster Faraday rotation, they derived an average rotation measure of $R_m \approx 130\text{m}^{-2}$ through the cluster center, which implies $l_B \gtrsim 20$ kpc; this is not inconsistent with the galactic wake model.

These limits on the strength and geometry of the intracluster B field are important to models for the X-ray emission for two reasons. First, the halo radio emission from a cluster depends on the product of the number of relativistic electrons and the magnetic field, and the relativistic electrons may heat the intracluster gas (Section 5.3.5). Second, the effectiveness of transport processes in the intracluster gas (such as thermal conduction) is determined by the geometry of the magnetic field because the gyroradius of electrons in even a very weak intracluster B field is very much less than the size of the cluster (Section 5.4.3).

3.7 21 cm line observations of clusters

The 21 cm hyperfine line in hydrogen allows one to measure the mass of neutral hydrogen in galaxies or clusters of galaxies. There are an enormous number of 21 cm observations of galaxies, which would require an entire review to discuss. The main results for clusters are negative, and I shall briefly summarize these. The H I observations of clusters and superclusters have been reviewed recently by Chincarini (1984).

The 21 cm line observations have shown that the galaxies that make up clusters are deficient in neutral hydrogen gas. First, many observations have shown that elliptical and S0 galaxies in clusters have very little neutral hydrogen gas (Section 2.10.2; Davies and Lewis, 1973; Krumm and Salpeter, 1976; Bieging and Bierman, 1977). Similarly, spiral galaxies in clusters generally have less neutral hydrogen than spirals found in the field (Huchtmeier *et al.*, 1976; Sullivan and Johnson, 1978; Helou *et al.*, 1979; Krumm and Salpeter 1979; Chamaraux *et al.*, 1980; Giovanelli *et al.*, 1981; Sullivan *et al.*, 1981), and the deficiency is stronger near the cluster center (Giovanelli *et al.*, 1982; Giovanardi *et al.*, 1983). As discussed in Section 2.10.2, these observations have been used to argue that stripping of gas from galaxies through ram pressure ablation or some other mechanism is an important process in clusters, and that S0 galaxies may be produced by this process.

Observations indicate that cD galaxies, including those accreting large amounts of intracluster gas (Sections 4.3.3 and 5.7.2), contain very little neutral hydrogen (less than $10^9 M_{\odot}$; Burns *et al.*, 1981a; Valentijn and Giovanelli, 1982). This indicates that the accreted gas is not all stored as neutral hydrogen gas.

Finally, 21 cm observations of clusters as a whole indicate that the missing mass component (Section

2.8) is not neutral hydrogen gas (Goldstein, 1966; Haynes *et al.*, 1978; Peterson, 1978; Baan *et al.*, 1978; Shostak *et al.*, 1980).

4 X-ray observations

4.1 Detections and identifications

The first extragalactic object to be detected as an X-ray source was M87 in the Virgo cluster (Byram *et al.*, 1966; Bradt *et al.*, 1967). Sources associated with the Perseus cluster (Fritz *et al.*, 1971; Gursky *et al.*, 1971a) and the Coma cluster (Meekins *et al.*, 1971; Gursky *et al.*, 1971b) were detected next. The idea that extragalactic X-ray sources were generally associated with groups or clusters of galaxies was suggested by Cavaliere *et al.* (1971). While the early detections were made with balloon- or rocket-borne detectors, a major advance in the study of X-ray clusters (and all of X-ray astronomy) came with the launch of the *Uhuru* X-ray satellite, which permitted more extended observations of individual sources and a complete survey of the sky in X-rays, the *Uhuru* Catalog (Giacconi *et al.*, 1972, 1974; Forman *et al.*, 1978a).

The early *Uhuru* observations established a number of properties of the X-ray sources associated with clusters. First, clusters of galaxies are the most common bright extragalactic X-ray sources. Second, clusters are extremely luminous in their X-ray emission, with luminosities $\approx 10^{43-45}$ ergs/s, and they have a wide range of luminosities. This makes clusters as a class the most luminous X-ray sources in the universe, with the exception of quasars. Third, the X-ray sources associated with clusters are extended (Kellogg *et al.*, 1972; Forman *et al.*, 1972); the sizes found from the *Uhuru* data range from about 200 to 3000 kpc. Fourth, the clusters have X-ray spectra that show no strong evidence for low energy photoabsorption, unlike the spectra of the compact sources associated with discrete sources either in the nuclei of galaxies or stellar sources within our own galaxy. Fifth, the X-ray emission from clusters is not time variable, as is the emission from many point sources of X-rays in our galaxy or in the nuclei of other galaxies (Elvis, 1976). These last three results suggest that the emission is truly diffuse, and not the result of one or many compact sources. Many of the early *Uhuru* results are presented in several review papers by Kellogg (1973, 1974, 1975).

The original identifications of clusters with *Uhuru* sources were made by Gursky *et al.* (1972) and Kellogg *et al.* (1971, 1973). Other identifications of clusters as *Uhuru* sources were made by Bahcall (1974c), Disney (1974), Rowan-Robinson and Fabian (1975), Elvis *et al.* (1975), Melnick and Quintana (1975), Vidal (1975a,b), Bahcall *et al.* (1976), Ives and Sanford (1976), Pye and Cooke (1976), Lugger (1978), and Johnston *et al.* (1981). The situation on the identifications of clusters was summarized by Bahcall and Bahcall (1975), who argued that the many unidentified *Uhuru* sources at high galactic latitude were probably also clusters of galaxies. A systematic search for Abell cluster identifications in the *Uhuru* catalog was made by Kellogg *et al.* (1973). Cluster X-ray sources were also identified in surveys made with the *Ariel 5* satellite (Elvis *et al.*, 1975; Cooke and Maccagni, 1976; Maccacaro *et al.*, 1977; Mitchell *et al.*, 1977; McHardy, 1978a; Ricketts, 1978; McHardy *et al.*, 1981) and SAS-C satellite (Markert *et al.*, 1976, 1979). A combined sample of X-ray identification from the *Uhuru* and *Ariel* surveys was given by Jones and Forman (1978), and the pre-HEAO cluster identifications have been reviewed by Gursky and Schwartz (1977).

The next major advance in sensitivity came with the launch of the HEAO-1 X-ray observatory with proportional counters with a very large collecting area. Identifications of clusters with hard X-ray sources detected in sky surveys with the HEAO-1 A-2 instrument have been given by Marshall *et al.* (1979) and by Piccinotti *et al.* (1982) for high galactic latitude X-ray sources. Searches for X-ray emission from the richest Abell clusters were made by Pravdo *et al.* (1979). A soft X-ray survey of a few clusters was made by Reichert *et al.* (1981), and a complete soft X-ray catalog is given by Nugent *et al.* (1983). A statistically complete sample of Abell clusters was surveyed with the HEAO-1 A-2 instrument by McKee *et al.* (1980). A survey of a large sample (≈ 1900) of Abell clusters with the HEAO-1 A-1 was made by Ulmer *et al.* (1981) and Johnson *et al.* (1983), while a survey of the most distant Abell clusters with the HEAO-1 A-1 detector (Ulmer *et al.*, 1980b) detected 11 such clusters, suggesting that many of them are extremely luminous. The southern cluster catalog of Duus and Newell (1977) was surveyed for X-ray emission using the HEAO-1 A-1 detector by Kowalski *et al.* (1984); this paper includes a compilation of all HEAO-1 A-1 cluster detections and limits. Wood *et al.* (1984) is the complete HEAO-1 A-1 X-ray source catalog.

X-ray astronomy made a quantum leap forward with the launch of the *Einstein* observatory. This was the first satellite with focusing optics for extrasolar X-ray observing. Because of its focusing capability, the sensitivity of this instrument to small sources was orders of magnitude higher than that for any previous X-ray detector. The two major *Einstein* surveys of X-ray emission from clusters are Abramopoulos and Ku (1983) and Jones and Forman (1984) (see Table 2 below). Other X-ray cluster detections with *Einstein* include those of Jones *et al.* (1979), Henry *et al.* (1979, 1982), Burns *et al.* (1981c), Forman *et al.* (1981),

Maccagni and Tarenghi (1981), Perrenod and Henry (1981), White *et al.* (1981a,b, 1987), Bechtold *et al.* (1983), Soltan and Henry (1983), and Henry and Lavery (1984).

Compact, poor clusters were detected as X-ray sources by Schwartz *et al.* (1980a,b) and Kriss *et al.* (1980, 1981, 1983), and will be discussed in more detail in Section 4.7.

Murray *et al.* (1978) suggested that a number of sources in the 4U *Uhuru* catalog were associated intrinsically with superclusters (that is, there was more emission than could be accounted for simply by the sum of the emissions from the number of X-ray clusters one would have expected to find in the supercluster). Kellogg (1978) presented evidence that X-ray clusters were located in superclusters, but did not suggest that there was any intrinsic emission associated with the supercluster itself. Subsequent observations have not confirmed the detections of superclusters as a distinct class of X-ray sources (Ricketts, 1978; Pravdo *et al.*, 1979). Forman *et al.* (1978b) claimed to have detected very large and luminous haloes of X-ray emission about clusters of galaxies, and argued that the extra emission seen by Murray *et al.* from superclusters might really be the sum of the haloes of the clusters in the superclusters. In general, subsequent observations have not confirmed the presence of these luminous and extensive haloes (Pravdo *et al.*, 1979; Nulsen *et al.*, 1979; Ulmer *et al.*, 1980a; Nulsen and Fabian, 1980).

4.2 X-ray luminosities and luminosity functions

The number of clusters per unit volume with X-ray luminosities in the range L_x to $L_x + dL_x$ is defined as $f(L_x)dL_x$, where $f(L_x)$ is the X-ray luminosity function. In general, the luminosity function will depend on the method used to select the clusters. One can begin with a statistically complete catalog of optically detected clusters (such as the ‘statistical sample’ of Abell clusters; see Section 2.1), which is surveyed for X-ray emission. Alternatively, a complete catalog of X-ray sources can be examined optically to determine which sources are associated with clusters of galaxies. In addition to reproducing the observed statistics of X-ray cluster identifications, the X-ray luminosity function is subject to the additional constraints that the total number of X-ray clusters not exceed the total number of all clusters, and the total emissivity of X-ray clusters not produce a larger X-ray background than is observed. Most data on cluster luminosities have been fit to either an exponential or a power-law form of the luminosity function. Thus we define

$$f(L_x) = A_e h_{50}^5 \exp(-L_x/L_{xo}) (10^{44} \text{ ergs/s})^{-1} \text{ Mpc}^{-3}, \quad (4.1)$$

$$f(L_x) = A_p h_{50}^5 \left(\frac{L_x}{10^{44} h_{50}^{-2} \text{ ergs/s}} \right)^{-p} (10^{44} \text{ ergs/s})^{-1} \text{ Mpc}^{-3}. \quad (4.2)$$

All luminosities in this section are given for the photon energy range of 2 – 10 keV. It is convenient to define $L_{44} \equiv L_x/10^{44}$ ergs/s.

Schwartz (1978) derived an estimate of the luminosity function for a sample of 14 Abell clusters in distance class 3 or less, which were detected with the *Uhuru*, *Ariel 5*, or SAS-C satellites. More distant clusters are used only to give an upper limit to the luminosity function at high luminosities. This sample is only expected to be complete for $1 \leq L_{44} \leq 10$. Schwartz found that the best fit exponential luminosity function has $A_e = 4.5 \times 10^{-7}$ and $L_{xo} = 2.0 \times 10^{44} h_{50} \text{ ergs/s}$, although a power-law with $A_p = 7.9 \times 10^{-7}$ and $p \approx 2.45$ would also fit the data if suitably truncated at high and low luminosities.

McHardy (1978a) derived an X-ray luminosity function from the *Ariel 5* fluxes for Abell clusters of distance class 3 or less; he argued that the *Uhuru* fluxes are unreliable for weak sources. While he did not fit his numerical luminosity function to any analytic expression, a suitable fit is given by a power law with $A_p = 2.5 \times 10^{-7}$ and $p \approx 2$ for $0.2 \leq L_{44} \leq 20$.

The HEAO-1 satellite provided a much more extensive data base for determining the luminosity function of clusters. Both the A-1 and A-2 experiments were used to survey the Abell clusters. A luminosity function was derived for a significant portion of the statistical sample of Abell clusters (Section 2.1), using the A-1 data by Ulmer *et al.* (1981). Exponential and power-law fits to these data gave $A_e = 0.49 \times 10^{-7}$, $L_{xo} = 2.9 \times 10^{44} h_{50}^{-2} \text{ ergs/s}$, and $A_p = 1.1 \times 10^{-8}$, $p = 1.7$, respectively. When the sample was extended to all Abell clusters and luminosities, the normalization A_e and the characteristic luminosity L_{xo} both were roughly doubled.

The HEAO-1 A-2 data were used to derive a luminosity function both by surveying the Abell clusters (McKee *et al.*, 1980; Hintzen *et al.*, 1980) and by identifying a complete sample of X-ray sources in a flux-limited survey at high galactic latitude (Piccinotti *et al.*, 1982). The Abell cluster survey included all richness

classes and all distance classes less than five. The luminosity function could be fit adequately with either an exponential or power-law form, and the coefficients in equations (4.1) and (4.2) were $A_e \approx 2.5 \times 10^{-7}$, $L_{x_0} \approx 1.8 \times 10^{44} h_{50}^{-2}$ ergs/s, $A_p \approx 3.8 \times 10^{-7}$, and $p \approx 2.2$. The cluster luminosity from the high latitude survey was not well represented by an exponential; a power law fit gave $A_p \approx 3.6 \times 10^{-7}$ and $p \approx 2.15$, which agrees well with the A-2 Abell survey result.

Bahcall (1979b) has attempted to predict the X-ray luminosity function of clusters from their optical luminosity function by assuming a one-to-one correspondence between the optical and X-ray luminosity of clusters. She predicts that clusters have a luminosity function that can be represented by two intersecting power laws with $p \approx 2.5$ for $L_{44} \gtrsim 1$ and $p \approx 1.3$ for $L_{44} \lesssim 1$. While the HEAO-1 data do not show any clear evidence for a change in the slope of the luminosity function at $L_{44} \approx 1$, they are probably not inconsistent with such a change because they do not extend much below this luminosity.

One important application of cluster luminosity functions is in determining the contribution of clusters to the hard X-ray background (see Field, 1980, for a review of its properties). From the estimates of the luminosity function of clusters discussed above, it appears that clusters probably provide only about 3 to 10% of the X-ray background in the 2 – 10 keV photon energy band, assuming they do not evolve rapidly with time (Rowan-Robinson and Fabian, 1975; Gursky and Schwartz, 1977; McKee *et al.*, 1980; Hintzen *et al.*, 1980; Piccinotti *et al.*, 1982; Ulmer *et al.*, 1981).

4.3 X-ray spectra

Observations of the X-ray spectra of clusters of galaxies have played a critical role in establishing the primary emission mechanism (thermal emission from diffuse hot intracluster gas) and in testing models for the origin of this gas. Models in which the emission comes from diffuse thermal gas predict (1) that the spectrum will be roughly exponential (the intensity I_ν (erg/cm²-s-Hz) varies as $\approx \exp(-h\nu/kT_g)$ where T_g is the gas temperature); (2) that the gas temperature will be such that the thermal velocity of protons in the gas $\approx \sqrt{kT_g/m_p}$ be comparable to the velocity of the galaxies in the cluster, as both are bound by the same gravitational potential; (3) that there will be no strong low energy photoabsorption; and (4) that emission lines will be present if the gas contains a significant contamination of heavy elements like iron. Alternatively, models in which the emission is due to relativistic nonthermal electrons predict a power-law spectrum $I_\nu \propto \nu^{-\alpha}$, which implies an excess at low and high energies when compared to an exponential spectrum; no line emission would be expected for a nonthermal emission process. As another possibility, the emission might be thermal emission from a number of compact sources, such as galactic nuclei or the binary stellar X-ray sources which dominate the X-ray sky within our own galaxy; however, such sources are generally optically thick at low X-ray energies ($\lesssim 1$ keV). The theories for each of these classes of emission processes and the basis for these predictions are discussed in Section 5.1.

The first three of the predictions given above concern the broad-band form of the spectrum (the continuum), while the last prediction concerns lines. Accordingly, the properties of the continuum spectra will first be reviewed, and then those of the line spectra. Reviews devoted primarily to the observations of the X-ray spectra of clusters have been given recently by Canizares (1981) and Mushotzky (1980, 1984, 1985), while Holt and McCray (1982) review all of X-ray spectroscopy.

4.3.1 Continuum features in the spectrum

If the X-ray emission from clusters is due to a diffuse plasma of either thermal or nonthermal electrons, the optical depth of the gas should be quite low. On the other hand, compact X-ray sources (such as galactic nuclei or binary stellar X-ray sources) often contain significant quantities of relatively cool neutral gas, which absorbs soft X-rays through photoionization. Because the fluorescent yield of the light elements is low, the absorbed X-rays are not reemitted and are lost from the spectrum. This low energy photoabsorption occurs in a series of edges which correspond to the absorption edges of cosmically abundant elements. The opacity of a solar abundance, low density, cold neutral gas has been calculated, for example, by Brown and Gould (1970). It is conventional to parametrize the absorption observed in an X-ray spectrum by the column density of hydrogen N_H in a gas with assumed solar abundances required to produce the observed absorption. Typically, compact sources have $N_H \gtrsim 10^{22} \text{ cm}^{-2}$. Even the earliest X-ray spectra of clusters suggested that they had rather weak low energy absorption (Catura *et al.*, 1972; Kellogg, 1973; Davidsen *et al.*, 1975; Kellogg *et al.*, 1975; Margon *et al.*, 1975; Avni, 1976), with column densities $N_H \lesssim 10^{22} \text{ cm}^{-2}$, which were generally consistent with the amount of neutral hydrogen in our own galaxy along the line of sight to the cluster. This indicated that the emission from clusters comes from a diffuse, ionized plasma.

Initially, there were two competing models for the nature of this ionized plasma (see Section 5.1). It could be a hot, thermal plasma with a temperature $T_g \approx 10^8$ K, or it could be a relativistic, nonthermal plasma with a power-law electron energy distribution, such as the plasma responsible for the radio emission observed in clusters (see Section 3.1). In the first case, the X-ray continuum would be primarily due to thermal bremsstrahlung (see Section 5.1.3), with a spectrum given by equation (5.11). If the frequency variation of the Gaunt factor $g_{ff}(\nu, kT_g)$ is ignored and the gas is all at a single temperature, the spectrum is exponential $I_\nu \propto \exp(-h\nu/kT_g)$. In the second case, the emission is primarily due to the inverse Compton process (the scattering of low energy photons to X-ray energies by the relativistic electrons; see Section 5.1.1), and the expected spectrum is a power-law $I_\nu \propto \nu^{-\alpha}$.

Unfortunately, proportional counters have rather poor spectral resolution, and it is therefore difficult to distinguish between thermal and nonthermal spectra. Moreover, any sufficiently smooth and monotonic spectrum can be produced by the combination of the thermal spectra with varying temperatures, or nonthermal spectra with varying spectral indices α ; thus the distinction between thermal and nonthermal spectra cannot be made unambiguously. It is not surprising, therefore, that the early proportional counter spectra of clusters could be fit consistently by either thermal (exponential) or nonthermal (power-law) spectra (for example, Kellogg *et al.*, 1975). However, spectra over a large energy range were better fit by the thermal model (Davidsen *et al.*, 1975; Scheepmaker *et al.*, 1976).

The first large surveys of cluster spectra came from observations with OSO-8 and *Ariel 5*. These satellites observed individual clusters for longer periods of time than had been possible with previous sky survey instruments, and had detectors that were optimized for spectral resolution. The spectra of clusters observed with OSO-8 and *Ariel 5* were significantly better fit by the thermal bremsstrahlung model than by the nonthermal model (Mushotzky *et al.*, 1978; Mitchell *et al.*, 1979). The required temperatures for the cluster gas were found to range from about 2×10^7 to 2×10^8 K from cluster to cluster, and some of the clusters required gas at several temperatures to fit the spectrum. Recently, a more extensive survey of X-ray cluster spectra was made with the A-2 experiment on the HEAO-1 satellite (Mushotzky, 1980, 1984, 1985; Henriksen, 1985; Henriksen and Mushotzky, 1985, 1986).

The two properties that can be derived most easily from the continuum X-ray spectrum are the gas temperature T_g and the emission integral

$$EI \equiv \int n_p n_e dV, \quad (4.3)$$

where n_p is the proton density, n_e is the electron density, and V is the volume of the gas in the cluster. The X-ray luminosity of a cluster is proportional to EI (equation 4.11). The X-ray luminosity (or EI) and gas temperature are found to be strongly correlated (Mitchell *et al.*, 1977, 1979; Mushotzky *et al.*, 1978). The HEAO-1 A-2 sample (Figure 11) gives $L_x \propto T_g^3$ (Mushotzky, 1984). The OSO-8, *Ariel 5*, and HEAO-1 A-2 spectral surveys established a number of correlations between these X-ray spectral parameters and the optical properties of X-ray clusters, which are discussed in Section 4.6 below.

There was also some evidence from the OSO-8 survey that the gas in clusters was isothermal; that is, the range of temperatures within the gas in a single cluster was relatively small. However, in many cases the OSO-8 and *Ariel 5* temperatures were not in agreement within the errors; if these differences are real, they suggest that there are multiple temperature components to the emission. If that were the case, the OSO-8 and *Ariel 5* detectors, which have different spectral and spatial sensitivities, might give different weights to the different components, and produce different average temperatures.

The HEAO-1 A-2 detector has provided much more data on the spectra of clusters in the photon energy range 2-60 keV (Mushotzky, 1980, 1984, 1985; Henriksen, 1985; Henriksen and Mushotzky, 1985, 1986). There is now evidence that most clusters contain a range of gas temperatures, with typical values between $T_g \approx 2 \times 10^7$ and 8×10^7 K. These multiple temperature components appeared to be most significant in clusters with low X-ray luminosities, although it is possible that similar low-luminosity cool components might remain undetected if hidden in the spectrum of clusters with high-luminosity high-temperature emission. The information on the spatial distribution of the X-ray emission in clusters (Section 4.4) suggests two locations for this cool gas. First, in low luminosity clusters, the X-ray emission is often inhomogeneous, with clumps of emission being associated, in some cases, with individual galaxies. These clumps may contain cooler gas. Second, in some clusters there are enhancements in the X-ray surface brightness at the position of the cD or

other centrally located dominant galaxy in the cluster. X-ray line observations suggest that these are regions at which the hot intracluster gas is cooling and being accreted by the central dominant galaxy (Section 5.7).

The *Einstein* X-ray observatory had two instruments capable of providing information on the continuum spectra of clusters. First, there was the Imaging Proportional Counter (IPC), which provided low resolution spatial and spectral information. Initially there were problems with the calibration of the energy scale of the spectra due to gain variations. These problems have now apparently been resolved, and a few cluster spectra from this instrument are available at the present time (Fabricant *et al.*, 1980; Perrenod and Henry, 1981; Fabricant and Gorenstein, 1983; White *et al.*, 1987). The second instrument was the Solid State Spectrometer (SSS), which had considerably better spectral resolution, but had no spatial resolution and less sensitivity than the IPC. Because of its small field of view (6 arc min), it could only observe a small portion of nearby clusters. Thus it was used primarily to determine spectra for the central regions of nearby clusters. It provided strong evidence for the presence of cool gas at the centers of a number of clusters (Mushotzky, 1980, 1984, 1985; Mushotzky *et al.*, 1981; Lea *et al.*, 1982); these observations are discussed further below. One problem with *Einstein* as an instrument for X-ray cluster spectroscopy is that the telescope was only sensitive to photons with energies of about 0.1 – 4.0 keV. With the typical temperatures of the gas in clusters being $kT_g \approx 8$ keV, observations with *Einstein* could not determine the thermal structure in this hot gas. However, the *Einstein* detectors were very sensitive to the presence of low temperature components of the emission.

Detections or limits on the hard X-ray spectrum and flux of clusters have been useful in limiting the contribution of nonthermal processes to their luminosity. As mentioned above, spectra extending into the hard X-ray region ($h\nu > 20$ keV) gave the first direct, strong indication that the primary emission mechanism was thermal, rather than nonthermal (Davidsen *et al.*, 1975; Scheepmaker *et al.*, 1976). Subsequently, stronger limits on the hard X-ray emission have shown that nonthermal emission makes at most a very small contribution to the X-ray luminosity of clusters (Mushotzky *et al.*, 1977; Lea *et al.*, 1981). When combined with observations of the diffuse radio emission in the cluster (Section 4.4), these hard X-ray limits can be used to give lower limits on the magnetic field in the cluster, because the synchrotron radio emissivity is proportional to the product of the density of relativistic electrons and the magnetic field strength, while the inverse Compton X-ray emission depends only on the density of relativistic particles (see Section 5.1.1 for a more detailed discussion of this point). Typically, these limits are $B \gtrsim 10^{-7}$ G (Lea *et al.*, 1981; Primini *et al.*, 1981; Bazzano *et al.*, 1984).

In the Perseus cluster, a power-law hard X-ray component with $\alpha \approx 2.25$ has been detected; it varies on a time scale of about a year, and the X-ray variations are correlated with variations in the radio flux of the compact radio source at the nucleus of the galaxy NGC1275 (Primini *et al.*, 1981; Rothschild *et al.*, 1981). Much weaker power law sources may also have been detected in the M87/Virgo, A2142, and 3C129 clusters (Lea *et al.*, 1981; Bazzano *et al.*, 1984).

4.3.2 Line features—the 7 keV iron line

In many ways, the most significant observational discovery concerning X-ray clusters (following their identification as X-ray sources) was the detection of line emission due to highly ionized iron as a strong feature in their X-ray spectra. Immediately, this discovery established that the primary emission mechanism in X-ray clusters was thermal, and that the hot intracluster gas contained at least a significant portion of processed gas, which had at some point been ejected from stars. This line feature was first detected in the spectrum of the Perseus cluster by Mitchell *et al.* (1976) and shortly thereafter in the spectra of the Coma, Perseus, and Virgo clusters by Serlemitsos *et al.* (1977). It has subsequently been detected in the spectra of a total of about thirty clusters (Mitchell and Culhane, 1977; Mushotzky *et al.*, 1978; Malina *et al.*, 1978; Berthelsdorf and Culhane, 1979; Mitchell *et al.*, 1979; Mushotzky, 1980, 1984, 1985; Henriksen, 1985; Henriksen and Mushotzky, 1985, 1986). Figure 12 gives the HEAO-1 A-2 spectra of the Coma and Perseus clusters, showing these line features.

The line feature that was detected is actually a blend of lines from iron ions (mainly Fe^{+24} and Fe^{+25}) and weaker lines from nickel ions (see Section 5.2.3). These lines are mainly at photon energies between 6.5 and 7.0 keV; for convenience, this blend will be referred to as the ‘7 keV Fe line’. The resolution of this component structure and the measurement of the relative intensities of the various components can provide a wealth of diagnostic information on the physical state and environment of the X-ray emitting gas (Sarazin and Bahcall, 1977; Bahcall and Sarazin, 1978). Unfortunately, while the *Einstein* observatory contained

a number of high resolution spectrometers, none were sensitive to the 7 keV Fe lines because the mirror in *Einstein* was ineffective for photon energies greater than about 4 keV. The application of the *Einstein* spectrometers to lower energy lines is discussed below. However, the proportional counters on the HEAO-1 A-2 experiment had sufficient spectral resolution to resolve the $K\beta$ line³ of Fe^{+24} from the $K\alpha$ line in the Centaurus and Perseus clusters (Mitchell and Mushotzky, 1980; Mushotzky, 1980; see Figure 12). The observation of this line proves that the emission is thermal in nature, and not the result of the fluorescence of cold gas through photoionization by an X-ray continuum, because the fluorescent yield for the $K\beta$ line is rather small. In fact, the observed $K\beta$ lines are so strong that they require that the X-ray emission arise from a non-isothermal gas, with both cool and hot temperature components.

The line strengths are often given as ‘equivalent widths’ or EW . The equivalent width of any line feature is defined as

$$EW \equiv \int \left(\frac{I_\nu - I_\nu^o}{I_\nu^o} \right) d(h\nu), \quad (4.4)$$

where I_ν is the observed intensity including the line as a function of frequency ν and I_ν^o is the continuum intensity without the line. The details of the emission processes for this feature are discussed in Section 5.2; here we note that the emissivity of the line is proportional to the square of the density and to the abundance of iron, and depends significantly on the electron temperature. Because the thermal bremsstrahlung emissivity also is proportional to the square of the density (Section 5.1.3), the equivalent width EW of the line is independent of density as long as the iron is well mixed in the gas (see Section 5.4.5 for a discussion of this point). If the shape of the X-ray continuum spectrum of a cluster is used to derive a temperature or a range of temperatures for the gas in the cluster, then the equivalent width of the 7 keV Fe line gives a measure of the abundance of iron in the gas (Section 5.2.3; Figure 35 below). The abundances by number of atoms determined from the observations of clusters are all roughly $\text{Fe}/\text{H} \approx 2 \times 10^{-5}$, which is about one-half of the solar value (see the detection references above, as well as Bahcall and Sarazin, 1977). The upper limits on iron abundances in clusters without line detections are also generally consistent with this value. Figure 13 gives the derived iron abundances for clusters from the HEAO-1 A-2 sample, plotted as a function of the cluster X-ray luminosity. The uniformity of the iron abundance suggests that the intracluster gas in clusters has a similar origin in all clusters, regardless of the dynamical state of the cluster.

The strong 7 keV Fe line emission observed from clusters is very difficult to reconcile with any model for the origin of the X-ray emission except the thermal intracluster gas model (Section 5.1.3; Mitchell *et al.*, 1976; Serlemitsos *et al.*, 1977; Bahcall and Sarazin, 1977). This line emission occurs naturally in the intracluster gas model if the abundances of heavy elements are roughly solar (Section 5.2.2). However, it is not at all expected in the IC model (Section 5.1.1) and unlikely in the individual stellar source model because such sources are generally optically thick (Section 5.1.2). Only a small portion of the X-ray luminosity of a typical cluster is emitted in these lines, so it is possible, in principle, that the lines might come from a different source than the majority of the X-ray emission. Several considerations show that this is extremely unlikely. First of all, the required abundances in the intracluster gas are nearly solar, and thus very high abundances would be needed if the lines were to come from gas that provided only a small fraction of the continuum X-ray emission. Second, the abundances derived for all the observed clusters are essentially the same within the errors (Section 4.3.2), although the clusters span a wide range in optical and X-ray properties. A very odd coincidence would be required to produce the appearance of constant abundances in such varied clusters, if the X-ray lines and continuum came from two distinct sources.

As discussed in Sections 5.1.3 and 5.10, the nearly solar iron abundance in the intracluster gas suggests that a significant portion of this gas has been processed in and ejected from stars.

The 7 keV Fe line can be used to determine the redshift of a cluster with moderate accuracy from even low resolution X-ray spectra (Boldt, 1976; Bahcall and Sarazin, 1978). Over a wide range of temperatures, the equivalent width of the 7 keV Fe line is at least an order of magnitude larger than that of any other feature in the spectrum of a hot plasma. Thus the detection of a strong feature at photon energies below

³This notation gives the principal quantum number n of the lower level of the transition and the change in the principal quantum number $\Delta n \equiv n' - n$, where n' is the principal quantum number of the upper level of the transition. K indicates that the lower level is in the K-shell ($n = 1$), L indicates the lower level is in the L-shell ($n = 2$), and so on, while α indicates that $\Delta n = 1$, β indicates that $\Delta n = 2$, etc.

7 keV, coupled with the failure to detect a feature at higher energies up to 7 keV, permits the immediate identification of the feature with the 7 keV Fe line and the determination of a single-line redshift. This application of the 7 keV line has not been useful thus far because the *Einstein* X-ray observatory could not detect hard X-rays and had too little sensitivity to measure spectra from high redshift clusters. However, it promises to be of great importance in the future (see Chapter 6).

4.3.3 Lower energy lines

In addition to the 7 keV Fe line complex, the X-ray spectrum of a solar abundance low density plasma contains a large number of lower energy lines (Sarazin and Bahcall, 1977; Figure 34 below). These include the K lines of the common elements lighter than iron, such as C, N, O, Ne, Mg, Si, S, Ar, and Ca, as well as the L lines of Fe and Ni. Although the Fe L line complex was tentatively identified in proportional counter spectra of the M87/Virgo cluster by Fabricant *et al.* (1978) and Lea *et al.* (1979), the capability for detecting these lines was greatly increased with the launch of the *Einstein* X-ray observatory with its moderate resolution Solid State Spectrometer (SSS) and its high resolution Focal Plane Crystal Spectrometer (FPCS); see Giacconi *et al.* (1979) for a description of the satellite and its capabilities.

The SSS has detected the K-lines from Mg, Si, and S and the L-lines from Fe in the spectra of M87/Virgo, Perseus, A496, and A576 (Mushotzky, 1980, 1984; Mushotzky *et al.*, 1981; Lea *et al.*, 1982; Nulsen *et al.*, 1982; Rothenflug *et al.*, 1984). Figure 14 shows the SSS spectrum from Virgo. In general, line emission from both the helium-like and hydrogenic ions of Si and S is seen, indicating that the emission occurs at relatively low temperatures $T_g \approx 2 \times 10^7$ K. The observations in M87/Virgo are consistent with nearly solar abundances of Si, S, and Mg, while the observation in A576 may require lower abundances. Observations of SSS spectra away from the center of M87/Virgo show that the heavy element abundances are roughly constant throughout the gas (Lea *et al.*, 1982).

In Perseus, the SSS observations of the Fe L lines imply that the iron abundance is about one-half of the solar value (Mushotzky *et al.*, 1981), which agrees with the abundance derived from the 7 keV Fe line in proportional counter spectra. However, the SSS observations were made with a very small (≈ 6 arc min) aperture centered on the cluster center, while the proportional counter observations determine the spectrum of the entire cluster. The approximate agreement of the two abundances suggests that the iron is well mixed throughout the cluster, and not just concentrated in the cluster core.

In general, the SSS spectral line observations seem to imply that many clusters contain cooler gas than was required to explain their continuum spectra. Because the SSS has a small field of view and was centered on the cluster center in these observations, this cool gas must be concentrated at the center of the cluster; in the case of the Virgo and Perseus clusters, the center coincides with the central dominant galaxies M87 and NGC1275. In Perseus, this cool gas has a cooling time (see Section 5.3.1) of less than 2×10^9 yr, which is considerably less than the probable age of the cluster. It therefore seems likely that the cool gas observed is part of a steady-state cooling flow (see Section 5.7 for a discussion of the theory of such flows). From the observed line intensities, Mushotzky *et al.* (1981) determined that about $300M_\odot$ per year of gas must currently be cooling onto NGC1275 in the Perseus cluster, and Nulsen *et al.* (1982) found that about $200M_\odot$ per year must be accreting onto the cD in A496. These rates assume that the gas is not being heated.

The FPCS has also provided strong evidence for the cooling and accretion of gas onto M87 in the Virgo cluster and NGC1275 in the Perseus cluster (Canizares *et al.*, 1979, 1982; Canizares, 1981). In M87, the FPCS has detected the O^{+7} K α line, as well as blends of the $Fe^{+(16-23)}$ L lines and the Ne^{+9} K α line. Figure 15 shows the FPCS detection of the O^{+7} K α line in M87. The ratio of the abundance of oxygen to iron is apparently 3 – 5 times higher than the solar ratio. The relative strengths of the various Fe L line blends cannot result from gas at any single temperature. Apparently, a range of temperatures is necessary, with the X-ray luminosity originating from gas in any range of temperature dT_g being roughly proportional to dT_g . This is just what is predicted if the cool gas results from the cooling and accretion of hotter gas onto the center of M87 (Cowie, 1981). Canizares *et al.* (1979, 1982) and Canizares (1981) show that the spectra are consistent with radiative accretion at a rate of $\approx 3 - 10M_\odot$ per year. Similar results were found for the Perseus cluster, except that the required accretion rate is very large $\approx 300M_\odot$ per year (Canizares, 1981), in agreement with the results from the SSS. The SSS spectra of about a half dozen other clusters also show evidence for such accretion flow, with rates between those of the M87/Virgo cluster and the Perseus cluster (Fabian *et al.*, 1981b; Mushotzky, 1984).

Thus the two primary observational results of the *Einstein* spectrometers are these: first, the intracluster

gas contains the heavy elements oxygen, magnesium, silicon, and sulfur, as well as iron; second, gas is cooling and being accreted onto central dominant galaxies in many clusters at rather high rates ($3 - 400 M_{\odot}/\text{yr}$). Such accretion had been predicted by Cowie and Binney (1977), Fabian and Nulsen (1977), and Mathews and Bregman (1978); models for these cooling flows are discussed in Section 5.7. The rates of cooling are so high that if they have persisted for the age of the cluster, the entire mass of the inner portions of the central dominant galaxies might be due to accretion. Models for central dominant galaxies based on this idea have been given by Fabian *et al.* (1982a) and Sarazin and O'Connell (1983), who argue that the majority of the accreted gas is converted into low mass stars.

X-ray line observations have established that the primary emission mechanism of X-ray clusters is thermal emission from hot, diffuse intracluster gas. They have also shown that at least part of that gas has been ejected from stars and presumably from galaxies. Apparently, some of this intracluster gas is now completing the cycle and returning to the central galaxies, and possibly being formed into stars!

4.4 The spatial distribution of X-ray emission

4.4.1 X-ray centers, sizes, and masses

Early *Uhuru* observations indicated that the X-ray sources were centered on the cluster center, as determined from the galaxy distribution, or on an active galaxy in the cluster (Bahcall, 1977a). In many of the X-ray clusters, the cluster center also corresponds with the position of a cD or other central dominant galaxy.

The earliest X-ray observations showed that the X-ray sources in clusters were extended. With the nonimaging proportional counters used in these observations, the spatial resolution of the detectors was usually determined by mechanical collimators in front of the detectors and was therefore relatively crude (generally an angular resolution of $\approx 1^\circ$). Since this is comparable to the sizes of the X-ray emission regions in the nearest clusters, the distribution of emission could not generally be observed in any detail. Only estimates of the size could be determined by convolving a model for the distribution of the emission with the resolution of the detector and comparing the result to the observations.

Lea *et al.* (1973), Kellogg and Murray (1974), and Abramopoulos and Ku (1983) derived sizes for cluster X-ray sources, assuming that the gas had the density distribution given by the King approximation to a self-gravitating isothermal sphere (equations 2.9, 2.13)

$$\rho_g \propto \left[1 + \left(\frac{r}{r_x} \right)^2 \right]^{-3/2}. \quad (4.5)$$

Here, ρ_g is the gas density and r_x is the X-ray core radius. While this model is not physically consistent (see Section 5.5.1), it does provide a convenient fitting form for comparison to the galaxy distribution, which is often fit by the same function. Values of r_x were derived from the X-ray distribution in nine clusters by Lea *et al.* (1973) and Kellogg and Murray (1974). More recently, equation (4.5) has been fit to the distribution of X-ray emission from 53 clusters detected in an extensive survey of clusters with the *Einstein* X-ray observatory (Abramopoulos and Ku, 1983; Table 2). In general, these studies found that the X-ray core radii were significantly larger than the core radii of the galaxies (Section 2.7). These results were fairly uncertain because of the difficulties in determining either of these radii accurately.

More physically consistent hydrostatic models for the intracluster gas have also been used to fit the observed X-ray distributions (Lea, 1975; Gull and Northover, 1975; Cavaliere and Fusco-Femiano, 1976; Bahcall and Sarazin, 1977; Cavaliere, 1980). One model that has been used extensively is the hydrostatic isothermal model for the intracluster gas (Cavaliere and Fusco-Femiano, 1976, 1981; Bahcall and Sarazin 1977, 1978; Sarazin and Bahcall, 1977; Gorenstein *et al.*, 1978; Jones and Forman, 1984; Section 5.5.1). In this model, both the galaxies and the intracluster gas are assumed to be isothermal, bound to the cluster, and in equilibrium. The galaxies are assumed to have an isotropic velocity dispersion. However, the gas and galaxies are not assumed to have the same velocity dispersion; the square of the ratio of the galaxy-to-gas velocity dispersions is

$$\beta \equiv \frac{\mu m_p \sigma_r^2}{k T_g}, \quad (4.6)$$

where μ is the mean molecular weight in amu, m_p is the mass of the proton, σ_r is the one-dimensional velocity dispersion, and T_g is the gas temperature. Then, the gas and galaxy densities vary as $\rho_g \propto \rho_{gal}^\beta$. If

the galaxy distribution is taken to be a King analytical form for the isothermal sphere (equation 2.13), then the X-ray surface brightness $I_x(b)$ at a projected radius b varies as

$$I_x(b) \propto \left[1 + \left(\frac{b}{r_c} \right)^2 \right]^{-3\beta+1/2}, \quad (4.7)$$

where r_c is the galaxy core radius (Section 2.7). The self-gravitating isothermal model (equation 4.5) has the same surface brightness distribution if one makes the replacements $\beta = 1$ and $r_c = r_x$. These and other models for the distribution of the intracluster gas will be discussed in detail in Section 5.5.

The earliest determinations of the extent of the intracluster gas used low spatial resolution proportional counters. Observations with somewhat higher spatial resolution were made using detectors with modulation collimators (Schwarz *et al.*, 1979) or one-dimensional imaging detectors (Gorenstein *et al.*, 1973). Some clusters were found to contain compact X-ray sources associated with the central dominant galaxies in the cluster (Schnopper *et al.*, 1977), such as NGC1275 in the Perseus cluster (Wolff *et al.*, 1974, 1975, 1976; Cash *et al.*, 1976; Malina *et al.*, 1976; Helmken *et al.*, 1978). In Perseus, this point source contributes $\approx 20\text{--}25\%$ of the X-ray luminosity at moderate photon energies ($\approx \text{keV}$); at very high energies $\gtrsim 20 \text{ keV}$ the point source is dominant (Primini *et al.*, 1981; Rothschild *et al.*, 1981). The X-ray surface brightness of the Perseus cluster was shown to be extended in the east–west direction (Wolff *et al.*, 1974, 1975, 1976; Malina *et al.*, 1976; Cash *et al.*, 1976); this is the same direction as the line of bright galaxies seen optically in this L cluster. Later high resolution observations of the center of the cluster by Branduardi-Raymont *et al.* (1981) show a smaller elongation. A similar but smaller elongation was observed in the X-ray emission from the Coma cluster (Johnson *et al.*, 1979; Gorenstein *et al.*, 1979).

The *Einstein* X-ray telescope has allowed much more accurate determinations of the distribution of the X-ray emitting gas in clusters, using the moderate resolution ($\approx 1 \text{ arc min}$) Imaging Proportional Counter (IPC) or the High Resolution Imager (HRI, $\approx 8 \text{ arc sec}$). Moderate resolution *Einstein* IPC images of the X-ray emission in 46 clusters have been fit by equation (4.7) (Jones and Forman, 1984). Because galaxy core radii, velocity dispersions, and X-ray temperatures are poorly determined for most of these clusters, both the core radius and β were derived from the observed X-ray surface brightness. Table 2 gives the X-ray luminosities (L_x), X-ray core radii r_x , and temperature parameters β derived by Jones and Forman (the entries without an *). Note that the X-ray luminosity is only for gas within a projected radius of 0.5 Mpc of the cluster center. Abramopoulos and Ku (1983) fit IPC observations of 53 clusters to equation (4.5); that is, they assumed $\beta = 1$. Table 2 also contains the results of their fits (the entries marked with an *) for the clusters with X-ray detections which were not studied by Jones and Forman. The values of the total X-ray luminosities of Abramopoulos and Ku were converted to luminosities within 0.5 Mpc using their best fit X-ray distribution. Of course, the value of β for all the Abramopoulos and Ku fits is listed as one. It is worth noting that in nearly every case where both Abramopoulos and Ku, and Jones and Forman studied the same cluster, Jones and Forman found best fit values of β which were much smaller than the $\beta = 1$ assumed by Abramopoulos and Ku, and Jones and Forman could rule out the value $\beta = 1$ with high statistical confidence. Thus the validity of the results of Abramopoulos and Ku is questionable because of the value of β assumed.

Figure 16 shows the X-ray surface brightness as a function of radius from the center of the X-ray emission (in minutes of arc) for A400, A2063, and A1795 from Jones and Forman (1984). The solid histogram is the data; the dots are the best fit using equation (4.6). About two-thirds of the clusters were fit quite well with this formula; however, one-third appeared to have an excess of emission near the cluster center. For those clusters with a central excess, the central surface brightness points were excluded for the fits given in Table 2. In Figure 16, A400 and A2063 were reasonably fit by equation (4.6), while A1795 was not. The second panel on A1795 shows the fit to equation (4.6) if the central several minutes of arc are ignored. Jones and Forman suggest that this excess central X-ray emission correlates with the radio luminosity of the cluster, and that the excess emission may be due to a cooling flow in the cluster core (Section 5.7). A1795 does in fact have a very large cooling flow (Table 4 below).

A wide range of core radii ($0.07 - 0.9 h_{50}^{-1} \text{ Mpc}$) were derived by Jones and Forman. They found a strong anticorrelation between the presence of a dominant cluster galaxy in the core and the size of the X-ray core radius. (Unfortunately, Abramopoulos and Ku (1983) found the opposite effect with their sample of clusters

observed with *Einstein*.) The average value of β found by Jones and Forman from the X-ray distributions is $\beta = 0.65$, which implies that the gas is considerably more extensively distributed than the mass in the cluster. Unfortunately, the average value of β determined by applying equation (4.6) to those clusters with measured velocity dispersions and X-ray temperatures is $\beta = 1.1$. It is not clear whether this discrepancy results from errors in the measured cluster properties, velocity anisotropies, or a failure of the isothermal model (see Section 5.5.1).

While in most clusters the X-ray emission was found to be as broadly distributed as the galaxy distribution, an exception was Virgo/M87. Here the soft X-ray emission comes from a small region around M87, while the galaxy core radius of this irregular cluster, although hard to define, is certainly much larger. While early observations suggested the existence of weaker hard X-ray emission originating from a larger region of the cluster (Davison, 1978; Lawrence, 1978), recent observations indicate that this emission actually is due to the nucleus of M87 (Lea *et al.*, 1981, 1982).

The thermal X-ray emission from intracluster gas is proportional to the emission integral EI (equation 4.3), while the mass of intracluster gas is proportional to $\int n_p dV$. All other things being equal, the mass of the intracluster gas is then $M_g \propto (L_x r_x^3)^{1/2}$. Thus, if the size and distribution of the X-ray emitting gas can be determined from the X-ray surface brightness, the mass in intracluster gas can be estimated. Unfortunately, the estimate is very uncertain because the X-ray emission falls off rapidly as the density decreases in the outer parts of the cluster, where a large fraction of the mass of the intracluster gas may be located. Early estimates of the mass of the intracluster gas based on the self-gravitating isothermal model (equation 4.7) were given in Lea *et al.* (1973) and Kellogg and Murray (1974). They found that the total mass of intracluster gas determined from the X-ray observations is similar to or somewhat larger than the total mass of the galaxies in the cluster estimated from their total luminosity and a typical galaxy mass-to-light ratio. However, based on these mass estimates, the intracluster gas is still only 5–15% of the total virial mass of the cluster, and thus the discovery of the intracluster gas has not resolved the missing mass problem in clusters (Section 2.8). Table 2 gives the intracluster gas masses (here within 3.0 Mpc radius of the cluster center) from the *Einstein* observations of Abramopoulos and Ku (1983), and Jones and Forman (1984). Recall that Abramopoulos and Ku assumed that $\beta = 1$, which is generally larger than the best fit values from Jones and Forman. Reducing the values of β greatly increases the gas mass for a fixed X-ray luminosity, and the gas mass actually diverges at large radii for $\beta \leq 1$. Thus the gas masses found by Jones and Forman are generally considerably larger than those found by Abramopoulos and Ku or the earlier studies. As similar result, using the HEAO-1 A-2 X-ray spectra as well as the *Einstein* X-ray images, was found by Henriksen and Mushotzky (1985). Because of uncertainties in the gas and virial mass densities at large radii, it is difficult to give an accurate determination of the mass fraction of the intracluster gas. It does appear likely that, on average, at least 10% of the virial mass must be in the form of intracluster gas. The fraction of the total mass in intracluster gas could in some cases be as high as 30–60%, if the intracluster gas is as extended as the results of Henriksen and Mushotzky suggest.

4.4.2 X-ray images of clusters and the morphology of the intracluster gas

A tremendous increase in our understanding of the distribution of the X-ray emitting gas in clusters has come about through imaging of the two-dimensional X-ray surface brightness. The launch of the *Einstein* X-ray observatory satellite made it possible to image X-ray clusters routinely, although a rocket-borne X-ray mirror was used to image the X-ray emission from the Coma, Perseus, and M87/Virgo clusters prior to the launch of *Einstein* (Gorenstein *et al.*, 1977, 1978, 1979). The results of the X-ray imaging observations of clusters have been extensively reviewed recently by Forman and Jones (1982).

Forman and Jones (1982; also Jones *et al.*, 1979, and Jones and Forman 1984) propose a two-dimensional classification scheme for the X-ray morphology of galaxies, which they relate to the evolutionary state of the cluster as determined by its optical properties (Section 2.9, 2.10). This scheme is presented in Table 3, which is taken in large part from Forman and Jones (1982). In Figure 17 the X-ray surface brightness distributions from the Imaging Proportional Counter (IPC) on the *Einstein* X-ray observatory are shown for representative examples of clusters of each morphological type. The X-ray brightness distribution is shown as a contour plot, in which the lines are loci of constant X-ray surface brightness. The contours are superimposed on optical photographs of the cluster for comparison.

First of all, Forman and Jones classify the clusters as being irregular ('early') or regular ('evolved'), based on their overall X-ray distribution. The early clusters have irregular X-ray surface brightnesses, and often

show small peaks in the X-ray surface brightness, many of which are associated with individual galaxies in the cluster. Their X-ray luminosities L_x and X-ray spectral temperatures T_g are low. Optically, these clusters tend to be irregular clusters (Table 1). That is, they have irregular galaxy distributions with subclustering and with low central concentration. They are not generally very rich, and tend to be Bautz–Morgan types II to III and Rood–Sastry types F and I. They generally have low velocity dispersions σ_r . They are often, but not always, spiral rich.

On the other hand, the evolved clusters have regular, centrally condensed X-ray structures (Forman and Jones, 1982). The X-ray distribution is smooth, and X-ray emission peaks are not found associated with individual galaxies, except possibly with a central dominant galaxy at the cluster center. The evolved clusters have high X-ray luminosities and gas temperatures. Optically, these are regular, symmetric clusters, generally of Bautz–Morgan type I to II and Rood–Sastry types cD, B, L, or C. They are rich and have a high central concentration of galaxies. They have high velocity dispersions and are spiral poor in their galaxy composition.

The second determinant of the X-ray morphology of clusters is the presence or absence of a central, dominant galaxy in the cluster. The X-ray emission from a cluster tends to peak at the position of such a galaxy. Clusters containing such central dominant galaxies are classified by Forman and Jones (1982) as X-ray dominant (XD); those without such a galaxy are classified as non-X-ray dominant (nXD). The nXD clusters have larger X-ray core radii $r_x \approx 500/h_{50}$ kpc. There is no strong X-ray emission associated with any individual galaxy in these systems.

The XD clusters have small X-ray core radii $r_x \approx 250/h_{50}$ kpc. The X-ray emission is strongly peaked on the central dominant galaxy. In many cases, spectral observations indicate that gas is cooling in this central peak and being accreted onto the central galaxy (Sections 4.3.3 and 5.7). The data of Jones and Forman (1984) suggest that *all* cooling flows are centered on the dominant galaxies in XD clusters. Regular XD clusters have, on average, the highest X-ray luminosities of any clusters (Forman and Jones, 1982; Jones and Forman, 1984).

Examples of irregular nXD clusters include A1367 (Jones *et al.*, 1979; Bechtold *et al.*, 1983), A194 (Forman and Jones, 1982), A566 (Harris *et al.*, 1982), and A2069 (Gioia *et al.*, 1982). A1367, which is shown in Figure 17a, is the prototype of this class.

Virgo/M87 is the prototype of the irregular XD clusters, of which A262 (Forman and Jones, 1982), A347 (Maccagni and Tarenghi, 1981), A2147 (Jones *et al.*, 1979), and A2384 (Ulmer and Cruddace, 1982) are also members. Figure 17b shows the X-ray image of A262.

Regular nXD clusters probably include Coma (A1656; Abramopoulos *et al.*, 1981; see Figure 19 below) A576 (White and Silk, 1980), A1763 (Vallee, 1981), A2255, A2256 (Jones *et al.*, 1979; Forman and Jones, 1982), and CA0340-538 (Ku *et al.*, 1983). Figure 17c shows the X-ray image of A2256. Although the statistics are poor, it appears that clusters with radio halo emission (see Section 3.4) are primarily of this class.

The regular XD clusters include Perseus (A426, Branduardi-Raymont *et al.*, 1981; Fabian *et al.*, 1981a), A85, A1413, A1795 (Jones *et al.*, 1979), A478 (Schnopper *et al.*, 1977), A496 (Nulsen *et al.*, 1982), A2029 (Johnson *et al.*, 1980), A2218 (Boynton *et al.*, 1982), and possibly A2319 (White and Silk, 1980). Figure 17d shows the X-ray image of the cD cluster A85.

The prototypes of each of these classes (A1367, Virgo/M87, Coma, and Perseus) are discussed individually in Section 4.5 below.

Forman and Jones (1982) and Jones *et al.* (1979) argue that this classification scheme represents a sequence of cluster evolution, just as the sequence of cluster optical morphology (Section 2.5) was related to the dynamical evolution of the cluster in Section 2.9. Specifically, the evolution of the overall cluster distribution may be due to violent relaxation during the collapse of the cluster. This occurs on the dynamical time scale of the cluster, which depends only on its density (equation 2.32). On the other hand, the formation of central dominant galaxies may be due to mergers of galaxies; as demonstrated in Section 2.10.1, this mechanism favors compact but poor regions. Thus, while both a regular overall distribution of a cluster and the presence of a central dominant galaxy may indicate that the cluster has undergone dynamical relaxation, the processes are different and depend in different ways on the size and mass of the region. This provides a possible qualitative explanation for this two-dimensional classification system of X-ray morphology.

Double peaked X-ray emission has been seen in a number of clusters, of which the best studied example

is A98 (Henry *et al.*, 1981; Forman *et al.*, 1981). Other possible examples include A115, A1750, SC0627-54 (Forman *et al.*, 1981), A1560, A2355-A2356 (Ulmer and Crudace, 1982), A982, A2241 (Bijleveld and Valentijn, 1982), CA0329-527 (Ku *et al.*, 1983), and possibly A399/401 (Ulmer *et al.*, 1979; Ulmer and Crudace, 1981) and A2204-2210 (Ulmer *et al.*, 1985). In Figure 18, the X-ray distributions are shown for four double clusters observed by Forman *et al.* (1981). In most of these systems, the galaxy distribution is also double peaked (Dressler, 1978c; Henry *et al.*, 1981; Beers *et al.*, 1983), and the two peaks correspond to slightly different radial velocities (Faber and Dressler, 1977; Beers *et al.*, 1982). The small velocity differences indicate that the two subclusters are bound, and would be expected to collapse together and merge in a time of typically a few billion years.

These double systems may represent an intermediate stage in the evolution of clusters from irregular to regular distributions (Tables 1 and 3). In fact, numerical N-body simulations of cluster formation often show an intermediate bimodal stage to the galaxy distributions (see Figure 5c; White, 1976c; Ikeuchi and Hirayama, 1979; Carnevali *et al.*, 1981). With the current (rather poor) statistics, the fraction of clusters detected in this double phase ($\approx 10\%$) is consistent with the relatively short lifetime of the phase.

In the irregular and double clusters, the gas distribution is correlated with the galaxy distribution. The gas is probably roughly in hydrostatic equilibrium with the cluster gravitational potential (Section 5.5), which is primarily influenced by the distribution of the dynamically dominant missing mass component (Section 2.8). The fact that the X-ray surface brightness and galaxy distribution correlate suggests that the galaxies and the missing mass have a similar distribution in clusters; other evidence favoring this viewpoint was given in Section 2.8.

4.5 Individual clusters

4.5.1 Coma

Coma is the prototype of the nXD regular cluster, and is often used as a general prototype for comparing models for relaxed clusters. Its optical image is given in Figure 1b. It is worthwhile noting that clusters as rich and compact as Coma are actually rather uncommon. Coma is also unusual in having the most prominent radio halo observed among clusters (Section 3.4). Coma and A1367 (see below) appear to be portions of a large supercluster system (Gregory and Thompson, 1978).

The X-ray image of Coma, which is shown in Figure 19, shows that the X-ray emission is somewhat elongated (Johnson *et al.*, 1979; Gorenstein *et al.*, 1979; Helfand *et al.*, 1980) in the same direction as the galaxies in the cluster (Section 2.7). The X-ray emission shows a central uniform core (Helfand *et al.*, 1980; Abramopoulos *et al.*, 1981) and falls off with radius more slowly than the galaxy distribution ($\beta \approx 0.8$ in equation 4.7). Coma has an unusually high X-ray temperature for its galaxy velocity dispersion (Mushotzky *et al.*, 1978), although this is consistent with its small value for β . The X-ray surface brightness from Coma is very smooth, and there is no apparent excess emission associated with the two central galaxies (Forman and Jones, 1982; Bechtold *et al.*, 1983).

4.5.2 Perseus

The Perseus cluster is one of the most luminous X-ray clusters known, and has an unusually high radio luminosity as well (Gisler and Miley, 1979). It may represent an extreme in the evolution of the gas component in clusters.

Optically, Perseus is an L cluster (Bahcall, 1974a); the brightest galaxies form a chain oriented roughly east–west (see Figure 1c). The very prominent and unusual galaxy NGC1275 (Figure 20) is located near the east end of this chain. Together, Perseus, the clusters A262 and A347, and some smaller groups form the Perseus supercluster, which is also elongated in an east–west direction (Gregory *et al.*, 1981). On a moderately large scale, the X-ray emission from Perseus also appears to be elongated in the same direction as the galaxy distribution (Wolff *et al.*, 1974, 1976; Cash *et al.*, 1976; Malina *et al.*, 1976; Branduardi-Raymont *et al.*, 1981). The X-ray emission is peaked on NGC1275 and becomes more spherically symmetric near this galaxy (see Figures 21 and 22). Oddly, the center of the extended X-ray emission in Perseus appears to lie slightly to the east of NGC1275, while the galaxy distribution is centered to the west (Figure 21).

The X-ray emission from Perseus has been observed out to large distances ($\approx 2.5^\circ$) from the cluster core (Nulsen *et al.*, 1979; Nulsen and Fabian, 1980; Ulmer *et al.*, 1980a), although the brightness which is observed is that expected from models of the inner X-ray emission. Unlike Coma, Perseus does not appear to have a significant radio halo (Gisler and Miley, 1979; Hanisch and Erickson, 1980; Birkinshaw, 1980). In

the outer portions of the cluster, the radio galaxy NGC1265 is the classic example of a head-tail radio galaxy (see Section 4.3). The distortion in the radio morphology of the source indicates that it is passing at high velocity through moderately dense intracluster gas (Owen *et al.*, 1978).

Perseus was the first cluster to have an X-ray line detected in its spectrum (Mitchell *et al.*, 1976). The gas temperature in Perseus, derived from its spectrum (Mushotzky *et al.*, 1981), is smaller than one would expect, given the very large line-of-sight velocity dispersion in the cluster (equation 4.10) or the extent of the gas distribution observed (equation 4.6). However, recent work on the galaxy spatial and velocity distribution by Kent and Sargent (1983) has reduced the discrepancy considerably.

The galaxy NGC1275 in the core of the Perseus cluster is one of the most unusual objects in the sky; it occupies a role in extragalactic astronomy similar to the position of the Crab Nebula (which it visually resembles) in galactic astronomy. The visual appearance of the galaxy (Figure 20) is dominated by a complex network of filaments, which show an emission line spectrum (Kent and Sargent, 1979). These filaments form two distinct velocity systems (Rubin *et al.*, 1977, 1978): a ‘low velocity’ system with the same velocity as the stars in NGC1275, and a ‘high velocity’ system with a radial velocity about 3000 km/s higher. Strangely enough, 21 cm absorption line measurements indicate that the high velocity system lies between us and NGC1275 (Rubin *et al.*, 1977). Thus the velocity difference cannot be explained as a difference in the Hubble velocity due to differences in the distance to the two systems. At the moment, the leading suggestion appears to be that the high velocity gas is in a spiral or irregular galaxy that is accidentally superposed on NGC1275 and that just happens to be moving towards the center of the Perseus cluster at 3000 km/s. It is possible that this intervening galaxy is in the outer parts of the cluster or supercluster and is falling into the cluster on a nearly radial orbit. Alternatively, Hu *et al.* (1983) suggest that the high velocity system is a spiral galaxy that is actually colliding with the X-ray emitting gas around NGC1275 (Section 5.7.3). Unfortunately, no stellar component of this possible intervening galaxy has ever been convincingly detected (Kent and Sargent, 1979; but see Adams, 1977), and the galaxy seems to have very little neutral hydrogen, given its strong line emission (van Gorkom and Ekers, 1983).

Although the spatial distribution of stellar light from NGC1275 is that of a giant elliptical galaxy (Oemler, 1976), the galaxy has very blue colors and an A-star stellar spectrum (Kent and Sargent, 1979), suggesting ongoing or recent star formation. The nucleus of the galaxy contains a very compact, highly variable, powerful nonthermal source of radiation with a spectrum that extends from radio to hard X-rays. Because of its line emission, filaments, blue color, and nuclear emission, NGC1275 was classified as a Seyfert galaxy, although Seyferts are generally spiral galaxies and are not usually located in compact cluster cores.

The diffuse cluster X-ray emission from Perseus is very strongly peaked on the position of NGC1275 (Fabian *et al.*, 1981a; Branduardi-Raymont *et al.*, 1981; Figures 21 and 22). This is in addition to the highly variable, very hard X-ray point source associated with the nucleus of NGC1275 (Primini *et al.*, 1981; Rothschild *et al.*, 1981). The gas around NGC1275 has a positive temperature gradient $dT_g/dr > 0$ (Ulmer and Jernigan, 1978), which is not expected if the gas is hydrostatic and not cooling. High resolution X-ray spectra of the region about NGC1275 show the presence of significant quantities of cool gas in this region (Canizares *et al.*, 1979; Mushotzky *et al.*, 1981). Together, the X-ray surface brightness, X-ray spectra, and temperature gradients are best explained if gas is cooling onto NGC1275 at a rate of $\dot{M} \approx 400 M_\odot/\text{yr}$ (Section 5.7).

This cooling flow can provide an explanation of all the unusual properties of NGC1275 except the high velocity filaments. As the gas continues to cool through lower temperatures, it could produce the optical line emission seen in the low velocity filaments. The line emission becomes filamentary because the cooling is thermally unstable (Fabian and Nulsen, 1977; Mathews and Bregman, 1978; Cowie *et al.*, 1980; Section 5.7.3). If a small fraction of the accreted mass reaches the nucleus, it could power the nonthermal nuclear emission. Over the age of the cluster of $\approx 10^{10}$ yr, roughly $3 \times 10^{12} M_\odot$ would have been accreted. There is no reasonable reservoir for this much mass except in low mass star formation, which may be favored under the physical conditions in accretion flows (Fabian *et al.*, 1982b; Sarazin and O’Connell, 1983). Ongoing star formation would explain the blue color and A-star spectrum of NGC1275. Moreover, if accretion has been going on for the lifetime of the cluster, the entire luminous mass of NGC1275 might be due to accretion (Fabian *et al.*, 1982b; Sarazin and O’Connell, 1983; Wirth *et al.*, 1983).

4.5.3 M87/Virgo

Virgo is the nearest rich cluster to our galaxy. The cluster is quite irregular and spiral rich (de Vau-

couleurs, 1961; Abell, 1975; Bautz and Morgan, 1970), although the X-ray emission comes from an elliptical-rich core surrounding the galaxy M87 (Figure 1a).

M87 is classified as a peculiar elliptical galaxy. It has fairly extended optical emission (de Vaucouleurs and Nieto, 1978), but is not a cD. This galaxy is one of the brightest radio sources in the sky. Nonthermal radio emission is observed from the nucleus and from the prominent optical jet (see Figure 23), and both the nucleus and jet also produce nonthermal X-ray emission (Schreier *et al.*, 1982; Figure 25). There is also a larger scale radio halo source with a size comparable to that of the entire galaxy (Andernach *et al.*, 1979; Hanisch and Erickson, 1980).

The X-ray emission from this cluster is strongly concentrated to the region around M87, and is much less broadly distributed than the galaxies in the cluster (Malina *et al.*, 1976; Gorenstein *et al.*, 1977; Fabricant and Gorenstein, 1983). Figure 24 shows the *Einstein* IPC X-ray image of this cluster; Figure 25 is the HRI image. The top panel shows the outer contours, while the bottom panel is an expanded view of the center, showing X-ray emission from the jet. The emission is also considerably cooler ($T_g \approx 2.5$ keV) than is usually associated with cluster emission (Lea *et al.*, 1982). Nearly all of the emission comes from a 1° region around M87, in which more than 95% of the optical emission comes from M87. As a result, it seems reasonable to assume that the bulk of the emission is associated with M87 itself, rather than the cluster as a whole.

Similar arguments led Bahcall and Sarazin (1977) and Mathews (1978b) to suggest that the X-ray emitting gas is gravitationally bound to M87 itself. Bahcall and Sarazin showed that M87 could only bind the gas if it had a very massive halo, with a total mass of $1 - 6 \times 10^{13} M_\odot$. This estimate was based on early low resolution X-ray observations. More recent *Einstein* observations (Fabricant *et al.*, 1980; Fabricant and Gorenstein, 1983; Stewart *et al.*, 1984a) appear to support this model and provide a more accurate estimate of the mass of $3 - 6 \times 10^{13} M_\odot$. These observations measure, at least approximately, the temperature gradient in M87 and allow a direct determination of the mass profile of the galaxy. The massive halo must extend out to roughly 1° from M87. The optical surface brightness of the galaxy is very low in this region, and thus the mass-to-light ratio of the material making up this halo must be rather large. These observations suggest that at least this one elliptical galaxy possesses a massive, dark missing mass halo of the type discussed in Section 2.8. Whether this is typical of giant ellipticals or whether it is a consequence of M87's position at the center of the Virgo cluster is uncertain.

Alternatively, Binney and Cowie (1981) have suggested that the mass of M87 might actually be rather small. They argue that the cooler ($T_g \approx 2.5$ keV) gas providing the bulk of the X-ray emission is confined by the pressure of a hotter ($T_g \approx 8$ keV), lower density, intracluster medium.

A key prediction of the Binney–Cowie model is the existence of a significant amount of hot gas ($T_g \approx 8$ keV) surrounding M87. The density of this gas is determined by the requirement of pressure equilibrium with the cooler gas in M87. Early observations (Davison, 1978; Lawrence, 1978) suggested that there was extended, hard X-ray emission in this cluster. More recent observations (Mushotzky *et al.*, 1977; Ulmer *et al.*, 1980a; Lea *et al.*, 1981, 1982) have not confirmed its existence, and it has been suggested that the previously observed hard X-ray emission is entirely from the nucleus of M87. The observed temperature gradient in M87 is apparently not consistent with the Binney–Cowie model (Fabricant *et al.*, 1980; Fabricant and Gorenstein, 1983; Stewart *et al.*, 1984a).

A host of X-ray lines have been detected from M87 (see Section 4.3). Both Fe L and K lines have been observed (Serlemitsos *et al.*, 1977; Lea *et al.*, 1979), as well as lines from lighter elements. The observations indicate that the abundance of these elements is reasonably uniform within M87 (Fabricant *et al.*, 1978; Lea *et al.*, 1982). The gas temperature appears to be roughly constant at projected radii of $\gtrsim 5$ arc min, and decreases rapidly within this radius. In the inner regions, large amounts of emission by quite cool gas are observed in the higher resolution spectra (Canizares *et al.*, 1979, 1982; Lea *et al.*, 1982). The X-ray surface brightness is also strongly peaked in this region.

The presence of a range of temperatures of cool gas and the central peak in the X-ray surface brightness both suggest that the gas in M87 is radiatively cooling and being accreted onto the center of the galaxy (Gorenstein *et al.*, 1977; Mathews, 1978b). Comparisons between models for the accretion (Mathews and Bregman, 1978; Binney and Cowie, 1981) and the observations suggest that the accretion rate is $\dot{M} \approx 3 - 20 M_\odot / \text{yr}$. Like NGC1275 in Perseus, M87 has optical line emitting filaments in its inner regions (Ford and Butcher, 1979; Stauffer and Spinrad, 1979), which may be produced by thermal instabilities in the cooling gas (Fabian and Nulsen, 1977; Mathews and Bregman, 1978). As discussed above for NGC1275, the

radio source may be powered by further accretion of a small fraction of the gas onto the nucleus (Mathews, 1978a), while the bulk of the accreted material may form low mass stars (Sarazin and O'Connell, 1983). Both the line emission and halo radio emission are concentrated to the north side of M87. De Young *et al.* (1980) suggest that M87 is moving at ≈ 200 km/s relative to the intracluster gas it is accreting, but Dones and White (1985) show that the thermodynamic structure of the gas is inconsistent with this motion.

Recently, Tucker and Rosner (1983) have suggested a hydrostatic (no accretion) model for M87. The gas in the outer portions of the galaxy is heated by the nonthermal electrons that produce the halo radio emission. Heat is conducted from these hotter outer regions into the cooler central regions, where it is radiated away; unlike Binney and Cowie (1981), Tucker and Rosner assume full thermal conductivity. The heating by nonthermal electrons balances the cooling, and the gas is in thermal equilibrium and hydrostatic. The radio source is itself powered by accretion; thus they argue that the behavior of the system is episodic, with alternating periods of accretion and nonthermal heating.

Many other galaxies in the central regions of the Virgo cluster have been detected as X-ray sources with luminosities in the range $L_x \approx 10^{39-41}$ ergs/s (Forman *et al.*, 1979). Although no spectra are available for these galaxies, it is likely that the X-rays are due to thermal emission from hot gas. Figure 26 shows the X-ray emission from the very optically luminous galaxy M84 and the galaxy M86, which has an X-ray plume extending away from M87. Fabian *et al.* (1980) suggest that M86 contains hot gas, which is currently being stripped by ram pressure as it moves into the core of the cluster (see Sections 2.10.2 and 5.9). The interpretation of the X-ray emission from M87 and the other galaxies in the Virgo cluster is discussed in more detail in Section 5.8.

4.5.4 A1367

A1367 is an irregular, BM type II-III cluster (Bautz and Morgan, 1970; Carter and Metcalfe, 1980) with a moderately high spiral fraction (Bahcall, 1977c). It has a low X-ray luminosity $L_x \approx 4 \times 10^{43} h_{50}^{-2}$ erg/s and a temperature $T_g \approx 3 \times 10^7$ K (Mushotzky, 1984). The X-ray emission is extended and elongated (Figure 27a); Bechtold *et al.* (1983) found an X-ray core radius (equation 4.7) of $r_x = (0.8 \times 0.4) h_{50}^{-1}$ Mpc for the semimajor and semiminor axes of the distribution. In addition, 8 point sources and 13 resolved peaks in the X-ray emission (with typical sizes of an arc minute) were observed at higher spatial resolution (Figure 27b), 11 of which were associated with cluster galaxies. These galaxies have X-ray luminosities in the range $10^{40-42} h_{50}^{-2}$ erg/s. 21 cm observations indicate that some of these galaxies also contain neutral hydrogen (Chincarini *et al.*, 1983). A1367 is the only irregular cluster that appears to have an extended radio halo (Hanisch, 1980).

4.6 X-ray-optical correlations

A number of correlations between the optical and X-ray properties of clusters have been found (Bahcall, 1977a). The optical cluster properties that have been used to study X-ray clusters include the richness (Section 2.3), morphology (RS or BM type; Section 2.5), the galactic content (cD galaxies and spiral fractions; Section 2.10), the core radius r_c (or other radii; Section 2.7), the velocity dispersion σ_r (Section 2.6), and the central galaxy density \bar{N}_o of Bahcall (1977b; Section 2.7). The largest X-ray surveys (Section 4.2) provide only X-ray fluxes (and thus luminosities L_x) for a given X-ray photon energy range. In addition, there are now smaller samples of clusters with X-ray surface brightness determinations, giving r_x or β (Section 4.4), and samples with X-ray spectra, yielding T_g and EI (Section 4.3).

Solinger and Tucker (1972) first suggested that the X-ray luminosity of a cluster correlates with its velocity dispersion $L_x \propto \sigma_r^4$, based on the small sample of known X-ray clusters at that time. This sample included M87/Virgo, in which the X-ray emission comes from the galaxy M87 rather than the entire cluster, and several clusters having multiple velocity components, which probably cause the velocity dispersion to be overestimated. Solinger and Tucker gave a simple model to explain the correlation based on the assumption that the emission comes from intracluster gas. This correlation has been reexamined a number of times and other explanations of its physical significance have been given (Yahil and Ostriker, 1973; Katz, 1976; Silk, 1976). As the sample of X-ray clusters has grown, the correlation has become less convincing (McHardy, 1978a), and certainly there is a great deal of scatter about any such correlation. However, Quintana and Melnick (1982) find basically the same relationships for larger data samples from the HEAO-1 and *Einstein* observatories. Their relationship for the HEAO-1 data is shown as Figure 28; roughly, the correlation is

$$L_x(2-10\text{ keV}) \approx 4.2 \times 10^{44} \text{ ergs/s} \left[\frac{\sigma_r}{10^3 \text{ km/s}} \right]^4. \quad (4.8)$$

For the lower energy *Einstein* observations, the power in equation (4.8) is closer to three (Quintana and Melnick, 1982; Abramopoulos and Ku, 1983).

In some sense, the richness of a cluster (the number of galaxies in the cluster) and L_x measure the mass of stars and of diffuse gas in the cluster, respectively, and one might expect these to be related. Such a relationship does indeed appear in the data (Bahcall, 1974b; Jones and Forman, 1978; McHardy, 1978a; McKee *et al.*, 1980; Ulmer *et al.*, 1981; Abramopoulos and Ku, 1983; Johnson *et al.*, 1983). It is difficult to give a quantitative measure of the correlation because the richness class 0 clusters are incomplete in the Abell catalog, Abell richnesses are binned in the original catalog, and the higher richness clusters are generally more distant. Abramopoulos and Ku (1983) find that the low energy X-ray luminosity increases with richness to the 1.2 power. Recently this correlation has been shown to extend to the very richest Abell clusters by Soltan and Henry (1983; see also Pravdo *et al.*, 1979). Low luminosity X-ray clusters are much less common among the richest clusters.

There appears to be a correlation between the optical morphology and X-ray luminosity of clusters. Bahcall (1974b), Owen (1974), Mushotzky *et al.* (1978), McKee *et al.* (1980), and Abramopoulos and Ku (1983) found that the more regular Rood–Sastry types (cD, B) in general have higher X-ray luminosities than the less regular clusters. The reality of this correlation has been disputed by Jones and Forman (1978) and Lugger (1978), although it does appear in the larger HEAO-1 and *Einstein* data samples. Bautz–Morgan type I clusters are found to be more luminous than the less regular clusters, although the correlation does not seem to continue to less regular BM types (McHardy, 1978a; McKee *et al.*, 1980; Ulmer *et al.*, 1981; Abramopoulos and Ku, 1983; Johnson *et al.*, 1983). Clusters that contain optically dominant galaxies near the cluster center tend to be stronger X-ray sources than those that do not (Bahcall, 1974b; Jones and Forman, 1984). (In the Jones–Forman classification scheme these are XD clusters; see Section 4.4).

Jones and Forman (1978) have argued that all of these X-ray–optical correlations are due to the correlation of X-ray luminosity and richness *plus* the tendency of rich clusters to be more regular (lower BM type, etc.). As the sample of X-ray clusters has enlarged, it has become possible to test this possibility by considering subsamples of fixed richness, and correlations with optical morphology appear in these samples as well (e.g., McHardy, 1978a).

Bahcall (1977b) showed that the X-ray luminosity of clusters was well-correlated with the projected central galaxy density parameter \bar{N}_o (Section 2.7). This correlation is tighter than the richness correlation; an explanation of this may be that thermal X-ray emission depends on the square of the density of the gas, and thus is most sensitive to the deepest portion of the cluster potential. This correlation has been confirmed by Mushotzky *et al.* (1978), Mitchell *et al.* (1979), Abramopoulos and Ku (1983), and Mushotzky (1984). The correlation in the HEAO-1 A-2 sample (Figure 29) is consistent with $L_x \propto (\bar{N}_o)^{3.5}$, and is one power weaker in the lower energy *Einstein* sample.

Bahcall (1977c) showed that the X-ray luminosity of a cluster correlated with its galactic content, in that luminous X-ray clusters generally have a small proportion of spiral galaxies. This correlation has been confirmed by McHardy (1978a), Mushotzky *et al.* (1978), Tyler and Vidal (1979), and Abramopoulos and Ku (1983). Melnick and Sargent (1977) showed that the spiral fraction increased with radius in X-ray clusters, and that the spiral fraction was inversely correlated to the velocity dispersion. These correlations are consistent with the theory that spiral galaxies formed in clusters are stripped of their gas content through interactions with the intracluster gas and become S0 galaxies (Sections 2.10.2 and 5.9), although they certainly do not prove that this has occurred. Let f_{Sp} be the fraction of cluster galaxies that are spirals. This spiral fraction is plotted against the X-ray luminosity of clusters in Figure 30. Bahcall (1977c) showed that the correlation could be represented as

$$f_{Sp} \approx 0.37 - 0.26 \log \left(\frac{L_x}{10^{44} \text{ ergs/s}} \right), \quad (4.9)$$

which she derived based on a simple model for stripping of spirals in a cluster. An important counterexample to this correlation is A194, which has a low spiral fraction $f_{Sp} \approx 0.27$ (Oemler, 1974; Dressler, 1980a). Yet this cluster has a very low X-ray luminosity (Jones and Forman, 1984) and low velocity dispersion, which make it unlikely that the spirals were stripped by the ram pressure of intracluster gas.

Kellogg and Murray (1974) suggested a correlation between the sizes of the X-ray sources in clusters and the galaxy core radii, $r_x \approx 2r_c$, but very few clusters have accurately determined galaxy core radii (Section

2.7). Ulmer *et al.* (1981) found that the X-ray luminosity of clusters in the HEAO-1 A-1 survey correlated with the cluster radius R_{LV} as given by Leir and van den Bergh (1977; Section 2.7), with $L_x \propto R_{LV}^2$. This correlation was confirmed in the *Einstein* data by Abramopoulos and Ku (1983).

The OSO-8, *Ariel 5*, and HEAO-1 A-1 spectral surveys established a number of correlations between the X-ray spectral parameters and the optical properties of X-ray clusters (Mitchell *et al.*, 1977, 1979; Mushotzky *et al.*, 1978; Smith *et al.*, 1979a; Mushotzky 1980, 1984). The two properties that can be derived most easily from a continuum X-ray spectrum are the gas temperature T_g and the emission integral EI (equation 4.3). The X-ray luminosity of a cluster is proportional to EI (equation 4.11).

An important correlation was found between the gas temperatures T_g , determined from the X-ray spectra, and the velocity dispersion of the galaxies in the cluster (see Section 2.6). The older surveys (Mushotzky *et al.*, 1978; Mitchell *et al.*, 1979; Smith *et al.*, 1979a) had found that, when cD clusters were excluded, the temperatures varied roughly as

$$T_g \approx 6 \times 10^7 \text{ K} \left(\frac{\sigma_r}{10^3 \text{ km/s}} \right)^2. \quad (4.10)$$

As will be demonstrated in Section 5.5, such a correlation would be expected if the gas were gravitationally bound to the cluster, because the velocity dispersion measures the depth of the cluster potential (equation 2.24). Moreover, this correlation between T_g and σ_r^2 supported the view that clusters must contain large masses of unseen material (the missing mass problem, Section 2.8). The fact that the galaxy and gas velocity dispersions were similar and proportional to one another suggested that both gas and galaxies were bound by the same gravitational potential, which required a very large mass for the cluster. The cD clusters showed less of a correlation, which may indicate that in these systems there is a significant amount of gas bound to the cD itself, rather than to the entire cluster. Unfortunately, the HEAO 1 A-2 data do not show a very tight correlation and give $T_g \propto \sigma_r$ (Mushotzky, 1984; Figure 31).

The gas temperature correlates strongly with the central density \bar{N}_o of Bahcall (1977b), with $T_g \propto \bar{N}_o$ (Mushotzky *et al.*, 1978; Mitchell *et al.*, 1979; Mushotzky, 1984). This correlation is shown in Figure 32.

For thermal bremsstrahlung emission (Section 5.1.3), the emission integral EI , X-ray luminosity L_x , and gas temperature T_g are related by

$$L_x \propto EI T_g^{1/2}, \quad (4.11)$$

and thus the correlations of L_x and T_g with optical properties imply correlations of EI with these properties as well.

4.7 Poor clusters

X-ray emission has also been detected associated with poor clusters of galaxies (Schwartz *et al.*, 1980a,b; Kriss *et al.*, 1980, 1981, 1983). Of special interest are the poor clusters that contain D or cD galaxies, lists of which have been given by Morgan *et al.* (1975) and Albert *et al.* (1977); clusters from these two lists are identified as MKW and AWM clusters, respectively. The optical properties of these cD galaxies were discussed in Section 2.10.1, where it was concluded that the cDs in poor clusters were similar to cDs in rich clusters, except that they lacked the very extended, low surface brightness envelopes seen in rich cluster cDs (Thuan and Romanishin, 1981; Oemler, 1976; van den Bergh, 1977a). This suggests that the main bodies of cDs are formed by a process, such as mergers, which is not strongly dependent on richness, while the halos are formed by a process, such as tidal interactions, which does depend strongly on richness.

In general, the optical properties of the poor clusters appear to be simple extensions of the properties of rich clusters to lower richness (Stauffer and Spinrad, 1978; Thomas and Batchelor, 1978; Schild and Davis, 1979; Bahcall, 1980; Malumuth and Kriss, 1986). The radio properties of the poor cluster cDs are also very similar to those of rich clusters (White and Burns, 1980; Burns *et al.*, 1980, 1981b).

The poor clusters containing cD galaxies are generally observed to be X-ray sources with luminosities of $\approx 10^{42-44}$ erg/s (Schwartz *et al.*, 1980a,b; Kriss *et al.*, 1980, 1981, 1983; Malumuth and Kriss, 1986). Kriss *et al.* (1983) surveyed 16 MKW and AWM clusters, and detected X-ray emission from 12. In the brighter clusters, the X-ray emission was found to be smoothly distributed, relatively symmetrical, and fairly extended (out to radii of ≈ 1 Mpc). Kriss *et al.* found that the X-ray temperatures in these poor clusters were fairly low $T_g \approx 1 - 5$ keV, in keeping with their low velocity dispersions (see equation 4.10). The X-ray emission is strongly peaked on the position of the cD galaxy (Canizares *et al.*, 1983; Malumuth and Kriss, 1986; see

the image of AWM4 in Figure 33). This suggests that the cDs in poor clusters are located at the bottoms of cluster potential wells, as was shown to be the case for rich cluster cDs (Section 2.10.1). In many ways, the bright poor cluster X-ray sources resemble the regular XD clusters discussed above. The X-ray emission in a number of cases is elongated in the same direction as the long axis of the cD galaxy (Kriss *et al.*, 1983; a similar effect is seen in rich clusters, as discussed in Section 2.10.1 and 4.5).

From the distribution of the gas in the cluster, the cluster gravitational potential and mass can be derived (Kriss *et al.*, 1983; Malumuth and Kriss, 1986). When these are compared to the optical luminosity of the clusters, a considerable missing mass problem is found. The mass-to-light ratios in the inner parts of these clusters are found to be roughly $M/L_V \approx 100h_{50} M_\odot/L_\odot$. The X-ray emitting gas makes up roughly 15% of the total cluster mass in these inner regions. Because the cD galaxy contributes a large fraction of the luminosity in these poor clusters, this dark matter can also be thought of as a massive halo around the central cD, as in the M87/Virgo cluster.

The X-ray emission is strongly peaked on the position of the cD in these poor clusters (Canizares *et al.*, 1983; Kriss *et al.*, 1983; Malumuth and Kriss, 1986), and the cooling times in the central portions of the gas are generally estimated to be less than the probable age of the cluster (the Hubble time). This suggests that the X-ray emitting gas forms a cooling accretion flow onto the cD (Canizares *et al.*, 1983), as has been observed in many rich clusters. The accretion rates in MKW4, MKW3s, AWM4, and AWM7 are estimated to be in the range of 5–100 M_\odot/yr (Canizares *et al.*, 1983; Malumuth and Kriss, 1986). It is possible that star formation from the accretion flow may provide a portion of the optical luminosity of the cD (Fabian *et al.*, 1982b; Sarazin and O'Connell, 1983). Very little cold gas (neutral hydrogen) is detected in these cDs (Burns *et al.*, 1981a; Valentijn and Giovanelli, 1982).

Poor clusters containing head-tail radio sources have also been detected in X-rays (Burns and Owen, 1979; Holman and McKee, 1981). Head-tail radio sources are believed to be produced by galaxy motions through intracluster gas (Sec 4.3).

4.8 High redshift clusters and X-ray cluster evolution

Observations of high redshift X-ray clusters (taken here to mean clusters with $z \geq 0.2$) offer the possibility of studying the formation and evolution of clusters. At the moment this study is hampered by the small sample of clusters that have been observed and the difficulty in obtaining detailed information (spectra and spatial distributions) for these distant sources. Samples of high redshift X-ray clusters have been given by Henry *et al.* (1979, 1982), Helfand *et al.* (1980), Perrenod and Henry (1981), White *et al.* (1981a,b, 1987), and Henry and Lavery (1984). These samples are based on lists of clusters detected initially through optical or radio emission.

One purpose of this study is to determine whether X-ray clusters have evolved in any detectable way over the cosmological time span represented by the redshift. Henry *et al.* (1979) compared a sample of six high redshift clusters with more nearby clusters, and concluded that they were similar in many ways. They found marginal evidence that X-ray luminosities increase with time in accordance with the models of Perrenod (1978b) and Sarazin (1979), but the sample was too small to make any strong statements. Henry *et al.* (1982) and Henry and Lavery (1984) studied the evolution of the X-ray luminosity function at redshifts $z < 0.5$. No significant evidence for evolution was found.

Perrenod and Henry (1981) estimated gas temperatures for a sample of seven high redshift clusters observed with the *Einstein* observatory. With one exception (the remarkable cluster 0016+16 discussed below), the six other clusters with $z > 0.3$ all had $T_g < 4$ keV. Nearby clusters of similar X-ray luminosity have average temperatures of $T_g \approx 7$ keV. Thus it appeared that gas temperatures in clusters might be increasing with time. Since the temperature of gas in a cluster reflects in part the depth of the cluster potential well (Section 5.5), this suggested that clusters might be growing through the merger of smaller clusters, in accord with models developed by Perrenod (1978b). White *et al.* (1981a) also detected a fairly distant cluster (SC2059-247) with a very high X-ray luminosity and a low X-ray temperature. However, White *et al.* (1987) observed a sample of 10 high redshift clusters; of the five detected cluster X-ray sources, all have reasonably high X-ray temperatures $T_g \geq 6$ keV. They suspect that the previous evidence for low X-ray temperatures by Perrenod and Henry was an artifact of the small size of the sample being discussed. They also point out that some high luminosity, low temperature clusters, such as SC2059-247, may be extreme examples of clusters with cooling accretion flows at their centers.

Several of the high redshift clusters that have been observed as X-ray sources (Henry *et al.*, 1979, 1982;

Helfand *et al.*, 1980) are Butcher–Oemler clusters (Butcher and Oemler, 1978a). As discussed in Section 2.10.2, these are high redshift clusters that appear to contain blue galaxies with colors similar to those of spirals in nearby clusters. Prior to the X-ray observations, the most straightforward explanation of these clusters was that they did not contain enough gas to strip spiral galaxies effectively by ram pressure ablation (Norman and Silk, 1979; Sarazin, 1979). Obviously, the detection of large quantities of hot gas in these clusters has made that explanation untenable. Recently optical observations indicate that the blue galaxies in these clusters are not normal spirals and are unlike any class of present day galaxies (Dressler and Gunn, 1982).

The most spectacular high redshift cluster to be detected so far is probably the $z = 0.54$ cluster 0016+16 (White *et al.*, 1981b). Unfortunately, the cluster field also contains a foreground cluster at $z = 0.30$ (Ellis *et al.*, 1985), which probably affected early estimates of the richness of the $z = 0.54$ cluster and of the colors of its galaxies (Koo, 1981). Assuming the emission is entirely due to the higher redshift cluster, 0016+16 has one of the highest X-ray luminosities observed $L_x = 3 \times 10^{45} h_{50}^{-2}$ erg/s, and may have a high temperature, although this is very uncertain. The cluster is probably about as rich as Coma (Ellis *et al.*, 1985; Koo, 1981). The cluster 0016+16 may also be similar to nearby rich clusters in that it appears to have predominately red galaxies (it does not show the Butcher–Oemler effect), although the foreground cluster makes this conclusion somewhat uncertain. A microwave decrement (Section 4.5) of -1.4 mK has been detected from the cluster (Birkinshaw *et al.*, 1981a). This effect has not been detected in many nearby clusters, and if the detection were confirmed it would demonstrate the extreme luminosity and temperature of 0016+16.

The tentative conclusion at this point is that the X-ray properties of clusters do not appear to evolve dramatically to redshifts of $z \approx 0.5$, and that evolution of the X-ray properties is probably not the explanation of the Butcher–Oemler effect.

5 Theoretical progress

5.1 Emission mechanisms

When clusters of galaxies were found to be an important class of X-ray sources, there were a number of suggestions as to the primary X-ray emission mechanism. The three most prominent ideas were that the emission resulted from thermal bremsstrahlung from a hot diffuse intracluster gas (Felten *et al.*, 1966), or that the emission resulted from inverse Compton scattering of cosmic background photons up to X-ray energies by relativistic electrons within the cluster (Brecher and Burbidge, 1972; Bridle and Feldman, 1972; Costain *et al.*, 1972; Perola and Reinhardt, 1972; Harris and Romanishin, 1974; Rephaeli, 1977a,b), or that the emission was due to a population of individual stellar X-ray sources, like those found in our galaxy (Katz, 1976; Fabian *et al.*, 1976). Subsequent observations have provided a great deal of support for the thermal bremsstrahlung model, and have generally not supported the other two suggestions. These three emission mechanisms will be briefly reviewed, and a few of the arguments in favor of thermal bremsstrahlung and against the other two models will be cited.

5.1.1 Inverse Compton emission

As noted in Section 3.2, many X-ray clusters also have strong radio emission, and cluster radio emission is distinguished by having a large spectral index α_r . This led to the idea that the X-ray emission might be produced by the same relativistic electrons that produce the synchrotron radio emission. Very high energy electrons, with very short lifetimes, would be required to produce the emission by synchrotron emission, but much lower energy electrons could scatter low energy background photons up to X-ray energies (Felten and Morrison, 1966). If the electrons are extremely relativistic, with energies $E_e = \gamma m_e c^2$ for $\gamma \gg 1$, and the initial frequency of the background photon is ν_b , then on average the frequency after the scattering will be

$$\nu_x = \frac{4\gamma^2\nu_b}{3}. \quad (5.1)$$

Because the cluster radio spectra vary as power-laws, it is generally assumed that the relativistic electrons have a power-law energy distribution

$$N_e(\gamma)d\gamma = N\gamma^{-p}d\gamma \quad \gamma_l < \gamma < \gamma_u, \quad (5.2)$$

where $N_e d\gamma$ is the number of electrons with energies between γ and $\gamma + d\gamma$. Then, the resulting inverse Compton (IC) radiation has a spectrum given by

$$\frac{dL_x}{d\nu_x} = 3\sigma_T c h 2^p N \frac{p^2 + 4p + 11}{(p+3)^2(p+1)(p+5)} \left[\int \nu_b^{\alpha_x} n_b(\nu_b) d\nu_b \right] \nu_x^{-\alpha_x}, \quad (5.3)$$

where L_x is the X-ray luminosity, n_b is the number density of background photons as a function of frequency ν_b , σ_T is the Thomson cross section, and the X-ray spectral index is

$$\alpha_x = \frac{p - 1}{2}. \quad (5.4)$$

A source of low energy photons that is present everywhere and dominates the overall photon density in the universe is the 3 K cosmic background radiation. If the background is taken to be a blackbody radiation field at a temperature T_r , then

$$\frac{dL_x}{d\nu_x} = \frac{3\pi\sigma_T}{h^2 c^2} b(p) N (kT_r)^3 \left(\frac{kT_r}{h\nu_x} \right)^{\alpha_x} \quad (5.5)$$

for frequencies in the range $\gamma_l^2 \ll (h\nu_x/kT_r) \ll \gamma_u^2$. Here

$$b(n) = \frac{2^{p+3}(p^2 + 4p + 11)\Gamma[(p+5)/2]\zeta[(p+5)/2]}{(p+3)^2(p+1)(p+5)}, \quad (5.6)$$

and Γ and ζ are the gamma and Riemann zeta functions.

Now, the synchrotron radio luminosity L_r produced by the same electron population is given by

$$\frac{dL_r}{d\nu_r} = \frac{8\pi^2 e^2 \nu_B N}{c} a(p) \left(\frac{3\nu_B}{2\nu_r} \right)^{\alpha_r}, \quad (5.7)$$

where $\nu_B \equiv (eB/2\pi m_e c)$ is the gyrofrequency in the magnetic field B , and

$$a(p) = 2^{(p-7)/2} \left(\frac{3}{\pi} \right)^{1/2} \frac{\Gamma(\frac{3p-1}{12}) \Gamma(\frac{3p+19}{12}) \Gamma(\frac{p+5}{4})}{(p+1)\Gamma(\frac{p+7}{4})}. \quad (5.8)$$

Equation (5.8) assumes that the electron distribution is isotropic and that the frequency is in the range $\gamma_l^2 \ll (\nu_r/\nu_B) \ll \gamma_u^2$. The radio spectral index α_r is equal to the X-ray spectral index α_x .

Thus the synchrotron radio and IC X-ray emission from the same relativistic electrons will have the same spectral shape, and the luminosities in the two spectral regimes are also very simply related:

$$\frac{L_x}{L_r} = \frac{U_r}{U_B}, \quad (5.9)$$

where U_r and $U_B = (B^2/8\pi)$ are the energy densities of the background radiation and the magnetic field, respectively. Because fluxes (f_x, f_r) at single frequencies are easier to measure than integrated luminosities, it is useful to note that (in cgs units)

$$\left(\frac{f_x}{f_r} \right) \left(\frac{\nu_x}{\nu_r} \right)^{\alpha_x} = \frac{2.47 \times 10^{-19} T_r^3 b(p)}{Ba(p)} \left(\frac{4960 T_r}{B} \right)^{\alpha_x}. \quad (5.10)$$

Since the temperature of the cosmic background radiation is known, the ratio of radio and IC X-ray fluxes determines the magnetic field strength (Harris and Romanishin, 1974). If the X-ray emission from clusters were due to inverse Compton emission, the magnetic field would typically be $B \approx 0.1 \mu\text{G}$. Fields at least ten times larger than this are expected if the relativistic electrons and the magnetic field have equal energy densities; the small magnetic fields required by the IC model would increase the energy requirements for cluster radio sources by about two orders of magnitude. On the other hand, if the X-rays are not due to IC emission, than this value is a lower limit to the magnetic field strength in the radio emitting region (Section 4.3.1).

For a magnetic field of $B \approx 1 \mu\text{G}$, radio photons at a frequency of 1 GHz are produced by electrons with $\gamma \approx 2 \times 10^4$. Similarly, IC X-ray emission at a photon energy of 1 keV is produced by electrons with $\gamma \approx 10^3$. Thus the electrons that could produce the observed X-ray emission would produce very low frequency radio emission, typically at a frequency of 20 MHz if $B \approx 1 \mu\text{G}$, and at an even lower frequency if B is lower as required in this model. Such low frequency radio emission generally cannot be detected from the Earth. However, one argument in favor of the IC model of cluster X-ray emission was that cluster radio sources have steep spectra (Section 3.1), indicating that they have many lower energy relativistic electrons, as required to explain their X-ray emission.

There is now considerable evidence against the IC model. In this model, one would expect a very strong correlation between the low frequency radio flux and the X-ray flux (equation 5.10). The reality of any radio–X-ray correlation is questionable (Section 3.2), and recent larger samples of X-ray clusters from HEAO-1 and *Einstein* do not support such a strong correlation. In the IC model, the radio and X-ray emission would come from identical spatial regions and would therefore have identical distributions on the sky. However, while the X-ray emission is extended and diffuse (Section 4.4), the radio emission comes primarily from individual radio galaxies. Only a very small fraction of clusters appear to have significant diffuse radio halo emission (Section 3.4). The IC model predicts that clusters have power-law X-ray spectra, which is not consistent with the best X-ray spectral observations (Section 2.3.1); one clear distinction is that a power-law provides much more high energy X-ray emission than an exponential thermal spectrum (Section 5.1.3), and such emission is either not seen or very weak. The IC model would not produce any of the X-ray line emission that is seen universally in clusters that have been observed with reasonable sensitivity (Sections 4.3.2 and 4.3.3). The magnetic field is required to be much weaker than would be favored by radio observations. The distorted radio morphologies in clusters would not be explained in this model. Finally, most of the X-ray – optical correlations could not be accounted for naturally in the IC model.

5.1.2 Individual stellar X-ray sources

Katz (1976) suggested that the X-ray emission from clusters was produced by a large number of individual cluster stellar X-ray sources, located either in galactic haloes or in intracluster space. He suggested that these sources might be similar to the binary X-ray sources found in our galaxy, and assumed that the X-ray luminosity to virial mass ratio would be the same for our galaxy and for clusters of galaxies. This predicts too little X-ray emission from clusters (see also Felten *et al.*, 1966); however, globular clusters within our galaxy have a much higher X-ray luminosity to mass ratio, and so they might be used as a model for cluster X-ray sources (Fabian *et al.*, 1976). In fact, M87 and some other central dominant galaxies are known to possess very extensive systems of globular star clusters.

Katz argued that the observed X-ray luminosities of clusters were consistent (in 1976) with a fixed X-ray luminosity to virial mass ratio; the current data would appear to rule out such a correlation (Section 4.6). It is difficult to see how this model could produce the variation in X-ray emission properties of clusters, or the optical – X-ray and radio – X-ray correlations. Of course, if the properties of the individual stellar X-ray sources are not constrained in any manner, any X-ray observations could be explained but the model has no predictive power. The granularity in the cluster X-ray emission produced by these individual sources should have been detected in M87/Virgo, but has not been seen (Schreier *et al.*, 1982). Finally, we note that luminous stellar X-ray sources are generally very optically thick. As a result, they do not show any X-ray line emission except for the fluorescence line of low ionization iron (which is probably produced by reprocessing of X-ray radiation in other parts of the binary star system). In particular, they do not show lines from highly ionized iron (Fe^{+24} and Fe^{+25}) or from lighter elements (for which the fluorescent yields are low). Clusters, on the other hand, do show strong lines from Fe^{+24} and Fe^{+25} , as well as lines from lighter elements (Sections 4.3.2 and 4.3.3).

In short, there is no real evidence in favor of the individual stellar source model, and considerable evidence against it.

5.1.3 Thermal bremsstrahlung from intracluster gas

Felten *et al.* (1966) first suggested that the X-ray emission from clusters (the Coma cluster in particular) was due to diffuse intracluster gas at a temperature of $T_g \approx 10^8 \text{ K}$ and an atomic density of $n \approx 10^{-3} \text{ cm}^{-3}$. (Unfortunately, the early X-ray detection of Coma that they sought to explain was spurious.) At such temperatures and densities, the primary emission process for a gas composed mainly of hydrogen is thermal bremsstrahlung (free–free) emission. The emissivity at a frequency ν of an ion of charge Z in a plasma with an electron temperature T_g is given by

$$\epsilon_{\nu}^{ff} = \frac{2^5 \pi e^6}{3m_e c^3} \left(\frac{2\pi}{3m_e k} \right)^{1/2} Z^2 n_e n_i g_{ff}(Z, T_g, \nu) T_g^{-1/2} \exp(-h\nu/kT_g) \quad (5.11)$$

where n_i and n_e are the number density of ions and electrons, respectively. The emissivity is defined as the emitted energy per unit time, frequency, and volume V ,

$$\epsilon_{\nu} \equiv \frac{dL}{dV d\nu}. \quad (5.12)$$

The Gaunt factor $g_{ff}(Z, T_g, \nu)$ corrects for quantum mechanical effects and for the effect of distant collisions, and is a slowly varying function of frequency and temperature given in Karzas and Latter (1961) and Kellogg *et al.* (1975). If the intracluster gas is mainly at a single temperature, then equation (5.11) indicates that the X-ray spectrum should be close to an exponential of the frequency. In fact, the observed X-ray spectra are generally fit fairly well by this equation (Section 4.3.1), with gas temperatures of 2×10^7 to 10^8 K . This equation predicts that the emission from clusters should fall off rapidly at high frequencies, as is observed.

As first noted by Felten *et al.* (1966), if the intracluster gas either came out of galaxies or fell into the cluster from intergalactic space, it would have a temperature such that the typical atomic velocity was similar to the velocity of galaxies in the cluster. That is,

$$\frac{kT_g}{\mu m_p} \approx \sigma_r^2, \quad (5.13)$$

where μ is the mean molecular weight in amu, m_p is the proton mass, and σ_r is the line-of-sight velocity dispersion of galaxies in the cluster. If the gas came out of galaxies, it would have the same energy per unit mass as the matter in galaxies. If it fell into clusters, it has roughly the same velocity dispersion because it responds to the same gravitational potential as the galaxies. Finally, if the gas were heated so that its initial energy per mass were much greater than that given by equation (5.13), it would not be bound to the cluster and would escape as a wind. Even in this situation, most of the extra energy would go into the kinetic energy of the wind (Section 5.6), and the gas temperature would stay within a factor of 3 or so of equation (5.13). Thus this model predicts

$$T_g \approx 7 \times 10^7 \text{ K} \left(\frac{\sigma_r}{1000 \text{ km/s}} \right)^2 \quad (5.14)$$

for a solar abundance plasma, in reasonable agreement with the temperatures required to explain the X-ray spectra (see equation 4.10).

If the intracluster gas is in hydrostatic equilibrium in the cluster potential, and if the gas temperature is given roughly by equation (5.14), then the spatial distribution of the gas and galaxies will be similar. In fact, the observed X-ray surface brightness of clusters is similar to the projected distribution of galaxies, with the gas distribution being slightly more extended in the inner regions (Section 4.4). The thermal emission model thus explains the extent of the X-ray emission from clusters (more detailed models and comparisons are given in Section 5.5). Moreover, gas at these temperatures and densities has a very long cooling time, and the time for sound waves to cross a cluster is much less than its probable age. These conditions suggest that the gas that the gas distribution will be smooth (at least, the pressure will be smooth), as is observed (Jones and Forman, 1984).

Given the extent of the X-ray emission, the gas temperature derived from the spectrum, and the observed X-ray flux, equation (5.11) gives the atomic number density if the emission is assumed to come mainly from hydrogen. As discussed in Section 4.3, the required mass of gas is less than the total virial mass of the cluster, as required for consistency with this model for the emission. Moreover, the derived densities are very similar to those required to explain the distortions of radio sources in clusters of galaxies (Section 3.3) as the result of ram pressure by intracluster gas.

Whether the intracluster gas came out of galaxies or fell into the cluster, the mass of this gas should increase as the total mass of the cluster increases, and as a result the X-ray luminosity should increase with the virial mass. Solinger and Tucker (1972) first pointed out that the X-ray luminosity (L_x) does appear to increase with the velocity dispersion of clusters, in a way consistent with a constant fraction of the total cluster mass being intracluster gas. They found $L_x \propto \sigma_r^4$, and were able to predict X-ray emission successfully in a number of clusters based on this relationship. As discussed in Section 4.6, there are many other correlations observed between L_x or T_g and optical properties such as richness, velocity dispersion, and projected central galaxy density \bar{N}_o . These correlations are generally consistent with the hypothesis that the mass of intracluster gas increases with the virial mass of the cluster, and that the temperature of the gas increases with the velocity dispersion (depth of the cluster gravitational potential well). There is also an inverse correlation between L_x and the fraction of spirals in the cluster, which may reflect the stripping of spiral galaxies by the intracluster gas. All of these optical – X-ray correlations have natural explanations if the cluster X-ray emission is due to intracluster gas, but not if it is due to the other two mechanism discussed above.

The clearest evidence in favor of the thermal bremsstrahlung model is the detection of strong X-ray line emission from clusters (Sections 4.3.2 and 4.3.3). The strong 7 keV Fe line emission observed from clusters is very difficult to reconcile with any nonthermal model for the origin of the X-ray emission (Mitchell *et al.*, 1976; Serlemitsos *et al.*, 1977; Bahcall and Sarazin, 1977). This emission occurs naturally in the intracluster gas model if the abundances of heavy elements are roughly solar (see Section 5.2.2 below), while nonthermal emission processes would not directly produce any significant line emission (Sections 5.1.1 and 5.1.2). It is possible, of course, that the line emission arises from some different source than the generally distributed cluster X-ray emission; the line observations that have so far been made do not have sufficient spatial resolution to determine the location of the line emission region (Ulmer and Jernigan, 1978). However, the fact that the required *abundance* of iron is roughly constant from cluster to cluster (Section 4.3.2 and Figure 13) argues strongly against such a possibility. The clusters that have been observed differ widely in

their properties, ranging from regular to irregular clusters, high X-ray luminosity to low luminosity clusters, clusters with strong radio sources to clusters without such sources, clusters with hard X-ray spectra to clusters with softer spectra, and so on. In fact, the strengths of the lines also vary widely; it is only the required abundances that are nearly constant. It is very unlikely that two independent sources for the line and continuum would vary simultaneously from cluster to cluster so as to maintain the appearance of a constant Fe abundance. Moreover, if the 7 keV Fe line were from a source that did not contribute significantly to the continuum X-ray emission, the iron abundance in this source would have to be vastly greater than solar. Such large iron abundances are certainly not common in astrophysical systems. Thus it appears very unlikely that the line emission and continuum emission could come from different sources; since the line emission is thermal, the continuum emission must be as well.

The nearly solar iron abundance in the intracluster gas, which has been derived from the 7 keV iron line, suggests that a significant portion of this gas has been processed in and ejected from stars. The total mass of X-ray emitting gas is very large, probably at least as large as the total mass in galaxies in a typical X-ray cluster (see Sec 4.4.1). Because of this large mass, the intracluster gas was initially believed to be primordial gas, which had never been bound in stars or even in galaxies (Gunn and Gott, 1972). During the formation of the universe (the big bang), hydrogen and helium were formed, but it is generally believed that no heavier elements could have been produced because of the lack of any stable isotopes with atomic weights of 5 or 8 (see, for example, Weinberg, 1972). The only sources which has been suggested for the formation of iron involve processing in stars. Moreover, the observed abundance in the intracluster gas is nearly the same as that in the solar system. The solar system is a second generation stellar system (the Sun is a Population I star). Thus it is possible that much of the intracluster gas has been processed through stars.

At the present epoch, there is no significant population of stars known that are not bound to galaxies. This may indicate that part of the intracluster gas may originally have been located within the galaxies. Alternatively, it is possible that there was a ‘pregalactic’ generation of stars (sometimes referred to as Population III stars; see, for example, Carr *et al.*, 1984). Now, the mass of intracluster gas implied by the X-ray observations is nearly as large as the total mass of the galaxies in the cluster (see Section 4.4.1). If the intracluster gas did originate in part in galaxies, there has been a considerable exchange of gas between stars, galaxies, and the intracluster medium. Thus we are led by the X-ray line observations to a much more complicated picture of galaxy formation and evolution than we might have envisioned without this crucial piece of information. Theories of the origin of the intracluster gas and the formation of galaxies and clusters are discussed in more detail in Section 5.10.

5.2 Ionization and X-ray emission from hot, diffuse plasma

The ionization state and X-ray line and continuum emission from a low density ($n \approx 10^{-3} \text{ cm}^{-3}$), hot ($T \approx 10^8 \text{ K}$) plasma will now be discussed. Several simple assumptions will be made. First, the time scale for elastic Coulomb collisions between particles in the plasma is much shorter than the age or cooling time of the plasma, and thus the free particles will be assumed to have a Maxwell–Boltzmann distribution at the temperature T_g (Section 5.4.1). This is the kinetic temperature of electrons, and therefore determines the rates of all excitation and ionization processes. Second, at these low densities collisional excitation and de-excitation processes are much slower than radiative decays, and thus any ionization or excitation process will be assumed to be initiated from the ground state of an ion. Three-body (or more) collisional processes will be ignored because of the low density. Third, the radiation field in a cluster is sufficiently dilute that stimulated radiative transitions are not important, and the effect of the radiation field on the gas is insignificant. Fourth, at these low densities, the gas is optically thin and the transport of the radiation field can therefore be ignored. These assumptions together constitute the ‘coronal limit’. Under these conditions, ionization and emission result primarily from collisions of ions with electrons, and collisions with ions can be ignored. Finally, the time scales for ionization and recombination are generally considerably less than the age of the cluster or any relevant hydrodynamic time scale, and the plasma will therefore be assumed to be in ionization equilibrium.

In nearly all astrophysical plasmas, hydrogen is the most common element and helium is the next commonest, with all the heavier elements being considerably less abundant. For example, this is true of the abundances of elements observed on the surface of the Sun. It is conventional to use these solar abundances as a standard of comparison when studying other astrophysical systems. Since most of the electrons originate in hydrogen and helium atoms, and they are fully ionized under the conditions considered here, the electron

density is nearly independent of the state of ionization and is given by $n_e = 1.21n_p$, where n_p is the density of hydrogen.

5.2.1 Ionization equilibrium

In equilibrium, the ionization state is determined by the balance between processes that produce or destroy each ion:

$$[C(X^i, T_g) + \alpha(X^{i-1}, T_g)] n(X^i) n_e = C(X^{i-1}, T_g) n(X^{i-1}) n_e + \alpha(X^i, T_g) n(X^{i+1}) n_e. \quad (5.15)$$

Here $n(X^i)$ is the number density of the ion X^i (X is the element), T_g is the electron temperature, and $C(X^i, T_g)$ and $\alpha(X^i, T_g)$ are the rate coefficients for collisional ionization out of ion X^i and recombination into ion X^i , respectively.

The collisional ionization rate is the sum of two processes: direct collisional ionization and collisional excitation of inner shell electrons to autoionizing levels which decay to the continuum. This last process is often referred to as autoionization. Recombination is also the sum of two processes, radiative and dielectronic recombination. Recent compilations of ionization and recombination rates and discussions of their accuracy include Mewe and Gronenschild (1981), Shull and Van Steenberg (1982), and Hamilton *et al.* (1983).

The electron density dependence drops out of equation (5.15), and the equilibrium ionization state of a diffuse plasma depends only on the electron temperature. Tables of ionization fractions of various elements are given by Shull and Van Steenberg (1982). Generally, each ionization fraction reaches a maximum at a temperature that is some fraction of its ionization potential. At the temperatures which predominate in clusters, iron is mainly in the fully stripped, hydrogenic, or heliumlike stages.

5.2.2 X-ray emission

The X-ray continuum emission from a hot diffuse plasma is due primarily to three processes, thermal bremsstrahlung (free-free emission), recombination (free-bound) emission, and two-photon decay of metastable levels. The emissivity for thermal bremsstrahlung is given by equation (5.11) above. The radiative recombination (bound-free) continuum emissivity is usually calculated by applying the Milne relation for detailed balance to the photoionization cross sections, which gives (Osterbrock, 1974)

$$\epsilon_\nu^{bf}(X^i) d\nu = n(X^{i+1}) n_e \sum_l \frac{\omega_l(X^i)}{\omega_{gs}(X^{i+1})} a_\nu^l(X^i) \times \frac{h^4 \nu^3}{c^2} \left[\frac{2}{\pi(m_e k T_g)^3} \right]^{1/2} \exp \left\{ - \left[\frac{h\nu - \chi_l(X^i)}{k T_g} \right] \right\} d\nu. \quad (5.16)$$

Here, l sums over all of the energy levels of the ion X^i , gs refers to the ground state of the recombining ion X^{i+1} , ω are the statistical weights of the levels, a_ν^l is the photoionization cross section, and $\chi_l(X^i)$ is the ionization potential for each energy level in the ion.

The two-photon continuum comes from the metastable $2s$ states of hydrogenic and heliumlike ions. These levels are excited by the same processes, discussed below, that excite line emission from less forbidden transitions. For hydrogenic ions, the spectral distribution of two-photon emission is given by Spitzer and Greenstein (1951).

At the high temperatures which predominate in clusters (outside of accretion flows), thermal bremsstrahlung is the predominant X-ray emission process. For solar abundances, the emission is primarily from hydrogen and helium.

Processes that contribute to the X-ray line emission from a diffuse plasma include collisional excitation of valence or inner shell electrons, radiative and dielectronic recombination, inner shell collisional ionization, and radiative cascades following any of these processes. The emissivity due to a collisionally excited line is usually written (Osterbrock, 1974)

$$\int \epsilon_\nu^{line} d\nu = n(X^i) n_e \frac{h^3 \nu \Omega(T_g) B}{4 \omega_{gs}(X^i)} \left[\frac{2}{\pi^3 m_e^3 k T_g} \right]^{1/2} e^{-\Delta E / k T_g}, \quad (5.17)$$

where $h\nu$ is the energy of the transition, ΔE is the excitation energy above the ground state of the excited level, B is the branching ratio for the line (the probability that the upper state decays through this transition), and Ω is the ‘collision strength’, which is often a slowly varying function of temperature. Recent compilations of emissivities for X-ray lines and continua include Kato (1976), Raymond and Smith (1977), Mewe and Gronenschild (1981), Shull (1981), Hamilton *et al.* (1983), and Gaetz and Salpeter (1983). Lines and line ratios that are particularly suited for determining the temperature, ionization state, and elemental abundances in the intracluster gas are described in Bahcall and Sarazin (1978).

Shapiro and Bahcall (1980) and Basko *et al.* (1981) have suggested that X-ray absorption lines due to intracluster gas might be observed in the spectra of background quasars.

5.2.3 Resulting spectra

All of these emission processes give emissivities that increase in proportion with the ion and electron densities, and otherwise depend only on the temperature, so that

$$\epsilon_\nu = \sum_{X,i} \Lambda_\nu(X^i, T_g) n(X^i) n_e, \quad (5.18)$$

where Λ is the emission per ion at unit electron density. If $n(X)$ is the total density of the element X , then in equilibrium the ionization fractions $f(X^i) \equiv n(X^i)/n(X)$ depend only on the temperature, and equation (5.18) becomes

$$\epsilon_\nu = n_p n_e \sum_{X,i} \frac{n(X)}{n(H)} [f(X^i, T_g) \Lambda_\nu(X^i, T_g)]. \quad (5.19)$$

As previously noted (equation 4.3), it is useful to define the emission integral EI as

$$EI \equiv \int n_p n_e dV, \quad (5.20)$$

where V is the volume of the cluster. Then the shape of the spectrum depends only on the abundances of elements $n(X)/n(H)$ and the distribution of temperatures $d(EI)/dT_g$. The normalization of the spectrum (the overall level or luminosity) is set by EI .

Detailed calculations of the X-ray spectra predicted by different models of the intracluster gas have been given by Sarazin and Bahcall (1977) and Bahcall and Sarazin (1977, 1978). Figure (34) gives the X-ray spectrum for isothermal (T_g constant) models for the intracluster gas at a variety of different temperatures, showing the continuum and X-ray emission. The emission integral for these models was taken to be $6.3 \times 10^{-7} \text{ cm}^{-6} \text{ Mpc}^3$.

In these models most of the X-ray emission is thermal bremsstrahlung continuum, and the strongest lines (highest equivalent width) are in the 7 keV iron line complex. This line complex is a blend of K lines from many stages of ionization, although Fe^{+24} and Fe^{+25} predominate at typical cluster temperatures. As noted in Section 4.3.2, the 7 keV iron line is indeed the strongest line feature observed from clusters. Weaker lines at lower energies from lighter elements, such as oxygen, silicon, and sulfur, as well as from L shell transitions in less ionized iron were also predicted to be present in the spectra of clusters, particularly at lower temperatures. Such low energy lines have recently been detected (Section 4.3.3).

Because the line intensities depend on the abundances of heavy elements, while the continuum intensity is mainly due to hydrogen, the line-to-continuum ratio of a line is proportional to the abundance of the element responsible. This ratio is given by the equivalent width (equation 4.4). Figure 35 gives the equivalent width of the iron 7 keV line complex as a function of temperature in a gas with solar abundances. Comparison of these models to the observed strengths of the lines from clusters leads to the determination that the iron abundances are roughly one-half of solar (Section 4.3.2).

The spectral observations of clusters also indicate that in a number of cases the low energy X-ray lines are stronger than would be expected based on these hydrostatic models. This indicates that gas is cooling at the cluster center (Section 4.3.3).

5.3 Heating and cooling of the intracluster gas

In this section, processes that heat or cool the intracluster gas are reviewed. Only processes that affect the total energy of the gas are considered here, while processes (such as heat conduction or mixing) that redistribute the gas energy are discussed in Section 5.4.

5.3.1 Cooling

The primary cooling process for intracluster gas is the emission of radiation by the processes discussed in Section 5.2.2 above. At temperatures $T_g \gtrsim 3 \times 10^7$ K, the main emission mechanism is thermal bremsstrahlung, for which the total emissivity is

$$\begin{aligned}\epsilon^{ff} &= 1.435 \times 10^{-27} \bar{g} T_g^{1/2} n_e \sum_i Z_i^2 n_i \text{ ergs cm}^{-3} \text{ s}^{-1} \\ &\approx 3.0 \times 10^{-27} T_g^{1/2} n_p^2 \text{ ergs cm}^{-3} \text{ s}^{-1},\end{aligned}\quad (5.21)$$

where \bar{g} is the integrated Gaunt factor, and Z_i and n_i are the charge and number density of various ions i . The second equation follows from assuming solar abundances and $\bar{g} = 1.1$ in a fully ionized plasma. For $T_g \lesssim 3 \times 10^7$ K, line cooling becomes very important. Raymond *et al.* (1976) give the cooling rate at lower temperatures; a very crude approximation is (McKee and Cowie, 1977)

$$\epsilon \approx 6.2 \times 10^{-19} T_g^{-0.6} n_p^2 \text{ ergs cm}^{-3} \text{ s}^{-1} \quad 10^5 \text{ K} < T_g < 4 \times 10^7 \text{ K}. \quad (5.22)$$

In assessing the role of cooling in the intracluster gas, it is useful to define a cooling time scale as $t_{cool} \equiv (d \ln T_g / dt)^{-1}$. For the temperatures that apply for the intracluster gas in most clusters, equation (5.21) gives a reasonable approximation to the X-ray emission. If the gas cools isobarically, the cooling time is

$$t_{cool} = 8.5 \times 10^{10} \text{ yr} \left(\frac{n_p}{10^{-3} \text{ cm}^{-3}} \right)^{-1} \left(\frac{T_g}{10^8 \text{ K}} \right)^{1/2}, \quad (5.23)$$

which is longer in most clusters than the Hubble time (age of the universe). Thus cooling is not very important in these cases. However, at the centers of some clusters the cooling time is shorter than the Hubble time, and these clusters are believed to have cooling flows (Section 5.7).

5.3.2 Infall and compressional heating

The heating of the intracluster gas will now be considered. The major point of this discussion is that, although the gas is quite hot, no major ongoing heating of the gas is generally necessary. This is true because the cooling time in the gas is long (equation 5.23), and the thermal energy in the gas is comparable to or less than its gravitational potential energy. Almost any method of introducing the gas into the cluster, either from outside the cluster or from galaxies within the cluster, will heat it to temperatures on the order of those observed.

Heating of the gas due to infall into the cluster and compression will be considered here. First, imagine that the cluster was formed before the intracluster gas fell into the cluster, and that the intracluster gas makes a negligible contribution to the mass of the cluster (Section 4.4). If the gas was initially cold, and located at a large distance from the cluster, then its initial energy can be ignored. If the cluster potential remains fixed while the gas falls into the cluster and the gas neither loses energy by radiation nor exchanges its energy with other components of the cluster, then the total energy of the gas will remain zero. After falling into the cluster, the gas will collide with other elements of gas, and its kinetic energy will be converted to thermal energy. Thus infall and compression can produce temperatures on the order of

$$\frac{3}{2} \frac{k T_g}{\mu m_p} \approx -\phi, \quad (5.24)$$

where ϕ is the gravitational potential in the cluster. At the center of an isothermal cluster, $\phi \approx -9\sigma_r^2$, where σ_r is the line-of-sight velocity dispersion of the cluster. If this is substituted in equation (5.24), the derived temperature is

$$T_g \approx 5 \times 10^8 \text{ K} \left(\frac{\sigma_r}{10^3 \text{ km/s}} \right)^2, \quad (5.25)$$

which is a factor of 5–10 times larger than the observed temperatures (equation 4.10). Of course, the gas that falls into a cluster was presumably bound to the cluster before it fell in, so that equation (5.25) overestimates the temperature. A similar calculation for gas bound to the cluster is given in Shibasaki *et al.* (1976).

The temperature may also be lower because of cooling during the infall, or because the gas fell in at the same time that the cluster was forming and thus experienced a smaller potential on average. If the gas fell in at the same time that the cluster collapsed and was heated by the rapid variation of the potential during violent relaxation (Section 2.9.2), then it might have the same energy per unit mass as the matter in galaxies,

$$\frac{3}{2} \frac{kT_g}{\mu m_p} \approx \frac{3}{2} \sigma_r^2, \quad (5.26)$$

which gives equation (5.14) for the temperature. This is in reasonable agreement with the intracluster gas temperatures determined from X-ray spectra (equation 4.10).

These crude estimates are meant only to illustrate the point that the observed gas temperatures are consistent with heating due to infall into the cluster. More detailed models for infall are discussed in Section 5.10.1.

5.3.3 Heating by ejection from galaxies

The presence of a nearly solar abundance of iron in the intracluster gas (Sections 4.3.2 and 5.2.3) suggests that a reasonable fraction of the gas may have come from stars in galaxies within the clusters. The gas ejected from galaxies is heated in two ways. First, the gas may have some energy when it is ejected. Let ϵ_{ej} be the total energy per unit mass of gas ejected from a galaxy in the rest frame of that galaxy, but not including the cluster gravitational potential, and define $3kT_{ej}/2 \equiv \mu m_p \epsilon_{ej}$. Second, the gas will initially be moving relative to the cluster center of mass at the galaxy's velocity. The ejected gas will collide with intracluster gas and thermalize its kinetic energy. On average this will give a temperature

$$kT_g \approx \mu m_p \sigma_r^2 + kT_{ej}. \quad (5.27)$$

If the ejection energy can be ignored, the temperature is given by equation (5.14), in reasonable agreement with the observations (equation 4.10).

In a steady-state wind outflow from a galaxy, one expects $kT_{ej} \gtrsim \mu m_p \sigma_*^2$, where $\sigma_* \approx 200$ km/s is the velocity dispersion of stars within the galaxy. If the ejection temperature is near the lower limit given by this expression, then this form of heating will not be very important because $\sigma_*^2 \ll \sigma_r^2$. However, the ejection temperature could be considerably higher. For example, supernovae within galaxies could both produce the heavy elements seen in cluster X-ray spectra and heat the gas in galaxies until it was ejected. Supernovae eject highly enriched gas at velocities of $v_{SN} \approx 10^4$ km/s. The highly enriched, rapidly moving supernova ejecta would collide with the interstellar medium in a galaxy and heat the gas. If M_{SN} is the mass of ejecta from a supernova and M_{ej} is the resulting total gas mass ejected from the galaxy, then $T_{ej} \approx 2 \times 10^9$ K($v_{SN}/10^4$ km/s) $^2(M_{SN}/M_{ej})$, which will be significant if the supernova ejecta are diluted by less than a factor of about 100.

5.3.4 Heating by galaxy motions

Although ongoing heating of the intracluster gas may not be necessary to account for the observed gas temperatures, the estimates given above and the history of the gas are sufficiently uncertain that one cannot rule out ongoing heating as an important processes. One way in which intracluster gas could be heated would be through friction between the gas and the galaxies that are constantly moving throughout the cluster (Ruderman and Spiegel, 1971; Hunt, 1971; Yahil and Ostriker, 1973; Schipper, 1974; Livio *et al.*, 1978; Rephaeli and Salpeter, 1980). The calculation of the magnitude of this drag force and of the consequent heating of the intracluster gas is complicated by the following problems. First, the motion of an average cluster galaxy through the intracluster medium is likely to be just transonic $M \approx 1$, where $M \equiv v/c_s$ is the Mach number, v is the galaxy velocity, and $c_s = 1480(T_g/10^8 \text{ K})^{1/2}$ km/s is the sound speed in the gas. If equation (4.10) for the observed gas temperatures is assumed, and the average galaxy velocity is taken to be $\sqrt{3}\sigma_r$, then the average Mach number is $\langle M \rangle \approx 1.5$. Thus the galaxy motion cannot be treated as being either highly supersonic (strong shocks, etc.) or very subsonic (incompressible, etc.). In some cases shocks will be formed by the motion, and in some cases no shocks form. Second, the mean free path λ_i of ions in the intracluster medium due to Coulomb collisions (equation 5.34) is similar to the radius of a galaxy $R_{gal} \approx 20$ kpc (Nulsen, 1982). Thus it is unclear whether the intracluster gas should be treated as a collisionless gas or as a fluid, and the role of transport processes such as viscosity (Section 5.4.4) is uncertain. For example,

the Reynolds number of the flow about an object of radius R is $Re \approx 3(R/\lambda_i)M$ (equation 5.46 below), and thus is somewhat larger than unity if $R = R_{gal}$. It is therefore uncertain whether the flow will be laminar or turbulent. The magnetic field can affect transport processes (Section 5.4.3), but the coherence length of the field l_B estimated from Faraday rotation observations is also comparable to the size of a galaxy (Section 3.6). Finally, the nature of the drag force depends on whether the galaxy contains interstellar gas or not. If the galaxy contains no gas, it affects the intracluster medium only through its gravitational field. If the galaxy contains high density gas, it can give the galaxy an effective surface. For example, a gasless galaxy in supersonic motion probably will not produce a bow shock, while a gas-filled galaxy may (Ruderman and Spiegel, 1971; Hunt, 1971; Gisler, 1976).

It is convenient to write the rate of energy loss by the galaxy and the heating rate of the intracluster medium as

$$\frac{dE}{dt} = \pi R_D^2 \rho_g v^3, \quad (5.28)$$

where ρ_g is the intracluster gas density and R_D is the effective radius of the galaxy for producing the drag force. First, assume that the intracluster gas is collisionless. Then the drag is given by the dynamical friction force of equation (2.34) and

$$R_D^2 = R_A^2 \ln(\Lambda) [\text{erf}(x) - x \cdot \text{erf}'(x)], \quad (5.29)$$

where $R_A \equiv 2Gm/v^2$ is the accretion radius (Ruderman and Spiegel, 1971), m is the galaxy mass, Λ is given by equation (2.29), and $x \equiv \sqrt{5/6}M$. By the virial theorem applied to the galaxy, $R_A \approx R_{gal}(\sigma_*/\sigma_r)^2$, and the accretion radius is typically much smaller than the galaxy radius, since the galaxy velocity dispersion is smaller than that of a cluster. In the limit of hypersonic motion $M \gg 1$, the term in brackets reduces to unity. In this limit, Rephaeli and Salpeter (1980) have shown that equation (5.29) gives the drag force for any value of the mean free path or viscosity if no gas is present in the galaxy. The definition of Λ must be slightly modified (Ruderman and Speigel, 1971).

For a $10^{11} M_\odot$ galaxy moving at 1000 km/s through intracluster gas with a proton number density of 10^{-3} cm^{-3} , the rate of heating of the gas is $\approx 10^{41} \text{ erg/s}$. If the cluster contained 1000 such galaxies, the total heating rate would be $\approx 10^{44} \text{ erg/s}$. While not trivial, this heating rate is too small to heat the intracluster gas in a Hubble time (Schipper, 1974; Rephaeli and Salpeter, 1980). If the total mass of intracluster gas is $\gtrsim 10^{14} M_\odot$ and the gas temperature is $\approx 6 \times 10^7 \text{ K}$, then the time required to heat the gas at this rate would be $\gtrsim 7 \times 10^{11} \text{ yr}$.

The drag force can be considerably increased if the galaxy contains gas or magnetic fields that prevent the penetration of the galaxy by the intracluster gas (Ruderman and Spiegel, 1971; Yahil and Ostriker, 1973; Livio *et al.*, 1978; Shaviv and Salpeter, 1982; Gaetz *et al.*, 1987). However, the main effect of the drag force may be to remove the gas from the galaxy and heat it to the temperature determined by the galaxy kinetic energy per mass (Section 5.3.3). The stripping of gas from a galaxy is shown in Section 5.9 to be very efficient, and as discussed in Sections 2.10.2 and 4.6, galaxies in X-ray clusters are known to be very deficient in gas. It is possible that galaxies may retain a core of gas produced by stellar mass loss within the galaxy. Gaetz *et al.* (1987) give useful analytic fitting formulas for drag coefficients for galaxies with stellar mass loss, based on two-dimensional hydrodynamic simulations. Yahil and Ostriker (1973) and Livio *et al.* (1978) have suggested very large heating rates produced by having high rates of gas output in galaxies. Livio *et al.* assume a rather large cross section for galaxies and a rather low intracluster gas density. Both of these papers argue that the heating rate due to galaxy motions is so large that the intracluster gas is heated beyond the escape temperature from the cluster, and a cluster wind results (see Section 5.6).

There is a simple argument against the great importance of heating due to drag forces from the motions of galaxies. The mass associated with intracluster gas in a typical X-ray cluster is comparable to or greater than the mass associated with galaxies (Section 4.4.1). The average thermal velocities in the gas are comparable to or greater than the typical galaxy velocities (Sections 4.6, 5.3.2, and 5.5). Thus the total thermal energy in the gas is greater than or comparable to the total kinetic energy in the galaxies. It is difficult to believe that the galaxies heated the gas under these circumstances. If they did, then the massive galaxies would have lost most of their initial kinetic energy in the process; this should have produced quite extreme mass segregation, which is not observed (Section 2.7).

5.3.5 Heating by relativistic electrons

Clusters of galaxies often contain radio sources having steep spectral indices (Section 3.1); the radio emission from these sources is believed to arise from synchrotron emission by relativistic electrons. These electrons (particularly the lower energy ones) can interact with the intracluster gas and may heat this gas (Sofia, 1973; Lea and Holman, 1978; Rephaeli, 1979; Scott *et al.*, 1980). A relativistic electron passing through a plasma loses energy through Coulomb interactions with electrons within a Debye length of the particle, and through interactions with plasma waves on larger scales. The rate of loss by an electron with total energy $\gamma m_e c^2$ is

$$-\frac{d\gamma}{dt} = \frac{\omega_p^2 r_e}{c} \left[\ln \left(\frac{m_e c^2 \gamma^{1/2}}{\hbar \omega_p} \right) + 0.2 \right], \quad (5.30)$$

where ω_p is the plasma frequency ($\omega_p^2 \equiv 4\pi n_e e^2 / m_e$), $r_e \equiv e^2 / (m_e c^2)$ is the classical electron radius, and n_e is the electron density in the plasma. The term in square brackets is ≈ 40 for most values of γ and n_e of interest, and does not vary significantly with either parameter. Numerically, the heating rate is $\approx 10^{-18} n_e$ erg/s, ignoring the variation of the term in square brackets. The total heating rate is determined by multiplying this rate per electron by the total number of relativistic electrons in the intracluster gas. If the relativistic electrons have a power-law spectrum (equation 5.2), then the heating rate can be determined from the synchrotron radio emission rate given by equation (5.7). The heating rate is

$$\frac{dE}{dt} = \frac{1.2 \times 10^9 n_e}{a(p)\alpha_r} \left(\frac{dL_r}{d\nu_r} \nu_r^{\alpha_r} \right) \left(\frac{2.4 \times 10^{-7}}{B} \right)^{\alpha_r+1} \gamma_l^{-2\alpha_r} \text{ ergs/s.} \quad (5.31)$$

Here L_r is the radio luminosity at a frequency ν_r , α_r is the radio spectral index, B is the intracluster magnetic field, $a(p)$ is the function given by equations (5.4) and (5.8), and γ_l is the lower limit to the electron spectrum (equation 5.2). Cgs units are to be used for L_r and B . The first quantity in parentheses is independent of the frequency ν_r at which the radio source is observed. Equation (5.31) includes only electrons and should be increased to include the heating by ions, which do not produce observable radio emission. While the radio flux and spectrum can be measured directly, the magnetic field strength must be estimated from the radio observations (Section 3.6), and the lower limit to the electron spectrum is generally unknown.

This heating rate is significant only for radio sources with steep spectra which extend to very, very low frequencies. Lea and Holman (1978) used low frequency radio observations of clusters to determine the radio flux and spectral index, and found that the electron spectrum must extend down to $\gamma_l \approx 10(B/\mu\text{G})^{-1}$ if the heating rate is to be comparable to the X-ray luminosity of the cluster. These low energy electrons would produce radio emission at a frequency of about 400 Hz, which is about 10^5 times too low to be observed. Extrapolating the radio spectrum from the lowest observed frequencies (≈ 26 MHz) down to these low frequencies increases the total number of relativistic electrons and the corresponding heating rate by about 10^5 . Thus the hypothesis that relativistic electrons provide significant heating to the intracluster gas according to equation (5.31) requires an enormous and untestable extrapolation of cluster radio properties.

Several authors have argued that the heating rate of equation (5.31) should be increased by collective plasma interactions between the relativistic electrons and the intracluster gas (Lea and Holman, 1978; Rephaeli, 1979; Scott *et al.*, 1980). In these models, it is assumed that the relativistic electrons are streaming away from a powerful radio source at the center of the cluster. Rephaeli (1979) assumed that the streaming speed is limited by the Alfvén velocity, as has generally been argued (Jaffe, 1977; Section 3.4). Then he finds that the relativistic electrons excite Alfvén waves and lose energy at a rate of $-d\gamma/dt \approx v_A \gamma / L_e$, where L_e is the scale length of the relativistic electron distribution. For $B \approx 1 \mu\text{G}$, $L_e \approx 1 \text{ Mpc}$, and $n_e \approx 10^{-3} \text{ cm}^{-3}$, this gives a heating rate about $10^{-2} \gamma_l$ of that in equation (5.31), which is never very significant.

Alternatively, Scott *et al.* (1980) have assumed that the relativistic electrons stream at nearly the speed of light (Holman *et al.*, 1979). As discussed in Section 3.4, this hypothesis is controversial. They discuss a number of plasma instabilities that greatly increase the heating rate under these circumstances. The increase could be as much as a factor of 10^5 , which would allow the electrons that produce the observed radio emission in clusters to heat the intracluster gas.

Models in which relativistic electrons heat the intracluster gas suffer from two general problems. First, the total energy requirements of 10^{63-64} erg are extreme for a single radio source, although it is possible that many cluster sources contribute to the heating over the lifetime of the cluster. Second, the radio sources generally occupy only a small fraction of the cluster; cluster-wide radio haloes are rare (Section 3.4). It is

difficult to see how several discrete radio sources would heat the intracluster gas without producing very strong, observable variations in the X-ray surface brightness, which are not seen.

Vestrand (1982) has suggested that heating of the intracluster gas by relativistic electrons is important only in clusters having radio haloes. He noted that Coma, which has a prominent halo, also has an unusually high gas temperature for its velocity dispersion (Section 4.5.1).

5.4 Transport processes

Processes that redistribute energy, momentum, or heavy elements within the intracluster gas will now be reviewed.

5.4.1 Mean free paths and equilibration time scales

The mean free paths of electrons and ions in a plasma without a magnetic field are determined by Coulomb collisions. As in the stellar dynamical case, it is important to include distant as well as nearby collisions. The mean free path λ_e for an electron to suffer an energy exchanging collision with another electron is given by (Spitzer, 1956)

$$\lambda_e = \frac{3^{3/2} (kT_e)^2}{4\pi^{1/2} n_e e^4 \ln \Lambda}, \quad (5.32)$$

where T_e is the electron temperature, n_e is the electron number density, and Λ is the ratio of largest to smallest impact parameters for the collisions. For $T_e \gtrsim 4 \times 10^5$ K, this Coulomb logarithm is

$$\ln \Lambda = 37.8 + \ln \left[\left(\frac{T_e}{10^8 \text{ K}} \right) \left(\frac{n_e}{10^{-3} \text{ cm}^{-3}} \right)^{-1/2} \right], \quad (5.33)$$

which is nearly independent of density or temperature. Equation (5.32) assumes that the electrons have a Maxwellian velocity distribution at the electron temperature. The equivalent mean free path of ions λ_i is given by the same formula, replacing the electron temperature and density with the ion temperature T_i and density, dividing by the ion charge to the fourth power, and slightly increasing $\ln \Lambda$. In the discussion that follows the ions will generally be assumed to be protons, and the diffusion of heavy elements will be discussed in Section 5.4.5 below. Numerically,

$$\lambda_e = \lambda_i \approx 23 \text{ kpc} \left(\frac{T_g}{10^8 \text{ K}} \right)^2 \left(\frac{n_e}{10^{-3} \text{ cm}^{-3}} \right)^{-1} \quad (5.34)$$

assuming that $T_e = T_i = T_g$.

In general, these mean free paths are shorter than the length scales of interest in clusters (≈ 1 Mpc), and the intracluster medium can be treated as a collisional fluid, satisfying the hydrodynamic equations. Note that the mean free paths are comparable to the size of a galaxy, however, and in the interaction between intracluster gas and individual galaxies the gas may be nearly collisionless, as mentioned previously (Section 5.3.4).

If a homogeneous plasma is created in a state in which the particle distribution is non-Maxwellian, elastic collisions will cause it to relax to a Maxwellian distribution on a time scale determined by the mean free paths (Spitzer, 1956, 1978). Electrons will achieve this equilibration (isotropic Maxwellian velocity distribution characterized by the electron temperature) on a time scale set roughly by $t_{eq}(e, e) \equiv \lambda_e / \langle v_e \rangle_{rms}$, where the denominator is the rms electron velocity,

$$t_{eq}(e, e) = \frac{3m_e^{1/2} (kT_e)^{3/2}}{4\pi^{1/2} n_e e^4 \ln \Lambda} \approx 3.3 \times 10^5 \text{ yr} \left(\frac{T_e}{10^8 \text{ K}} \right)^{3/2} \left(\frac{n_e}{10^{-3} \text{ cm}^{-3}} \right)^{-1}. \quad (5.35)$$

The time scale for protons to equilibrate among themselves is $t_{eq}(p, p) \approx (m_p/m_e)^{1/2} t_{eq}(e, e)$, or roughly 43 times longer than the value in equation (5.35). Following this time, the protons and ions would each have Maxwellian distributions, but generally at different temperatures. The time scale for the electrons and ions to reach equipartition $T_e = T_i$ is $t_{eq}(p, e) \approx (m_p/m_e) t_{eq}(e, e)$, or roughly 1870 times the value in equation

(5.35). For heavier ions, the time scales for equilibration are generally at least this short if the ions are nearly fully stripped, because the increased charge more than makes up for the increased mass. For $T_g \approx 10^8$ K and $n_e \approx 10^{-3}$ cm $^{-3}$, the longest equilibration time scale is only $t_{eq}(p, e) \approx 6 \times 10^8$ yr. Since this is shorter than the age of the cluster or the cooling time, the intracluster plasma can generally be characterized by a single kinetic temperature T_g . Under some circumstances, plasma instabilities may bring about a more rapid equilibration than collisions (McKee and Cowie, 1977).

5.4.2 Thermal conduction

In a plasma with a gradient in the electron temperature, heat is conducted down the temperature gradient. If the scale length of the temperature gradient $l_T \equiv T_e / |\nabla T_e|$ is much longer than the mean free path of electrons λ_e , then the heat flux is given by

$$\mathbf{Q} = -\kappa \nabla T_e, \quad (5.36)$$

where the thermal conductivity for a hydrogen plasma is (Spitzer, 1956)

$$\begin{aligned} \kappa &= 1.31 n_e \lambda_e k \left(\frac{kT_e}{m_e} \right)^{1/2} \\ &\approx 4.6 \times 10^{13} \left(\frac{T_e}{10^8 \text{ K}} \right)^{5/2} \left(\frac{\ln \Lambda}{40} \right)^{-1} \text{ ergs s}^{-1} \text{ cm}^{-1} \text{ K}^{-1}. \end{aligned} \quad (5.37)$$

Because of the inverse dependence on the particle mass, thermal conduction is primarily due to electrons. This equation includes a correction for the self-consistent electric field set up by the diffusing electrons. If the very weak dependence of $\ln \Lambda$ on density is ignored, then κ is independent of density but depends very strongly on temperature.

If the scale length of the thermal gradient l_T is comparable to or less than the mean free path of electrons, then equation (5.36) overestimates the heat flux, since it would imply that the electrons are diffusing at a speed greater than their average thermal speed. Under these circumstances the conduction is said to ‘saturate’, and the heat flux approaches a limiting value Q_{sat} . Cowie and McKee (1977) calculate this saturated heat flux by assuming that the electrons have a Maxwellian distribution, and an infinitely steep temperature gradient, and that the correction for a self-consistent electric field is the same as in the unsaturated case. They find

$$Q_{sat} = 0.4 n_e k T_e \left(\frac{2kT_e}{\pi m_e} \right)^{1/2}. \quad (5.38)$$

A general expression for the heat flux, which interpolates between the two limits of equations (5.36) and (5.38), is then

$$\mathbf{Q} \approx -\frac{\kappa T_e}{l_T + 4.2 \lambda_e} \frac{\nabla T_e}{|\nabla T_e|}. \quad (5.39)$$

The mean free path of electrons in the intracluster gas (equation 5.34) is typically small compared to the cluster dimensions, and heat conduction within the intracluster gas itself is probably unsaturated. However, the mean free path is comparable to the size of a galaxy, and saturated heat conduction may be important in evaporation from or accretion to galaxies (Sections 5.7 and 5.9).

Within the intracluster medium, thermal conduction will act to transport heat from hot to cold regions and, in the absence of any competing effect, to make the temperature spatially constant (isothermal). Assuming equal ion and electron temperatures, the temperature in a Lagrangian element of the intracluster gas will vary as

$$\frac{3}{2} \frac{\rho_g k}{\mu m_p} \frac{dT_g}{dt} - \frac{kT_g}{\mu m_p} \frac{d\rho_g}{dt} = -\nabla \cdot \mathbf{Q}, \quad (5.40)$$

where ρ_g is the gas density. It is useful to define a conduction time scale as $t_{cond} \equiv -(d \ln T_e / dt)^{-1}$, which is on the order of $|t_{cond}| \approx (n_e l_T^2 k) / \kappa$. As a specific example, consider a cluster in which the gas is hydrostatic, adiabatic (isentropic), and extends to very large distances but is not in contact with any intercluster gas; such models are discussed in some detail in Section 5.5.2 below. If the cluster potential is given by the

analytic King form (equation 5.59 below), the gas is assumed to cool isobarically (at constant pressure), and the variation of $\ln \Lambda$ is ignored, then the conduction time at a radius r is given by

$$\frac{1}{t_{\text{cond}}} = \frac{2\mu m_p \kappa_o}{5\rho_{g,o} r_c^2 k} g(r/r_c), \quad (5.41)$$

where κ_o and $\rho_{g,o}$ are the conductivity and gas density at the cluster center, and r_c is the cluster core radius. The function $g(x)$ is

$$g(x) \equiv (x^2 + 1)^{-3/2} - \frac{5}{2x^2 f(x)} \left[f(x) - (x^2 + 1)^{-1/2} \right]^2, \quad (5.42)$$

where $f(x) \equiv \phi(r)/\phi_o$ is the ratio of the cluster potential to its central value and is given by equation (5.59) below. The function $g(x)$ is plotted in Figure 36. Because conduction only transports heat, the average temperature of the gas is not changed; in the inner parts the gas is cooled and in the outer parts the gas is heated. However, because the X-ray emission is proportional to the square of the density, temperatures determined from X-ray spectra are mainly affected by the innermost gas and are lowered by conduction. As is clear from Figure 36, heat conduction is most effective in the cluster core, and $|t_{\text{cond}}|$ increases very rapidly with radius. Since $g(0) = 1$, the central value to the conduction time scale $t_{\text{cond}}(0)$ is given by the first term in equation (5.41), or

$$t_{\text{cond}}(0) = 3.3 \times 10^8 \text{ yr} \left(\frac{n_o}{10^{-3} \text{ cm}^{-3}} \right) \left(\frac{T_e}{10^8 \text{ K}} \right)^{-5/2} \left(\frac{r_c}{0.25 \text{ Mpc}} \right)^2 \left(\frac{\ln \Lambda}{40} \right), \quad (5.43)$$

where n_o is the central proton density, and solar abundances have been assumed. Thus heat conduction may be relatively effective in the core of a cluster. At radii $r \gtrsim 2r_c$, the conduction time is typically a factor of ≈ 100 longer, and conduction is only marginally effective in the outer parts of the cluster. The conduction time may be increased further by the presence of a magnetic field in the cluster.

5.4.3 Effects of the magnetic field

Charged particles gyrate around magnetic field lines on orbits with a radius (the gyroradius) of $r_g = (mv_\perp c/ZeB)$, where m is the particle mass, v_\perp is the component of its velocity perpendicular to the magnetic field, Z is the particle charge, and B is the magnetic field strength. If $v_\perp = \sqrt{2kT_g/m}$, which is the rms value in a thermal plasma, then

$$r_g = \frac{3.1 \times 10^8 \text{ cm}}{Z} \left(\frac{T_g}{10^8 \text{ K}} \right)^{1/2} \left(\frac{m}{m_e} \right)^{1/2} \left(\frac{B}{1 \mu\text{G}} \right)^{-1} \quad (5.44)$$

which is much smaller than any length scale of interest in clusters, and is also much smaller than the mean free path of particles due to collisions $r_g \ll \lambda_e$. Then, the effective mean free path for diffusion perpendicular to the magnetic field is only on the order of r_g^2/λ_e (Spitzer, 1956). Because they have larger gyroradii, the ions are most effective in transport processes perpendicular to the magnetic field. In practice, the gyroradii are so small that diffusion perpendicular to the magnetic field can be ignored in the intracluster gas.

Consider the effect of the magnetic field on thermal conduction, when the temperature gradient lies at an angle θ to the local magnetic field direction. Only the component of the gradient parallel to the field is effective in driving a heat flux, and only the component of the resulting heat flux in the direction of the temperature gradient transports any net energy. If the conduction is unsaturated, the heat flux parallel to the thermal gradient is thus reduced by a factor of $\cos^2 \theta$. If the conduction is saturated, the heat flux is independent of the temperature gradient; the appropriate factor is just $\cos \theta$ (Cowie and McKee, 1977).

Observations suggest that the magnetic field in clusters may be tangled, so that the direction θ varies throughout the intracluster gas (Section 3.6). Let l_B be the coherence length of the magnetic field, so that the field will typically have changed direction by $\approx 90^\circ$ over this distance. Let l_T be the temperature scale height (Section 5.4.2 above) and let λ_e be the mean free path of electrons. Consider first the case where $\lambda_e \ll l_T$, so that the conduction is unsaturated. Then, if $l_B \gtrsim l_T$, the magnetic field is ordered over the scales of interest in the cluster, and the value of $\cos^2 \theta$ depends on the geometry of the magnetic field. For example, for a cluster with a radial temperature gradient and a circumferential magnetic field, thermal conduction

would be suppressed. Alternatively, if $l_T \gg l_B \gg \lambda_e$, then the field direction can be treated as a random variable, and the heat flux is reduced by $\approx \langle \cos^2 \theta \rangle = 1/3$. If the coherence length of the magnetic field is less than the mean free path $l_B \ll \lambda_e$, the conductivity depends on the topology of the magnetic field (i.e., whether it is connected over distances greater than l_B). In general, the effective mean free path for diffusion will always be at least as small as l_B , and could be as small as l_B^2/λ_e if the magnetic field is disconnected on the scale of l_B . In this limit, the thermal conduction would be very significantly reduced.

The Faraday rotation observations (Section 3.6) suggest that $l_B \approx \lambda_e \approx 20$ kpc. Thus thermal conduction will probably be reduced by a factor of at least 1/3, and the conductivity time scale (equations 5.41 and 5.43) should be increased by at least this factor.

5.4.4 Viscosity

If there are shears in the velocity in a fluid, then viscosity will produce forces that act against these shears. A unit volume of the fluid will be subjected to a force given by

$$\mathbf{F}_{vis} = \eta \left(\nabla^2 \mathbf{v} + \frac{1}{3} \nabla \nabla \cdot \mathbf{v} \right), \quad (5.45)$$

where η is the dynamic viscosity, and the bulk viscosity has been assumed to be zero. For an ionized plasma without a magnetic field, η is given by

$$\begin{aligned} \eta &\approx \frac{1}{3} m_i n_i \langle v_i \rangle_{rms} \lambda_i \\ &\approx 5500 \text{ gm cm}^{-1} \text{ s}^{-1} \left(\frac{T_e}{10^8 \text{ K}} \right)^{5/2} \left(\frac{\ln \Lambda}{40} \right)^{-1}, \end{aligned} \quad (5.46)$$

where m_i , n_i , $\langle v_i \rangle_{rms}$, and λ_i are the mass, number density, rms velocity, and mean free path of ions, respectively. Like the thermal conductivity, the dynamic viscosity is independent of density and depends strongly on the temperature. However, because of the dependence on the particle mass, the viscosity is primarily due to ions, not electrons. The Reynolds number for flow at a speed v past an object of size l is defined as $Re \equiv \rho_g v l / \eta$, which can be written as

$$Re \approx 3M \left(\frac{l}{\lambda_i} \right), \quad (5.47)$$

where $M \equiv v/c_s$ is the Mach number and c_s is the sound speed. As noted in Section 5.3.4, this indicates that the flow around moving galaxies in a cluster is probably laminar, but not certainly so. The viscosity affects the rate of heating of the intracluster gas by galaxy motions (Section 5.3.4), and the rate of stripping by interstellar gas in cluster galaxies (Section 5.9).

As with thermal conduction, one can define a velocity scale length l_v so that the absolute value of the term in parentheses in equation (5.45) is $\langle v_i \rangle_{rms} / l_v^2$. Then, if the ion mean free path λ_i is shorter than l_v , equation (5.45) applies. However, if $l_v < \lambda_i$, then equation (5.45) requires that the viscous stresses exceed the ion pressure and that momentum be transported faster than the thermal speed of the ions. This is not possible, and the viscous stresses must saturate at a value comparable to the ion pressure. To my knowledge, this effect has not been included in calculations of astrophysical fluid flows. It should be particularly important in flows around galaxies, where previous calculations, particularly those involving Kelvin–Helmholtz instabilities, have applied equation (5.45) to flows with very large shears.

5.4.5 Diffusion and settling of heavy ions

At the same temperature, heavy ions will move more slowly than light ions, and all ions will move more slowly than electrons. As a result, the heavy ions will tend to settle towards the center of the cluster. This settling is generally halted when it results in an electric field large enough to balance the extra gravitational force on the heavier ions. Models for the distribution of heavy elements in clusters, assuming that settling has occurred rapidly enough to reach this equilibrium, are discussed in Section 5.5.6 below. Here, the rate of settling will be considered.

Heavy ions that diffuse toward the cluster center rapidly reach a drift velocity at which gravitational acceleration is balanced by the slowing effect of collisions (Fabian and Pringle, 1977; Rephaeli, 1978). Using

the expression for the mean free path of a slow moving heavy ion in a plasma (Spitzer, 1956), we obtain the drift velocity \mathbf{v}_D ,

$$Am_p \mathbf{g} = \mathbf{v}_D \frac{16\sqrt{\pi}Z^2 e^4 m_p^{1/2}}{3(2kT_g)^{3/2}} \sum_i Z_i^2 A_i^{1/2} n_i \ln \Lambda_i, \quad (5.48)$$

where A and Z are the heavy ion atomic number and charge, \mathbf{g} is the gravitational acceleration, T_g is the gas temperature, and the sum is over the various ions in the plasma. Assuming cosmic abundances, one finds

$$\begin{aligned} v_D &= 2.9 \left(\frac{A}{56} \right) \left(\frac{Z}{26} \right)^{-2} \left(\frac{|g|}{3 \times 10^{-8} \text{ cm s}^{-2}} \right) \left(\frac{T_g}{10^8 \text{ K}} \right)^{3/2} \\ &\times \left(\frac{n_p}{10^{-3} \text{ cm}^{-3}} \right)^{-1} \left(\frac{\ln \Lambda}{40} \right)^{-1} \text{ km/s}. \end{aligned} \quad (5.49)$$

The values of A and Z in equation (5.49) are those for fully ionized iron. The X-ray lines of iron are the strongest lines observed from clusters, and have been detected in a large number of clusters (Section 4.3.2). These lines play an important role in considerations of the origin and energetics of the intracluster gas (Section 5.10), and thus the distribution of iron is particularly important. As noted by Rephaeli (1978), the value of the drift velocity in equation (5.49) is considerably smaller than the value given by Fabian and Pringle (1977) because they failed to include helium and heavier elements in the slowing time.

According to equation (5.49), the drift velocity depends on the cluster gravitational potential and the density and temperature of the intracluster gas. If the cluster potential is given by the King analytic approximation to the isothermal distribution (equations 2.13, 5.59), then the largest value of g occurs at roughly the core radius r_c (actually $r = 1.027r_c$), and

$$g \leq 2.0 \times 10^{-8} \left(\frac{\sigma_r}{10^3 \text{ km/s}} \right)^2 \left(\frac{r_c}{0.25 \text{ Mpc}} \right)^{-1} \text{ cm s}^{-2}, \quad (5.50)$$

where σ_r is the radial velocity dispersion of the cluster. The time to drift in a small distance dr is then $dt = dr/v_D$. If one sets $dr \approx r_c$, then the drift time is

$$\begin{aligned} t_D &\approx 1.2 \times 10^{11} \text{ yr} \left(\frac{A}{56} \right)^{-1} \left(\frac{Z}{26} \right)^2 \left(\frac{T_g}{10^8 \text{ K}} \right)^{-3/2} \left(\frac{n_o}{10^{-3} \text{ cm}^{-3}} \right) \left(\frac{\ln \Lambda}{40} \right) \\ &\times \left(\frac{\sigma_r}{10^3 \text{ km/s}} \right)^{-2} \left(\frac{r_c}{0.25 \text{ Mpc}} \right)^2. \end{aligned} \quad (5.51)$$

If the intracluster gas is adiabatic (constant entropy per particle; Section 5.5), then $n_p \propto T_g^{3/2}$, and the variation of these two quantities does not affect the drift time. If the gas is isothermal, then the reduction of density with distance from the cluster center shortens the drift time, although the increase in the distance to be traversed and the decrease in the gravitational acceleration more than make up for this. Rephaeli (1978) has integrated the drift time for a variety of equations of state for the gas, and finds that the drift times from radii greater than four core radii in to one core radius is always greater than 10^{11} yr times the cluster parameters in equation (5.51), with the gas density and temperature evaluated at the cluster center. Since this means the drift time is generally at least ten times the probable age of the cluster, the settling of iron into the cluster center is unlikely to occur.

As noted by Rephaeli (1978), the presence of a magnetic field in the intracluster gas would inhibit the sedimentation of iron even further. The situation is similar to that for thermal conduction (Section 5.4.3); iron ions could only drift along the direction of magnetic field lines. Unless the magnetic field were predominately radial, settling of iron would be very strongly inhibited. The effect is even larger than that on thermal conduction, because the sedimentation of iron ions requires that the individual ions drift over large distances, while heat can be conducted from the inner to the outer portions of a cluster without having an individual electron traverse these distances. For example, loop structures in the magnetic field would tend to prevent any sedimentation beyond the bottom of the loop.

Although it appears unlikely that heavy elements like iron would have settled to the cluster center, this does not mean that the abundance of heavy elements is necessarily uniform. Heavy elements might be concentrated in the cluster core if they were injected into the intracluster gas in the core. If the heavy elements were produced in stars within galaxies, as is usually assumed (Sections 4.3.2 and 5.10), then this gas might have been predominantly deposited in the cluster core because either ram pressure or collisional stripping will occur more easily there. Of course, if all the intracluster gas were produced by this mechanism, then no abundance gradients would be produced if all galaxies in the cluster contained gas with the same average abundances. If, however, the intracluster medium were partly due to heavy element containing gas stripped from galaxies, and partly due to heavy element lacking intergalactic gas that fell into the cluster, and if the stripping were concentrated at the cluster core, then a gradient in the abundance of heavy elements could result (Nepveu, 1981b). Subsequent mixing of the intracluster gas (Section 5.4.6 below) might still destroy such abundance variations.

Such an abundance gradient could greatly reduce the amount of iron needed to produce the observed iron lines. If the iron were concentrated at the cluster core, it would all be located in regions of high electron density, so that the iron line emissivity per iron ion would be higher than the average emissivity per hydrogen ion (Section 5.2). The amount of iron needed to produce the observed iron lines could be reduced by as much as a factor of 20 (Abramopoulos *et al.*, 1981). Since the production of a nearly solar abundance of iron is one of the major problems with galaxy formation in clusters and the origin of the intracluster gas, the possibility that the real abundance might be much lower is very important. Unfortunately, all of the available spectra of the 7 keV iron line in clusters have been made with low spatial resolution detectors, which do not provide sufficient information on the location of the iron. The observation of spatially resolved spectra of clusters including the 7 keV iron line are critically important to our understanding of clusters, and will be possible if the AXAF satellite is launched (Chapter 6).

5.4.6 Convection and mixing

The diffusive transport processes discussed so far tend to cause the gas in a cluster to be isothermal, tend to damp out fluid motions, and tend to cause heavy elements to settle to the cluster center. On the other hand, mixing processes due to turbulent motions of the intracluster gas tend to make the specific entropy (the entropy per atom) equal within the cluster, tend to make the composition of the gas homogeneous, and drive fluid motions. It is conventional to describe a gas in which the entropy per atom is constant as an ‘adiabatic’ gas. If mixing occurs on a time scale that is rapid compared to the age of the cluster or the time scale for diffusive transport, the resulting intracluster gas distribution will tend to be adiabatic and to have constant heavy element abundances.

One possible source of mixing motions in the gas is convection. If the intracluster gas were hydrostatic but had a steep temperature gradient

$$-\frac{d \ln T_g}{dr} > -\frac{2}{3} \frac{d \ln n_p}{dr}, \quad (5.52)$$

it would be unstable to convective mixing. If the temperature gradient in equation (5.52) were exceeded by a significant amount, mixing would occur within several sound crossing times in the cluster. Since this is a rather short time (equation 5.54 below), it is reasonable to assume that the temperature gradient is smaller than that in equation (5.52).

Motions of galaxies through the intracluster gas may mix the gas. This process has not been treated in any great detail in the literature. It might be reasonable to assume that each galaxy mixes the gas within a wake of radius R_W . Very roughly, the gas within the whole cluster would be mixed on a time scale given by

$$t_{mix} \approx \frac{1}{n_{gal} \sigma_r \pi R_W^2} \left(\frac{r_c}{R_W} \right)^2, \quad (5.53)$$

where n_{gal} is the number density of galaxies, σ_r is their velocity dispersion, and r_c is the cluster core radius. Depending on whether the galaxy contained any interstellar medium or not, the wake radius might be ≈ 10 kpc, or only as large as the accretion radius (Section 5.3.4). The resulting value of t_{mix} is generally much longer than 10^{11} yr for any reasonable values of the cluster parameters, and mixing due to galaxy motions is probably not very important. Nepveu (1981b) discussed the mixing of gas ejected from cluster galaxies and showed that the mixing was not effective unless the galaxy motions were highly subsonic, which is not the case. His calculations indicated that the gas remained inhomogeneous on both small and large scales.

Large scale hydrodynamic motions during the formation of the cluster may be effective in mixing the intracluster gas. Similarly, mixing may occur when subclusters merge within the cluster.

5.5 Distribution of the intracluster gas-hydrostatic models

In general, the elastic collision times for ions and electrons (equations 5.32, 5.33) in the intracluster gas are much shorter than the time scales for heating or cooling (Section 5.3) or any dynamical process, and the gas can be treated as a fluid. The time required for a sound wave in the intracluster gas to cross a cluster is given by

$$t_s \approx 6.6 \times 10^8 \text{ yr} \left(\frac{T_g}{10^8 \text{ K}} \right)^{-1/2} \left(\frac{D}{\text{Mpc}} \right), \quad (5.54)$$

where T_g is the gas temperature and D is the cluster diameter. Since this time is short compared to the probable age of a cluster of 10^{10} yr, the gas will be hydrostatic and the pressure will be a smoothly varying function of position unless the cluster gravitational potential varies on a shorter time scale or the gas is heated or cooled more rapidly than this. The cooling time due to thermal bremsstrahlung (equation 5.23) is much longer than the sound crossing time, and the same is true for the time scales for heating by any of the processes that have been suggested (Section 5.3). Thus the gas distribution is usually assumed to be hydrostatic:

$$\nabla P = -\rho_g \nabla \phi(r), \quad (5.55)$$

where $P = \rho_g k T_g / \mu m_p$ is the gas pressure, ρ_g is the gas density, and $\phi(r)$ is the gravitational potential of the cluster. Now, the sound crossing time condition given above ensures that the pressure in the gas be uniform on small scales, but the same need not be completely true of the entropy. In the rest of this section, we will assume that the intracluster gas is locally homogeneous.

If, in addition, the cluster is assumed to be spherically symmetric, equation (5.55) reduces to

$$\frac{1}{\rho_g} \frac{dP}{dr} = -\frac{d\phi}{dr} = -\frac{GM(r)}{r^2}, \quad (5.56)$$

where r is the radius from the cluster center and $M(r)$ is the total cluster mass within r . If the contribution of the intracluster gas to the gravitational potential is ignored, then the distribution of the intracluster gas is determined by the cluster potential $\phi(r)$ and the temperature distribution of the gas $T_g(r)$. Further, then equation (5.56) is a linear equation for the logarithm of the gas density; the central density (or any other value) can be specified in order to determine the full run of densities. If self gravity is included, then the density scale, temperature variation, and cluster potential cannot be given independently.

In most models, the cluster potential is assumed to be that of a self-gravitating isothermal sphere (Section 2.7). For convenience, the analytic King approximation to the isothermal sphere is often assumed (equation 2.13), so that the cluster total density, mass, and potential are given by

$$\rho(r) = \rho_o (1 + x^2)^{-3/2}, \quad (5.57)$$

$$M(r) = 4\pi \rho_o r_c^3 \left\{ \ln \left[x + (1 + x^2)^{1/2} \right] - x(1 + x^2)^{-1/2} \right\}, \quad (5.58)$$

$$\begin{aligned} \phi(r) &= -4\pi G \rho_o r_c^2 \frac{\ln [x + (1 + x^2)^{1/2}]}{x}, \\ x &\equiv \frac{r}{r_c}. \end{aligned} \quad (5.59)$$

Here, ρ_o is the central density and r_c is the core radius of the cluster (Section 2.7). The central density and core radius are related to the line-of-sight velocity dispersion σ_r (Section 2.6) of an isothermal cluster by

$$\sigma_r^2 = \frac{4\pi G \rho_o r_c^2}{9}, \quad (5.60)$$

which follows from equation (2.9). Thus the central value of the cluster gravitational potential is

$$\phi_o = -9\sigma_r^2. \quad (5.61)$$

It is perhaps worth noting that although the analytic King model is a good fit to the inner portions of an isothermal sphere, the analytic King potential has a finite depth (given by equation 5.61), while the isothermal sphere potential is infinitely deep. The total mass associated with the analytic King distribution diverges as the radius increases, and it is usual to cut the distribution off in some way (equations 2.11 and 2.12). The effect of these cutoffs on the cluster potential and on the resulting hydrostatic gas models is discussed in Sarazin and Bahcall (1977) and Bahcall and Sarazin (1978).

In calculating the cluster potential, we have ignored the effect of individual galaxies. The potential wells associated with individual galaxies are considerably shallower than those associated with clusters. This is shown by the fact that the velocity dispersions of stars in galaxies ($\sigma_* \approx 300$ km/s; Faber and Gallagher, 1979) are much smaller than the velocity dispersions of galaxies in clusters ($\sigma_r \approx 1000$ km/s); the gravitational potentials increase with the square of the velocity dispersions. Thus an individual galaxy will not significantly perturb the distribution of intracluster gas (Gull and Northover, 1975).

5.5.1 Isothermal distributions

The simplest distribution of gas temperatures would be an isothermal distribution, with T_g being constant. The intracluster gas would become isothermal if thermal conduction were sufficiently rapid (see equation 5.41 for the relevant time scale). Alternatively, the gas may have been introduced into the cluster with an approximately constant temperature, and its thermal distribution be unchanged since that time. Lea *et al.* (1973) fit the gas distributions in the Coma, Perseus, and M87/Virgo clusters, assuming that the gas distribution were self-gravitating and isothermal. If $\phi(r)$ is due only to ρ_g , then equation (5.56) and Poisson's equation for the gravitational potential are equivalent to equation (2.9), the equation of an isothermal sphere. Note, however, that this is not trivially true, since equation (2.9) was derived from the stellar dynamical equation for a collisionless gas, while equation (5.56) is the hydrostatic equation for a collisionally dominated fluid. Lea *et al.* therefore used equation (5.57) to fit the gas density ρ_g . This is not consistent, since the gas masses derived from these fits were generally less than 20% of the virial mass of the cluster, and the core radii derived were larger than those of the galaxies (Section 4.4).

More consistent isothermal models can be derived if the gravitational potential of the cluster is not assumed to come only from the gas (Cavaliere and Fusco-Femiano, 1976, 1978; Sarazin and Bahcall, 1977). Equation (5.56) can be written

$$\frac{d \ln \rho_g}{dr} = -\frac{\mu m_p}{k T_g} \frac{d \phi(r)}{dr}. \quad (5.62)$$

If the potential is given by equations (5.59) and (5.61), then the gas distribution is given by

$$\rho_g(r) = \rho_{go} \left[1 + \left(\frac{r}{r_c} \right)^2 \right]^{-3\beta/2}. \quad (5.63)$$

Here

$$\beta \equiv \frac{\mu m_p \sigma_r^2}{k T_g} = 0.76 \left(\frac{\sigma_r}{10^3 \text{ km/s}} \right)^2 \left(\frac{T_g}{10^8 \text{ K}} \right)^{-1}, \quad (5.64)$$

where the numerical value follows if solar abundances are assumed so that $\mu = 0.63$. Another way to derive equation (5.63) is to note that if the galaxy velocity dispersion is isotropic, the galaxy density ρ_{gal} in a cluster will be given by equation (5.62), replacing $kT_g/\mu m_p$ with σ_r^2 . Then eliminating the potential between the equations for ρ_g and ρ_{gal} yields $\rho_g \propto \rho_{gal}^\beta$, and equation (5.63) follows if the galaxy distribution is given by equation (5.57) (Cavaliere and Fusco-Femiano, 1976).

This self-consistent isothermal model (equation 5.63) assumes that the gas and galaxy distributions are both static and isothermal and that the galaxy and total mass distributions are identical. While none of these assumptions is fully justified, and the gas is probably not generally isothermal, this model has the advantage that the resulting gas distribution is analytic and that basically all the integrals needed to compare the model to the observations of clusters are also analytic. For example, the total gas mass and emission integral (equation 5.20) are

$$\begin{aligned} M_g &= \pi^{3/2} \rho_{go} r_c^3 \frac{\Gamma[3(\beta-1)/2]}{\Gamma(3\beta/2)} \quad (\beta > 1) \\ &= 3.15 \times 10^{12} M_\odot \left(\frac{n_o}{10^{-3} \text{ cm}^{-3}} \right) \left(\frac{r_c}{0.25 \text{ Mpc}} \right)^3 \frac{\Gamma[3(\beta-1)/2]}{\Gamma(3\beta/2)}, \end{aligned} \quad (5.65)$$

$$EI = \pi^{3/2} \left(\frac{n_e}{n_p} \right) n_o^2 r_c^3 \frac{\Gamma(3\beta - 3/2)}{\Gamma(3\beta)} \quad (\beta > 1/2) \quad (5.66)$$

$$= 3.09 \times 10^{66} \text{ cm}^{-3} \left(\frac{n_o}{10^{-3} \text{ cm}^{-3}} \right)^2 \left(\frac{r_c}{0.25 \text{ Mpc}} \right)^3 \frac{\Gamma(3\beta - 3/2)}{\Gamma(3\beta)},$$

where n_o is the central proton density, Γ is the gamma function, and solar abundances have been assumed in deriving the numerical values. The values of β in the parentheses give the limits such that the appropriate integrals converge at large radii. Similarly, the X-ray surface brightness at a projected radius b is proportional to the emission measure EM , defined as

$$EM \equiv \int n_p n_e dl, \quad (5.67)$$

where l is the distance along the line-of-sight through the cluster at a projected radius b . Then, for the self-consistent isothermal model, the emission measure is

$$EM = \sqrt{\pi} \left(\frac{n_e}{n_p} \right) n_o^2 r_c \frac{\Gamma(3\beta - 1/2)}{\Gamma(3\beta)} (1 + x^2)^{-3\beta+1/2} \quad (\beta > 1/6) \quad (5.68)$$

where $x \equiv b/r_c$. The microwave diminution (Section 3.5) at low frequencies is given by

$$\frac{\Delta T_r}{T_r} = -\frac{2\sqrt{\pi}\sigma_T k T_g}{m_e c^2} \left(\frac{n_e}{n_p} \right) n_o r_c \frac{\Gamma(3\beta/2 - 1/2)}{\Gamma(3\beta/2)} (1 + x^2)^{-(3\beta-1)/2} \quad (\beta > 1/3) \quad (5.69)$$

where σ_T is the Thomson cross section, and T_r is the cosmic background radiation temperature. Numerically,

$$\Delta T_r = 0.10 \text{ mK} \left(\frac{n_o}{10^{-3} \text{ cm}^{-3}} \right) \left(\frac{T_g}{10^8 \text{ K}} \right) \left(\frac{r_c}{0.25 \text{ Mpc}} \right) \times \left(\frac{T_r}{2.7 \text{ K}} \right) \frac{\Gamma(3\beta/2 - 1/2)}{\Gamma(3\beta/2)} (1 + x^2)^{-3\beta/2+1/2}. \quad (5.70)$$

Equation (5.68) has been used extensively to model the surface brightness $I(b)$ of the X-ray emission from clusters (Gorenstein *et al.*, 1978; Branduardi-Raymont *et al.*, 1981; Abramopoulos and Ku, 1983; Jones and Forman, 1984; Section 4.4.1). The most accurate data have come from the *Einstein* X-ray satellite, which was sensitive only to low energy X-rays ($h\nu < 4 \text{ keV}$). For high temperature gas ($T_g \gtrsim 3 \times 10^7 \text{ K}$), the low energy X-ray emissivity is nearly independent of temperature, and thus $I(b) \propto EM$ even if the gas temperature varies. Large surveys of X-ray distributions fit by equation (5.68) have been made by Abramopoulos and Ku (1983), who set $\beta = 1$ (equal gas and galaxy distributions), and Jones and Forman (1984), who allowed β to vary. Figure 16 shows the data on the X-ray surface brightness of three clusters from Jones and Forman (1984) and their best fit models using equation (5.68). Equation (5.68) is a good fit to the majority of clusters, but fails in the central regions of some clusters, possibly because these clusters contain cooling accretion flows (Sections 4.3.3, 4.4.1, and 5.7). The average value of β determined by fits to the X-ray surface brightness of a large number of clusters was found to be (Jones and Forman, 1984)

$$\langle \beta_{fit} \rangle = 0.65. \quad (5.71)$$

Thus the X-ray surface brightness and implied gas density vary on average as

$$I_x(b) \propto \left[1 + \left(\frac{b}{r_c} \right)^2 \right]^{-3/2}, \quad (5.72)$$

$$\rho_g(r) \propto \left[1 + \left(\frac{r}{r_c} \right)^2 \right]^{-1}. \quad (5.73)$$

This indicates that the gas density should fall off less rapidly with radius than the galaxy density (in agreement with many other observations, such as Abramopoulos and Ku, 1983), and that the energy per unit mass is higher in the gas than in the galaxies (Jones and Forman, 1984). For this average value of β , the total X-ray luminosity converges, but the total gas mass given by equation (5.65) does not.

Unfortunately, this does not agree with the determinations of the X-ray spectral temperatures and the galaxy velocity dispersions of clusters (Mushotzky, 1984). For example, the observed correlation between σ_r and T_g in equation (4.10) implies that the average value β determined by gas temperatures and galaxy velocity dispersions is $\langle \beta_{spect} \rangle \approx 1.3$. From a sample of clusters with well-determined spectra, Mushotzky (1984) finds $\langle \beta_{spect} \rangle \approx 1.2$, which he notes is about twice the value determined from observations of the X-ray surface brightness. While Jones and Forman (1984) argue that their values of β_{fit} are in excellent agreement with the determinations from spectral observations, in fact their data show that the two values do not agree to within the errors in the majority of cases. For the clusters they studied $\langle \beta_{spect} \rangle \approx 1.1$. Thus the general result seems to be that

$$\langle \beta_{spect} \rangle \approx 1.2 \approx 2\langle \beta_{fit} \rangle. \quad (5.74)$$

A number of suggestions have been made as to the origin of this discrepancy. First, the gas may very well not be isothermal. However, Mushotzky (1984) has argued that the same problem occurs for other thermal distributions in the gas. Second, it may be that the line-of-sight velocity dispersion does not represent accurately the energy per unit mass of the galaxies. Equation (5.63) assumes that the galaxy velocity distribution is isotropic. The distribution could be anisotropic, either if the cluster is highly flattened (Section 2.9.3) or if the cluster is spherical but galaxy orbits are largely radial (rather than having a uniform distribution of eccentricities; see Section 2.8). A detailed study (Kent and Sargent, 1983) of the positions and velocities of galaxies in the Perseus cluster has produced a more accurate description of the cluster potential and significantly reduced the discrepancy, although it still is significant. Third, it may be that many of the galaxy velocity dispersions measured for clusters are contaminated by foreground or background groups (Geller and Beers, 1982). The velocity dispersions may also be affected by subclustering or nonvirialization of the cluster. All of these effects will cause the data to overestimate the actual velocity dispersion (the cluster potential), and thus to overestimate β_{spect} . Perhaps one indication that such systematic errors in the velocity dispersion might be occurring is that Coma, the best studied regular cluster, does not show a β discrepancy. Another possible solution to the β discrepancy would be if the gas dominated the total cluster mass at large radii (Henriksen and Mushotzky, 1985; Section 5.5.5); this would invalidate the assumption that the galaxies and gas were test particle distributions in the missing mass potential.

Finally, let me point out an utterly trivial possible explanation of the β discrepancy. The particular form of the gas distribution in the self-consistent isothermal model as given above (equations 5.63 through 5.70) depends on assuming that the cluster potential is fit by the *King approximation* to the isothermal sphere (equations 5.59 and 5.61). This approximation breaks down at large radii, where the King model density varies as $\rho \propto r^{-3}$ while a real isothermal sphere density varies as $\rho \propto r^{-2}$. Of course, it is the gas distribution at large radii which has the greatest leverage in affecting the fit to equation 5.63. If the mass density in clusters is *really* isothermal, then equation 5.63 will not fit the observed gas distribution for the correct value of β . As an alternative, let us assume that the cluster mass distribution can be fit by the simple analytic form $\rho(r) = \rho_o(1 + x^2)^{-1}$, which is similar to equation (5.57) but has the correct asymptotic form for an isothermal sphere. Then all the formulae for the self-consistent isothermal sphere remain unchanged if we substitute $\beta \rightarrow (2/3)\beta$ in equations (5.63) through (5.70). Retaining the present definition of β , this is equivalent to $\beta_{fit} = (2/3)\beta_{spect}$, which is essentially consistent with the observations. The observed $\langle \beta_{fit} \rangle \approx 2/3$ implies $\langle \beta_{spect} \rangle \approx 1$ which is consistent with the spectral observations. Another way of expressing all this is to say that the observed gas distribution (equation 5.73) is essentially that of an isothermal sphere.

This possible discrepancy between the globally determined X-ray temperatures of clusters and their surface brightness will probably only be resolved when spatially resolved X-ray spectra are available, allowing a simultaneous determination of the spatial variation of the gas density and temperature (Chapter 6).

One of the derivations of the self-consistent isothermal model involves noting that the galaxy distribution also solves the hydrostatic equation if the galaxy velocity dispersion is isotropic (see discussion following equation (5.64)). If the galaxy distribution is spherical, but the galaxy velocity dispersions in the directions parallel to the cluster radius (σ_r) and transverse to that direction (σ_t) are different, this derivation fails.

However, for a constant (isothermal) galaxy velocity dispersion and a constant anisotropy of the velocity dispersion $\eta \equiv 1 - \sigma_t^2/\sigma_r^2$, the gas density is given by $\rho_g \propto (\rho_{gal} r^{2\eta})^\beta$ (White, 1985).

Finally, Henry and Tucker (1979) have pointed out the existence of a simple relationship $(L_x r_c)^{2/5} \propto T_g$ between X-ray temperatures, luminosities, and core radii in clusters based on the isothermal model.

5.5.2 Adiabatic and polytropic distributions

The intracluster gas will be isothermal if thermal conduction is sufficiently rapid (Section 5.4.2). On the other hand, if thermal conduction is slow, but the intracluster gas is well-mixed, then the entropy per atom in the gas will be constant (Section 5.4.6). In an adiabatic gas, the pressure and density are simply related,

$$P \propto \rho^\gamma, \quad (5.75)$$

where γ is the usual ratio of specific heats and is $\gamma = 5/3$ for a monatomic ideal gas. While the value of $5/3$ would be expected to apply if the intracluster gas were strictly adiabatic, equation (5.75) is often used to parametrize the thermal distribution of the intracluster gas, with γ taken to be a fitting parameter. For example, $\gamma = 1$ implies that the gas distribution is isothermal. Intracluster gas models with an arbitrary value of γ are often referred to as ‘polytropic’ models, and γ is called the polytropic index. Intracluster gas models with the polytropic index $\gamma > 5/3$ are convectively unstable (equation 5.52), and thus hydrostatic polytropic models must have $1 \leq \gamma \leq 5/3$.

Adiabatic and polytropic intracluster gas distributions were introduced by Lea (1975), Gull and Northover (1975), and Cavaliere and Fusco-Femiano (1976). Given equation (5.75), the hydrostatic equation (5.55) can be rewritten by noting that

$$\frac{1}{\rho_g} \nabla P = \frac{\gamma}{\gamma - 1} \frac{k}{\mu m_p} \nabla T_g, \quad (5.76)$$

so that

$$\begin{aligned} \frac{T_g}{T_{go}} &= 1 + (\alpha - 1) \left[1 - \frac{\phi(\mathbf{r})}{\phi_o} \right], \\ \frac{\rho_g}{\rho_{go}} &= \left(\frac{T_g}{T_{go}} \right)^{1/(\gamma-1)}. \end{aligned} \quad (5.77)$$

Here, ϕ_o , T_{go} , and ρ_{go} are the central values of the cluster gravitational potential, the intracluster gas temperature, and the density, respectively. In nonspherical clusters, ‘central’ means the lowest point in the cluster gravitational potential.

From equation (5.77), it is clear that the intracluster gas temperature will always decrease with increasing distance from the cluster center in adiabatic or polytropic models.

The integration constant α is defined as

$$\alpha \equiv \frac{T_{go}}{T_{g\infty}}, \quad (5.78)$$

where $T_{g\infty}$ is the gas temperature at infinity. If $\alpha > 0$, the gas distribution extends to infinity, and the gas is not gravitationally bound to the cluster. If $\alpha < 0$, the gas extends only to a finite distance at which $\phi(\mathbf{r})/\phi_o = \alpha/(\alpha-1)$, and the gas is gravitationally bound to the cluster. In the $\alpha > 0$ models, the intracluster gas connects with and is confined by intercluster gas, whose temperature is given by $T_{g\infty}$. In general, the central gas temperature is given by

$$T_{go} = -\frac{\gamma - 1}{\gamma} \frac{\mu m_p \phi_o}{k(1 - \alpha)}. \quad (5.79)$$

The enthalpy per particle in these models is a constant, both spatially and temporally, and is given by $h = \alpha[\gamma/(\gamma - 1)]kT_{go}$. Thus models with $\alpha \approx 0$ cool very rapidly (Bahcall and Sarazin, 1978). Reasonable bound, intracluster gas models therefore usually assume $\alpha \lesssim 0.1$.

If equation (5.59) is used for the cluster potential, then the gas density and temperature are analytic functions of radius in the cluster, but are sufficiently complex that the integrals for emission measures, masses, and so on are not analytic. Moreover, since the gas is not isothermal, the X-ray surface brightness and the emission measure are not simply related. Spectra, X-ray surface brightness profiles, masses, microwave diminutions, and a number of other quantities for these models are given in Sarazin and Bahcall (1977) and Bahcall and Sarazin (1978). Fits of these models to the surface brightness and spectra of a number of clusters are given in Bahcall and Sarazin (1977) and Mushotzky *et al.* (1978).

5.5.3 More complicated distributions

The models discussed above assume that clusters are spherical, that the gas is hydrostatic, that the cluster potential is known in advance, and that the entropy distribution in the gas is given by a very simple polytropic distribution. Intracluster gas models have been calculated which attempt to generalize each of these assumptions. Here, some more complicated hydrostatic models are reviewed.

Stimpel and Binney (1979) (see also Binney and Stimpel, 1978) showed how spheroidal models for the intracluster gas distribution could be derived, using the observed galaxy distribution to determine the shape of the cluster potential. These models were fit to the galaxy counts in the Coma cluster, and models for the distribution of X-ray emission and microwave diminution (Section 3.5) were derived. There is an uncertainty in the shape of the cluster potential because the galaxy counts themselves cannot determine whether the cluster shape is more nearly prolate or oblate. However, Stimpel and Binney show that the resulting X-ray distributions in the two cases are considerably different, and thus X-ray observations can be used to determine the true shapes of elongated clusters (Chanan and Abramopoulos, 1984).

In general, the entropy distribution in a cluster will depend on the origin of the intracluster gas and the history of cluster. A number of authors have solved the hydrodynamic equations for simple models for the origin of the intracluster gas and of the subsequent history of the cluster; these calculations will be reviewed in Section 5.10. It is perhaps not surprising that these calculations do not generally lead to entropy distributions of the very simple sort (isothermal, adiabatic, or polytropic) considered above.

A number of authors have attempted to model these more detailed entropy distributions by allowing the polytropic index or the isothermal parameter β to vary with radius. One example is to allow the gas to be isothermal in the cluster core, where conduction is effective (Section 5.4.2), but adiabatic in the outer parts (Cavaliere and Fusco-Femiano, 1978; Cavaliere, 1980). Cavaliere and Fusco-Femiano (1978) also included the effect of the gas density on the cluster potential.

5.5.4 Empirical gas distributions derived by surface brightness deconvolution

The gas distributions in clusters can be derived directly from observations of the X-ray surface brightness of the cluster, if the shape of the cluster is known and if the X-ray observations are sufficiently detailed and accurate. This method of analysis also leads to a very promising method for determining cluster masses (Section 5.5.5). The X-ray surface brightness at a photon frequency ν and at a projected distance b from the center of a spherical cluster is

$$I_\nu(b) = \int_{b^2}^{\infty} \frac{\epsilon_\nu(r) dr^2}{\sqrt{r^2 - b^2}}, \quad (5.80)$$

where ϵ_ν is the X-ray emissivity. This Abel integral can be inverted to give the emissivity as a function of radius,

$$\epsilon_\nu = -\frac{1}{2\pi r} \frac{d}{dr} \int_{r^2}^{\infty} \frac{I_\nu(b) db^2}{\sqrt{b^2 - r^2}}. \quad (5.81)$$

Because of the quantized nature of the observations of the X-ray surface brightness (photon counts per solid angle) and the sensitivity of integral deconvolutions to noise in the data, the X-ray surface brightness data are often smoothed, either by fitting a smooth function to the observations or by applying these equations to the surface brightness averaged in rings about the cluster center.

Now, the emissivity is given by equation (5.19) and depends on the elemental abundances, the density of the gas, and the gas temperature. Thus the distribution of these three properties in the cluster could be determined from observations of the X-ray surface brightness $I_\nu(b)$. Basically, the continuum emission is due mainly to free-free emission (equation 5.11) and is relatively insensitive to the heavy element abundances, while the line emission measures these abundances. For a given set of abundances, the emissivity can be written as

$$\epsilon_\nu = n_e^2 \Lambda_\nu(T_g), \quad (5.82)$$

where $\Lambda_\nu(T_g)$ is n_p/n_e times the sum on the right side of equation (5.19). Thus $\epsilon_\nu(r)$ can be found from observations of the X-ray surface brightness as a function of photon frequency, and the frequency dependence and magnitude of ϵ_ν give the local gas temperature and density.

Unfortunately, observations of $I_\nu(b)$ are really not available for clusters. These observations require an instrument with good spatial *and* spectral resolution. Most of the X-ray spectra of clusters have been taken

with instruments having very poor spatial resolution (comparable to the size of the cluster; Section 4.3.1). High spatial resolution observations of clusters have primarily come from the *Einstein* X-ray observatory (Section 4.4). The imaging instruments on *Einstein* had only limited spectral resolution. Moreover, the optics in *Einstein* could only focus soft X-rays ($h\nu \lesssim 4$ keV). At typical gas temperatures in clusters ($kT_g \approx 6$ keV), most of the X-ray emissivity is due to thermal bremsstrahlung, and the emissivity is nearly independent of frequency for $h\nu \lesssim kT_g$. Thus the *Einstein* surface brightness distributions cannot be used directly to determine the local temperature. If the limited spectral resolution of the *Einstein* imagers is ignored, their observations provide $\langle I_x(b) \rangle$, the surface brightness averaged over the sensitivity of the detector as a function of photon frequency. From $\langle I_x(b) \rangle$, the sensitivity averaged emissivity

$$\langle \epsilon_x(r) \rangle = n_e^2 \langle \Lambda_x(T_g) \rangle \quad (5.83)$$

can be found from equation (5.81). Unfortunately, even if the elemental abundances are assumed to be known, this equation provides only one quantity at each radius, and it is impossible to determine both n_e and T_g .

In many analyses of X-ray cluster observations, a second equation for the density and temperature has been provided by assuming that the intracluster gas is hydrostatic and that the cluster potential is known. The hydrostatic equation in a spherical cluster (5.56) can be rewritten as

$$\frac{d \ln(n_e T_g)}{d \ln r} = -\frac{GM(r)\mu m_p}{rkT_g}, \quad (5.84)$$

where $M(r)$ is the total cluster mass and the mean atomic weight $\mu \approx 0.63$ is assumed to be independent of radius. Combining this with equation (5.83) gives

$$\left[1 - \frac{1}{2} \frac{d \ln(\Lambda_x)}{d \ln T_g} \right] \frac{d \ln T_g}{d \ln r} = -\frac{GM(r)\mu m_p}{rkT_g} - \frac{1}{2} \frac{d \ln(\epsilon_x)}{d \ln r}. \quad (5.85)$$

This is an ordinary differential equation for $T_g(r)$, which can be integrated given a boundary condition. This has been taken to be the central temperature (White and Silk, 1980) or the intergalactic pressure (Fabian *et al.*, 1981a). Given $T_g(r)$, equation (5.83) gives the density profile.

Various versions of this method have been used to determine gas distributions in a large number of clusters using data from *Einstein* (White and Silk, 1980; Fabricant *et al.*, 1980; Fabian *et al.*, 1981; Nulsen *et al.*, 1982; Fabricant and Gorenstein, 1983; Canizares *et al.*, 1983; Stewart *et al.*, 1984a,b). In some cases, spectral information from *Einstein* or low spatial resolution spectra have been used to further constrain the temperature profiles or to determine the form of the cluster potential necessary for a consistent fit (Section 5.5.5). These analyses have provided information on the mass distribution in clusters and central galaxies (M87, in particular), the gas distributions in clusters, and the prevalence of cooling accretion flows in clusters. These topics will be discussed in more detail later.

Gas temperature and density profiles could be derived more directly if high spatial and spectral resolution data at photon energies up to 10 keV were available. The proposed AXAF satellite will have these capabilities (Chapter 6).

5.5.5 Total masses and mass distributions in clusters—the hydrostatic method

Masses for individual galaxies or for clusters of galaxies can be derived from the distribution of their X-ray emitting gas if this gas is in hydrostatic equilibrium. This is a reasonable approximation as long as the cluster is stationary (the gravitational potential does not change on a sound crossing time), forces other than gas pressure and gravity (magnetic fields, for example) are not important, and any motions in the gas are subsonic. Estimates of intracluster magnetic fields based on radio observations (Chapter 4) show that they are much too small to have a significant dynamical effect. In Section 5.6, we will show that supersonic expansion of the gas (cluster winds) is unlikely. Some clusters show evidence for a slow settling of the intracluster gas due to cooling at the cluster center (a cooling flow; Sections 4.3, 4.4, and 5.7), but these motions are very subsonic except possibly very near the cluster center ($r \lesssim 1$ kpc; equation 5.105).

Under these circumstances, the gas obeys the hydrostatic equation (5.56), and the cluster mass $M(r)$ can be determined if the density and temperature of the intracluster gas are known. This method of determining

the mass has a number of advantages over the use of the virial theorem (Section 2.8) or any other method which uses the galaxies as test particles. First, the gas is a collisional fluid, and the particle velocities are isotropically distributed. On the other hand, galaxies in clusters (or stars in galaxies) are collisionless, and uncertainties in the velocity anisotropy can significantly affect mass determinations. Second, the hydrostatic method gives the mass as a function of radius, rather than the total mass alone as given by the virial method. Third, the statistical accuracy of this method is not limited by the number of galaxies in the cluster; the statistical accuracy can be improved by lengthening the observation time. Fourth, better statistics in the X-ray measurements means that it is easier to avoid problems with background contamination, and to resolve possible uncertainties due to subclustering (Geller and Beers, 1982). Finally, hydrostatic mass determinations are not very sensitive to the shape of the cluster (Strimpel and Binney, 1979; Fabricant *et al.*, 1984).

The first applications of this method to determine mass distributions were by Bahcall and Sarazin (1977) and Mathews (1978). The method has been developed extensively by Fabricant *et al.* (1980, 1984) and Fabricant and Gorenstein (1983). Applications of the method have been reviewed by Sarazin (1986b). Ideally, one would measure the spatially and spectrally resolved X-ray surface brightness $I_\nu(b)$ to directly deconvolve the gas density and temperature as a function of the radius (Section 5.5.4; equations 5.80 through 5.83). The mass is then given by the hydrostatic equation, which can be written as

$$M(r) = -\frac{kT_g(r)r}{\mu m_p G} \left(\frac{d \ln n_e}{d \ln r} + \frac{d \ln T_g}{d \ln r} \right).$$

Note that the mass depends only weakly on the gas density n_e (only its logarithmic derivative enters), but depends strongly on the gas temperature.

Unfortunately, as discussed in Section 5.5.4, the limited spectral response of the *Einstein* X-ray observatory has prevented the direct determination of temperature profiles for the intracluster gas. Accurate profiles of the gas density are known. In order to apply the hydrostatic method to clusters, some simple assumption must be made about the temperature distribution $T_g(r)$. Unfortunately, because the mass is strongly affected by T_g , this means that the resulting mass profiles will be very uncertain. Several analyses (Vallee, 1981; Fabricant *et al.*, 1984) have assumed that gas is isothermal (Section 5.5.1) or that the gas temperature and density are related by a simple polytropic equation (Section 5.5.2). These analyses give a somewhat smaller total cluster mass than previous virial estimates, and as noted in Section 5.5.1, somewhat higher gas masses.

As discussed in Section 4.3, excellent global cluster X-ray spectra exist from the HEAO-1 A-2 detectors. These spectra generally cannot be fit by emission at a single temperature (Henriksen, 1985). The spectra can be used to determine the amount of gas (or, more precisely, the amount of $EI = n_p n_e V$, where V is the volume (equation 4.3)) as a function of temperature, but cannot tell us where the gas is located because of their poor spatial resolution. The *Einstein* imaging observations give $n_e(r)$ (which can be integrated to give $n_p n_e V$), but give no information about the temperature structure. However, the comparison of these two results ((EI vs. T_g) and (EI vs. r)) allow the determination of (T_g vs. r), if we assume that T_g is a monotonic function of the radius r (Henriksen, 1985; Henriksen and Mushotzky, 1986; Cowie *et al.*, 1987). Since the observed gas densities vary at large distances like an isothermal sphere $\rho_g \propto r^{-2}$ (equation 5.73), while these mass determinations require that the gas temperature decrease with radius, the total mass density will always decrease with radius more rapidly than the gas density. Since the total density is the sum of the gas density, the galaxy density, and the missing mass density, these determinations suggest that the missing mass is concentrated towards the cluster center. For the Coma cluster, this method gives a *missing mass density which is more centrally concentrated than that of the galaxies*, which is in turn more centrally concentrated than that of the intracluster gas. Values of the mass-to-light ratio of the entire cluster of $M/L_V \approx 100h_{50}$ are found, with the intracluster gas contributing about 30% of the mass.

If these monotonic temperature mass determinations are correct, one is lead to a different picture of the missing mass than that presented in Section 2.8. If the missing mass is more concentrated than the visible matter in the cluster, it suggests that the missing mass has undergone dissipation. Combined with the smaller ratio of missing mass to visible mass, this suggests that the missing mass is baryonic matter, and not some weakly interacting species. The major uncertainty in these analyses is the assumption of a monotonic temperature gradient. While this seems quite plausible, there is no compelling physical argument

requiring that this be true. While most detailed models for the ejection or infall of intracluster gas give monotonic gradients (Section 5.10), some do not.

When spatially and spectrally resolved X-ray surface brightness measurement are available, it will be possible to directly determine the mass profiles of clusters of galaxies and determine the distribution of the missing mass. The same method can be applied to the X-ray emitting gas in elliptical galaxies (Section 5.8). This capability should become available with the launch of AXAF (Chapter 6). Until that time, the present results on the distribution of the missing mass must be regarded as tantalizing but tentative.

5.5.6 Chemically inhomogeneous equilibrium models

The heavier an ion is, the slower its thermal motion in the intracluster gas will be. As a result, heavy ions will tend to drift toward the center of the cluster. In Section 5.4.5 the drift rate was calculated and shown to be rather slow, although the precise value depends on cluster parameters. Further, the intracluster magnetic field may be very effective in retarding any sedimentation of heavy elements. Nonetheless, it is possible that heavier elements may have drifted to the cluster center in some cases, or been introduced into the intracluster gas quite near the center. Eventually, the motion of heavy ions towards the cluster center will stop when the resulting charge separation produces an electric field that cancels the effect of the gravitational field.

Abramopoulos *et al.* (1981) have calculated chemically inhomogeneous models for the intracluster gas in thermodynamic equilibrium. The temperature of the gas is therefore constant, both spatially and among the various ions and electrons. These models satisfy the hydrostatic equation (5.55) for the total pressure, but this does not uniquely define the chemically inhomogeneous model, since any value of the total pressure can be attributed to an infinite number of different combinations of partial pressures of different ions. In thermodynamic equilibrium, the number density of any ion is given by the Boltzmann equation

$$n_i(\mathbf{r}) = n_{io} \exp \left[-\frac{A_i m_p \phi(\mathbf{r}) + Z_i e \phi_E(\mathbf{r})}{k T_g} \right], \quad (5.86)$$

where A_i , Z_i , and n_{io} are the mass, charge, and central density of the ion i . The cluster gravitational potential ϕ and electrical potential ϕ_E are defined in this equation to be zero at the cluster center. The gravitational potential is assumed to be fixed and known, as in most of the hydrostatic models discussed in the previous sections. The electrical potential is given as a solution of Poisson's equation

$$\nabla^2 \phi_E(\mathbf{r}) = -4\pi \sum_i n_i(\mathbf{r}) Z_i e, \quad (5.87)$$

where the sum includes electrons as one of the ionic species. Because the coupling constant for electrical forces exceeds that of gravitational forces by $\approx 10^{40}$, while the fraction of the cluster mass due to intracluster gas is much greater than the inverse of this, the electrical potential will generally be quite small $\phi_E \approx -\phi m_p/e$, and a sufficient approximation to the solution of equation (5.87) is to assume that the gas is very nearly electrically neutral:

$$\sum_i n_i Z_i = 0. \quad (5.88)$$

The electric potential is then given by requiring a simultaneous solution of equations (5.86) and (5.88), and requiring that the total abundance $\int n_i dV$ of each of the ions be fixed.

In the case where the plasma can be assumed to consist only of protons and electrons, with the abundances of all heavy elements being so low that they do not contribute to the potential, the solution for the ion and electron densities is the same as the isothermal hydrostatic solution given above (equation 5.63), the electric potential is $\phi_E = -m_p \phi / 2e$, and the densities of any trace heavy elements will be

$$n_i = n_{io} \left(\frac{n_p}{n_{po}} \right)^{2A_i - Z_i}. \quad (5.89)$$

Unfortunately, the abundance of helium certainly can not be treated as insignificant, and other heavier elements may contribute to the electric potential near the cluster core. Equations (5.86) and (5.88) have been

solved numerically by Abramopoulos *et al.* (1981), assuming an analytic King potential for the cluster, taking the gas temperature to be $\beta = 2/3$ (equation 5.64), and assuming a range of heavy element abundances from pure hydrogen and helium to solar. They find that the heavy elements are strongly concentrated to the cluster core. To balance the resulting increase in the charge density, the electrons also must be centrally condensed.

The concentration of heavy elements to the cluster core, where the electron density is high, results in an increase in the X-ray line intensity for a given set of heavy element abundances. As a result, Abramopoulos *et al.* (1981) argue that the observed line strengths could be produced by gas with iron abundances of only 1/20 of solar.

5.6 Wind models for the intracluster gas

If the rate of heating of the intracluster medium is sufficiently large, the gas will be heated to the escape temperature from the cluster, and will form a wind leaving the cluster. Since the sound crossing time across a cluster is relatively short (equation 5.54), the cluster would be emptied of gas rapidly unless gas were constantly being added to the cluster. Yahil and Ostriker (1973) suggested that gas is being injected into the intracluster medium at rates of $10^3 - 10^4 M_{\odot}/\text{yr}$, and is sufficiently heated to produce a steady-state outflow of gas. These mass ejection rates are considerably larger than those expected from stellar mass loss from visible stars in galaxies (Section 5.10).

Let $\dot{\rho}$ be the rate of mass input per unit volume into the intracluster gas, and let $h(r)$ be the heating rate per unit volume. If the cluster is assumed to be spherically symmetric, with a gravitational potential $\phi(r)$, and the gas is injected into the cluster with no net velocity, but with the velocity dispersion of the galaxies, then the hydrodynamic equations for a steady-state wind are

$$\begin{aligned} \frac{1}{r^2} \frac{d}{dr} (r^2 \rho_g v) &= \dot{\rho}, \\ \dot{\rho}v + \rho_g v \frac{dv}{dr} + \frac{dP}{dr} + \rho_g \frac{d\phi(r)}{dr} &= 0, \\ \frac{1}{r^2} \frac{d}{dr} \left\{ r^2 \rho_g v \left[\frac{v^2}{2} + \frac{5}{2} \frac{P}{\rho_g} + \phi(r) \right] \right\} &= h(r) + \dot{\rho} \left[\frac{3}{2} \sigma_r(r)^2 + \phi(r) \right], \end{aligned} \quad (5.90)$$

where ρ_g , P , and v are, respectively, the gas density, pressure, and velocity, and σ_r is the one-dimensional velocity dispersion of the galaxies in the cluster.

Yahil and Ostriker argued that the source of mass input into the cluster was mass loss by stars in cluster galaxies, and that

$$\dot{\rho}(r) = \alpha_* \rho(r), \quad (5.91)$$

where ρ is the total mass density of the cluster and α_*^{-1} is the characteristic time scale for gas ejection from galaxies. They argued that $\alpha_*^{-1} \approx 10^{12}$ yr.

Yahil and Ostriker considered two mechanisms that might heat the gas. First, if the gas were ejected by supernovae it could be heated within galaxies before being ejected into the cluster (Section 5.3.3; equation 5.27); they referred to this as the HIG (heating in galaxies) model. Defining $\lambda \equiv 3kT_{ej}/(2\mu m_p \sigma_r^2)$, the heating rate due to ejection from galaxies becomes

$$h(r) = \lambda \sigma_r^2 \dot{\rho}. \quad (5.92)$$

Alternatively, Yahil and Ostriker suggested that the heating might be due to the motion of galaxies through the cluster (Section 5.3.4); they referred to this as the HBF (heating by friction) model. Then, the heating rate is

$$h(r) = \pi \langle R_D^2 \Delta v^3 \rangle \rho_g \left[\frac{\rho(r)}{m_{gal}} \right], \quad (5.93)$$

where R_D is the drag radius (equation 5.28), $m_{gal} \equiv \rho/n_{gal}$ is the total cluster mass per galaxy, Δv is the velocity of the galaxy relative to the gas (which is moving at v), and the average is over the Gaussian distribution of galaxy velocities. It is useful to define $\lambda(r)$ such that $h(r) \equiv \lambda(r) \sigma_r^2 \dot{\rho}$, so that λ is constant in the HIG model, and $\lambda(r) \approx \pi R_D^2 \rho_g \sigma_r / \alpha_* m_{gal}$ in the HBF model.

Let the cluster density be given by $\rho = \rho_o g(x)$, where ρ_o is the central density, $x \equiv r/r_c$, r_c is the cluster core radius, and $g(x)$ is a dimensionless function. Similarly, let the cluster potential be written as $\phi(r) = \phi_o \bar{\phi}(x)$, where ϕ_o is the central potential. If the mass distribution is assumed to be isothermal, so that σ_r is constant and is related to the cluster central potential ϕ_o and the central density ρ_o by equations (5.60) and (5.61), then the equation of continuity becomes

$$x^2 \rho_g v = \alpha_* \rho_o r_c G(x), \quad (5.94)$$

$$G(x) \equiv \int_0^x g(x') (x')^2 dx'.$$

If the King analytic form of the cluster density is assumed, then g , G , and $\bar{\phi}$ are given by the functions on the right-hand sides of equations (5.57), (5.58) and (5.59), respectively. The energy equation can also be integrated to give

$$\frac{1}{2} v^2 + \frac{5}{2} \frac{P}{\rho_g} = \sigma_r^2 \left[\Lambda(x) + \frac{3}{2} + 9\bar{\phi} - 9\Phi \right], \quad (5.95)$$

where Λ and Φ are the integrated heating rate and potential energy

$$\Lambda(x) \equiv \frac{\int_0^x g(x') \lambda(x') (x')^2 dx'}{G(x)}, \quad (5.96)$$

$$\Phi(x) \equiv \frac{\int_0^x g(x') \bar{\phi}(x') (x')^2 dx'}{G(x)}.$$

The left-hand side of equation (5.95) is positive definite, which implies that there is a minimum heating rate at any radius

$$\Lambda(x) > 9\Phi - 9\bar{\phi} - \frac{3}{2}. \quad (5.97)$$

If the King analytic model for the potential is assumed, then the maximum value for the right-hand side of equation (5.97) is 1.35, and this is the minimum value of λ in the HIG model (equation 5.92). Similarly, the central temperature is given by

$$kT_o = \frac{2}{5} \sigma_r^2 \mu m_p \left[\Lambda(0) + \frac{3}{2} \right], \quad (5.98)$$

so that the minimum central temperature in the HIG model is

$$T_o(\text{HIG}) > 8.7 \times 10^7 \text{ K} \left(\frac{\sigma_r}{1000 \text{ km/s}} \right)^2. \quad (5.99)$$

Similar arguments for the HBF model indicate that $\pi R_D^2 \rho_{go} \sigma_r / (\alpha_* m_{gal}) \gtrsim 3/2$, and that the central temperature is

$$T_o(\text{HBF}) \gtrsim 3.4 \times 10^8 \text{ K} \left(\frac{\sigma_r}{1000 \text{ km/s}} \right)^2. \quad (5.100)$$

These analytic results were given in Bahcall and Sarazin (1978). The temperatures are lower than those in Yahil and Ostriker (1973) because they assumed $\mu = 1$, which is not valid for an ionized plasma.

The observed cluster temperatures (Section 4.3.1) are generally below those required by equations (5.99) or (5.100), so it does not seem that the gas is heated sufficiently to produce a wind. The energy required to support such a wind also seems to be prohibitively large, since it would require that considerably more than the whole thermal content of the gas (more because of the kinetic energy of the flow) be produced in a flow time, which for a transonic wind is comparable to a sound crossing time. In the HIG model, this would require a very large supernova rate or other energy source; in the HBF model, it would require that the galaxies have delivered to the gas much more kinetic energy than they currently possess (Section 5.3.4). As a result, it seems unlikely that wind models fit the distribution of the intracluster gas at all radii, but it remains possible that gas is flowing out from the outer portions of clusters.

Livio *et al.* (1978) argue that viscous drag between intracluster and interstellar gas can produce a very large heating rate due to galaxy motions if the intracluster gas is sufficiently hot, so that the Reynolds number is small (equations 5.46 and 5.47). As discussed in Section 5.3.4, this requires that galaxies maintain large amounts of intracluster gas; it seems more likely that the drag forces will strip the gas from the galaxies. Unless the galaxies have a large rate of mass outflow, Rephaeli and Salpeter (1980) have shown that the drag does not increase for small Reynolds number (Section 5.4.4). Again, there is the problem that the heating required to support a strong wind is greater than the total kinetic energy content of the galaxies.

Lea and Holman (1978) discussed winds driven by heating by relativistic electrons (Section 5.3.5). They derived analytic heating limits like those of equation (5.59). To produce a steady-state wind, the heating rate must be much greater than that necessary to heat the gas in a Hubble time, and as a result the required values of γ_l are very low. They find that even in the Perseus cluster, which is one of the strongest radio clusters known, $\gamma_l \ll (B/\mu G)^{-(\alpha_r+1)/2\alpha_r}$ is required for a steady-state wind. Moreover, this mechanism suffers the problem that extended halo radio sources, of the sort necessary to heat the intracluster gas, are relatively rare.

In summary, it seems unlikely that the intracluster gas in most clusters is involved in steady-state outflow, although gas may be flowing out of the outer portions of clusters.

5.7 Cooling flows and accretion by cDs

5.7.1 Cooling flows

If the rate of cooling in the intracluster gas is sufficiently rapid, gas may cool and flow into the center of the cluster. Lea *et al.* (1973), Silk (1976), Cowie and Binney (1977), and Fabian and Nulsen (1977) noted that the cooling times t_{cool} (equation 5.23) in the more luminous X-ray clusters are comparable to the Hubble time t_H (the age of the universe and approximate age of clusters). This might be a coincidence, but it would also be a natural consequence of the cooling of the intracluster gas. Imagine that a certain amount of gas was introduced into the cluster when it formed. Initially, the cooling time in this gas might be much longer than the age of the cluster, and cooling would be unimportant. However, as long as cooling was more rapid than heating, and the gas was not removed from the cluster, the cooling time would not increase, and eventually the age of the cluster would exceed the cooling time. Once this occurs, the gas will cool in the cluster core, and the pressure of the surrounding gas will cause the cool gas to flow into the cluster center. The surrounding hot intracluster gas will always have $t_{cool} \approx t_H$. Silk (1976) and Cowie and Binney (1977) showed that this cooling condition could explain a number of the observed correlations between the optical and X-ray properties of clusters, particularly the $L_x - \sigma_r$ relation (equation 4.8). The subject of these cooling flows has been reviewed recently by Fabian *et al.* (1984b).

Silk (1976) suggested that the intracluster gas was introduced into the cluster when it formed, and that cooling has reduced the amount of intracluster gas so as to maintain $t_c \approx t_H$. Cowie and Binney (1977), on the other hand, argued that the intracluster gas is constantly being replenished by ejection from cluster galaxies. Then, the density of gas in the cluster center would increase until the age exceeded the cooling time. After this point, a stable steady-state would be achieved, in which the rate of ejection of gas by galaxies into the cluster was balanced by the rate of removal of gas through cooling in the cluster core (Cowie and Binney, 1977; Cowie and Perrenod, 1978). As long as the mass ejection rate is not too low, cooling would regulate the inflow, and the densest hot gas at the cluster center would always have $t_{cool} \approx t_H$.

The equations for a steady-state cooling flow in a spherically symmetric cluster are identical to those for a wind (equation 5.90), except that the heating rate $h(r)$ is replaced by the cooling rate ϵ (equations 5.21 and 5.22). Because the cooling rate is proportional to the square of the density, it is useful to define $\epsilon \equiv \rho_g^2 \Lambda(T_g)$. Then,

$$\begin{aligned} \frac{1}{r^2} \frac{d}{dr} (r^2 \rho_g v) &= \dot{\rho}, \\ \dot{\rho}v + \rho_g v \frac{dv}{dr} + \frac{dP}{dr} + \rho_g \frac{d\phi(r)}{dr} &= 0, \\ \frac{1}{r^2} \frac{d}{dr} \left\{ r^2 \rho_g v \left[\frac{v^2}{2} + \frac{5}{2} \frac{P}{\rho_g} + \phi(r) \right] \right\} &= -\rho_g^2 \Lambda(T_g) + \dot{\rho} \left[\frac{3}{2} \sigma_r(r)^2 + \phi(r) \right], \end{aligned} \quad (5.101)$$

and the other symbols are defined following equation (5.90). If the gas is injected continuously into the cluster ($\dot{\rho} \neq 0$), then the boundary conditions are set by requiring no inflow from outside the cluster. If

the gas is not constantly being added to the cluster ($\dot{\rho} = 0$), then no cluster-wide steady-state flow will be possible. However, these equations will still apply approximately within the radius defined by setting the cooling time equal to the age of the cluster (the ‘cooling surface’). In order to match to the hydrostatic intracluster gas distribution beyond this radius, the proper density and temperature is specified on this surface. Because the inflow is generally extremely subsonic except in the innermost parts, the outer parts of these cooling flows are very nearly hydrostatic (the middle equation in 5.101 reduces to 5.55). The velocity is determined by the inflow rate $\dot{M} \equiv 4\pi\rho_g vr^2$, which must be constant in a steady-state inflow without sources or sinks for mass.

Thermal conduction (Section 5.4.2) has not been included in the cooling flow equations. Generally, the existence of cooling flows requires that conduction not be very important, since otherwise cooling in the cluster core will be balanced by heat transported inwards by conduction, rather than heat convected inwards by the cooling flow (Mathews and Bregman, 1978; Takahara and Takahara, 1979; Binney and Cowie, 1981; Nulsen *et al.*, 1982). In most cases, conduction must be much slower than the Spitzer rate (equation 5.37) for a nonmagnetic plasma; the suppression of conduction might be due to a very tangled or circumferential magnetic field (Section 5.4.3).

These equations have been solved for cooling in clusters by Cowie and Binney (1977), assuming a King model for the cluster density (equation 2.13), mass ejection by galaxies as given by Yahil and Ostriker (1973; equation 5.91), pure bremsstrahlung cooling (equation 5.21), and subsonic flow (dropping the quadratic velocity terms in equations 5.101). Figure 37 shows the density and temperature variations in a typical model. The density rises continuously into the center; this contrasts with hydrostatic models, in which it levels off (Section 5.5). The temperature has a maximum at about two core radii and declines into the cluster center. The behavior of these solutions in the inner regions can be understood approximately. Assuming that the mass flux is fixed ($\dot{M} = \text{const}$), that the flow is subsonic, that the gravitational potential is not important (so that the pressure is nearly constant and the flow is driven by pressure and cooling and not by gravity), and that the cooling function is a power-law in the temperature $\Lambda \propto T_g^\alpha$, the temperature and density are found to vary as $\rho_g \propto 1/T_g \propto r^{-3/(3-\alpha)}$ (Nulsen *et al.*, 1982). For $T_g \gtrsim 4 \times 10^7$, cooling is due to thermal bremsstrahlung (equation 5.21), $\alpha = 1/2$, and $\rho_g \propto r^{-6/5}$. At lower temperatures, $\alpha \approx -0.6$ (equation 5.22) and $\rho_g \propto r^{-5/6}$. When the gas has cooled sufficiently for the gravitational potential to be important (or if the potential gradient is increased by a central galaxy), then the temperature tends to vary as $kT_g/\mu m_p \approx GM(r)/r$, where $M(r)$ is the total mass within r (Fabian and Nulsen, 1977).

Evidence for the existence of such cooling flows in X-ray clusters includes the detection of peaks in the soft X-ray surface brightness at the cluster center (Branduardi-Raymont *et al.*, 1981; Fabian *et al.*, 1981a; Canizares *et al.*, 1983; Fabricant and Gorenstein, 1983; Jones and Forman, 1984; Stewart *et al.*, 1984a,b), the measurement of inverted temperature gradients $dT_g/dr > 0$ (Gorenstein *et al.*, 1977; Ulmer and Jernigan, 1978; White and Silk, 1980), and the observation of central X-ray surface brightnesses and temperatures that imply cooling times much less than the Hubble time (Canizares *et al.*, 1983; Stewart *et al.*, 1984b). The strongest evidence comes from the detection of soft X-ray line emission from low ionization stages produced at temperatures of $T_g \approx 10^6 - 10^7$ K, coming from the cluster center (Canizares *et al.*, 1979, 1982; Canizares, 1981; Mushotzky *et al.*, 1981; Lea *et al.*, 1982; Nulsen *et al.*, 1982; Culver *et al.*, 1983; Mushotzky, 1984; Section 4.3.3).

Theoretical models for cooling flows can be used to estimate the rates of cooling in clusters from these X-ray observations. A simple estimate may also be derived by noting that for steady-state, isobaric cooling, the luminosity emitted in any temperature range dT_g is

$$dL(T_g) = \frac{5}{2} \frac{\dot{M}}{\mu m_p} k dT_g \quad (5.102)$$

if the gravitational potential does not change significantly during the cooling (Cowie, 1981). Here, μ is the mean atomic mass. This equation can be used to estimate the luminosity in any spectral feature produced in the cooling flow by integrating $dL(T_g)$ over the fraction of the total emission in that feature as a function of temperature. A more accurate semi-empirical method of determining cooling rates for clusters has been derived Canizares *et al.* (1983), Stewart *et al.* (1984b), and Fabian *et al.* (1984b). This method is essentially that of Section 5.5.4. First, the X-ray surface brightness is deconvolved to give the X-ray emissivity as a

function of position (equation 5.81). Second, the hydrostatic equation is used to derive the gas temperature as a function of position, assuming some form for the gravitational potential (equation 5.85). (Note that the hydrostatic equation holds for the outer portions of these cooling flows because they are very subsonic; see Section 5.7.2.) Third, the cooling time is calculated in the gas, and the gas is assumed to form a steady-state cooling flow within the cooling surface at which $t_{cool} \approx t_H$. Finally, the steady-state energy equation (the last of equations 5.101) is used to determine the cooling rate \dot{M} .

Table 4 gives a list of clusters showing evidence for cooling flows, with estimates for the cooling rate \dot{M} . The values for \dot{M} range from $3 - 1000 M_\odot/\text{yr}$. The Perseus cluster has one of the largest observed cooling rates (see also Section 4.5.2). A very small cooling rate of $3 - 20 M_\odot/\text{yr}$ is found in M87 in the Virgo cluster; such a small cooling flow would be difficult to observe in more distant clusters. It is particularly interesting that cooling flows have now been detected in poor clusters (Canizares *et al.*, 1983), as well as in a large number of rich clusters (Stewart *et al.*, 1984b).

It should perhaps be pointed out that all of the above evidence shows that there is *cool* gas in the cores of these clusters, but it does not directly show that there is *cooling* gas flowing into the cluster core. That is, the motion of the gas has not been detected directly, as X-ray spectral observations have insufficient resolution to detect Doppler shifts produced by the slow inflow. It has occasionally been argued that the gas might actually be in hydrostatic and thermal equilibrium, with the extra emission at the cluster center being due to some central heat source (see, for example, Tucker and Rosner, 1983). Since the cooling rate of a gas at constant pressure decreases with temperature at all temperatures of interest for X-ray emission ($10^5 \lesssim T_g \lesssim 10^9 \text{ K}$), the higher the heating rate (or the nearer the cluster center), the cooler the gas would be in thermal equilibrium. However, such hydrostatic, thermal equilibrium models (in which one heats the gas in order to cool it) can generally be shown to be thermally unstable on the cooling time scale (Stewart *et al.*, 1984a; Fabian *et al.*, 1984b). The gas must either heat up and expand out of the core, or cool down and form a cooling flow on this time scale. Moreover, such a thermal equilibrium model for the X-ray emission would not explain the observed association of cooling flows with optical line emitting filaments, as discussed below.

5.7.2 Accretion by central galaxies

Many clusters of galaxies have luminous galaxies located at their centers (Section 2.10.1), and these galaxies appear to be moving relatively slowly compared to the average cluster galaxy (Quintana and Lawrie, 1982). In fact, such a luminous galaxy is found at the center of every cooling flow that has been observed (Jones and Forman, 1984). The cooling intracluster gas may, then, be accreted by the central galaxy in the cluster when its temperature has fallen to the point where it can be bound to the galaxy, and the gas can cool further and flow into the galaxy center (Silk, 1976; Cowie and Binney, 1977; Fabian and Nulsen, 1977; Mathews and Bregman, 1978).

It is unlikely that the presence of the central galaxy causes the cooling flow (that is, increases the cooling rate significantly) because the gravitational potential associated with a galaxy is much smaller than that associated with a cluster (cluster velocity dispersions are much larger than those of galaxies). Thus the presence of a central galaxy will not cause a large perturbation in the density of hydrostatic intracluster gas (Section 5.5) and will not affect the cooling rate significantly. Initially, the inflow of cooling gas is driven by the pressure of the surrounding hot medium; only when the gas has cooled significantly does the gravitational potential of the galaxy influence the flow. If the mass of the central galaxy is small, the flow may be pressure-driven over most of its extent (Fabian and Nulsen, 1977; Binney and Cowie, 1981).

In order to accrete cooling gas, a galaxy must be moving slowly through the intracluster medium. The gas must be able to cool before the galaxy has moved away. Thus the velocity is limited to

$$v_{gal} \lesssim \frac{R_{gal}}{t_{cool}} \approx 100 \text{ km/s} \left(\frac{R_{gal}}{100 \text{ kpc}} \right) \left(\frac{t_{cool}}{10^9 \text{ yr}} \right)^{-1}. \quad (5.103)$$

Thus a typical cluster galaxy with $v_{gal} \approx \sigma_r \approx 10^3 \text{ km/s}$ will have great difficulty accreting. Central dominant galaxies are expected to have velocities of at least $\approx 100 \text{ km/s}$ on average, due to gravitational encounters with other galaxies (Section 2.9.1). Based on the asymmetry in the nonthermal radio emission and optical line filaments around the galaxy M87 in the Virgo cluster, De Young *et al.* (1980) argued that this accreting galaxy was moving about 200 km/s to the north. Dones and White (1985) showed that this

was inconsistent with the observed temperature structure in the cooling flow; that is, it would strongly violate this cooling limit.

Models for the accretion of cooling gas onto central galaxies have been given by Fabian and Nulsen (1977), Mathews and Bregman (1978), and Binney and Cowie (1981). All of these calculations assume the $\dot{\rho} = 0$ in equations (5.101); the sources of the gas are external to the central galaxy, so that \dot{M} is constant within the galaxy. Mathews and Bregman integrate equations (5.101) inward through a sonic radius r_s at which the inflow becomes supersonic. The structure of the flow is determined by \dot{M} and the gas temperature at the sonic radius T_s . The sonic radius is a solution of the following equation (Mathews and Bregman, 1978):

$$r_s = \frac{3\mu m_p}{10kT_s} \left[GM(r_s) + \frac{\dot{M}\Lambda(T_s)\mu m_p}{10\pi kT_s} \right], \quad (5.104)$$

where $M(r)$ is the total galactic and cluster mass within the radius r . This equation can have multiple solutions, although usually only one of these at most corresponds to an astrophysically interesting density. For reasonable values of \dot{M} and T_s , one finds $r_s \approx 0.1 - 2$ kpc (Mathews and Bregman, 1978). A simple estimate of the sonic radius can be derived in a number of ways. First, if the inflow is driven by cooling rather than gravity, then the second term in equation (5.104) will be more important. This is essentially equivalent to arguing that the inflow time and cooling time must be roughly equal. Second, even if gravity is important to the inflow, $kT_g/\mu m_p \approx GM(r)/r$ (see discussion after equations 5.101), and the first term in equation (5.104) only increases r_s by several times. Thus an estimate for r_s is

$$r_s \approx 0.1 \text{ kpc} \left[\frac{\dot{M}}{100 M_\odot / \text{yr}} \right] \left[\frac{T_s}{10^7 \text{ K}} \right]^{-2.6}, \quad (5.105)$$

where the cooling function in equation (5.22) has been used. Note that r_s is proportional to \dot{M} .

On the other hand, Fabian and Nulsen (1977) and Cowie *et al.* (1980) argued that the gas cannot flow in as far as r_s because the inflowing gas must have at least a small amount of angular momentum. Let $\mathbf{l} \equiv \mathbf{r} \times \mathbf{v}$ be the angular momentum per unit mass of the incoming gas, and assume that this is conserved. Then the inflow will stop and the gas will form a rotating disk at a stagnation radius r_{st} given by

$$r_{st} = \frac{l^2}{GM(r_{st})} = 2.3 \text{ kpc} \left[\frac{l}{10^3 \text{ kpc} \cdot \text{km/s}} \right]^2 \left[\frac{M(r_{st})}{10^{11} M_\odot} \right]^{-1}. \quad (5.106)$$

Ideally, the angular momentum of inflowing gas in a spherical cluster would be zero, but realistically the gas must have some nonradial velocity. Cowie *et al.* argued that the gas was ejected by cluster galaxies, and that the angular momentum of the central cooling gas would be determined by the rms residual velocity of galaxies in the cluster core,

$$l \approx \frac{\sigma_r}{\sqrt{N_{gal}}} r_c \approx 2 \times 10^4 \text{ kpc} \cdot \text{km/s}, \quad (5.107)$$

where σ_r is the cluster velocity dispersion, $N_{gal} \approx 100$ is the number of galaxies in the cluster core, and r_c is the cluster core radius. Alternatively, if the gas had no nonradial motion relative to the cluster, l might be due to the central galaxy orbital velocity, which for $v_{gal} \approx 100$ km/s and a galaxy orbital radius of ≈ 30 kpc gives $l \approx 3 \times 10^3$ kpc · km/s. However, both of these estimates imply gas velocities relative to the central galaxy that exceed the cooling time scale limit (equation 5.103). Taking this limit at a cooling radius of 200 kpc gives $l \lesssim 4 \times 10^3$ kpc · km/s. Thus, unless the inflow is much more radial than these estimates suggest, it is likely that the flow will stagnate before becoming supersonic ($r_{st} \geq r_s$).

Viscosity (Section 5.4.4) might transport angular momentum out of the flow and reduce l . If ionic viscosity is effective, thermal conduction will probably also be important and may suppress the cooling flow (Cowie *et al.*, 1980). However, the magnetic field may couple the inner and outer parts of the flow and provide an effective viscosity; a circumferential field could both transport angular momentum out of the inflow, and suppress thermal conduction. Alternatively, if the magnetic field were too weak to be dynamically important and the ionic viscosity were small, the Reynolds number for the flow would be large and the flow could become turbulent. Turbulent viscosity might transport angular momentum out of the flow (Nulsen

et al., 1984), and, at the same time, the turbulence could tangle the magnetic field and inhibit thermal conduction.

Cowie *et al.* (1980) treated the inflow as radial (which is reasonable for $r \gg r_{st}$), and included the effect of angular momentum by adding a repulsive centrifugal potential $\phi_{cent} = l^2/2r^2$ to the gravitational potential ϕ in equations (5.101). Moreover, they assumed that $r_{st} \gg r_s$, and therefore they dropped all quadratic or higher order terms in the flow velocity v .

If the accretion rate in the flow decreases inward ($d\dot{M}/dr > 0$), the radial flow can continue into the center of the galaxy without passing through a sonic radius (Nulsen *et al.*, 1984). There is some evidence that \dot{M} does decrease inwards in M87/Virgo and NGC1275/Perseus (Fabian *et al.*, 1984b; Stewart *et al.*, 1984a). It may be that some of the cooling gas is being converted into stars (Section 5.7.4 below).

Many central galaxies in clusters with cooling flows have nuclear nonthermal radio sources, which may be powered by the further accretion of a small portion of the cooling gas onto the central ‘engines’ (black holes?) of these sources (Burns *et al.*, 1981b; Bijleveld and Valentijn, 1983; Valentijn and Bijleveld, 1983; Jones and Forman, 1984; Nulsen *et al.*, 1984). There is evidence for a correlation between the radio luminosity and accretion rate in these galaxies (Valentijn and Bijleveld, 1983; Jones and Forman, 1984).

5.7.3 Thermal instability and optical filamentation

As the gas cools, any inhomogeneities in the gas density will tend to be amplified (Fabian and Nulsen, 1977). In a region of higher than average density, the temperature will tend to be lower, to preserve pressure equilibrium. Both higher density and lower temperature will speed up the cooling rate, and lowering the temperature will increase the density contrast. Mathews and Bregman (1978) analyzed the growth of density inhomogeneities and the thermal instability of the cooling gas. They considered only radial comoving, isobaric perturbations (no change in the velocity or pressure as compared to the unperturbed flow at the same position), and assumed \dot{M} is constant. They found that the gas is thermally unstable if

$$\Upsilon(T_g) \equiv \frac{2\Lambda(T_g)}{kT_g} - \frac{d\Lambda(T_g)}{dkT_g} > 0. \quad (5.108)$$

This is true for any interesting gas temperature ($T_g \gtrsim 10^5$ K). Assuming that there is an initial perturbation of relative amplitude $(\Delta\rho/\rho)_i$ at some large radius r_i , the amplified perturbation at an inner radius $r < r_i$ is

$$\left(\frac{\Delta\rho}{\rho}\right) = \left(\frac{\Delta\rho}{\rho}\right)_i \exp\left[\frac{2}{5} \frac{\mu m_p}{4\pi} \int_r^{r_i} \Upsilon(T_g) \dot{M} \frac{dr}{(vr)^2}\right] \quad (5.109)$$

in the linear regime ($(\Delta\rho/\rho) \ll 1$). Mathews and Bregman found that amplification factors of 10^{3-4} were likely for flows into the sonic radius. White and Sarazin (1987a) generalized the perturbation analysis to arbitrary perturbations and flows with sinks for gas ($\dot{M} \neq \text{const}$), and showed that the flow is always unstable if inequality 5.108 is satisfied, and that the fastest growing mode is radial, comoving, and isobaric, and is amplified according to equation 5.109. Thermal conduction could suppress the thermal instability of cooling flows, unless conduction is itself suppressed by magnetic fields.

Cowie *et al.* (1980) generalized this analysis by considering the motion of finite sized ‘blobs’ with a finite density perturbation, without assuming that the blobs comove with the flow. They argue that the motion of blobs relative to the unperturbed gas (buoyancy) will stabilize some perturbations, but that the gas will still be unstable for some blob sizes and densities. Because they assume the flow stagnates, the inflow must become very unstable for $r \approx r_{st}$.

Given that the gas in cooling flows is thermally unstable and that the growth time for the instability is comparable (or somewhat shorter than) the cooling time, these flows should become very inhomogeneous unless the inflowing intracluster gas were very smooth. Thus it may not be reasonable to model the flows with a single phase; multiphase models for cooling flows are discussed in Section 5.7.5.

Optical emission line filaments are often seen near the centers of clusters having cooling flows, often within the central galaxies that are accreting the cooling gas (Ford and Butcher, 1979; Kent and Sargent, 1979; Stauffer and Spinrad, 1979; Heckman, 1981; Fabian *et al.*, 1981b, 1982a; Cowie *et al.*, 1983; Hu *et al.*, 1983, 1985; van Breugel *et al.*, 1984). The size of these extended filament systems ranges roughly from 1–100 kpc. These filaments emit the Balmer lines of hydrogen, as well as forbidden lines from heavier elements.

Their spectra are similar to those seen in astrophysical shocks. The emission from these filaments is generally believed to be the result of the same cooling flows; the gas is visible in optical line emission as it cools through the temperature range $T_g \approx 10^4$ K. The clumpy nature of these filaments is due, at least in part, to thermal instability in the cooling gas as discussed above. Figure 20 shows the optical filaments around NGC1275 in the Perseus cluster, which is one of the best studied examples of this optical filamentation (Lynds, 1970). Clusters with cooling flows having optical filamentation are listed in Table 4.

Cowie *et al.* (1980, 1983) have derived the total luminosity in the H α line expected in a cooling flow; they find

$$L(\text{H}\alpha) \approx 5 \times 10^{40} \text{ ergs/s} \left(\frac{\dot{M}}{100 M_\odot/\text{yr}} \right). \quad (5.110)$$

This is somewhat larger than would be derived from equation (5.102), which assumes isobaric cooling. Thermally unstable blobs of gas probably cool nearly isobarically until $T_g \approx 10^6$ K; at lower temperatures the cooling time is shorter than the sound crossing time for the filaments, and the gas initially cools isochorically (at constant density). The cool clumps reach pressure equilibrium with the surrounding hot gas by the passage of repressurizing shocks (Mathews and Bregman, 1978; Cowie *et al.*, 1980). The H α flux may also be increased by photoionization by X-rays and ultraviolet radiation from the surrounding hot gas or from the nuclear source in the accreting galaxy.

Cowie *et al.* (1983) have suggested, based on extensive observations of optical line emission, that these filament systems consist of two components: a cluster core component, which is 20–100 kpc in size, and a galaxy component, which is smaller. In some clusters both components are present; in some, only one or the other is observed. The larger cluster components tend to consist of highly elongated filaments, which may be stretched by tidal effects in the cluster or along the magnetic field lines. The galaxy components are more homogeneous, and are also elongated. Cowie *et al.* suggest that they are gas disks at the stagnation radius. In many clusters the filament systems show velocity shears consistent with centrifugally supported disks (Hu *et al.*, 1985), although velocity measurements in NGC1275 indicate that the gas is not rotating fast enough in this galaxy (Hu *et al.*, 1983). These emission line gas disks can be used to estimate the masses of the inner portions of the cD galaxies, just as the rotation velocities of spiral galaxy disks are used to determine their masses (Hu *et al.*, 1985).

Hu *et al.* (1985) find that the presence of optical emission line filaments in the central galaxies of clusters correlates very strongly with the central density of the intracluster gas. The emission lines are only found in clusters with central proton number densities $n_o \gtrsim 5 \times 10^{-3} h_{50}^{1/2} \text{ cm}^{-3}$. This is just what is expected if the emission lines are due to cooling flows, since a lower limit on the density implies an upper limit on the cooling time. For this density limit and an intracluster gas temperature of 7×10^7 K, equation (5.23) gives $t_{cool} \lesssim 10^{10} h_{50}^{-1/2} \text{ yr}$, which implies that the gas could cool in the lifetime of the cluster.

The filament system around NGC1275 in the Perseus cluster is the most luminous such system known, and is considerably brighter than would be expected given the accretion rate determined from the X-ray measurements. Hu *et al.* (1983) argue that it is also too luminous to be due to photoionization by the nuclear source in NGC1275. They show that the optical filaments at the velocity of NGC1275 are elongated at the same position and in the same direction as the filaments due to the spiral galaxy that is moving towards NGC1275 at 3000 km/s (see Section 4.5.2). They suggest that this galaxy is actually colliding with the base of the accretion flow onto NGC1275, and that the kinetic energy in this collision powers the emission line filaments. Two problems with this model, which are discussed by Hu *et al.*, are the lack of intermediate velocity gas and the difficulty of a gas-rich galaxy penetrating into the core of the Perseus cluster without having its gas stripped (Sections 2.10.2 and 5.9).

5.7.4 Accretion-driven star formation

What is the duration of these cooling flows in clusters? Both optical and X-ray clusters have been observed at moderately high redshifts, and the gas distributions in clusters are relaxed and smooth; both suggest that the intracluster medium has been present in clusters for a significant fraction of the Hubble time. As described above, the cooling times at the centers of these flows are less than the Hubble time, which suggests that the flows have persisted for a significant fraction of the cluster lifetime. Thus the total

mass of gas that has cooled and been accreted by the central galaxy in the cluster should be

$$M_{acc} \approx 10^{12} M_\odot \left(\frac{\dot{M}}{100 M_\odot / \text{yr}} \right). \quad (5.111)$$

This assumes that the accreted mass is not ejected from the cluster center.

Although this is a small fraction of the total gas mass in a cluster ($\approx 10^{14} M_\odot$), it is comparable to the mass in luminous material in a very large (cD) galaxy. It is important to understand where this gas goes after it cools. Many possibilities can be ruled out. First, the gas could remain gaseous but cooler. However, the observations of the X-ray and optical line emission give upper limits on the total amount of ionized gas well below the total accreted mass. Observations of the 21 cm hydrogen line from central galaxies in clusters with cooling flows (Section 3.7) give upper limits on the total amount of neutral atomic hydrogen, typically $M_{HI} \leq 10^9 M_\odot$ (Burns *et al.*, 1981a; Valentijn and Giovanelli, 1982). Because of the instrumental sensitivity, current observations do not rule out the possibility that the accreted gas could be molecular hydrogen, although it seems likely that the large amounts of molecular gas required would lead to an efficient conversion into stars. As mentioned above, many of the accreting galaxies are radio sources (Burns *et al.*, 1981b), and this may indicate that some of the accreted gas flows into the galactic nucleus and is used to power the central engine of the radio source. However, these sources require only $\lesssim 10^{-2} M_\odot / \text{yr}$ of gas to provide their observed radio power (Burns *et al.*, 1981b), so it is likely that the fraction of the gas accreted by the central engine is small. Moreover, there could not be as much as $10^{12} M_\odot$ in the nucleus of these galaxies, given their observed stellar velocity dispersions (Sarazin and O'Connell, 1983).

Generally, conversion of gas into stars is quite efficient within galaxies, and this conversion could provide both a stable reservoir for the accreted gas and a partial explanation for the existence of central dominant galaxies. Burns *et al.* (1981a) argued that star formation cannot be occurring at a high rate in these galaxies, because they would then contain more neutral hydrogen than is observed (see above). However, Fabian *et al.* (1982b) and Sarazin and O'Connell (1983) showed that the cooling times for neutral hydrogen were short enough that all the accreted gas could be cooling through this temperature range and forming stars. Within the disk of our own galaxy, stars are formed with a very wide range of masses extending up to $\approx 100 M_\odot$ (Salpeter, 1955). Stars more massive than $\approx 10 M_\odot$ produce Type II supernovae when they die, and the rate of these supernovae would probably heat the accreted gas sufficiently to prevent the formation of cooling flows (Wirth *et al.*, 1983). Moreover, if the spectrum of stellar masses formed from the cooling gas (the ‘initial mass function’) were similar to that in our own galaxy, the central galaxies in the accretion flows would be considerably bluer and brighter than they are observed to be (Fabian *et al.*, 1982b, 1984a; Sarazin and O'Connell, 1983). Burns *et al.* (1981a) gave similar arguments and concluded that star formation cannot be the ultimate reservoir for the cooling gas.

However, there is really no reason why the initial mass function for star formation in these cooling flows should be the same as that in the disk of our galaxy. If the forming stars had low masses $\lesssim 1 M_\odot$, these stars would not be very different from the stars found in typical elliptical galaxies (Cowie and Binney, 1977). Since star formation is very poorly understood and there is no successful quantitative theory for this process, one cannot calculate the initial mass function directly. However, Fabian *et al.* (1982b) and Sarazin and O'Connell (1983) have given a simple plausibility argument as to why the initial mass function for star formation in cooling flows might be limited to low mass stars; a similar argument for elliptical galaxies in general was given by Jura (1977). It is assumed that stars form eventually from the thermally unstable clouds of gas that are seen as optical filaments. Star formation is assumed to start when these clouds become gravitationally unstable and can no longer support themselves against their own gravity and the pressure of the surrounding medium. Clouds become gravitationally unstable when their mass exceeds the Jeans’ mass, which for a spherical, static, nonmagnetic isothermal cloud of temperature T immersed in a low density medium of pressure P is given by (Spitzer, 1978)

$$\begin{aligned} M_J &= 1.2 \left[\left(\frac{kT}{\mu m_p} \right)^4 \frac{1}{G^3 P} \right]^{1/2} \\ &= 0.64 M_\odot \left(\frac{T}{10 \text{ K}} \right)^2 \left(\frac{\mu}{\text{amu}} \right)^{-2} \left(\frac{P}{10^7 k \text{ cm}^{-3} \text{ K}} \right)^{-1/2}. \end{aligned} \quad (5.112)$$

Once a cloud starts to collapse, the pressure within the cloud will increase and the Jeans' mass may be reduced; this can cause the cloud to fragment and result in lower mass stars being formed. It is difficult to produce stars more massive than M_J , however, because before a suitably massive cloud could be assembled, it would become unstable and collapse. Thus it is possible that the Jeans' mass may provide an upper limit on the mass of the largest stars that form. In the disk of our galaxy, the interstellar medium typically has a pressure of $P \approx 2 \times 10^3 k \text{ cm}^{-3} \text{ K}$, and equation (5.112) gives $M_J \approx 50 M_\odot$. In the cooling flows in clusters, the pressures derived from models for the X-ray emission or determined directly from the optical line emitting filaments are 10^{3-4} times larger ($P \approx 10^{6-7} k \text{ cm}^{-3} \text{ K}$), and thus the Jeans' mass is $M_J \approx 1 M_\odot$. Thus it is possible that only low mass stars are formed from the cooling gas in clusters. (A similar argument was given earlier for low mass star formation in normal elliptical galaxies by Jura (1977).) In Fabian *et al.* (1982) and Sarazin and O'Connell (1983), this conclusion is shown to be unaffected by the temperature dependence in equation (5.112).

Fabian *et al.* (1982b) also point out that star formation in cooling flows may be different than in the disk of our galaxy because the star forming regions in these flows are unlikely to contain dust grains. In star forming regions in our galaxy, most of the refractory heavy elements are in the form of solid dust grains, and these grains absorb starlight, emit infrared radiation, and act as a heat source for the gas. Dust grains are destroyed in high temperature gas. Since the gas in cooling flows is initially very hot, any grains would have evaporated (Cowie and Binney, 1977; Fabian *et al.*, 1982b), and it is very difficult for grains to form in low density gas, even if it is cool. Thus it is unlikely that grains will be present in the cooling flows, even in the coolest, star forming clouds. The lack of grains probably lowers the gas temperature, which also tends to reduce M_J . Further, any attempt to estimate the star forming rate in these galaxies from the infrared emission of dust (Wirth *et al.*, 1983) is likely to greatly underestimate the real rate.

Sarazin and O'Connell (1983) have calculated the expected colors and optical spectra of central galaxies assuming that their stellar populations are a mixture of a normal giant elliptical population with a continuously forming population due to accretion-driven star formation. A variety of values for the upper mass limit and the shape of the initial mass function were used. They found that the accreting galaxies should have spectra and colors measurably different from those of nonaccreting giant ellipticals (color differences of typically $\Delta(U - V) \approx -0.3$ mag). They also found that with accretion rates of typically $\dot{M} \gtrsim 100 M_\odot/\text{yr}$, the entire stellar population of the central galaxies in many clusters could be due to accretion-driven star formation.

Valentijn (1983) has attempted to measure the colors of the stellar populations in 7 cD galaxies and their spatial variations by surface photometry in two colors (B and V). He finds very large color gradients within these galaxies of typically $\Delta(B - V) \approx 0.4$ mag, with the galaxy centers being extremely red. Valentijn argues that these gradients are the result of accretion-driven star formation, and that as the pressure increases inwards in the cooling flow, the Jeans' mass is lowered (equation 5.112) and the stellar population becomes redder. One problem is that the innermost regions of these galaxies are so red that the required stellar population would have a very small mass-to-light ratio and could not provide enough light (for the observed accretion rates) to account for the observed galaxy luminosity. The color gradients observed by Valentijn are very large (larger than Sarazin and O'Connell predicted), and it is very important that they be confirmed by further observations. Valentijn's photometry appears to disagree, in some cases, with that of other observers (Hoessel, 1980; Malumuth and Kirshner, 1985).

Color gradients might also result from dust extinction, abundance gradients in an old stellar population, or mergers of galaxies having differing colors. A more direct way to detect a stellar population due to accretion is to observe absorption features due to that population in the spectrum of the cD galaxy. The galaxy NGC1275 in the Perseus cluster (Figure 20; Section 4.5.2) is particularly interesting in this regard. Its stellar surface brightness distribution is similar to that of a typical giant elliptical galaxy (Oemler, 1976). However, the stellar population is very blue, and has an A-star absorption spectrum (Kent and Sargent, 1979), whereas typical giant elliptical galaxies are dominated by K stars. Sarazin and O'Connell (1983) show that the colors and spectrum of this galaxy can be understood if the luminous portion of the galaxy is entirely due to accretion-driven star formation at a rate of $\approx 300 M_\odot/\text{yr}$ as given by the X-ray observations, and the upper mass cutoff of the stars formed is $2.8 M_\odot$. (By contrast, Wirth *et al.* (1983) present a model for NGC1275 in which the initial mass function for star formation is similar to that in our galaxy, and very massive O stars are formed. However, in this model the star formation rate is more than an order of

magnitude less than the observed rate of accretion onto NGC1275.)

One concern with NGC1275 is the presence of the foreground spiral galaxy. Hu *et al.* (1983) have suggested that this galaxy is colliding with the cooling flow and that this collision powers the optical line emission. It is also possible that this collision might affect the rate and initial mass function of star formation in the galaxy. It is thus very important to observe the spectra of other accreting central dominant galaxies, and see if a younger stellar population can be detected in them. Recently, O'Connell *et al.* (1987) obtained spectra for the inner regions of a number of accreting galaxies. The cD in A1795, which has a very large accretion rate $\dot{M} \approx 400 M_{\odot}/\text{yr}$ (Table 4), has an F-star stellar spectrum, consistent with the entire galaxy being the result of accretion-driven star formation with an upper mass cutoff of about $1.5 M_{\odot}$.

Central dominant cluster galaxies appear, in many cases, to have a very large number of globular star clusters (spherical clusters of $\approx 10^{5-6}$ stars) associated with them (Harris *et al.*, 1983b). Fabian *et al.* (1984b) have suggested that these globular clusters might be produced by accretion-driven star formation. They note that the mass of a globular cluster is similar to the Jeans' mass of gas in the cooling flows at a temperature of 10^4 K, the temperature at which thermally unstable clouds are repressurized. However, Fall and Rees (1985) showed that the cooling time was much shorter than the free-fall time for Jeans' unstable clumps at this temperature, preventing the gravitational instability of the gas. The cooling time depends on the abundance of heavy elements, which may have been considerably lower when galaxies first formed (Section 5.10). Thus Fall and Rees argue that globular clusters formed out of cooling flows during the formation of galaxies.

5.7.5 Cooling flow models with star formation

If cooling gas is being converted into stars in cooling flows, then terms representing the loss of gas should be added to the equations for the flow (equations 5.101). Regardless of the ultimate fate of the cooling gas, the fact that the gas is thermally unstable means that it is not reasonable to treat the cooling flow as homogeneous. In homogeneous models for cooling flows, the gas remains hot enough to emit X-rays until it is within the sonic radius (Section 5.7.2). However, in an inhomogeneous flow, thermal instability will cause denser lumps of gas to cool below X-ray emitting temperatures while the more diffuse gas is still quite hot. Thus, even if star formation were not occurring, thermal instabilities would reduce the amount of *hot* gas as it moves towards the center of the galaxy.

If the mass flow rate \dot{M} of hot gas decreases with distance from the center of the flow, this should result in an observable reduction in the amount of X-ray emission near the center of the flow. Such a reduction does appear to be required by the X-ray surface brightness data of cooling flow clusters. Using the semi-empirical method to determine \dot{M} from X-ray surface brightness profiles (Section 5.7.1), Fabian *et al.* (1984b) and Stewart *et al.* (1984a) found that the cooling rate \dot{M} increased with radius in M87/Virgo and NGC1275/Perseus. Unfortunately, their method of determining \dot{M} is inconsistent if \dot{M} is not constant (White and Sarazin, 1986a,b). First, their particular form of the energy equation (last of equations 5.101) requires that \dot{M} be constant. Second, they assumed that only the hot diffuse gas contributed to the X-ray emission. But the gas being removed from the flow by thermal instabilities cools radiatively, producing X-ray emission. Thus the X-ray emissivity should include a term proportional to the rate of gas loss through thermal instabilities. Semi-empirical methods to determine $\dot{M}(r)$ including more consistent treatments of mass loss have been given by Fabian *et al.* (1985, 1986a) and White and Sarazin (1987a,b). These studies indicate that $\dot{M}(r)$ increases with increasing r in the best studied cooling flow clusters. Fabian *et al.* argue that the variation of $\dot{M}(r)$ is well represented by $\dot{M}(r) \propto r$, while White and Sarazin contend that the variation of $\dot{M}(r)$ is too sensitive to the assumed form of the gravitational potential to allow any strong statements to be made.

If the gas in cooling flows is inhomogeneous, it is much more difficult to model the dynamics of the flows. In principle, in a correct treatment of the flow the gas would be represented by a continuous range of densities ρ_g . Rather than giving a single set of thermodynamic variables, say ρ_g , T_g , and v , as a function of position r , one should specify a distribution of densities. For example, $f(\rho_g, r)$ might be the fraction of the volume or mass in the flow at r which is in the form of gas at a density ρ_g . Correspondingly, $T_g(\rho_g, r)$ and $v_g(\rho_g, r)$ would be the mean temperature and velocity of gas having a density ρ_g located at r . Obviously, this would vastly increase the complexity of the hydrodynamical modeling of the flow. For one thing, there is no clear physical argument which specifies the boundary conditions on the inhomogeneities (for example, the value of $f(\rho_g)$ at the cooling radius r_{cool}). Perturbations in the flow probably cool isobarically until

they are too cool to produce a significant amount of X-ray emission ($T_g \lesssim 10^6$ K), so that it is probably reasonable to assume that all density phases have the same pressure at each position. Then the temperature is just $T_g(\rho_g, r) = P(r)/\rho_g$. However, the possibility that the different densities phases would have different velocities still is an enormous complication, since the hydrodynamical interactions between lumps of differing density and velocity would be extremely complex.

Two opposite approximations have been made to deal with this problem. First, Fabian *et al.* (1985, 1986a) assumed that all the density phases comove, so that both v and P are functions only of position r and not of density ρ_g . As noted above (Section 5.7.3), the fastest growing linear perturbations ($\Delta\rho_g/\rho_g$) $\ll 1$ in a homogeneous flow do comove, which supports this idea. However, one might expect that once perturbations grow nonlinearly ($\Delta\rho_g/\rho_g \gg 1$) they might drop out of the flow. This led White and Sarazin (1987a,b) to an opposite approximation. They argued that the isobaric cooling time decreases quite rapidly with decreasing temperature. Thus once a lump has cooled significantly below the average temperature, it will cool below X-ray emitting temperatures rapidly. As the density of the lump increases, its surface area will decrease and it can decouple from the flow and fall ballistically. Since the flow time is determined by the cooling time of the diffuse gas, the cooling and decoupling of higher density lumps can occur before the flow has moved inwards by a significant amount. In this limit, the cooling of dense lumps of gas can be treated as a local sink for the diffuse gas, and the flow equations revert to equations 5.101 with loss terms. Numerical models for cooling flow including loss terms which are proportional to either the cooling time or the growth time of thermal instabilities have been given by White and Sarazin (1986a,b).

These inhomogeneous models for cooling flows can be used to predict the X-ray surface brightness profiles and spectral variations of cooling flows (White and Sarazin, 1986b). If clumps of gas cool rapidly from X-ray emitting temperatures to $T_g \approx 10^4$ K, they can also predict the surface brightness of optical line emission. If these cooling condensates form stars quickly or if the cool lumps are decoupled from the flow and falling ballistically (and form stars eventually), these models can give the predicted distribution of these stars. In Section 5.7.4, attempts to detect the presence of a younger stellar population due to accretion in the optical spectra of central galaxies in cooling flows were discussed. It is also very important to study the spatial distribution of this population. As discussed in Section 2.10.1, central dominant galaxies in clusters appear to be composed of an extended giant elliptical interior, surrounded in the case of rich cluster cDs by a very extended halo. These cDs may also have dark, missing mass haloes (Section 5.8.1). Which of these components could be the result of accretion-driven star formation? If accretion produces the giant elliptical interiors, why do these resemble the stellar distributions in other nonaccreting giant ellipticals? Giant ellipticals have light distributions that are reasonably fit by de Vaucouleurs or Hubble profiles (Section 2.10.1), and recent numerical studies have suggested that these form naturally in violent relaxation (Section 2.9.2). Accretion-driven star formation is a slow process; will it give a similar distribution?

As discussed above, semi-empirical determinations of the hot gas inflow rates by Fabian *et al.* (1985, 1986a) are consistent with $\dot{M} \propto r$, although White and Sarazin (1986a,b) have argued that $\dot{M}(r)$ is extremely uncertain. If the cooling lumps form stars rapidly and if the orbits of the newly formed stars are not affected by the galaxy potential, this might imply that the density of the new stars varied as $\rho_* \propto r^{-2}$. This is much flatter than the density distribution of the luminous stars in elliptical galaxies. This is just the density distribution of an isothermal sphere, and is similar to the density distributions inferred for the missing mass haloes of spiral galaxies (Section 2.8). This led Fabian *et al.* (1986a,b) to suggest that the missing mass is very low mass stars formed in cooling flows; this requires that the initial mass function for star formation in cooling flows produce mainly very low mass stars ($M_* \lesssim 0.1 M_\odot$).

On the other hand, White and Sarazin (1986a,b) found that the predicted stellar distributions in their cooling flow models with star formation were very similar to those of the light from giant elliptical galaxies. In calculating the stellar distributions, they self-consistently included the effects of the stellar orbits in the galaxy and cluster gravitational potential.

Accretion-driven star formation may result in different stellar orbits in cD galaxies than in nonaccreting giant ellipticals. If the flows have little angular momentum (see above) and are radial, the resulting stellar orbits may be radial. If the processes of clumping and star formation impart significant random velocities to the star forming regions, the orbits might be isotropic. If the flow stagnates and forms a disk, stellar disks may be found within cD galaxies.

5.7.6 Evolution of cooling flows and active galaxies

One important question is what effect the evolution of clusters has on these cooling flows. An important type of evolution is the merging of subclusters; the observed double clusters (Section 4.4.2) may be systems undergoing this process. Such a merger will probably heat up the gas and disrupt any existing cooling flows (McGlynn and Fabian, 1984). Along these lines, Stewart *et al.* (1984b) speculate on the possibility that the Coma cluster, with its pair of D galaxies, is the result of such a merger of two subclusters. Perhaps each of the subclusters had a cooling flow, and accretion-driven star formation produced the two D galaxies. The merger would disrupt the flow, which would explain why Coma apparently lacks a cooling flow despite its high X-ray luminosity. The heating of the gas during the merger might explain the unusually high temperature of the intracluster medium in Coma (Section 4.5.1). In this way, accretion could have formed the central galaxies in clusters that do not currently have accretion flows.

Although this review is concerned with clusters, elliptical galaxies in irregular clusters and groups also appear to have considerable quantities of X-ray emitting gas, which may also form cooling flows (Section 5.8.3). These cooling flows are probably powered by stellar mass loss within the galaxies, rather than accretion of intracluster gas. For giant ellipticals, stellar mass loss rates of $\approx 1M_{\odot}/\text{yr}$ are expected. In the past, rates of stellar mass loss in elliptical galaxies were probably higher than they are today (Section 5.10.1). Since cooling flows in cluster centers were probably weaker in the past, this suggests that individual early type galaxies may make an important contribution to the emission from clusters at high redshift (Fabian *et al.* 1986a).

As was noted above, accreting central galaxies are often strong radio sources, indicating that some small portion of the cooling gas is accreted by the nucleus of the galaxy. Quasars appear to be galaxies with extremely luminous nonthermal nuclei. It is possible that they are also powered by accretion through cooling flows (Sarazin and O'Connell, 1983; Fabian *et al.*, 1986a). It is possible that the evolution of cooling flows in clusters or individual galaxies may help to determine the evolution of active galactic nuclei (Section 5.10). Active galaxies often show strong optical line emission; perhaps the lower excitation line emission seen in some active galaxies may be due to thermal instabilities in cooling flows, the mechanism that produces the line emission in nearby cooling flow galaxies.

In summary, it appears possible that most of the accreted gas in cooling flows is converted into low mass stars, and that this accretion-driven star formation may provide much of the luminous or nonluminous mass of central galaxies in clusters with cooling flows. This process may compete with mergers and tidal effects as a mechanism for the formation of cD galaxies (Section 2.10.1). It may also affect the evolution of active galaxies.

5.8 X-ray emission from individual galaxies

A number of topics concerning the X-ray emission from hot gas associated with individual galaxies in clusters will now be discussed. Material on the accretion of gas by central galaxies in clusters with cooling flows was reviewed in the previous section.

5.8.1 Massive haloes around M87 and other central galaxies

The X-ray emission from the M87/Virgo cluster is considerably different from the emission from richer and more regular clusters (Section 4.5.3). The gas is much cooler and less extended, and the X-ray luminosity is rather small. Initially, this was explained by noting that the X-ray luminosity and temperature of clusters appeared to correlate with the cluster velocity dispersion (equations 4.8 and 4.10), although this argument was somewhat circular, since M87/Virgo was one of the strongest pieces of evidence in favor of these correlations.

Bahcall and Sarazin (1977) and Mathews (1978b) suggested an alternative; the gas in M87/Virgo might be hydrostatic and bound to the galaxy M87, rather than to the cluster as a whole. To bind the gas, the galaxy would have to have a much deeper gravitational potential well than it would appear to have from its optical emission. Based on the best observations at that time, Bahcall and Sarazin showed that M87 must have a massive halo, with a total mass of $\approx 3 \times 10^{13} M_{\odot}$ extending out to ≈ 100 kpc from the galaxy center. Bahcall and Sarazin assumed a number of reasonable equations of state for the gas (adiabatic, polytropic, etc.; see Section 5.5). Mathews (1978b) reached a similar conclusion, although he assumed that the gas was exactly isothermal and that the mass distribution was the King analytic form for an isothermal distribution (5.57). Because such isothermal gas models diverge unless the gas is cool compared to the potential (equations 5.65 and 5.66), Mathews required much more mass, $\geq 10^{14} M_{\odot}$. This high mass limit is very sensitive to the assumption that the gas is exactly isothermal (Bahcall and Sarazin, 1978; Fabricant *et*

al., 1980).

There is considerable evidence that M87/Virgo has a cooling flow (Section 5.7). However, at large distances from the galaxy center the flow is highly subsonic, and the hydrostatic equation applies. Thus the cooling flow does not alter the requirement that the mass be large.

Binney and Cowie (1981) suggested that the mass of M87 might be much lower ($\approx 3 \times 10^{11} M_{\odot}$). They argued that the gas around M87 is bound by the *pressure* of surrounding, hotter gas, rather than by the *gravity* of M87. This hotter gas was in turn assumed to be extended and bound by the cluster gravitation potential of Virgo. The hot, diffuse, confining, intracluster gas is required to have a density and temperature of $n_p \approx 10^{-3} \text{ cm}^{-3}$ and $T_g \approx 10^8 \text{ K}$. The cool gas is produced by a cooling flow, which is drawn from the hot gas.

A key point in the Binney and Cowie model is that there must be enough hot, diffuse gas to provide a pressure that confines the cool gas around M87. Because this gas would be less dense than the cooler gas, it would have a lower surface brightness, but it should be detectable in hard X-rays. Davison (1978) and Lawrence (1978) found a hard X-ray component from M87/Virgo, which they suggested was extended. However, more recent observations indicate that this hard component is not extended, has a power-law spectrum, and is centered on M87 (Lea *et al.*, 1981, 1982). This hard source is probably nonthermal emission from the nucleus of M87. The present limits on the amount of extended hard X-ray emission are, at best, marginally consistent with the Binney and Cowie model (Fabricant and Gorenstein, 1983).

The *Einstein* satellite observations of the X-ray emission from M87 have been analyzed in detail by Fabricant *et al.* (1980), Fabricant and Gorenstein (1983), and Stewart *et al.* (1984a). All of these papers used basically the method of analysis outlined in Sections 5.5.4 and 5.5.5; that is, the observed surface brightness was deconvolved to give the variation of emissivity with radius from the galaxy center. At the time of the first paper, the spectral response of the IPC on *Einstein* was poorly calibrated, and the density was derived from the emissivity assuming that the gas was isothermal. From the hydrostatic equation, the total mass $M(r)$ interior to a radius r is

$$M(r) = -\frac{kT_g r}{G\mu m_p} \left(\frac{d\log \rho_g}{d\log r} + \frac{d\log T_g}{d\log r} \right). \quad (5.113)$$

Beyond 3 arc min from the center of M87, the surface brightness is well represented as a power-law $\propto r^{-1.62}$ (Fabricant and Gorenstein, 1983). Thus, as long as the temperature gradient is not significant, the density derivative in this equation is a constant, and the mass increases with the radius. Fabricant *et al.* (1980) found a total mass of $2 - 4 \times 10^{13} M_{\odot}$ within a radius of 230 kpc.

Binney and Cowie (1981) analyzed the same *Einstein* data and found consistency with their low mass model for M87. They argued that the disparity between their conclusion and that of Fabricant *et al.* (1980) was due to two differences. First, Fabricant *et al.* assumed that the surface brightness approached zero far from M87, whereas Binney and Cowie invoke extended intracluster gas. More importantly, in the Binney and Cowie cooling flow model the gas temperature decreases rapidly towards the center, $d\log T_g/d\log r \approx 0.8$. This large temperature gradient nearly cancels the density gradient (i.e., the pressure is nearly constant), and the resulting mass is significantly reduced.

The *Einstein* data were reanalyzed by Fabricant and Gorenstein (1983) and Stewart *et al.* (1984a), who found that the mass of M87 is $3 - 6 \times 10^{13} M_{\odot}$ within a radius of 260 kpc. By this time, the spectral response of the IPC had been calibrated and the temperature gradient could be derived directly from the data, albeit with large errors. (Note that this is possible for M87/Virgo and not for most cluster sources because the temperature in M87/Virgo is low enough to be measured with the soft X-ray sensitivity of *Einstein*). In addition, the temperature gradient was constrained by the wide field proportional counter observations from previous satellites, and by data from the SSS and FPCS X-ray spectrometers on *Einstein*. The observed temperature gradient appears to be inconsistent with that required by Binney and Cowie. These data appear to rule out the Binney and Cowie model.

Thus M87 appears to have a very massive halo, which extends well beyond the region where the galaxy is luminous. Like the massive haloes around spiral galaxies (Section 2.8), the mass appears to increase roughly in proportion to the radius in the outer portions of M87. However, spiral galaxies have rotation velocities nearly independent of radius, on the order of 300 km/s (Faber and Gallagher, 1979). The inferred circular orbital velocity in the halo of M87 is much higher, about 750 km/s. This is also much higher than the orbital

velocities of stars in the visible portion of M87, and thus the orbital velocities in M87 are not independent of radius over the span from the luminous portions of the galaxy to the outer halo.

One important question is whether other elliptical galaxies have massive dark haloes. This question will be addressed in Section 5.8.3.

5.8.2 Other models for M87 and other central galaxies

In Sections 5.7 and 5.8.1, a standard model for M87 (and other X-ray emitting central galaxies) has been given, in which there is a cooling flow leading to accretion by M87. The outer parts of the flow are very subsonic, and the gas is nearly hydrostatic and bound either by the massive halo of M87 or by the pressure of surrounding gas. One feature of this standard model is that thermal conduction must be suppressed in order to ensure that the losses due to cooling in the center of the galaxy are balanced by the enthalpy flux of inflowing gas rather than by heat conducted inwards. Two alternative models are now discussed in which conduction is not suppressed.

Takahara and Takahara (1979, 1981) suggested that gas was actually being thermally evaporated and flowing out from M87, rather than flowing inward. They argued that the gas in the evaporative flow was supplied by mass loss from stars in M87, at a rate of about $1 M_{\odot}/\text{yr}$. This gas is heated by thermal conduction from surrounding, extended hot intracluster gas, in which the galaxy is assumed to be immersed.

One problem with this model is that the range of temperatures in the gas is rather small, and the gas is generally hotter than 2×10^7 K. This model cannot produce the very strong soft X-ray line emission seen in M87 (Canizares *et al.*, 1979, 1982; Lea *et al.*, 1982; Stewart *et al.*, 1984a). It does not provide a very good fit to the X-ray surface brightness in the inner regions of the galaxy, and does not explain the origin of the line emitting filaments in M87 (Section 5.7.3).

Tucker and Rosner (1983) suggested that the gas in M87 (and other central dominant galaxies) is actually hydrostatic (not moving). In their model, the cooling of the gas is balanced by heating from thermal conduction in the outer regions and by heating by the relativistic electrons associated with the radio source in the inner parts. As discussed in Section 5.7.1, the extra heating by relativistic electrons actually results in thermal equilibrium being attained at a lower temperature, since cooling increases rapidly as the temperature is reduced. Such a model is necessarily thermally unstable (Stewart *et al.*, 1984a), and might form a cooling flow due to the growth of thermal fluctuations. This model requires a fairly large rate of heating from relativistic electrons (Scott *et al.*, 1980; Section 5.3.5), and there is considerable uncertainty about such large heating rates. The model does have the advantage that the similarity in the morphology of the X-ray and diffuse radio emission in M87 is explained.

In Section 5.7.2 the suggestion was made that the radio emission of central cluster galaxies was powered by having a small portion of the accreting gas reach the nucleus. In the Tucker and Rosner model, the gas is static. Thus there is no inflow to power the radio source. Tucker and Rosner suggest that these sources might be episodic. Initially, gas cooling in the core of the cluster might produce an accretion flow on the central galaxy, and start the radio source. The radio source would produce large numbers of relativistic electrons, which would heat the gas and turn off the accretion flow. This would stop the radio source, and the heating would decline as the relativistic electrons lost their energy. This would allow the accretion flow to restart, and the process might oscillate in this fashion indefinitely.

5.8.3 X-ray emission from noncentral cluster galaxies

Recent analyses of the X-ray observation of early-type galaxies made with the *Einstein* X-ray observatory show that many non-cD ellipticals which are not at the centers of rich clusters are moderately strong X-ray sources (Nulsen *et al.*, 1984; Forman *et al.*, 1985; Trinchieri and Fabbiano 1985; Canizares *et al.*, 1987; Trinchieri *et al.*, 1986). The X-ray luminosities range from $L_x \approx 10^{39} - 10^{42}$ erg/s. Many of these X-ray emitting elliptical and S0 galaxies are in the Virgo cluster (Forman *et al.*, 1979; Forman and Jones, 1982; Section 4.5.3). A number of peaks in the X-ray surface brightness of A1367 correspond to the positions of galaxies (Bechtold *et al.*, 1983; Section 4.5.4). There is a strong correlation between the optical and X-ray luminosities of these galaxies, with $L_x \propto L_B^{1.5-2.0}$, where L_B is the optical (blue) luminosity (Forman *et al.*, 1985; Trinchieri and Fabbiano 1985). The X-ray emission is spatially extended, with a typical maximum radius for the brighter galaxies being $R_x \approx 50$ kpc, and the X-ray surface brightness is reasonably fit by $I_x \propto r^{-2}$ in the outer parts, where r is the radius from the galaxy center (Forman *et al.*, 1985).

There are a number of strong arguments, given in the papers just cited, which indicate that the X-rays arise as thermal emission from hot, diffuse gas. In the lower luminosity ellipticals, emission by binary

X-ray sources may also play a role. The crude X-ray spectral information available suggests that the gas temperatures are on the order of $T_g \approx 10^7$ K (Forman *et al.*, 1985). The observed radial variation of the X-ray surface brightness requires that the gas density vary roughly as $\rho_g \propto r^{-3/2}$. The required gas densities in the inner parts are $\gtrsim 0.1$ atom/cm³. The total mass of hot gas in these galaxies is roughly $M_g \approx 10^9 - 10^{11} M_\odot$ (Forman *et al.*, 1985). Assuming a stellar mass-to-light ratio of $(M/L_B)_\star \approx 8M_\odot/L_\odot$, the ratio of gas mass to stellar mass is $M_g/M_\star \approx 0.02$.

Previous to the detection of this hot gas, ellipticals were believed to be gas-poor systems. It was assumed that the gas ejected by stars in ellipticals was effectively heated by supernovae, forming a strong galactic wind which quickly removed the gas from the galaxy (Mathews and Baker 1971). The X-ray observations show that gas is *not* being removed rapidly from these galaxies; apparently, the galactic winds do not generally occur, at least at the present time. The expected X-ray luminosities from winds are $\lesssim 10^{-4}$ of those observed in elliptical galaxies.

A simple model which fits the properties of these X-ray emitting galaxies is that the bulk of the gas forms a cooling flow, in which gas lost by stars in the galaxy is heated by supernovae, the motions of the gas-ejecting stars, and adiabatic compression, and then cools and slowly flows into the galaxy center (Nulsen *et al.*, 1984; Fabian *et al.* 1986b; Canizares *et al.*, 1987; Sarazin 1986b,c). There are several simple arguments which support this cooling flow model. First, the amount of gas seen is consistent with the current rates of stellar mass loss in ellipticals. The rate of stellar mass loss is proportional to the stellar density ρ_* , so that $\dot{\rho}_g = \alpha_* \rho_*$. For the stellar population in ellipticals, α_* is calculated to be $\alpha_* \approx 4.7 \times 10^{-20} \text{ s}^{-1} = 1.5 \times 10^{-12} \text{ yr}^{-1}$ (Faber and Gallagher 1976). Thus, over $\approx 10^{10}$ years, the total mass of gas ejected at the present rate is $M_g \approx 0.015 M_*$, which is consistent with the observed gas masses. Second, the cooling times in the gas are quite short. Near the centers of the observed X-ray images, the gas densities and temperatures derived from the observations indicate that $t_{cool} \lesssim 10^7$ yr, and the cooling times typically reach 10^{10} yr only at the very edge of the observed X-ray images. Thus, any gas introduced into the galaxy will tend to cool. Even if the gas could be heated sufficiently to balance the average cooling rate, gas at temperature $\approx 10^7$ K is very thermally unstable (Section 5.7.3), and it would form clumps which would cool rapidly. Third, cooling flow models may lead to a natural explanation for the correlation observed between the X-ray and optical luminosities of elliptical galaxies (Nulsen *et al.*, 1984; Canizares *et al.*, 1987; Sarazin, 1986b,c). Finally, the observed radial variation of the X-ray surface brightness of ellipticals can be understood under the cooling flow hypothesis (Sarazin, 1986b,c).

One important aspect of these models is that the X-ray luminosities and their correlation with the optical luminosities can only be understood if the heating of the gas is primarily due to the motions of the gas-losing stars and to adiabatic compression during the inflow, and not due to supernova heating (Canizares *et al.*, 1987; Sarazin, 1986b,c). This implies that the supernova rates in elliptical galaxies are much smaller than had previously been thought. The low supernova rates would also explain why these galaxies lack galactic winds.

In Section 5.5.5, a method was described which allows the mass in galaxies or clusters to be determined if they contain hot, hydrostatic gas. One very important application of this method would be to measure the masses of elliptical galaxies out to large distances. At present it is unclear whether elliptical galaxies possess extended dark matter. X-ray observations of M87 at the center of the Virgo cluster show that it does have such a halo (Bahcall and Sarazin 1977; Mathews 1978; Fabricant *et al.*, 1980; Fabricant and Gorenstein 1983; Section 5.8.1), but it is unclear whether this mass is associated with the cluster center or with M87. Dark haloes have been deduced for spiral galaxies from the rotational velocities of the neutral hydrogen in the disks of the galaxies far outside of their optically luminous regions. Elliptical galaxies do not possess much neutral hydrogen, and thus this technique cannot be applied to them. The masses of elliptical galaxies can be determined from the orbital velocities of their stars (actually the line-of-sight component of the stellar velocity dispersion; see Dressler, 1979, 1981). Unfortunately, these observations are very difficult and cannot be done in the outermost portions of the galaxies. Moreover, the masses are uncertain because the shapes of the orbits of the stars are not known (they may be radial or isotropic, for example; see Section 2.8). The orbital velocities of globular star clusters can also be used to determine the masses of elliptical galaxies (Hesser *et al.*, 1984). Masses can be derived from studies of binary galaxies, but here the orbital characteristics are even more uncertain (Faber and Gallagher, 1979).

The use of X-ray emitting gas to measure the masses of elliptical galaxies out to large distances has a

number of important advantages (Section 5.5.5). For example, this gas can be observed out to very large distances from the center of the galaxy. Also, the orbits of gas particles are known to be isotropic because the gas is a collisional fluid. Thus X-ray measurements can give important information on the possible existence of massive haloes around elliptical galaxies.

Forman *et al.* (1985) attempted to derive mass profiles by applying the hydrostatic method to the X-ray observations of these normal elliptical galaxies. They concluded that these galaxies do indeed have heavy haloes. However, the errors in the temperature determinations for these galaxies are quite large, and the derived masses are very strongly affected by errors in the temperatures (Section 5.5.5). Trinchieri *et al.* (1986) find that the temperature errors are so large that the masses cannot be determined with sufficient accuracy to decide whether normal elliptical have missing mass haloes. In any case, this is an ideal problem for a future X-ray observatory, such as AXAF (Chapter 6).

The environment of an early type galaxy may also affect the distribution of its X-ray emitting gas. Intracluster gas might either aid in the retention of gas in a galaxy by providing a confining pressure, or aid in the removal of the gas through ram pressure ablation and other stripping processes (Section 5.9). Of particular interest in this regard are M84 and M86, which are two of the most X-ray luminous ellipticals in the Virgo cluster (Figure 26), and the X-ray emitting galaxies in A1367. In A1367, 11 galaxies were detected in X-rays by Bechtold *et al.* (Section 4.5.4; Figure 27b), of which 8 were found to be spatially extended. The luminosities of these galaxies range from $L_x \approx 1 - 7 \times 10^{41}$ erg/s. It is possible that pressure confinement plays a role in the X-ray emission from these galaxies (Forman *et al.*, 1979; Fabian *et al.*, 1980; Bechtold *et al.*, 1983). The galaxies detected in A1367 have unusually high X-ray luminosities for their optical luminosities, and do not show the correlation between X-ray and optical luminosities seen in other elliptical galaxies. This suggests that the gas is not gravitationally bound, but rather is confined by the intracluster gas in A1367 (Bechtold *et al.*, 1983). However, the density of the intracluster gas in A1367 is comparable to that in the Virgo cluster, where the galaxies have normal X-ray luminosities. Alternatively, Canizares *et al.* (1987) have suggested that the clumps of X-ray emission attributed by Bechtold *et al.* to galaxies are really just fluctuations in the intracluster gas, and are only projected near galaxies by coincidence.

In Virgo, the X-ray emission from M86 is more extended than the emission from M84, and forms a plume extending to the north of the galaxy (Figure 26). Based on their radial velocities, M84 appears to be moving fairly slowly, while M86 is moving rapidly. A galaxy that is moving slowly through the intracluster medium might be able to retain more of the gas produced by stellar mass loss, while a more rapidly moving galaxy would tend to lose its gas due to stripping (Section 5.9; Forman *et al.*, 1979; Fabian *et al.*, 1980; Takeda *et al.*, 1984). Because its velocity is considerably larger than the average in the Virgo cluster, Forman *et al.* and Fabian *et al.* suggest that the orbit of M86 will carry it far outside the cluster core, where the intracluster gas density is low. In these outer regions of the cluster, stripping of gas may be ineffective, and the galaxy will accumulate gas (Takeda *et al.*, 1984). Forman *et al.* and Fabian *et al.* estimate an orbital period for M86 of roughly 5×10^9 yr. Thus the galaxy could amass roughly $5 \times 10^9 M_\odot$ of gas during each orbit. This gas would be stripped during each passage through the cluster core. In this interpretation, the plume to the north of M86 is the trail of gas being stripped from M86 as it enters the core of the Virgo cluster (Forman *et al.*, 1979; Fabian *et al.*, 1982).

5.9 Stripping of gas from galaxies in clusters

The galaxies found in rich clusters generally have considerably less gas than galaxies in the field (Section 2.10.2). Their 21 cm radio line luminosities indicate that they have less neutral gas than more isolated galaxies (see Section 3.7 and the many references therein), and they also appear to have weaker optical line fluxes from ionized gas (Gisler, 1978). The galaxies in rich, regular, X-ray luminous clusters are predominantly ellipticals and S0s (Sections 2.5 and 2.10.1, Table 1), and the spirals that are seen in clusters often have weak ('anemic') spiral arms (van den Bergh, 1976) and tend to be found at large projected distances from the cluster center (Gregory, 1975; Melnick and Sargent, 1977). The fraction of spiral galaxies is anticorrelated with the X-ray luminosity of the cluster (Bahcall, 1977c; Melnick and Sargent, 1977; Tytler and Vidal, 1979; Figure 30; equation 4.9) and with the local galaxy density (Dressler, 1980b). A primary difference between spiral galaxies and ellipticals or S0s is that spirals contain large amounts of gas.

These observations suggest that gas in galaxies in clusters may be removed by some process. In Section 2.10.2 the possibility that the differences in the galaxian populations of clusters and the field might result from the stripping of gas from spiral galaxies was discussed. This hypothesis is very controversial, and it

appears unlikely that spiral stripping alone accounts for all the differences between different types of galaxies. Nonetheless, even within individual galaxy classes (Hubble types), galaxies in clusters appear to have less gas (Sullivan *et al.*, 1981; Giovanardi *et al.*, 1983; Section 3.7). This suggests that some process must remove gas from galaxies in clusters, even if this stripping does not solely determine the morphology of galaxies.

In addition to this indirect evidence for gas removal, the X-ray observations of M86 suggest that it is currently being stripped of gas (Forman *et al.*, 1979; Fabian *et al.*, 1980; Sections 4.5.3 and 5.8.3). There are also a number of cases where optical or 21 cm radio observations may show gas currently being stripped from galaxies (see, for example, Gallagher, 1978).

Mathews and Baker (1971) showed that, once an isolated elliptical or S0 galaxy had been stripped of its gas, supernova driven winds would keep it gas free. If the interstellar gas density is low, then the energy input from supernovae cannot be radiated away, and heats the gas until it flows out of the galaxy (Sections 5.3.3 and 5.6). If the galaxy is immersed in intracluster gas, this mechanism is less effective, since the wind must overcome the confining pressure of the intracluster gas.

Spitzer and Baade (1951) suggested that collisions between spiral galaxies in the cores of compact clusters remove the interstellar medium from the disks of these galaxies. Their basic argument was that the stellar components of galaxies could pass through one another, because two-body gravitational interaction between galaxies or their component stars is very ineffective (Section 2.9.1). However, the mean free paths of gas atoms or ions (Section 5.4.1) are relatively short, and in the center of mass frame for the collision, the kinetic energy of the gas will be thermalized. If the gas cannot cool rapidly, it will be heated to a temperature at which it is no longer bound to either galaxy, since galaxies within a cluster move much more rapidly than stars move within galaxies. If the gas cools rapidly, it will still be left at rest in the center of mass frame, while the galaxies move away from it. In either case, the gas is no longer bound to either galaxy, since it has more kinetic energy (either thermal or bulk) than the maximum galaxy binding energy.

Spitzer and Baade estimated the rate of gas removal in the Coma cluster assuming all the galaxies were equivalent, and that the galaxies all had radial orbits, and found that a galaxy should be stripped in less than 10^8 yr. This is probably an underestimate of the time scale, since galaxy orbits are probably not radial, and a typical off-center collision between two galaxies of dissimilar mass and density may not remove all of the gas from both galaxies. Moreover, subsequent increases in the extragalactic distance scale have had the effect of increasing the time scale for this process. Sarazin (1979) estimated the mean time scale for collisional stripping of galaxies in a cluster, averaging over the cluster velocity dispersion, the spatial distribution of galaxies, and the diameters and masses of galaxies, and found a rate about 100 times smaller.

Gunn and Gott (1972) suggested that galaxies lose their interstellar gas by ram pressure ablation because of the rapid motion of the galaxies through the intracluster gas. They were primarily concerned with stripping gaseous disks from spiral galaxies to make S0s. Based on a static force balance argument, they suggested that a gaseous disk would be removed when the ram pressure $P_r = \rho_g v^2$ exceeded the restoring gravitational force per unit area in the disk $2\pi G \sigma_D \sigma_{ISM}$ (where ρ_g is the intracluster gas density, v is the galaxy velocity, σ_D the surface density of the spiral disk, and σ_{ISM} the surface density of interstellar gas in the disk). Taking $v^2 = 3\sigma_r^2$ for a typical galaxy in a cluster with a line-of-sight velocity dispersion σ_r , and assuming the disk has a uniform surface density, with a radius r_D and mass M_D , this condition becomes

$$\left(\frac{n_g}{10^{-3} \text{ cm}^{-3}} \right) \left(\frac{\sigma_r}{10^3 \text{ km/s}} \right)^2 \gtrsim 3 \left(\frac{M_D}{10^{11} M_\odot} \right)^2 \left(\frac{r_D}{10 \text{ kpc}} \right)^{-4} \left(\frac{M_{ISM}}{0.1 M_D} \right), \quad (5.114)$$

where n_g is the number density of atoms in the intracluster medium and M_{ISM} is the mass of interstellar medium in the disk.

Tarter (1975) and Krtsuk (1983) have given a simple semianalytical treatment of the stripping of gas disks in spiral galaxies in clusters by integrating the net force due to ram pressure and gravity, assuming one-dimensional motion. For a given intracluster gas distribution, they found the smallest distance from the cluster center at which a galaxy could retain its gas. The calculations by Tarter indicated that most spirals should be stripped fairly easily in X-ray clusters, and that remaining spirals would only be found in the outer parts. On the other hand, Krtsuk suggested that molecular clouds might be very difficult to strip from spirals.

It is interesting to calculate the expected dependence of the spiral fraction and the typical distance of spirals from the cluster center on the cluster X-ray luminosity, if one assumes that all clusters originally

had the same spiral-rich galactic populations at all positions. If one also assumes that the reduction in the fraction of spirals is due to ram pressure ablation, that a spiral is stripped whenever the ram pressure exceeds a critical amount (given by equation 5.114 or something similar), that the galaxies in clusters have an analytic King isothermal distribution (equation 5.57), and that the gas is isothermal and hydrostatic (equation 5.63), then the spiral fraction varies as $f_{Sp} \approx A - B \log L_x$, where L_x is the X-ray luminosity and A and B are constants. This is just the empirical relationship found by Bahcall (1977c) (equation 4.9; Figure 30). The radius of a typical spiral varies as $r_{Sp}/r_c \propto (L_x)^{1/6\beta}$, where r_c is the core radius and β is the ratio of gas and galaxy temperatures (equation 5.64).

The majority of galaxies presently in X-ray clusters are ellipticals and S0s that have a more spherical stellar distribution. Sarazin (1979) gave a semi-analytic treatment of the stripping of gas from spherical galaxies, using basically the same formulation as Tarter (1975). He found that galaxies would be stripped if the ram pressure were greater than about $2GM_{gal}\sigma_{ISM}/R^2$, where M_{gal} and R are the galaxy mass and radius, respectively. Takeda *et al.* (1984) performed numerical hydrodynamic simulations and derived a critical ram pressure of $2\rho_{ISM}\sigma_*^2$. Assuming that the ram pressure is much greater than this limit, the time scale for ram pressure stripping of a galaxy, defined as $t_{rp} \equiv (d \ln M_{ISM}/dt)^{-1}$, is

$$\begin{aligned} t_{rp} &\approx \frac{R}{v} \left(\frac{2\rho_{ISM}}{\rho_g} \right)^{1/2} \\ &\approx 3 \times 10^7 \text{ yr} \left(\frac{\rho_{ISM}}{\rho_g} \right)^{1/2} \left(\frac{v}{10^3 \text{ km/s}} \right)^{-1} \left(\frac{R}{20 \text{ kpc}} \right), \end{aligned} \quad (5.115)$$

where ρ_{ISM} is the average interstellar gas density, ρ_g is the intracluster gas density, v and M_{gal} are the galaxy velocity and mass, respectively. The calculations of Tarter and Sarazin assume that there are no sources of gas in the galaxy.

Gisler (1976, 1979) pointed out that it would be more difficult to strip a galaxy if the stars in the galaxy were constantly resupplying it with gas. Then, the ram pressure must overcome the momentum flux due to this mass input, as well as the gravitational attraction of the galaxy. He derived an approximate analytical relationship for the ram pressure needed to strip gas out of a galaxy out to a projected radius b , ignoring pressure forces in the gas and assuming one-dimensional motion. If the rate of gas loss by stars in the galaxy is $\dot{\rho}_{ISM} = \alpha_* \rho_*$ where ρ_* is the mass density of stars in the galaxy, then he found that gas would be stripped at any projected radius b such that the ram pressure exceeded $\approx 8\alpha_* \Sigma_*(b)\sigma_*$. Here, $\Sigma_*(b)$ is the projected stellar mass density at b , and σ_* is the line-of-sight velocity dispersion of stars in the galaxy. For a typical giant elliptical galaxy in a typical X-ray cluster, Gisler found that all the gas is lost if the rate of input is less than $0.1 M_\odot/\text{yr}$, and that all the gas is retained if it is greater than about $3 M_\odot/\text{yr}$. For intermediate values, a core of interstellar gas is retained by the galaxy.

Gisler's treatment ignored pressure forces, which must be important in elliptical galaxies (Jones and Owen, 1979; Takeda *et al.*, 1984). Jones and Owen argued that extended gas in elliptical galaxies must be hot (the sound speed in the gas must be comparable to the velocity dispersion of stars), and that pressure forces in the gas cannot be ignored. A pressure gradient may be set up in the interstellar gas that opposes the ram pressure; they argued that this increases the region of a galaxy that is shielded from stripping by about a factor of ten over Gisler's result. Then elliptical galaxies will typically retain cores of interstellar gas with radii of about 10 kpc, and Jones and Owen argue that this gas can explain the existence of well-defined radio jets in the inner parts of head-tail radio sources (Section 3.3; Figure 9).

Takeda *et al.* (1984) recently rederived Gisler's condition for continuous stripping of gas from galaxies with stellar gas loss, including pressure forces. They find that the correct condition for stripping is that the ram pressure exceed

$$\rho_g v^2 \gtrsim \frac{2}{v} \int \alpha_* \rho_* |\phi_*| dl \approx \frac{16\alpha_* \Sigma_*(b)\sigma_*^2}{v}, \quad (5.116)$$

where ϕ_* is the galaxy gravitational potential. Equation (5.116) differs from Gisler's result by a factor of $\approx 2\sigma_*/v$.

Most of these calculations of stripping have been based on semianalytic estimates, which assume the motion to be one-dimensional, generally ignore the finite temperature and compressibility of the gases, and ignore any viscous forces. To avoid these assumptions and test the efficiency of ram pressure ablation, a

large number of numerical hydrodynamic simulations have been made by Gisler (1976), Lea and De Young (1976), Toyama and Ikeuchi (1980), and Nepveu (1981a). These studies have all taken the galaxy to be spherical, and the simulations have been two-dimensional, with the flows assumed to be axially symmetric. Lea and De Young started with a galaxy with a significant amount of interstellar gas, moving in the midst of intracluster gas. On the other hand, Gisler was mainly interested in the effects of stellar mass loss on ram pressure ablation, and he started with an empty galaxy. The numerical calculations indicate that stripping occurs even more easily than the simple analytic force balance arguments suggest. Unless the rate of mass input due to stars is high, a typical galaxy will be stripped almost completely during a single passage through the core of a cluster if the intracluster gas density exceeds roughly 10^{-4} atom/cm³.

Recently, Shaviv and Salpeter (1982) and Gaetz *et al.* (1985) made two-dimensional hydrodynamic calculations of ram pressure ablation for galaxies with stellar mass loss, including the cooling of the gas. Gaetz *et al.* also included star formation. They found that gas was ablated from the outer portions of the galaxy, but was retained in the inner portions and formed a cooling flow. One very useful feature of Gaetz *et al.* is that the hydro simulations were used to derive analytic fitting formulas for many physical quantities associated with the gas flows.

All of these numerical calculations start with the galaxy in the middle of a uniform cluster. Recently, Takeda *et al.* (1984) calculated the stripping of gas from a galaxy moving on a radial orbit from the outer parts of the cluster into the core. Stellar mass loss was assumed to add to the interstellar gas in the galaxy. The stripping was determined from two-dimensional hydrodynamic simulations. After the first passage through the cluster core, the behavior was periodic, with the galaxy accumulating interstellar gas when far from the core and losing nearly all of it when it passed through the core on each orbit. A large fraction of the accumulated interstellar gas was pushed out in a single ‘blob’ on each passage through the cluster core. These calculations may provide a model for the galaxy M86, other gas containing galaxies, and the galaxyless extended X-ray sources in A1367 (Sections 4.5.3, 4.5.4 and 5.8.3). The latter might be blobs released from stripped galaxies (Takeda *et al.*, 1984).

Under many circumstances, transport processes such as viscosity and turbulent mixing can be more important than ram pressure in removing gas from a galaxy in a cluster (Livio *et al.*, 1980; Nepveu, 1981b; Nulsen, 1982). Nulsen has calculated the rate of laminar viscous stripping and stripping due to turbulence, and finds a viscous stripping time scale

$$t_{vs} \approx \frac{4R}{3v} \left(\frac{\rho_{ISM}}{\rho_g} \right) \left(\frac{12}{Re} + 1 \right)^{-1}, \quad (5.117)$$

where Re is the Reynolds number (equation 5.47). In many cases, this is faster than the rate of ram pressure ablation given by equation (5.115). However, one caution is that the viscous stresses may saturate, since the mean free paths of ions are similar to the sizes of galaxies (Section 5.4.4).

Livio *et al.* (1980) suggest that the Kelvin–Helmholtz instability at the boundary between interstellar and intracluster gas can produce ‘spikes’ of interstellar gas protruding into the intracluster gas, which are sheared off, increasing the stripping rate. They estimate a mass loss rate roughly $\dot{M}_{KH} \approx \rho_{ISM} \pi R^2 \lambda_{KH} / t_{KH}$, where $\lambda_{KH} \approx 10^{21}$ cm and $t_{KH} \approx 10^6$ yr are the wavelength and growth time of the fastest growing modes. This stripping rate is considerably smaller than that given by equation (5.117). Moreover, Nulsen (1982) has argued that Livio *et al.* overestimated this rate because they ignored the effect of compressibility on the instability, and because they did not include the effect of the mass loss on the instability. Another problem is that the wavelength of the fastest growing mode is shorter than the ion mean free path $\lambda_{KH} \ll \lambda_i$, and the viscosity expression used is therefore not valid (Section 5.4.4). Nulsen showed that the instability was suppressed when viscosity is significant, and that the mass loss is not controlled by the fastest growing modes but by the induced velocities on the largest scales.

Another mechanism for removing gas from galaxies, which can operate even when the galaxies are moving slowly through the intracluster gas, is evaporation (Gunn and Gott, 1972; Cowie and McKee, 1977; Cowie and Songaila, 1977). Heat is conducted into the cooler galactic gas from the hotter intracluster gas, and if the rate of heat conduction exceeds the cooling rate, the galactic gas will heat up and evaporate. If cooling is assumed to be small, the evaporation rate with unsaturated conduction for a spherical galaxy

immersed in intracluster gas of temperature T_g is (Cowie and Songaila, 1977)

$$\begin{aligned}\dot{M}_{ev} &\approx \frac{16\pi\mu m_p \kappa R}{25k} \\ &\approx 700 M_\odot/\text{yr} \left(\frac{T_g}{10^8 \text{K}}\right)^{5/2} \left(\frac{R}{20 \text{kpc}}\right) \left(\frac{\ln \Lambda}{40}\right)^{-1},\end{aligned}\quad (5.118)$$

where κ is the thermal conductivity (Section 5.4.2; equation 5.37) and Λ is the Coulomb logarithm (equation 5.33). The stripping rate is a factor of $(2/\pi)$ smaller for a disk galaxy with the same radius. The evaporation rate will be significantly reduced if the conductivity saturates (Section 5.4.2), as is probably the case at least for disk galaxies. Unfortunately, thermal conductivity also depends critically on the magnetic field geometry (Section 5.4.3). If the conductivity is not suppressed by the magnetic field, this mechanism can play an important role in stripping gas from galaxies.

As pointed out by Nulsen (1982), there is a simple connection between mass loss by evaporation and mass loss by laminar viscosity (the Re term in equation 5.117). At low velocities, this term dominates, and the viscous stripping rate is nearly independent of velocity, because the Reynolds number is proportional to velocity. Since both thermal conduction and ionic viscosity are transport processes and the ion and electron mean free paths are essentially equal, the rates of stripping from these two processes are simply related:

$$\dot{M}_{ev} \approx 3.5 \dot{M}_{vs}. \quad (5.119)$$

While this expression was derived for unsaturated conduction and viscosity and no magnetic suppression of either, it probably will remain approximately true even when these effects are included, since both thermal conduction and viscosity are affected similarly.

5.10 The origin and evolution of the intracluster medium

Where does all of this X-ray emitting intracluster gas come from? Did it fall into clusters from intergalactic space, or was it ejected from galaxies? When did it first fill the great volumes of space between galaxies in clusters? What can we learn about the origin of galaxies and clusters from the intracluster gas? Theories of the origin of the intracluster medium attempt to explain the current properties of X-ray clusters, to make predictions about their past history that can be tested with observations of clusters at high redshift, and to relate the history of the intracluster gas to theories of the origin of structure in the universe.

There are two basic constraints on such theories from the observations of present day X-ray clusters. First, these clusters typically contain $\approx 10^{14} M_\odot$ of intracluster gas, an amount at least comparable to the mass of luminous material in galaxies and at least one-tenth of the total mass of the cluster. Second, this gas produces X-ray lines from iron which require that the abundance of that element be about one-half of its solar value if the gas is homogeneous.

At present, there are very few observations of high redshift X-ray clusters, and thus the constraints that can be imposed on theories of the evolution of X-ray clusters by these observations are fairly weak. The only safe statement one can make at the present time is that X-ray clusters do not show any evidence of very dramatic evolution out to redshifts of $z \approx 0.5$ (Section 4.8). Another constraint on the evolution of X-ray clusters comes from the diffuse X-ray background (Fabian and Nulsen, 1979); the total emission from high redshift clusters cannot exceed the background brightness.

The question of the origin of the intracluster gas is strongly tied to the question of the origin of clusters and galaxies. This is a major area of astronomical research, and I shall not even attempt to summarize the theories in this area. Instead, I shall deal very narrowly with theories of the intracluster medium and shall largely ignore their dependence on models for galaxy formation, except when the X-ray cluster models provide information that affects these theories. Discussions of the effects that X-ray cluster observations have on theories of galaxy formation include Silk (1978), Binney and Silk (1978), White and Rees (1978), Fabian and Nulsen (1979), Binney (1980), and Field (1980).

There is an increasingly large amount of observational information on the evolution of galaxies; here only a few points will be noted. First, galaxies contain stars with a range of ages and heavy element abundances. Elliptical and S0 galaxies contain mainly older stars, with ages of around 10^{10} yr (Fall, 1981). Although galaxies do contain some stars with very low heavy element abundances, there is no significant population of stars with *no* heavy elements, and the number of low abundance stars is less than might be expected if all the

heavy elements were formed in stars like the present stellar population (Carr *et al.*, 1984). No mechanism is known to produce heavy elements in any quantity outside of stars. This has led to the suggestion that there was an early (before 10^{10} yr ago) generation of stars, which produced the minimum level of heavy elements seen in stars today (Carr *et al.*, 1984). Second, there is some evidence that the population of galaxies has evolved in some clusters since redshifts of $z \approx 0.5$; this is the ‘Butcher–Oemler effect’ (Section 2.10.2). These clusters have an excess of blue galaxies, which suggests that they were undergoing fairly rapid star formation at that time. Finally, violent activity in the nuclei of galaxies appears to have been much more common in the past (Schmidt, 1978); this activity produces Seyfert galaxies, radio galaxies, and quasars. There were many more very luminous quasars at a time corresponding to a redshift of 2 than are seen around us today. On the other hand, since few quasars are known with redshifts greater than 3.5 (Osmer, 1978), something may have occurred then to start the violent nuclear activity in galaxies. Also, if quasars are all located in the nuclei of galaxies, then galaxies must have formed by a redshift of 3. Although this nuclear activity in galaxies is very poorly understood, most current theories require that gas be supplied to the nucleus (Rees, 1978); thus its variation may be related to the supply of intracluster gas and the stripping of gas from galaxies.

5.10.1 Infall models

Gunn and Gott (1972) suggested that the intracluster medium was primordial intergalactic gas that had fallen into clusters. This intergalactic gas was never associated with stars or galaxies, and thus could be expected to have no heavy elements. They also noted that some of the intracluster medium could come from ram pressure stripping of interstellar gas out of galaxies. Assuming that clusters were immersed in uniform intergalactic medium, they estimated the amount that would fall into clusters. By comparing this to the amount of intracluster gas deduced from the early X-ray observations of clusters, they could give an upper limit on the density of the intergalactic gas. (This is an upper limit because some of the intracluster gas could have come out of galaxies.) These limits are usually given in terms of ρ_c , the density of matter necessary to close the universe

$$\rho_c \equiv \frac{3H_o^2}{8\pi G} = 4.7 \times 10^{-30} h_{50}^2 \text{ gm/cm}^3, \quad (5.120)$$

where H_o is the Hubble constant. If ρ_{IG} is the density of intergalactic gas, then Gunn and Gott found $\rho_{IG}/\rho_c \lesssim 0.01$. They used a rather low value for the gas mass in clusters, and more recent calculations (for example, Cowie and Perrenod, 1978) give $\rho_{IG}/\rho_c \lesssim 0.2$. Gunn and Gott also noted that infall would heat the intracluster gas to about the observed temperatures (Section 5.3.2).

To determine the final configuration and evolution of the intracluster gas in the infall model, hydrodynamic simulations of the infall have been done by a number of authors (Gull and Northover, 1975; Lea, 1976; Takahara *et al.*, 1976; Cowie and Perrenod, 1978). These calculations all assumed that the cluster potential was fixed; the gas was assumed to fall into the cluster after the cluster itself had collapsed. All of these calculations were one-dimensional simulations of spherical clusters, although a number of different techniques were used to solve the hydrodynamic equations. With the exception of Lea’s calculations, these simulations have given similar results.

As the gas first collapses into the core, its density increases and a shock propagates outward from the cluster center and heats the gas. This shock passes through the cluster in $\approx 10^9$ yr, essentially the sound crossing time for the cluster (equation 5.54). After the passage of the shock, the hot intracluster gas is nearly hydrostatic, and its further evolution is quasistatic. As the shock moves into the outer parts of the cluster, it weakens; less gas is added to the cluster, and the cluster luminosity is nearly constant. On the other hand, Lea found that the shock heating caused the gas pressure to increase until the inflow was reversed and the intracluster gas expanded. This cooled the gas adiabatically, lowered its pressure, and it collapsed again. This process repeated itself, producing a large number of pulsations with a period of about 5×10^9 yr. During these pulsations, the X-ray luminosity oscillated wildly between roughly 10^{41} and 10^{48} erg/s. The other calculations of the infall of intracluster gas have failed to find these oscillations (Gull and Northover, 1975; Takahara *et al.*, 1976; Cowie and Perrenod, 1978; Perrenod, 1978b), and they are probably an artifact of Lea’s numerical method. Such oscillations are in violent disagreement with the observed X-ray luminosity function of clusters (Schwartz, 1978).

Gull and Northover (1975) found that the shock strength was nearly constant as it propagated outward; they argued that this occurred because the shock speed was always about the free-fall speed in the cluster.

They found that the resulting intracluster gas distribution was nearly adiabatic (Section 5.5.2). On the other hand, more detailed calculations by Cowie and Perrenod (1978) and Perrenod (1978b) found that the cluster gas distributions were not well represented by any polytropic distribution, unless thermal conduction was so effective that the gas was isothermal.

In the absence of significant cooling or thermal conduction, Cowie and Perrenod (1978) showed that the infall models with a fixed cluster potential are characterized by a single parameter, which gives the depth of the cluster potential well, $K \equiv (\sigma_r/H_0 r_c)^2$ where σ_r and r_c are the cluster velocity dispersion and core radius, respectively. Models with significant cooling are also characterized by $B \equiv t_{cool} H_0$, where the cooling time is evaluated at the cluster center. If thermal conduction is important, the cluster evolution is also determined by the value of $C \equiv \kappa T_g / r_c \rho_g c_s^3$, where κ is the thermal conductivity, and T_g , ρ_g , and c_s are the gas temperature, density, and sound speed. When conduction saturates, the models are independent of C . In general, the gas temperatures in these models scale with σ_r^2 .

Cowie and Perrenod found that models without significant cooling or conduction showed a very small decrease in the X-ray luminosity with time, less than a 40% reduction from $z = 1$ to the present. This decrease in luminosity resulted from the slow reexpansion of the intracluster gas as the shock weakened. In models with significant cooling, the cluster evolved to a nearly steady-state cooling flow (Section 5.7). In models with conduction, the X-ray luminosity increased slowly with time, by about 40% from a redshift $z = 1$ to the present. This occurred because conduction lowered the temperature in the cluster core (Section 5.4.2). The core then contracted so that the increasing density could maintain the pressure in the core. Since the X-ray luminosity increases more rapidly with density than does the pressure (equation 5.21), the luminosity went up.

These models all assumed that the cluster potential was static; the cluster was assumed to collapse before any gas fell into it. Of course, there is no reason why intergalactic gas should wait until the cluster has formed before gaseous infall can occur. Perrenod (1978a,b) calculated the evolution of the intracluster gas in infall models in which the cluster potential varied in time. The cluster potential was taken from White's (1976c) N-body calculations of the collapse of a Coma-like cluster (Section 2.9; Figure 5). In White's models, the cluster first collapses with violent relaxation, then contracts slowly due to two-body interactions between galaxies. This contraction causes the cluster potential well to become deeper, and as a result the intracluster gas temperature and density increase with time. In contrast to the static potential models, Perrenod finds that the X-ray luminosity of his model clusters increases by about an order of magnitude from $z = 1$ to the present. The sizes of the gas distributions also shrink considerably. If thermal conduction is important, the further contraction in the gas distributions it produces makes them smaller than the observed sizes of X-ray clusters.

One interesting aspect of Perrenod's models is that many of the infall models have a temperature inversion ($dT_g/dr > 0$) in the cluster core, even if there is no significant cooling. This occurs because gas in the core has fallen through a shallower gravitational potential than gas further out. If such a temperature inversion were observed, it might be confused with a cooling flow (Section 5.7).

In White's N-body models, the cluster shows very strong subclustering at the beginning of its collapse, and forms two roughly equal subclusters, which merge as the cluster undergoes violent relaxation. Several double X-ray clusters are known (Section 4.4.2; Figure 18) that appear to be in just this stage of evolution (Forman *et al.*, 1981). Obviously, such subclustering cannot be treated in one-dimensional, spherical, hydrodynamic simulations. Gingold and Perrenod (1979) have made simplified three-dimensional hydro simulations of the evolution of clusters. When applied to the cluster potential from White's N-body models, these verified the previous one-dimensional calculations of Perrenod (1978b). They found that there was no significant enhancement of the X-ray emission from merging subclusters, beyond that predicted by single cluster models. Similar calculations were made by Ikeuchi and Hirayama (1979).

One major concern about all the Perrenod varying-potential models is the use of White's (1976c) N-body calculations for the cluster potential. In this particular set of models by White, the total virial mass of the cluster was assumed to reside in the individual galaxies. This gave the galaxies large masses, which increased their two-body interactions (Section 2.9.1), and caused the cluster core to contract rapidly. However, associating the missing mass in clusters with individual galaxies appears to produce more two-body relaxation in clusters than is observed (Sections 2.8 and 2.9.4); in fact, this was one of White's conclusions from his models. Thus it seems likely that Perrenod's calculations may significantly overestimate the increase

with time of the X-ray luminosity and gas temperature and the decrease in the gas core size.

Clusters of galaxies are the largest organized structures in the universe, and X-ray emission from them should be recognizable to large redshifts (Chapter 6). They might therefore be useful as probes of the cosmological structure of the universe. Several cosmological tests have been proposed using X-ray clusters (Schwartz, 1976; Silk and White, 1978); although some of these tests are relatively insensitive to X-ray cluster evolution, most are strongly affected. These models suggest that it will be difficult to apply any tests that require that X-ray clusters have remained unchanged since $z = 1$ (Perrenod, 1978b; Falle and Meszaros, 1980). On the other hand, in Perrenod's models the luminosity and size of X-ray clusters depend strongly on the density of material in the universe, since this determines the speed with which clusters contract. In principle, this dependence of cluster evolution on density might provide useful cosmological information; in practice, the evolution models are too uncertain to be used reliably for this purpose.

The models described so far have dealt with the evolution of single clusters. Perrenod (1980) has attempted to predict the evolution of the luminosity function of X-ray clusters (Section 4.2). He assumed that galaxies formed before clusters, and that clusters were formed by the gravitational attraction of galaxies. He argued that the merging of clusters tends to produce larger clusters with deeper potential wells, and as a result the average X-ray luminosity increases. White (1982) showed that this argument is incorrect; the increase in the depth of cluster potential wells is more than offset by the decrease in their characteristic densities. Perrenod found a very rapid evolution of the luminosity function to higher luminosities; he predicted that there should be few luminous X-ray clusters at redshifts $z \gtrsim 1/2$. This evolution depends strongly on the average density of matter in the universe, and Perrenod proposed using it as a cosmological test. However, his basic model for clustering is apparently incorrect (White, 1982).

5.10.2 Ejection from galaxies

The observation that the intracluster medium contains a significant abundance of heavy elements shows that it cannot be *entirely* due to the infall of primordial gas (Sections 4.3.2 and 5.2). The only way known for producing reasonable quantities of heavy elements is through nuclear reactions in stars. Since there is no significant luminous stellar population outside of galaxies at present, there are two possibilities. First, there may have been an early generation of pregalactic stars (Carr *et al.*, 1984), or second, it may be that some portion of the intracluster gas was ejected from galaxies. This section considers the second possibility.

Are the present rates of mass loss from stars in galaxies sufficient to produce the required amount of gas if all the stellar mass loss is added to the intracluster gas? Let us assume that the intracluster gas is chemically homogeneous (Section 5.4.5), so that the inferred heavy element abundance is about half of the solar value. Then, since this is comparable to the present abundances in the stars in elliptical galaxies in clusters, ejection from galaxies would have to supply a significant portion of the observed intracluster gas. The total mass of intracluster gas is at least as large as the total mass of stars in galaxies in a luminous X-ray cluster ($\gtrsim 10\%$ of the virial mass). Now the current rate of mass loss expected from the stellar populations seen in elliptical or S0 galaxies is

$$\alpha_* \approx 1.5 \times 10^{-12} / \text{yr.} \quad (5.121)$$

Thus, only a few per cent of the intracluster gas could be supplied in a Hubble time $\approx 10^{10}$ yr at the current rate.

In doing estimates of this sort, it is very important to remember that mass loss from stars is due to their evolution as they exhaust their nuclear fuel, and thus the mass loss from a stellar population is primarily due to the most luminous stars. In estimating the rate of mass loss, the rate per luminous star should be multiplied by the mass of luminous matter, and not the total (virial) mass. Another way of stating this is that the rate of mass loss is proportional to the luminosity of a stellar population and not to its mass; thus the rate of mass loss per unit mass varies inversely with the mass-to-light ratio (Section 2.8).

The rate of mass loss from stars in galaxies must have been higher in the past, if stellar mass loss has contributed significantly to the intracluster medium. One simple way this can have occurred is for elliptical and S0 galaxies to have contained more massive stars in the past, since massive stars have higher rates of mass loss. At present, these galaxies have only relatively low mass stars $M_* \lesssim 0.8M_\odot$ (Fall, 1981). Since stellar lifetimes decrease with mass, and the presently observed stars have lifetimes comparable to the Hubble time, any higher mass stars produced at the time of galaxy formation would no longer exist. High mass stars tend to die as supernovae which are very effective in producing and dispersing heavy elements, so these stars

might provide the heavy elements in the intracluster gas and in the stars seen in the galaxies today without leaving any very low abundance stars (Carr *et al.*, 1984). Moreover, the supernovae could aid in the removal of gas from the galaxies into the intracluster medium.

It is often suggested that all the stars in an elliptical or S0 galaxy formed at one time during the formation of the galaxy itself. Stars with a wide range of masses are usually assumed to have been made in protogalaxies, and the present stellar population is only those lower mass stars whose lifetimes exceed the age of the galaxy. Usually, the distribution of the masses of stars that form (the initial mass function or IMF) is taken to be a power-law, and the star formation rate is assumed to decline exponentially with the age of the galaxy:

$$\frac{\partial^2 N_*}{\partial M_* \partial t} \propto M_*^{-a} \exp(-t/t_*) \quad M_L \leq M_* \leq M_U \quad (5.122)$$

where N_* is the number of stars formed of mass M_* , the lower and upper limits to the IMF are M_L and M_U , and t_* is the time scale for star formation. A power law with $a = 2.35$ is called the ‘Salpeter IMF’ (Salpeter, 1955) and fits the current star formation in the disk of our galaxy. The time scale for star formation is often taken to be comparable to the dynamical time in a galaxy, $t_* \approx 3 \times 10^8$ yr. The resulting model for the gas loss from a galaxy depends on the IMF assumed and on the values of t_* and t_{st} , the time scale for the removal of gas from the galaxy.

Larson and Dinerstein (1975) calculated the properties of the intracluster gas based on this model. They assumed a Salpeter IMF. The only gas loss process they considered was supernova heating from the same stellar population; this may underestimate the ejected mass if collisions or ram pressure ablation also contribute. Based on this model, they found that the majority of the gas was not removed in galaxies more massive than about $10^{10} M_\odot$. The total gas mass ejected from galaxies in a cluster was about 30% of the mass in stars in galaxies, and the ejected mass had roughly solar abundances. Because these calculations preceded the detection of the iron X-ray lines in clusters, they successfully predicted that nearly solar abundances would be found.

Fairly similar results were found by Ikeuchi (1977), Biermann (1978), De Young (1978), and Sarazin (1979). Biermann assumed most of the galaxies were spirals that were stripped by collisions and ram pressure; this resulted in a more complete removal of interstellar gas, but over a longer time ($t_{st} \geq t_*$). Biermann found somewhat lower heavy element abundances of 0.1 to 0.5 of solar, with the gas mass ranging from 1 to 0.1 of that of the stars. De Young noted that the supernova energy was sufficient to unbind the gas produced by stellar mass loss in galaxies, and thus assumed that nearly all the gas was ejected quite rapidly ($t_{st} < t_*$). He also considered a wider range of IMFs, and generally found heavy element abundances that were larger, 1–3 times solar. This was primarily sensitive to the exponent a in the IMF. The ejected gas masses were 0.15 to 0.5 of the stellar mass. One exception to the agreement among these authors was Vigroux (1977), who claimed that galaxies could not make enough iron during the course of their normal evolution. The account he gave of his calculations was rather sketchy, so it is difficult to compare them to the others. With this exception, the general conclusion was that galaxies could eject an amount of gas about half the stellar mass with roughly solar abundances during their normal stellar evolution. If this gas were diluted with roughly an equal amount of unprocessed primordial gas, either within the forming galaxies or in the cluster, the observed mass and heavy element abundances in the intracluster gas would be reproduced.

In most of these models, the time scale of star formation is assumed to be short, $t_* \lesssim 10^9$ yr. During this time, most of the stars in the galaxy are formed, and the more massive and luminous stars live and die explosively in supernovae. As a result, it is expected that the newly formed galaxies would be very bright during this era (De Young, 1978; Bookbinder *et al.*, 1980).

The evolution of the intracluster gas in models with ejection from galaxies depends on the length of time it takes the newly enriched gas to be stripped, t_{st} . De Young (1978) noted that the energy input from the supernovae produced by a stellar population which would give the needed iron abundance would be sufficient to unbind the interstellar gas. This assumes that the supernova energy is efficiently converted to kinetic energy in the gas.

On the other hand, if the supernovae energy is radiated away, the gas may remain bound to the galaxy. Norman and Silk (1979) and Sarazin (1979) showed that this was likely to be the case, because the large quantity of intracluster gas in clusters implies that there was a large density of gas in protogalaxies. At high densities the gas cools rapidly, and individual supernova remnants radiate away their energy before

they overlap. Under these circumstances, the galaxies may retain their gas. Norman, Silk, and Sarazin suggested that galaxies retain much of their initial gas content as extended hot coronae. If this gas cannot be removed by supernovae, then collisions or ram pressure remain as stripping mechanisms (Section 5.9). They further assumed that galaxy formation is very efficient, in the sense that nearly all the gas in a cluster was initially contained in galaxies. Then, there would be very little intracluster gas at first, and ram pressure ablation would not be effective. The galaxies would first lose gas slowly through collisions, and when the intracluster density was high enough, ram pressure stripping would start. Because ram pressure ablation both increases the gas density and increases with increasing gas density, this leads to a runaway stripping of cluster galaxies. The evolution of the gas in a cluster would then occur in two extended stages, with a rapid transition between them. First, all the gas would be bound to galaxies. Then, it would be rapidly stripped and remain distributed in the intracluster medium after that time. This was proposed as an explanation of the Butcher–Oemler effect (Section 2.10.2); the Butcher–Oemler clusters were still in this first stage. Unfortunately, this model predicts that these clusters have very little intracluster gas, when in fact they were subsequently observed to be luminous X-ray sources (Section 4.8). Larson *et al.* (1980) argued that the disks of spiral and S0 galaxies are produced by infall from coronae of gas bound to galaxies, and that a spiral galaxy becomes an S0 when the corona is stripped and the gas supply to the disk stops. In this way, there would be a longer interval between the stripping of the corona and the cessation of star formation in the disk. Perhaps Butcher–Oemler clusters are within this interval. One problem with these models for gaseous coronae is that they require a rather delicate and unstable balance between supernova heating and cooling.

Biermann (1978) proposed a similar model in which gas produced by disk galaxies is stored in their disks, and is eventually stripped by collisions and ram pressure. Himmes and Biermann (1980) gave a somewhat more detailed model, in which elliptical galaxies in a cluster lose their interstellar gas rapidly by supernova heating and provide an initial amount of intracluster gas, which begins the process of ram pressure stripping of spiral galaxies. They argued that this model can reproduce the present intracluster gas masses, iron abundances, and dependence of the galactic population on X-ray luminosity (Section 4.6). In this model, the spiral fraction in cluster decreases continuously with time, and the variation from a redshift of $z \approx 0.4$ to the present is consistent with the Butcher–Oemler effect.

One general feature of these models in which gas is ejected from galaxies over a long period of time is that the luminosities of X-ray clusters are expected to increase with time. Unfortunately, this is the same prediction made by Perrenod's infall models with deepening cluster potentials, as discussed in the previous section.

A number of one-dimensional, spherically symmetric, hydrodynamic simulations have been made of the evolution of intracluster gas, including ejection from galaxies. Cowie and Perrenod (1978) calculated models with a fixed cluster potential and assumed that the rate of gas ejection from galaxies varied inversely with time $\alpha_* \propto 1/t$. There was no primordial intracluster gas in these models. The gas was ejected at zero temperature (Section 5.3.3) and assumed to mix immediately with the intracluster gas. When the gas ejection rate is large, the models evolve to steady-state cooling flows (Cowie and Binney, 1977; Section 5.7.1). In models with lower ejection rates, the X-ray luminosity either is roughly constant (no thermal conduction) or increases by about a factor of two (thermal conduction) from a redshift $z = 1$ to the present. Perrenod (1978b) calculated ejection models in a varying cluster potential; the results were very similar to those described above for infall models. He found a better fit to the present gas distributions with these models than with infall models, and the ejection models were less sensitive to the assumed initial conditions and model parameters. These models showed a rapid increase in X-ray luminosity with time, as did the infall models.

Ikeuchi and Hirayama (1980) ran hydro models with no primordial gas, in which gas is ejected from all the galaxies simultaneously and very rapidly ($t_{st} \lesssim 10^7$ yr). This seems rather unlikely, since this time scale is less than the sound crossing time for a single galaxy. Because of this assumption of rapid ejection, they chose the following initial conditions: at the start of their calculation, the ejected gas was placed in the cluster in a nonhydrostatic distribution determined by their ejection model, and then 'let go'. The gas then adjusted to the cluster potential on a sound crossing time (Section 5.5). These models have very large X-ray luminosities $\approx 10^{48}$ erg/s during this initial relaxation time $t \approx 3 \times 10^8$ yr. It seems very unlikely that such high luminosities would be realized, since the actual gas ejection must take considerably longer than was assumed.

It has generally been assumed that the ejected gas mixes rapidly (both chemically and thermally) with the intracluster gas (De Young, 1978). However, Nepveu (1981b) argues that this will not occur (although the only mechanism he considers is turbulent mixing), and that the ejected and intracluster gases must be treated as two separate fluids.

Hirayama (1978) and Nepveu (1981b) have given hydrodynamic models for the evolution of the intracluster medium including both gas ejected from galaxies and primordial gas. In both cases, the primordial gas is initially relaxed, and galaxy gas is injected at a constant rate. In both of these calculations, the ejected gas is concentrated to the cluster center ($R \lesssim 2$ Mpc), and there is a large gradient in the heavy element abundance across the cluster. As noted previously (Sections 5.4.5 and 5.5.6), such a concentration of heavy elements to the cluster center will increase the strengths of the X-ray lines from these elements. Thus the abundances derived from the X-ray spectra of clusters could overestimate the real abundances. However, because most of the X-ray emission in these models comes from radii of less than 2 Mpc, this effect is not very serious. Spatially resolved X-ray spectra across a cluster might detect such a gradient, and this would allow one to deduce the proportions of ejected and intracluster gas.

6 Prospects for the future and AXAF

What advances can be expected in the near future in the study of X-ray clusters? In this brief look at future prospects, I shall concentrate on new observational opportunities. At present, observational X-ray astronomy is in a rather quiet period. The *Einstein* X-ray satellite, which revolutionized the study of X-ray astronomy, is no longer operational. As this review was being written, the European Space Agency X-ray satellite EXOSAT was nearing the end of its operational life. EXOSAT is a somewhat less powerful imaging instrument than *Einstein*. What advances in the technology of X-ray astronomy are needed to answer the major questions we have about X-ray clusters, and what plans are there for the realization of these advances?

The basic data in the X-ray study of clusters consist of the surface brightness of X-rays I_ν , as a function of the photon frequency ν and the position on the sky. Given these data and a suitable assumption of symmetry for the cluster, the X-ray emissivity $\epsilon_\nu(r)$ as a function of position in the cluster can be derived by deconvolution of the surface brightness (Section 5.5.4). The emissivity varies as the square of the density ρ_g , and its frequency dependence is determined by the gas temperature T_g and by the abundances of heavy elements (equation 5.19). At the temperatures usually found in the intracluster gas, the heavy element abundances mainly affect the emission in several narrow line features, and the temperature produces an exponential falloff in the intensity for frequencies $h\nu > kT_g$. Thus, given suitable observations of the X-ray surface brightness I_ν , one can derive the gas density, temperature, and several heavy element abundances, all as a function of position in the cluster.

In relatively nearby clusters, the required instrument for these observations must measure the X-ray surface brightness with at least moderate spatial resolution (< 1 arc minute), and modest spectral resolution (better than about 15%), and must be sensitive to X-rays with photon energies $h\nu$ of at least 7 keV. Obviously, it must also have a sufficient sensitivity to detect the clusters. Unfortunately, no past or currently existing satellite has had all these capabilities. Proportional counter systems, such as the *Uhuru* satellite, have modest spectral resolution out to high X-ray energies, but have very poor spatial resolution. The *Einstein* satellite had excellent spatial resolution, fairly poor spectral resolution in the imaging detectors, and no sensitivity for photon energies $h\nu \gtrsim 4$ keV.

The Advanced X-ray Astronomy Facility (AXAF) would provide the new observational capabilities needed for the further study of X-ray clusters (Giacconi *et al.*, 1980). AXAF is a 1.2 meter diameter X-ray telescope, which would be carried into orbit by the Space Shuttle. As currently planned, AXAF would have roughly 100 times the sensitivity of the *Einstein* telescope for point sources and a considerably increased sensitivity for extended sources as well. Its mirrors would produce images with a spatial resolution of better than one second of arc, and would be sensitive to X-rays with photon energies of at least 8 keV. At least some of the imaging detectors being considered would have moderate spectral resolution (10–20% or better), and the satellite would have a number of higher resolution spectrometers. With its high spatial resolution, moderate spectral resolution, and sensitivity to harder X-rays, it would provide exactly the data on cluster X-ray surface brightnesses needed to derive their densities, temperatures, and abundances.

Given the run of density and temperature of the gas in a cluster or in an individual galaxy, the hydrostatic equation (5.56) allows one to determine the total mass in the galaxy or cluster as a function of position (Sections 5.5.5 and 5.8.1). These mass determinations are less uncertain than those based on the radial velocities of galaxies in clusters or stars in galaxies because the gas atoms are known to be moving isotropically. These mass distributions would provide very important information on the distribution and nature of the missing mass component in clusters and galaxies.

If measurements of the microwave diminutions of clusters can be made reliably, they can be combined with the determinations of the variation of the gas temperature and density to give distances to clusters that are independent of the Hubble constant (Section 3.5). This will provide a direct determination of the Hubble constant, independent of the usual extragalactic distance scale. If this method could be applied to high redshift clusters, it might allow the determination of the overall structure of the universe.

*From the variation of the gas temperature with density and with position in a cluster, the influence of the various heating, cooling, and energy transport processes (Sections 5.3, 5.4) can be deduced. As discussed in Section 5.5.1, the surface brightness distributions in nearby clusters from the *Einstein* satellite are consistent with isothermal gas distributions, although temperatures could not be determined directly. Unfortunately, the temperatures required by the surface brightness fits are generally not consistent with temperatures derived from the integrated spectra of the clusters. Given this discrepancy, we cannot claim*

to have any real understanding of the thermal processes in intracluster gas. Direct measurements of the temperature profiles of clusters are needed to resolve this problem.

The moderate spectral resolution of AXAF's imaging detectors and sensitivity to 7 keV X-rays will allow this instrument to map out the abundance of iron and possibly other heavy elements in clusters. The distribution of heavy elements in clusters must be known if accurate abundances are to be derived for them. As noted in Sections 5.4.5 and 5.5.6, if the iron in clusters is concentrated in the core, the iron abundances may have been significantly overestimated. These abundances are used to determine the amount of gas that must have been ejected from stars in galaxies, and affect models for the origin and early evolution of galaxies (Section 5.10.2). Moreover, the distributions of heavy elements provide information on the relative proportions of ejected galactic gas and primordial intergalactic gas in clusters.

The higher resolution spectrometers on AXAF can be used to determine the abundances of additional elements and give more precise information on the temperature structure. It is particularly useful that the 7 keV iron lines will be observable, as these are the strongest lines in the intracluster gas and have many fine structure components whose intensities are sensitive to temperature. High resolution line studies will be particularly useful in studying the physical conditions in cooling flows (Section 5.7). They may also permit the determination of redshifts for X-ray clusters for which optical data are not available, and will certainly resolve any ambiguities when several clusters at different redshifts are seen along the same line-of-sight. At the highest spectral resolution, it may be possible to determine directly the flow velocities in clusters, particularly those with cooling flows.

The high spatial resolution of AXAF will be very important to the study of cooling flows and of other gas associated with individual galaxies (Sections 5.7 and 5.1). The increased sensitivity and enhanced spectral response of AXAF should make it possible to get spectra of the gas in these individual galaxy sources, in order to test the hypothesis that the emission is from hot gas. Gas in individual galaxies in clusters has so far been studied only in relatively nearby clusters. Of particular interest is the interaction of this gas with the intracluster medium.

AXAF should detect X-ray clusters out to very high redshifts, $z \approx 1 - 4$. From the study of these clusters, we shall learn directly about the origin of the intracluster gas and its evolution in clusters. We may actually see the gas being ejected from galaxies. If galaxy morphologies are altered by the galaxy's environment, and the main mechanism is gas stripping by intracluster gas, the buildup of the intracluster gas should be related to the evolution of galaxy morphologies. With the Hubble Space Telescope (Hall, 1982), it should be possible to classify galaxies out to at least moderate redshifts. The variation in the heavy element abundances in clusters as a function of redshift should constrain models for the chemical evolution of galaxies. The variations in the temperature of the gas will allow us to assess the effects of heating and cooling.

It will also be interesting to see if there is any relationship between the evolution of X-ray clusters and that of quasars and other active galactic nuclei.

One problem with these studies of high redshift clusters is that few are currently known. Because AXAF is not primarily a survey instrument, it might not detect a very large number of previously unknown clusters. It is possible that deeper ground based optical surveys or studies with the Hubble Space Telescope will provide longer lists of cosmological clusters. It is also possible they may be found by studying high redshift radio galaxies and quasars with the morphological distortions normally associated with sources in clusters (Section 3.3). Another exciting possibility involves the Roentgen Satellite (ROSAT). This instrument will perform an all-sky soft X-ray survey, with a spatial resolution of about 1 minute of arc and a sensitivity limit similar to that of the *Einstein* satellite. A luminous X-ray cluster at a redshift of 1 might possibly be detected in this survey. Because an X-ray cluster at a redshift of 1 would have an angular size of about 1 minute of arc, it might appear as a resolved source. The most common extragalactic X-ray sources found in deep surveys are quasars, which are point sources. Thus most of the resolved high galactic latitude sources in the ROSAT survey should be clusters, and some of those should be at high redshifts. This survey may provide a valuable list of X-ray clusters for further study.

References

- Aarseth, S. J., and J. Binney, 1978, *Mon. Not. R. Astron. Soc.* **185**, 227.
- Abell, G. O., 1958, *Astrophys. J. Suppl.* **3**, 211.
- Abell, G. O., 1961, *Astron. J.* **66**, 607.
- Abell, G. O., 1962, in *Problems of Extra-Galactic Research*, edited by G. C. McVittie, p. 213. Chicago: University of Chicago.
- Abell, G. O., 1965, *Ann. Rev. Astron. Astrophys.* **3**, 1.
- Abell, G. O., 1975, in *Stars and Stellar Systems IX: Galaxies and the Universe*, edited by A. Sandage, M. Sandage, and J. Kristian, p. 601. Chicago: University of Chicago.
- Abell, G. O., 1977, *Astrophys. J.* **213**, 327.
- Abell, G. O., 1982, private communication.
- Abell, G. O., J. Neyman, and E. L. Scott, 1964, *Astron. J.* **69**, 529.
- Abramopoulos, F., G. Chanan, and W. Ku, 1981, *Astrophys. J.* **248**, 429.
- Abramopoulos, F., and W. Ku, 1983, *Astrophys. J.* **271**, 446.
- Adams, M. T., K. M. Strom, and S. E. Strom, 1980, *Astrophys. J.* **238**, 445.
- Adams, T. F., 1977, *Publ. Astron. Soc. Pac.* **89**, 488.
- Albert, C. E., R. A. White, and W. W. Morgan, 1977, *Astrophys. J.* **211**, 309.
- Allen, C. W., 1973, *Astrophysical Quantities*, p. 197. London: Athlone.
- Andernach, H., J. R. Baker, A. von Kap-herr, and R. Wielebinski, 1979, *Astron. Astrophys.* **74**, 93.
- Andernach, H., D. Schallwich, C. Haslam, and R. Wielebinski, 1981, *Astron. Astrophys. Suppl.* **43**, 155.
- Andernach, H., H. Waldhausen, and R. Wielebinsky, 1980, *Astron. Astrophys. Suppl.* **41**, 339.
- Arp, H., and J. Lorre, 1976, *Astrophys. J.* **210**, 58.
- Auriemma, C., G. Perola, R. Ekers, R. Fanti, C. Lari, W. Jaffe, and M. Ulrich, 1977, *Astron. Astrophys. Suppl.* **57**, 41.
- Austin, T. B., J. G. Godwin, and J. V. Peach, 1975, *Mon. Not. R. Astron. Soc.* **171**, 135.
- Austin, T. B., and J. V. Peach, 1974a, *Mon. Not. R. Astron. Soc.* **167**, 437.
- Austin, T. B., and J. V. Peach, 1974b, *Mon. Not. R. Astron. Soc.* **168**, 591.
- Avni, Y., 1976, *Astrophys. J.* **210**, 642.
- Avni, Y., and N. Bahcall, 1976, *Astrophys. J.* **209**, 16.
- Baan, W. A., A. D. Haschick, and B. F. Burke, 1978, *Astrophys. J.* **225**, 339.
- Bahcall, J. N., and N. A. Bahcall, 1975, *Astrophys. J. Lett.* **199**, L89.
- Bahcall, J. N., and C. L. Sarazin, 1977, *Astrophys. J. Lett.* **213**, L99.
- Bahcall, J. N., and C. L. Sarazin, 1978, *Astrophys. J.* **219**, 781.
- Bahcall, N. A., 1971, *Astron. J.* **76**, 995.
- Bahcall, N. A., 1972, *Astron. J.* **77**, 550.
- Bahcall, N. A., 1973a, *Astrophys. J.* **180**, 699.
- Bahcall, N. A., 1973b, *Astrophys. J.* **183**, 783.
- Bahcall, N. A., 1974a, *Astrophys. J.* **187**, 439.
- Bahcall, N. A., 1974b, *Astrophys. J.* **193**, 529.
- Bahcall, N. A., 1974c, *Nature* **252**, 661.
- Bahcall, N. A., 1975, *Astrophys. J.* **198**, 249.
- Bahcall, N. A., 1977a, *Ann. Rev. Astron. Astrophys.* **15**, 505.
- Bahcall, N. A., 1977b, *Astrophys. J. Lett.* **217**, L77.
- Bahcall, N. A., 1977c, *Astrophys. J. Lett.* **218**, L93.
- Bahcall, N. A., 1979a, *Astrophys. J.* **232**, 689.
- Bahcall, N. A., 1979b, *Astrophys. J. Lett.* **232**, L83.
- Bahcall, N. A., 1980, *Astrophys. J. Lett.* **238**, L117.
- Bahcall, N. A., 1981, *Astrophys. J.* **247**, 787.
- Bahcall, N. A., D. E. Harris, and R. G. Strom, 1976, *Astrophys. J. Lett.* **209**, L17.
- Bahcall, N. A., and W. L. W. Sargent, 1977, *Astrophys. J. Lett.* **217**, L19.
- Bahcall, N. A., and R. M. Soneira, 1982, *Astrophys. J.* **262**, 419.
- Baldwin, J. E., and P. F. Scott, 1973, *Mon. Not. R. Astron. Soc.* **165**, 259.
- Barnes, J., 1983, *Mon. Not. R. Astron. Soc.* **203**, 223.

- Basko, M. M., B. V. Komberg, and E. I. Moskalenko, 1981, *Sov. Astron.* **25**, 402.
- Bautz, L. P., and G. O. Abell, 1973, *Astrophys. J.* **184**, 709.
- Bautz, L. P., and W. W. Morgan, 1970, *Astrophys. J. Lett.* **162**, L149.
- Bazzano, A., R. Fusco-Femiano, C. La Padula, V. Polcaro, P. Ubertini, and R. Manchanda 1984, *Astrophys. J.*, **279**, 515.
- Bechtold, J., W. Forman, R. Giacconi, C. Jones, J. Schwarz, W. Tucker, and L. Van Speybroeck, 1983, *Astrophys. J.* **265**, 26.
- Beers, T. C., M. J. Geller, and J. P. Huchra, 1982, *Astrophys. J.* **257**, 23.
- Beers, T. C., M. J. Geller, J. P. Huchra, D. W. Latham, and R. J. Davis, 1984, *Astrophys. J.* **283**, 33.
- Beers, T. C., J. P. Huchra, and M. J. Geller, 1983, *Astrophys. J.* **264**, 356.
- Beers, T. C., and J. L. Tonry, 1986, *Astrophys. J.* **300**, 557.
- Begelman, M. C., M. J. Rees, and R. D. Blandford, 1979, *Nature* **279**, 770.
- Berthelsdorf, R. F., and J. L. Culhane, 1979, *Mon. Not. R. Astron. Soc.* **187**, 17p.
- Bieging, J. H., and P. Biermann, 1977, *Astron. Astrophys.* **60**, 361.
- Biermann, P., 1978, *Astron. Astrophys.* **62**, 255.
- Biermann, P., and P. Kronberg, 1983, *Astrophys. J.* **268**, L69.
- Biermann, P., P. P. Kronberg, and B. F. Madore, 1982, *Astrophys. J. Lett.* **256**, L37.
- Biermann, P., and S. L. Shapiro, 1979, *Astrophys. J. Lett.* **230**, L33.
- Bijleveld, W., and E. A. Valentijn, 1982, *Astron. Astrophys.* **111**, 50.
- Bijleveld, W., and E. A. Valentijn, 1983, *Astron. Astrophys.* **125**, 217.
- Binggeli, B., 1982, *Astron. Astrophys.* **107**, 338.
- Binney, J., 1977, *Mon. Not. R. Astron. Soc.* **181**, 735.
- Binney, J., 1980, in *X-ray Astronomy*, edited by R. Giacconi and G. Setti, p. 245. Dordrecht: Reidel.
- Binney, J., and L. L. Cowie, 1981, *Astrophys. J.* **247**, 464.
- Binney, J., and J. Silk, 1978, *Comm. Astrophys.* **7**, 139.
- Binney, J., and O. Strimpel, 1978, *Mon. Not. R. Astron. Soc.* **185**, 473.
- Birkinshaw, M., 1978, *Mon. Not. R. Astron. Soc.* **184**, 387.
- Birkinshaw, M., 1979, *Mon. Not. R. Astron. Soc.* **187**, 847.
- Birkinshaw, M., 1980, *Mon. Not. R. Astron. Soc.* **190**, 793.
- Birkinshaw, M., and S. F. Gull, 1984, *Mon. Not. R. Astron. Soc.* **206**, 359.
- Birkinshaw, M., S. F. Gull, and H. Hardebeck, 1984, *Nature* **309**, 34.
- Birkinshaw, M., S. F. Gull, and A. T. Moffet, 1981a, *Astrophys. J. Lett.* **251**, L69.
- Birkinshaw, M., S. F. Gull, and K. J. Northover, 1978, *Mon. Not. R. Astron. Soc.* **185**, 245.
- Birkinshaw, M., S. F. Gull, and K. J. Northover, 1981b, *Mon. Not. R. Astron. Soc.* **197**, 571.
- Blumenthal, G. R., S. M. Faber, J. R. Primack, and M. J. Rees, 1984, *Nature* **311**, 517.
- Bohlin, R. C., R. C. Henry, and J. R. Swandic, 1973, *Astrophys. J.* **182**, 1.
- Boldt, E., 1976, *Astrophys. J. Lett.* **208**, L15.
- Bookbinder, J., L. L. Cowie, J. H. Krolik, J. P. Ostriker, and M. Rees, 1980, *Astrophys. J.* **237**, 647.
- Bothun, G. D., M. J. Geller, T. C. Beers, and J. P. Huchra, 1983, *Astrophys. J.* **268**, 47.
- Boynton, P. E., S. J. Radford, R. A. Schommer, and S. S. Murray, 1982, *Astrophys. J.* **257**, 473.
- Bradt, H., W. Mayer, S. Narayan, S. Rappaport, and G. Spuda, 1967, *Astrophys. J. Lett.* **161**, L1.
- Braid, M. K., and H. T. MacGillivray, 1978, *Mon. Not. R. Astron. Soc.* **182**, 241.
- Branduardi-Raymont, G., D. Fabricant, E. Feigelson, P. Gorenstein, J. Grindlay, A. Soltan, and G. Zamorani, 1981, *Astrophys. J.* **248**, 55.
- Brecher, K., and G. R. Burbidge, 1972, *Astrophys. J.* **174**, 253.
- Bridle, A. H., and P. A. Feldman, 1972, *Nature Phys. Sci.*, **235**, 168.
- Bridle, A. H., and E. B. Fomalont, 1976, *Astron. Astrophys.* **52**, 107.
- Bridle, A. H., E. B. Fomalont, G. K. Miley, and E. A. Valentijn, 1979, *Astron. Astrophys.* **80**, 201.
- Bridle, A. H., and J. P. Vallee, 1981, *Astron. J.* **86**, 1165.
- Brown, R. L., and R. J. Gould, 1970, *Phys. Rev. D* **1**, 2252.
- Bruzual A., G., and H. Spinrad, 1978a, *Astrophys. J.* **220**, 1.
- Bruzual A., G., and H. Spinrad, 1978b, *Astrophys. J.* **222**, 1119.
- Bucknell, M. J., J. G. Godwin, and J. V. Peach, 1979, *Mon. Not. R. Astron. Soc.* **188**, 579.

- Burns, J. O., 1981, *Mon. Not. R. Astron. Soc.* **195**, 523.
- Burns, J. O., J. A. Eilek, and F. N. Owen, 1982, in *IAU Symposium 97: Extragalactic Radio Sources*, edited by D. Heeschen and C. Wade, p. 45. Dordrecht: Reidel.
- Burns, J. O., S. A. Gregory, and G. D. Holman, 1981c, *Astrophys. J.* **250**, 450.
- Burns, J. O., C. P. O'Dea, S. A. Gregory, and T. J. Balonek 1986, *Astrophys. J.* **307**, 73.
- Burns, J. O., and F. N. Owen, 1977, *Astrophys. J.* **217**, 34.
- Burns, J. O., and F. N. Owen, 1979, *Astron. J.* **84**, 1478.
- Burns, J. O., and F. N. Owen, 1980, *Astron. J.* **85**, 204.
- Burns, J. O., F. N. Owen, and L. Rudnick, 1978, *Astron. J.* **83**, 312.
- Burns, J. O., and M. P. Ulmer, 1980, *Astron. J.* **85**, 773.
- Burns, J. O., R. A. White, and R. J. Hanisch, 1980, *Astron. J.* **85**, 191.
- Burns, J. O., R. A. White, and M.P. Haynes, 1981a, *Astron. J.* **86**, 1120.
- Burns, J. O., R. A. White, and D. H. Hough, 1981b, *Astron. J.* **86**, 1.
- Burstein, D., 1979a, *Astrophys. J.* **234**, 435.
- Burstein, D., 1979b, *Astrophys. J.* **234**, 829.
- Butcher, H., and A. Oemler, Jr., 1978a, *Astrophys. J.* **219**, 18.
- Butcher, H., and A. Oemler, Jr., 1978b, *Astrophys. J.* **226**, 559.
- Butcher, H., and A. Oemler, Jr., 1984a, *Astrophys. J.* **285**, 426.
- Butcher, H., and A. Oemler, Jr., 1984b, *Nature* **310**, 31.
- Byram, E. T., T. A. Chubb, and H. Friedman, 1966, *Science* **152**, 66.
- Cane, H. V., W. C. Erickson, R. J. Hanisch, and P. J. Turner, 1981, *Mon. Not. R. Astron. Soc.* **196**, 409.
- Canizares, C. R., 1981, in *Proceedings of the HEAD Meeting on X-ray Astronomy*, edited by R. Giacconi, p. 215. Dordrecht: Reidel.
- Canizares, C. R., G. W. Clark, J. G. Jernigan, and T. H. Markert, 1982, *Astrophys. J.* **262**, 33.
- Canizares, C. R., G. W. Clark, T. H. Markert, C. Berg, M. Smedira, D. Bardas, H. Schnopper, and K. Kalata, 1979, *Astrophys. J.* **234**, L33.
- Canizares, C. R., G. Fabbiano, and G. Trinchieri, 1987, *Astrophys. J.* **312**, 503.
- Canizares, C. R., G. C. Stewart, and A. C. Fabian, 1983, *Astrophys. J.* **272**, 449.
- Capelato, H. V., D. Gerbal, G. Mathez, A. Mazure, E. Salvador-Sole, and H. Sol, 1980, *Astrophys. J.* **241**, 521.
- Carnevali, P., A. Cavaliere, and P. Santangelo, 1981, *Astrophys. J.* **249**, 449.
- Carr, B. J., J. R. Bond, and W. D. Arnett, 1984, *Astrophys. J.* **277**, 445.
- Carter, D., 1977, *Mon. Not. R. Astron. Soc.* **178**, 137.
- Carter, D., 1980, *Mon. Not. R. Astron. Soc.* **190**, 307.
- Carter, D., and J. G. Godwin, 1979, *Mon. Not. R. Astron. Soc.* **187**, 711.
- Carter, D., and N. Metcalfe, 1980, *Mon. Not. R. Astron. Soc.* **191**, 325.
- Cash, W., R. F. Malina, and R. S. Wolff, 1976, *Astrophys. J. Lett.* **209**, L111.
- Catura, R. C., P. G. Fisher, M. M. Johnson, and A. J. Meyerott, 1972, *Astrophys. J. Lett.* **177**, L1.
- Cavaliere, A., 1980, in *X-ray Astronomy*, edited by R. Giacconi and G. Setti, p. 217. Dordrecht: Reidel.
- Cavaliere, A., L. Danese, and G. deZotti, 1977, *Astrophys. J.* **217**, 6.
- Cavaliere, A., L. Danese, and G. deZotti, 1979, *Astron. Astrophys.* **75**, 322.
- Cavaliere, A., G. DeBiase, P. Santangelo, and N. Vittorio, 1983, in *Clustering in the Universe*, edited by D. Gerbal and A. Mazure, p. 15. Paris: Editions Frontieres.
- Cavaliere, A., and R. Fusco-Femiano, 1976, *Astron. Astrophys.* **49**, 137.
- Cavaliere, A., and R. Fusco-Femiano, 1978, *Astron. Astrophys.* **70**, 677.
- Cavaliere, A., and R. Fusco-Femiano, 1981, *Astron. Astrophys.* **100**, 194.
- Cavaliere, A., H. Gursky, and W. H. Tucker, 1971, *Nature* **231**, 437.
- Cavallo, G., and N. Mandolesi, 1982, *Astrophys. Lett.* **22**, 119.
- Chamaraux, P., C. Balkowski, and E. Gerard, 1980, *Astron. Astrophys.* **83**, 38.
- Chanan, G. A., and F. Abramopoulos, 1984, *Astrophys. J.* **287**, 89.
- Chandrasekhar, S., 1939, *An Introduction to the Study of Stellar Structure*, p. 155. Chicago: University of Chicago.
- Chandrasekhar, S., 1942, *Principles of Stellar Dynamics*, p. 231. Chicago: University of Chicago.

- Chandrasekhar, S., 1968, *Ellipsoidal Figures of Equilibrium*, New Haven: Yale University.
- Chincarini, G., 1984, *Adv. Space Res.* **3**, 393.
- Chincarini, G. L., R. Giovanelli, M. Haynes, and P. Fontanelli, 1983, *Astrophys. J.* **267**, 511.
- Chincarini, G., and H. J. Rood, 1976, *Astrophys. J.* **206**, 30.
- Chincarini, G., and H. J. Rood, 1977, *Astrophys. J.* **214**, 351.
- Chincarini, G., M. Tarenghi, and C. Bettis, 1978, *Astrophys. J.* **221**, 34.
- Chincarini, G., M. Tarenghi, and C. Bettis, 1981, *Astron. Astrophys.* **96**, 106.
- Christiansen, W. A., A. G. Pacholczyk, and J. S. Scott, 1981, *Astrophys. J.* **251**, 518.
- Ciardullo, R., H. Ford, F. Bartko, and R. Harms, 1983, *Astrophys. J.* **273**, 24.
- Coleman, G., P. Hintzen, J. Scott, and M. Tarenghi, 1976, *Nature* **262**, 476.
- Cooke, B. A., and D. Maccagni, 1976, *Mon. Not. R. Astron. Soc.* **175**, 65p.
- Cooke, J. A., D. Emerson, B. D. Kelly, H. T. MacGillivray, and R. J. Dodd, 1981, *Mon. Not. R. Astron. Soc.* **196**, 397.
- Corwin, H. G., 1974, *Astron. J.* **79**, 1356.
- Costain, C. H., A. H. Bridle, and P. A. Feldman, 1972, *Astrophys. J. Lett.* **175**, L15.
- Cowie, L. L., 1981, in *Proceedings of the HEAD Meeting on X-ray Astronomy*, edited by R. Giacconi, p. 227. Dordrecht: Reidel.
- Cowie, L. L., and J. Binney, 1977, *Astrophys. J.* **215**, 723.
- Cowie, L. L., A. C. Fabian, and P. E. Nulsen, 1980, *Mon. Not. R. Astron. Soc.* **191**, 399.
- Cowie, L. L., M. J. Henriksen, and R. Mushotzky, 1987, *Astrophys. J.* **317**, 593.
- Cowie, L., E. Hu, E. Jenkins, and D. York, 1983, *Astrophys. J.* **272**, 29.
- Cowie, L. L., and C. F. McKee, 1975, *Astron. Astrophys.* **43**, 337.
- Cowie, L. L., and C. F. McKee, 1977, *Astrophys. J.* **211**, 135.
- Cowie, L. L., and S. C. Perrenod, 1978, *Astrophys. J.* **219**, 254.
- Cowie, L. L., and A. Songaila, 1977, *Nature* **266**, 501.
- Crane, P., and J. A. Tyson, 1975, *Astrophys. J. Lett.* **201**, L1.
- Da Costa, L. N., and E. Knobloch, 1979, *Astrophys. J.* **230**, 639.
- Dagkesamansky, R. D., A. G. Gubanov, A. D. Kuzmin, and O. B. Slee, 1982, *Mon. Not. R. Astron. Soc.* **200**, 971.
- Danese, L., G. deZotti, and G. di Tullio, 1980, *Astron. Astrophys.* **82**, 322.
- Davidson, A., S. Bowyer, M. Lampton, and R. Cruddace, 1975, *Astrophys. J.* **198**, 1.
- Davidson, A., and W. Welch, 1974, *Astrophys. J. Lett.* **191**, L11.
- Davies, R. D., and B. M. Lewis, 1973, *Mon. Not. R. Astron. Soc.* **165**, 231.
- Davis, M., J. Huchra, D. Latham, and J. Tonry, 1982, *Astrophys. J.* **253**, 423.
- Davison, P. J., 1978, *Mon. Not. R. Astron. Soc.* **183**, 39p.
- Dawe, J., R. Dickens, and B. Peterson, 1977, *Mon. Not. R. Astron. Soc.* **178**, 675.
- Dennison, B., 1980a, *Astrophys. J.* **236**, 761.
- Dennison, B., 1980b, *Astrophys. J. Lett.* **239**, L93.
- Demoulin-Ulrich, M.-H., H. R. Butcher, and A. Boksenberg, 1984, *Astrophys. J.* **285**, 527.
- des Forets, G., R. Dominguez-Tenreiro, D. Gerbal, G. Mathez, Alain Mazure, and E. Salvador-Sole, 1984, *Astrophys. J.* **280**, 15.
- de Vaucouleurs, G., 1948a, *Ann. d'Astrophys.* **11**, 247.
- de Vaucouleurs, G., 1948b, *Contrib. Inst. Astrophys. Paris* **A**, no. 27.
- de Vaucouleurs, G., 1953, *Astron. J.* **58**, 30.
- de Vaucouleurs, G., 1956, *Mem. Mt. Stromlo Obs.* **15**, no. 13.
- de Vaucouleurs, G., 1961, *Astrophys. J. Suppl.* **6**, 213.
- de Vaucouleurs, G., 1975, in *Stars and Stellar Systems IX: Galaxies and the Universe*, edited by A. Sandage, M. Sandage, and J. Kristian, p. 557. Chicago: University of Chicago.
- de Vaucouleurs, G., 1976, *Astrophys. J.* **203**, 33.
- de Vaucouleurs, G., and A. de Vaucouleurs, 1970, *Astrophys. Lett.* **5**, 219.
- de Vaucouleurs, G., and J.-L. Nieto, 1978, *Astrophys. J.* **220**, 449.
- De Young, D. S., 1972, *Astrophys. J.* **173**, L7.
- De Young, D. S., 1978, *Astrophys. J.* **223**, 47.

- De Young, D. S., J. J. Condon, and H. Butcher, 1980, *Astrophys. J.* **242**, 511.
- Dickens, R. J., and C. Moss, 1976, *Mon. Not. R. Astron. Soc.* **174**, 47.
- Dickey, J. M., and E. E. Salpeter, 1984, *Astrophys. J.* **284**, 461.
- Disney, M. J., 1974, *Astrophys. J. Lett.* **193**, L103.
- Dones, L., and S. D. M. White, 1985, *Astrophys. J.* **290**, 94.
- Doroshkevich, A. G., S. F. Shandarin, and E. Saar, 1978, *Mon. Not. R. Astron. Soc.* **184**, 643.
- Dressler, A., 1978a, *Astrophys. J.* **222**, 23.
- Dressler, A., 1978b, *Astrophys. J.* **223**, 765.
- Dressler, A., 1978c, *Astrophys. J.* **226**, 55.
- Dressler, A., 1979, *Astrophys. J.* **231**, 659.
- Dressler, A., 1980a, *Astrophys. J. Suppl.* **42**, 565.
- Dressler, A., 1980b, *Astrophys. J.* **236**, 351.
- Dressler, A., 1981, *Astrophys. J.* **243**, 26.
- Dressler, A., 1984, *Ann. Rev. Astron. Astrophys.* **22**, 185.
- Dressler, A., and J. E. Gunn, 1982, *Astrophys. J.* **263**, 533.
- Dressler, A., J. E. Gunn, and D. P. Schneider, 1985, *Astrophys. J.* **294**, 70.
- Duus, A., and B. Newell, 1977, *Astrophys. J. Suppl.* **35**, 209.
- Eggen, O. J., D. Lynden-Bell, and A. Sandage, 1962, *Astrophys. J.* **136**, 748.
- Einasto, J., M. Joeveer, and E. Saar, 1980, *Mon. Not. R. Astron. Soc.* **193**, 353.
- Einasto, J., A. Kaasik, and E. Saar, 1974, *Nature* **250**, 309.
- Ellis, R. S., W. J. Couch, I. MacLaren, and D. C. Koo 1985, *Mon. Not. R. Astron. Soc.* **217**, 239.
- Elvis, M., 1976, *Mon. Not. R. Astron. Soc.* **177**, 7p.
- Elvis, M., B. A. Cooke, K. A. Pounds, and M. J. Turner, 1975, *Nature* **257**, 33.
- Elvis, M., E. Schreier, J. Tonry, M. Davis, and J. Huchra, 1981, *Astrophys. J.* **246**, 20.
- Erickson, W. C., T. A. Matthews, and M. R. Viner, 1978, *Astrophys. J.* **222**, 761.
- Fabbri, R., F. Melchiorri, and V. Natale, 1978, *Astrophys. Space Sci.* **59**, 223.
- Faber, S. M., and A. Dressler, 1976, *Astrophys. J. Lett.* **210**, L65.
- Faber, S. M., and A. Dressler, 1977, *Astron. J.* **82**, 187.
- Faber, S. M., and J. S. Gallagher, 1976, *Astrophys. J.* **204**, 365.
- Faber, S. M., and J. S. Gallagher, 1979, *Ann. Rev. Astron. Astrophys.* **17**, 135.
- Fabian, A. C., K. A. Arnaud, P. E. Nulsen, and R. F. Mushotzky, 1986a, *Astrophys. J.* **305**, 9.
- Fabian, A. C., K. A. Arnaud, P. E. Nulsen, M. G. Watson, G. C. Stewart, I. McHardy, A. Smith, B. Cook, M. Elvis, and R. F. Mushotzky, 1985, *Mon. Not. R. Astron. Soc.* **216**, 923.
- Fabian, A. C., K. A. Arnaud, and P. A. Thomas, 1986b, Preprint.
- Fabian, A. C., P. D. Atherton, K. Taylor, and P. E. Nulsen, 1982a, *Mon. Not. R. Astron. Soc.* **201**, L17.
- Fabian, A. C., E. M. Hu, L. L. Cowie, and J. Grindlay, 1981a, *Astrophys. J.* **248**, 47.
- Fabian, A. C., W. H. Ku, D. F. Malin, R. F. Mushotzky, P. E. Nulsen, and G. C. Stewart, 1981b, *Mon. Not. R. Astron. Soc.* **196**, 35p.
- Fabian, A. C., and P. E. Nulsen, 1977, *Mon. Not. R. Astron. Soc.* **180**, 479.
- Fabian, A. C., and P. E. Nulsen, 1979, *Mon. Not. R. Astron. Soc.* **186**, 783.
- Fabian, A. C., P. E. Nulsen, and K. A. Arnaud, 1984a, *Mon. Not. R. Astron. Soc.* **208**, 179.
- Fabian, A. C., P. E. Nulsen, and C. R. Canizares, 1982b, *Mon. Not. R. Astron. Soc.* **201**, 933.
- Fabian, A. C., P. E. Nulsen, and C. R. Canizares, 1984b, *Nature* **310**, 733.
- Fabian, A. C., and J. E. Pringle, 1977, *Mon. Not. R. Astron. Soc.* **181**, 5p.
- Fabian, A. C., J. E. Pringle, and M. J. Rees, 1976, *Nature* **263**, 301.
- Fabian, A. C., J. Schwarz, and W. Forman, 1980, *Mon. Not. R. Astron. Soc.* **192**, 135.
- Fabricant, D., and P. Gorenstein, 1983, *Astrophys. J.* **267**, 535.
- Fabricant, D., M. Lecar, and P. Gorenstein, 1980, *Astrophys. J.* **241**, 552.
- Fabricant, D., G. Rybicki, and P. Gorenstein, 1984, *Astrophys. J.* **286**, 186.
- Fabricant, D., K. Topka, F. R. Harnden, and P. Gorenstein, 1978, *Astrophys. J. Lett.* **226**, L107.
- Fall, S. M., 1981, in *The Structure and Evolution of Normal Galaxies*, edited by M. Fall and D. Lynden-Bell, p.1. Cambridge University Press.
- Fall, S. M., and M. J. Rees, 1985, *Astrophys. J.* **298**, 18.

- Falle, S.A., and P. Meszaros, 1980, *Mon. Not. R. Astron. Soc.* **190**, 195.
- Fanti, C., R. Fanti, L. Feretti, I. Gioia, G. Giovannini, L. Gregorini, B. Marano, L. Padrielli, P. Parma, P. Tomasi, and V. Zitelli, 1983, *Astron. Astrophys. Suppl.* **52**, 411.
- Farouki, R., G. L. Hoffman, and E. E. Salpeter, 1983, *Astrophys. J.* **271**, 11.
- Farouki, R., and E. E. Salpeter, 1982, *Astrophys. J.* **253**, 1.
- Farouki, R., and S. L. Shapiro, 1980, *Astrophys. J.* **241**, 928.
- Farouki, R., and S. L. Shapiro, 1981, *Astrophys. J.* **243**, 32.
- Feigelson, E. D., T. Maccacaro, and G. Zamorani, 1982, *Astrophys. J.* **255**, 392.
- Felten, J. E., R. J. Gould, W. A. Stein, and N. J. Woolf, 1966, *Astrophys. J.* **146**, 955.
- Felten, J. E., and P. Morrison, 1966, *Astrophys. J.* **146**, 686.
- Field, G. E., 1980, *Astron. Gesell. Mitt.* **47**, 7.
- Flannery, B. P., and M. Krook, 1978, *Astrophys. J.* **223**, 447.
- Fomalont, E., and A. H. Bridle, 1978, *Astrophys. J. Lett.* **223**, L9.
- Fomalont, E., and D. Rogstad, 1966, *Astrophys. J.* **146**, 52.
- Ford, H. C., and H. Butcher, 1979, *Astrophys. J. Suppl.* **41**, 147.
- Forman, W., J. Bechtold, W. Blair, R. Giacconi, L. Van Speybroeck, and C. Jones, 1981, *Astrophys. J. Lett.* **243**, L133.
- Forman, W., and C. Jones, 1982, *Ann. Rev. Astron. Astrophys.* **20**, 547.
- Forman, W., C. Jones, L. Cominsky, P. Julien, S. Murray, G. Peters, H. Tananbaum, and R. Giacconi, 1978a, *Astrophys. J. Suppl.* **38**, 357.
- Forman, W., C. Jones, S. Murray, and R. Giacconi, 1978b, *Astrophys. J. Lett.* **225**, L1.
- Forman, W., C. Jones, and W. Tucker, 1985, *Astrophys. J.* **293**, 102.
- Forman, W., E. Kellogg, H. Gursky, H. Tananbaum, and R. Giacconi, 1972, *Astrophys. J.* **178**, 309.
- Forman, W., J. Schwarz, C. Jones, W. Liller, and A. C. Fabian, 1979, *Astrophys. J. Lett.* **234**, L27.
- Forster, J. R., 1980, *Astrophys. J.* **238**, 54.
- Fritz, G., A. Davidsen, J. F. Meekins, and H. Friedman, 1971, *Astrophys. J. Lett.* **164**, L81.
- Gaetz, T. J., and E. E. Salpeter, 1983, *Astrophys. J. Suppl.* **55**, 155.
- Gaetz, T. J., E. E. Salpeter, and G. Shaviv, 1987, *Astrophys. J.* **316**, 530.
- Gallagher, J. S., 1978, *Astrophys. J.* **223**, 386.
- Gallagher, J. S., and J. P. Ostriker, 1972, *Astrophys. J.* **77**, 288.
- Gavazzi, G., 1978, *Astron. Astrophys.* **69**, 355.
- Gavazzi, G., G. C. Perola, and W. Jaffe, 1981, *Astron. Astrophys.* **103**, 35.
- Geller, M. J., and T. C. Beers, 1982, *Publ. Astron. Soc. Pac.* **94**, 421.
- Geller, M. J., and P. J. Peebles, 1976, *Astrophys. J.* **206**, 939.
- Giacconi, R. et al., 1979, *Astrophys. J.* **230**, 540.
- Giacconi, R. et al., 1980, *Advanced X-ray Astrophysics Facility—Science Working Group Report*, NASA Report No. TM-78285.
- Giacconi, R., S. Murray, H. Gursky, E. Kellogg, E. Schreier, T. Matilsky, D. Koch, and H. Tananbaum, 1974, *Astrophys. J. Suppl.* **27**, 37.
- Giacconi, R., S. Murray, H. Gursky, E. Kellogg, E. Schreier, and H. Tananbaum, 1972, *Astrophys. J.* **178**, 281.
- Gingold, R. A., and S. C. Perrenod, 1979, *Mon. Not. R. Astron. Soc.* **187**, 371.
- Gioia, I. M., M. J. Geller, J. P. Huchra, T. Maccacaro, J. E. Steiner, and J. Stocke, 1982, *Astrophys. J. Lett.* **255**, L17.
- Giovanardi, C., G. Helou, E. E. Salpeter, and N. Krumm, 1983, *Astrophys. J.* **267**, 35.
- Giovanelli, R., G. Chincarini, and M. P. Haynes, 1981, *Astrophys. J.* **247**, 383.
- Giovanelli, R., M. P. Haynes, and G. L. Chincarini, 1982, *Astrophys. J.* **262**, 422.
- Gisler, G. R., 1976, *Astron. Astrophys.* **51**, 137.
- Gisler, G. R., 1978, *Mon. Not. R. Astron. Soc.* **183**, 633.
- Gisler, G. R., 1979, *Astrophys. J.* **228**, 385.
- Gisler, G. R., and G. K. Miley, 1979, *Astron. Astrophys.* **76**, 109.
- Godwin, J. C., and J. V. Peach, 1977, *Mon. Not. R. Astron. Soc.* **181**, 323.
- Goldstein, S. J., 1966, *Science* **151**, 3706.

- Gorenstein, P., P. Bjorkholm, B. Harris, and F. Harnden, 1973, *Astrophys. J. Lett.* **183**, L57.
 Gorenstein, P., D. Fabricant, K. Topka, and F. Harnden, 1979, *Astrophys. J.* **230**, 26.
 Gorenstein, P., D. Fabricant, K. Topka, F. R. Harnden, and W. H. Tucker, 1978, *Astrophys. J.* **224**, 718.
 Gorenstein, P., D. Fabricant, K. Topka, W. Tucker, and F. Harnden, 1977, *Astrophys. J. Lett.* **216**, L95.
 Gott, J. R., 1977, *Ann. Rev. Astron. Astrophys.* **15**, 235.
 Gott, J. R., and T. X. Thuan, 1976, *Astrophys. J.* **204**, 649.
 Gould, R. J., and Y. Rephaeli, 1978, *Astrophys. J.* **219**, 12.
 Gregory, S. A., 1975, *Astrophys. J.* **199**, 1.
 Gregory, S. A., and L. A. Thompson, 1978, *Astrophys. J.* **222**, 784.
 Gregory, S. A., L. A. Thompson, and W. G. Tifft, 1981, *Astrophys. J.* **243**, 411.
 Gregory, S. A., and W. G. Tifft, 1976, *Astrophys. J.* **205**, 716.
 Grindlay, J. A., D. R. Parsignault, H. Gursky, A. C. Brinkman, J. Heise, and D. E. Harris, 1977, *Astrophys. J. Lett.* **214**, L57.
 Gudehus, D. H., 1973, *Astron. J.* **78**, 583.
 Gudehus, D. H., 1976, *Astrophys. J.* **208**, 267.
 Guindon, B., 1979, *Mon. Not. R. Astron. Soc.* **186**, 117.
 Guindon, B., and A. H. Bridle, 1978, *Mon. Not. R. Astron. Soc.* **184**, 221.
 Gull, S. F., and K. J. Northover, 1975, *Mon. Not. R. Astron. Soc.* **173**, 585.
 Gull, S. F., and K. J. Northover, 1976, *Nature* **263**, 572.
 Gunn, J. E., 1977, *Astrophys. J.* **218**, 592.
 Gunn, J. E., 1978, in *Observational Cosmology*, edited by A. Maeder, L. Martinet, and G. Tammann. Geneva: Geneva Observatory.
 Gunn, J. E., and J. R. Gott, 1972, *Astrophys. J.* **176**, 1.
 Gunn, J. E., and J. B. Oke, 1975, *Astrophys. J.* **195**, 255.
 Gunn, J. E., and B. M. Tinsley, 1976, *Astrophys. J.* **210**, 1.
 Gursky, H., E. Kellogg, C. Leong, H. Tananbaum, and R. Giacconi, 1971a, *Astrophys. J. Lett.* **165**, L43.
 Gursky, H., E. M. Kellogg, S. Murray, C. Leong, H. Tananbaum, and R. Giacconi, 1971b, *Astrophys. J. Lett.* **167**, L81.
 Gursky, H., and D. Schwartz, 1977, *Ann. Rev. Astron. Astrophys.* **15**, 541.
 Gursky, H., A. Solinger, E. Kellogg, S. Murray, H. Tananbaum, R. Giacconi, and A. Cavalieri, 1972, *Astrophys. J. Lett.* **173**, L99.
 Guthie, B. N., 1974, *Mon. Not. R. Astron. Soc.* **168**, 15.
 Hall, D. N., 1982, *The Space Telescope Observatory*, NASA Report No. CP-2244.
 Hall, A., and D. Sciama, 1979, *Astrophys. J. Lett.* **228**, L15.
 Hamilton, A. J., C. L. Sarazin, and R. A. Chevalier, 1983, *Astrophys. J. Suppl.* **51**, 115.
 Hanisch, R. J., 1980, *Astron. J.* **85**, 1565.
 Hanisch, R. J., 1982a, *Astron. Astrophys.* **111**, 97.
 Hanisch, R. J., 1982b, *Astron. Astrophys.* **116**, 137.
 Hanisch, R. J., and W. C. Erickson, 1980, *Astron. J.* **85**, 183.
 Hanisch, R. J., T. A. Matthews, and M. M. Davis, 1979, *Astron. J.* **84**, 946.
 Hanisch, R. J., and R. A. White, 1981, *Astron. J.* **86**, 806.
 Harris, D. E., P. Dewdney, C. Costain, H. Butcher, and A. Willis, 1983a, *Astrophys. J.* **270**, 39.
 Harris, D. E., V. K. Kapahi, and R. D. Ekers, 1980, *Astron. Astrophys. Suppl.* **39**, 215.
 Harris, D. E., and G. K. Miley, 1978, *Astron. Astrophys. Suppl.* **34**, 117.
 Harris, D. E., J. G. Robertson, P. E. Dewdney, and C. H. Costain, 1982, *Astron. Astrophys.* **111**, 299.
 Harris, D. E., and W. Romanishin, 1974, *Astrophys. J.* **188**, 209.
 Harris, E. W., M. G. Smith, and E. S. Myra, 1983b, *Astrophys. J.* **272**, 456.
 Hartwick, F. D., 1973, *Astrophys. J.* **219**, 345.
 Hartwick, F. D., 1976, *Astrophys. J. Lett.* **208**, L13.
 Haslam, C., P. Kronberg, H. Waldhausen, R. Wielebinski, and D. Schallwich, 1978, *Astron. Astrophys. Suppl.* **31**, 99.
 Hausman, M. A., and J. P. Ostriker, 1978, *Astrophys. J.* **224**, 320.
 Havlen, R. J., and H. Quintana, 1978, *Astrophys. J.* **220**, 14.

- Haynes, M. P., R. L. Brown, and M. S. Roberts, 1978, *Astrophys. J.* **221**, 414.
- Heckman, T. M., 1981, *Astrophys. J. Lett.* **250**, L59.
- Helfand, D., W. Ku, and F. Abramopoulos, 1980, *Highlights Astron.* **5**, 747.
- Helmkens, H., J. P. Delvaille, A. Epstein, M. J. Geller, H. W. Schnopper, and J. G. Jernigan, 1978, *Astrophys. J. Lett.* **221**, L43.
- Helou, G., and E. E. Salpeter, 1982, *Astrophys. J.* **252**, 75.
- Helou, G., E. E. Salpeter, and N. Krumm, 1979, *Astrophys. J. Lett.* **228**, L1.
- Henriksen, M. J., 1985, Ph.D. thesis University of Maryland.
- Henriksen, M. J., and R. F. Mushotzky, 1985, *Astrophys. J.* **292**, 441.
- Henriksen, M. J., and R. F. Mushotzky, 1986, *Astrophys. J.* **302**, 287.
- Henry, J. P., G. Branduardi, U. Briel, D. Fabricant, E. Feigelson, S. Murray, A. Soltan, and H. Tananbaum, 1979, *Astrophys. J. Lett.* **239**, L15.
- Henry, J. P., J. T. Clarke, S. Bowyer, and R. J. Lavery, 1983, *Astrophys. J.* **272**, 434.
- Henry, J. P., M. J. Henriksen, P. A. Charles, and J. R. Thorstensen, 1981, *Astrophys. J. Lett.* **243**, L137.
- Henry, J. P., and R. J. Lavery, 1984, *Astrophys. J.* **280**, 1.
- Henry, J. P., A. Soltan, U. Briel, and J. E. Gunn, 1982, *Astrophys. J.* **262**, 1.
- Henry, J. P., and W. Tucker, 1979, *Astrophys. J.* **229**, 78.
- Hesser, J. E., H. C. Harris, S. van den Bergh, and G. L. Harris, 1984, *Astrophys. J.* **276**, 491.
- Hickson, P., 1977, *Astrophys. J.* **217**, 16.
- Hickson, P., 1982, *Astrophys. J.* **255**, 382.
- Hickson, P., and P. J. Adams, 1979a, *Astrophys. J. Lett.* **234**, L87.
- Hickson, P., and P. J. Adams, 1979b, *Astrophys. J. Lett.* **234**, L91.
- Hill, J. M., J. R. Angel, J. S. Scott, D. Lindley, and P. Hintzen, 1980, *Astrophys. J. Lett.* **242**, L69.
- Hill, J. M., and M. S. Longair, 1971, *Mon. Not. R. Astron. Soc.* **154**, 125.
- Himmes, A., and P. Biermann, 1980, *Astron. Astrophys.* **86**, 11.
- Hintzen, P., G. O. Boeshaar, and J. S. Scott, 1981, *Astrophys. J. Lett.* **246**, L1.
- Hintzen, P., J. M. Hill, D. Lindley, J. S. Scott, and J. R. Angel, 1982, *Astron. J.* **87**, 1656.
- Hintzen, P., and J. S. Scott, 1978, *Astrophys. J. Lett.* **224**, L47.
- Hintzen, P., and J. S. Scott, 1979, *Astrophys. J. Lett.* **232**, L145.
- Hintzen, P., and J. S. Scott, 1980, *Astrophys. J.* **239**, 765.
- Hintzen, P., J. S. Scott, and J. D. McKee, 1980, *Astrophys. J.* **242**, 857.
- Hintzen, P., J. S. Scott, and M. Tarenghi, 1977, *Astrophys. J.* **212**, 8.
- Hintzen, P., J. Ulvestad, and F. Owen, 1983, *Astron. J.* **88**, 709.
- Hirayama, Y., 1978, *Prog. Theor. Phys.* **60**, 724.
- Hirayama, Y., Y. Tanaka, and T. Kogure, 1978, *Prog. Theor. Phys.* **59**, 751.
- Hoessel, J. G., 1980, *Astrophys. J.* **241**, 493.
- Hoessel, J. G., K. D. Borne, and D. P. Schneider, 1985, *Astrophys. J.* **293**, 94.
- Hoessel, J. G., J. E. Gunn, and T. X. Thuan, 1980, *Astrophys. J.* **241**, 486.
- Hoffman, A. A., and P. Crane, 1977, *Astrophys. J.* **215**, 379.
- Hoffman, G. L., and E. E. Salpeter, 1982, *Astrophys. J.* **263**, 485.
- Holman, G. D., J. A. Ionson, and J. S. Scott, 1979, *Astrophys. J.* **228**, 576.
- Holman, G. D., and J. D. McKee, 1981, *Astrophys. J.* **249**, 35.
- Holmberg, E. B., A. Lauberts, H. E. Schuster, and R. M. West, 1974, *Astron. Astrophys. Suppl.* **18**, 463.
- Holt, S., and R. McCray, 1982, *Ann. Rev. Astron. Astrophys.* **20**, 323.
- Hooley, T., 1974, *Mon. Not. R. Astron. Soc.* **166**, 259.
- Hu, E. M., L. L. Cowie, P. Kaaret, E. B. Jenkins, D. G. York, and F. L. Roesler, 1983, *Astrophys. J. Lett.* **275**, L27.
- Hu, E. M., L. L. Cowie, and Z. Wang, 1985, *Astrophys. J. Suppl.* **59**, 447.
- Hubble, E. P., 1930, *Astrophys. J.* **71**, 231.
- Huchra, J. P., and M. F. Geller, 1982, *Astrophys. J.* **257**, 423.
- Huchtmeier, W. K., G. A. Tammann, and H. J. Wendker, 1976, *Astron. Astrophys.* **46**, 381.
- Humason, M. L., N. U. Mayall, and A. R. Sandage, 1956, *Astron. J.* **61**, 97.
- Humason, M. L., and A. Sandage, 1957, *Carnegie Inst. Washington Yearb.* **56**, 62.

- Hunt, R., 1971, *Mon. Not. R. Astron. Soc.* **154**, 141.
- Ikeuchi, S., 1977, *Prog. Theor. Phys.* **58**, 1742.
- Ikeuchi, S., and Y. Hirayama, 1979, *Prog. Theor. Phys.* **61**, 881.
- Ikeuchi, S., and Y. Hirayama, 1980, *Prog. Theor. Phys.* **64**, 81.
- Ives, J. C., and P. W. Sanford, 1976, *Mon. Not. R. Astron. Soc.* **176**, 13p.
- Jaffe, W. J., 1977, *Astrophys. J.* **212**, 1.
- Jaffe, W. J., 1980, *Astrophys. J.* **241**, 924.
- Jaffe, W. J., 1982, *Astrophys. J.* **262**, 15.
- Jaffe, W. J., and G. C. Perola, 1973, *Astron. Astrophys.* **26**, 423.
- Jaffe, W. J., and G. Perola, 1975, *Astron. Astrophys. Suppl.* **21**, 137.
- Jaffe, W. J., and G. Perola, 1976, *Astron. Astrophys.* **46**, 273.
- Jaffe, W. J., G. C. Perola, and E. A. Valentijn, 1976, *Astron. Astrophys.* **49**, 179.
- Jaffe, W. J., and L. Rudnick, 1979, *Astrophys. J.* **233**, 453.
- Jenner, D. C., 1974, *Astrophys. J.* **191**, 55.
- Joeveer, M., J. Einasto, and E. Tago, 1978, *Mon. Not. R. Astron. Soc.* **185**, 357.
- Johnson, H. M., 1981, *Astrophys. J. Suppl.* **47**, 235.
- Johnson, H. M., H. W. Schnopper, and J. P. Delvalle, 1980, *Astrophys. J.* **236**, 738.
- Johnson, M. W., R. G. Cruddace, G. Fritz, S. Shulman, and H. Friedman, 1979, *Astrophys. J. Lett.* **231**, L45.
- Johnson, M. W., R. G. Cruddace, M. P. Ulmer, M. P. Kowalski, and K. S. Wood, 1983, *Astrophys. J.* **266**, 425.
- Johnston, M. D., H. V. Bradt, R. E. Doxsey, B. Margon, F. E. Marshall, and D. A. Schwartz, 1981, *Astrophys. J.* **245**, 799.
- Jones, C., and W. Forman, 1978, *Astrophys. J.* **224**, 1.
- Jones, C., and W. Forman, 1984, *Astrophys. J.* **276**, 38.
- Jones, C., E. Mandel, J. Schwarz, W. Forman, S. S. Murray, and F. R. Harnden, 1979, *Astrophys. J. Lett.* **234**, L21.
- Jones, T. W., and F. N. Owen, 1979, *Astrophys. J.* **234**, 818.
- Jura, M., 1977, *Astrophys. J.* **212**, 634.
- Just, K., 1959, *Astrophys. J.* **129**, 268.
- Karzas, W., and R. Latter, 1961, *Astrophys. J. Suppl.* **6**, 167.
- Karachentsev, I. D., and A. I. Kopylov, 1980, *Mon. Not. R. Astron. Soc.* **192**, 109.
- Kato, T., 1976, *Astrophys. J. Suppl.* **30**, 397.
- Katz, J. I., 1976, *Astrophys. J.* **207**, 25.
- Kellogg, E. M., 1973, in *X- and Gamma-Ray Astronomy*, edited by H. Bradt and R. Giacconi, p. 171. Dordrecht: Reidel.
- Kellogg, E. M., 1974, in *X-Ray Astronomy*, edited by R. Giacconi and H. Gursky, p. 321. Dordrecht: Reidel.
- Kellogg, E. M., 1975, *Astrophys. J.* **197**, 689.
- Kellogg, E. M., 1977, *Astrophys. J.* **218**, 582.
- Kellogg, E. M., 1978, *Astrophys. J. Lett.* **220**, L63.
- Kellogg, E. M., J. R. Baldwin, and D. Koch, 1975, *Astrophys. J.* **199**, 299.
- Kellogg, E. M., H. Gursky, C. Leong, E. Schreier, H. Tananbaum, and R. Giacconi, 1971, *Astrophys. J. Lett.* **165**, L49.
- Kellogg, E. M., H. Gursky, H. Tananbaum, R. Giacconi, and K. Pounds, 1972, *Astrophys. J. Lett.* **174**, L65.
- Kellogg, E. M., and S. Murray, 1974, *Astrophys. J. Lett.* **193**, L57.
- Kellogg, E. M., S. Murray, R. Giacconi, H. Tananbaum, and H. Gursky, 1973, *Astrophys. J. Lett.* **185**, L13.
- Kent, S. M., and J. E. Gunn, 1982, *Astron. J.* **87**, 945.
- Kent, S. M., and W. L. Sargent, 1979, *Astrophys. J.* **230**, 667.
- Kent, S. M., and W. L. Sargent, 1983, *Astron. J.* **88**, 697.
- Kiang, T., 1961, *Mon. Not. R. Astron. Soc.* **122**, 263.

- King, I. R., 1962, *Astron. J.* **67**, 471.
 King, I. R., 1966, *Astron. J.* **71**, 64.
 Kirshner, R. P., A. Oemler, and P. L. Schechter, 1978, *Astron. J.* **83**, 1549.
 Kirshner, R. P., A. Oemler, P. L. Schechter, and S. A. Schectman, 1981, *Astrophys. J. Lett.* **248**, L57.
 Klemola, A. R., 1969, *Astron. J.* **74**, 804.
 Knobloch, E., 1978a, *Astrophys. J.* **222**, 779.
 Knobloch, E., 1978b, *Astrophys. J. Suppl.* **38**, 253.
 Koo, D. C., 1981, *Astrophys. J. Lett.* **251**, L75.
 Kotanyi, C., J. H. van Gorkom, and R. D. Ekers, 1983, *Astrophys. J. Lett.* **273**, L7.
 Kowalski, M., M. P. Ulmer, and R. Cruddace, 1983, *Astrophys. J.* **268**, 540.
 Kowalski, M., M. P. Ulmer, R. Cruddace, and K. S. Wood, 1984, *Astrophys. J. Suppl.* **56**, 403.
 Kraan-Korteweg, R. C., 1981, *Astron. Astrophys.* **104**, 280.
 Kriss, G. A., C. E. Canizares, J. E. McClintock, and E. D. Feigelson, 1980, *Astrophys. J. Lett.* **235**, L61.
 Kriss, G. A., C. E. Canizares, J. E. McClintock, and E. D. Feigelson, 1981, *Astrophys. J. Lett.* **245**, L51
 erratum for above.
 Kriss, G. A., D. F. Cioffi, and C. R. Canizares, 1983, *Astrophys. J.* **272**, 439.
 Kristian, J., A. Sandage, and J. A. Westphal, 1978, *Astrophys. J.* **221**, 383.
 Kritsuk, A. G., 1983, *Astrophys.* **19**, 263.
 Kron, R. G., H. Spinrad, and I. R. King, 1977, *Astrophys. J.* **217**, 951.
 Krumm, N., and E. E. Salpeter, 1976, *Astrophys. J. Lett.* **208**, L7.
 Krumm, N., and E. E. Salpeter, 1979, *Astrophys. J.* **227**, 776.
 Krupp, E. C., 1974, *Publ. Astron. Soc. Pac.* **86**, 385.
 Ku, W., F. Abramopoulos, P. Nulsen, A. Fabian, G. Stewart, G. Chincarini, and M. Tarenghi, 1983, *Mon. Not. R. Astron. Soc.* **203**, 253.
 Lake, G., and R. B. Partridge, 1977, *Nature* **270**, 502.
 Lake, G., and R. B. Partridge, 1980, *Astrophys. J.* **237**, 378.
 Larson, R. B., and H. L. Dinerstein, 1975, *Publ. Astron. Soc. Pac.* **87**, 911.
 Larson, R. B., B. M. Tinsley, and C. N. Caldwell, 1980, *Astrophys. J.* **237**, 692.
 Lasenby, A. N., and R. D. Davies, 1983, *Mon. Not. R. Astron. Soc.* **203**, 1137.
 Lawler, J. M., and B. Dennison, 1982, *Astrophys. J.* **252**, 81.
 Lawrence, A., 1978, *Mon. Not. R. Astron. Soc.* **185**, 423.
 Lea, S. M., 1975, *Astrophys. Lett.* **16**, 141.
 Lea, S. M., 1976, *Astrophys. J.* **203**, 569.
 Lea, S. M., and D. S. De Young, 1976, *Astrophys. J.* **210**, 647.
 Lea, S. M., and G. D. Holman, 1978, *Astrophys. J.* **222**, 29.
 Lea, S. M., K. Mason, G. Reichert, P. Charles, and G. Reigler, 1979, *Astrophys. J. Lett.* **227**, L67.
 Lea, S. M., R. Mushotzky, and S. Holt, 1982, *Astrophys. J.* **262**, 24.
 Lea, S. M., G. Reichert, R. Mushotzky, W. A. Baity, D. E. Gruber, R. Rothschild, and F. A. Primini, 1981,
 Astrophys. J. **246**, 369.
 Lea, S. M., J. Silk, E. Kellogg, and S. Murray, 1973, *Astrophys. J. Lett.* **184**, L105.
 Lecar, M., 1975, in *Dynamics of Stellar Systems*, edited by A. Hayli, p. 161. Dordrecht: Reidel.
 Leir, A. A., and S. van den Bergh, 1977, *Astrophys. J. Suppl.* **34**, 381.
 Limber, D. N., and W. G. Mathews, 1960, *Astrophys. J.* **132**, 286.
 Livio, M., O. Regev, and G. Shaviv, 1978, *Astron. Astrophys.* **70**, L7.
 Livio, M., O. Regev, and G. Shaviv, 1980, *Astrophys. J. Lett.* **240**, L83.
 Lugger, P. M., 1978, *Astrophys. J.* **221**, 745.
 Lynden-Bell, D., 1967, *Mon. Not. R. Astron. Soc.* **136**, 101.
 Lynds, R., 1970, *Astrophys. J. Lett.* **159**, L151.
 Maccacaro, T., B. A. Cooke, M. J. Ward, M. V. Penston, and R. F. Haynes, 1977, *Mon. Not. R. Astron. Soc.* **180**, 465.
 Maccagni, D., and M. Tarenghi, 1981, *Astrophys. J.* **243**, 42.
 MacDonald, R., S. Kenderdine, and A. Neville, 1968, *Mon. Not. R. Astron. Soc.* **138**, 259.
 MacGillivray, H. T., and R. J. Dodd, 1979, *Mon. Not. R. Astron. Soc.* **186**, 743.

- MacGillivray, H. T., R. Martin, N. M. Pratt, V. C. Reddish, H. Seddon, L. W. Alexander, G. S. Walker, and P. R. Williams, 1976, *Mon. Not. R. Astron. Soc.* **176**, 649.
- Malina, R., M. Lampton, and S. Bowyer, 1976, *Astrophys. J.* **209**, 678.
- Malina, R. F., S. M. Lea, M. Lampton, and C. S. Bowyer, 1978, *Astrophys. J.* **219**, 795.
- Malumuth, E. M., and R. P. Kirshner, 1981, *Astrophys. J.* **251**, 508.
- Malumuth, E. M., and R. P. Kirshner, 1985, *Astrophys. J.* **291**, 8.
- Malumuth, E. M., and G. A. Kriss, 1986, *Astrophys. J.* **308**, 10.
- Malumuth, E. M., and D. O. Richstone, 1984, *Astrophys. J.* **276**, 413.
- Marchant, A. B., and S. L. Shapiro, 1977, *Astrophys. J.* **215**, 1.
- Margon, B., M. Lampton, S. Bowyer, and R. Cruddace, 1975, *Astrophys. J.* **197**, 25.
- Markert, T. H., C. R. Canizares, G. W. Clark, F. K. Li, P. L. Northridge, G. F. Sprott, and G. F. Wargo, 1976, *Astrophys. J.* **206**, 265.
- Markert, T., P. Winkler, F. Laird, G. Clark, D. Hearn, G. Sprott, F. Li, V. Bradt, W. Lewin, and H. Schnopper, 1979, *Astrophys. J. Suppl.* **39**, 573.
- Marshall, F. E., E. A. Boldt, S. S. Holt, R. F. Mushotzky, S. H. Pravdo, R. E. Rothschild, and P. J. Serlemitsos, 1979, *Astrophys. J. Suppl.* **40**, 657.
- Mason, K. O., H. Spinrad, S. Bowyer, G. Reichert, and J. Stauffer, 1981, *Astron. J.* **86**, 803.
- Materne, J., G. Chincarini, M. Tarenghi, and U. Hopp, 1982, *Astron. Astrophys.* **109**, 238.
- Mathews, T. A., W. W. Morgan, and M. Schmidt, 1964, *Astrophys. J.* **140**, 35.
- Mathews, W. G., 1978a, *Astrophys. J.* **219**, 408.
- Mathews, W. G., 1978b, *Astrophys. J.* **219**, 413.
- Mathews, W. G., and J. C. Baker, 1971, *Astrophys. J.* **170**, 241.
- Mathews, W. G., and J. N. Bregman, 1978, *Astrophys. J.* **224**, 308.
- Mathieu, R. D., and H. Spinrad, 1981, *Astrophys. J.* **251**, 485.
- Matilsky, T., C. Jones, and W. Forman, 1985, *Astrophys. J.* **291**, 621.
- McGlynn, T. A., and A. C. Fabian, 1984, *Mon. Not. R. Astron. Soc.* **208**, 709.
- McGlynn, T. A., and J. P. Ostriker, 1980, *Astrophys. J.* **241**, 915.
- McHardy, I. M., 1974, *Mon. Not. R. Astron. Soc.* **169**, 527.
- McHardy, I. M., 1978a, *Mon. Not. R. Astron. Soc.* **184**, 783.
- McHardy, I. M., 1978b, *Mon. Not. R. Astron. Soc.* **185**, 927.
- McHardy, I. M., 1979, *Mon. Not. R. Astron. Soc.* **188**, 495.
- McHardy, I. M., A. Lawrence, J. P. Pye, and K. A. Pounds, 1981, *Mon. Not. R. Astron. Soc.* **197**, 893.
- McKee, C. F., and L. L. Cowie, 1977, *Astrophys. J.* **215**, 213.
- McKee, J. D., R. F. Mushotzky, E. A. Boldt, S. S. Holt, F. E. Marshall, S. H. Pravdo, and P. J. Serlemitsos, 1980, *Astrophys. J.* **242**, 843.
- Meekins, J. F., F. Gilbert, T. A. Chubb, H. Friedman, and R. C. Henry, 1971, *Nature* **231**, 107.
- Melnick, J., and H. Quintana, 1975, *Astrophys. J. Lett.* **198**, L97.
- Melnick, J., and H. Quintana, 1981a, *Astron. Astrophys. Suppl.* **44**, 87.
- Melnick, J., and H. Quintana, 1981b, *Astron. J.* **86**, 1567.
- Melnick, J., and W. L. Sargent, 1977, *Astrophys. J.* **215**, 401.
- Melnick, J., S. D. M. White, and J. Hoessel, 1977, *Mon. Not. R. Astron. Soc.* **180**, 207.
- Merritt, D., 1983, *Astrophys. J.* **264**, 24.
- Merritt, D., 1984a, *Astrophys. J.* **276**, 26.
- Merritt, D., 1984b, *Astrophys. J. Lett.* **280**, L5.
- Merritt, D., 1985, *Astrophys. J.* **289**, 18.
- Mewe, R., and E. H. Gronenschild, 1981, *Astron. Astrophys. Suppl.* **45**, 11.
- Meyer, S. S., A. D. Jeffries, and R. Weiss, 1983, *Astrophys. J.* **271**, L1.
- Miley, G., 1973, *Astron. Astrophys.* **26**, 413.
- Miley, G., 1980, *Ann. Rev. Astron. Astrophys.* **18**, 165.
- Miley, G., and D. E. Harris, 1977, *Astron. Astrophys.* **61**, L23.
- Miley, G., G. Perola, P. van der Kruit, and H. van der Laan, 1972, *Nature* **237**, 269.
- Miller, G. E., 1983, *Astrophys. J.* **274**, 840.
- Mills, B. Y., 1960, *Aust. J. Phys.* **13**, 550.

- Minkowski, R., 1960, *Astrophys. J.* **132**, 908.
- Minkowski, R., 1961, *Astron. J.* **66**, 558.
- Minkowski, R., and G. O. Abell, 1963, in *Basic Astronomical Data*, edited by K. A. Strand, p. 481. Chicago: University of Chicago.
- Mitchell, R. J., P. A. Charles, J. L. Culhane, P. J. Davison, and A.C. Fabian, 1975, *Astrophys. J. Lett.* **200**, L5.
- Mitchell, R. J., and J. L. Culhane, 1977, *Mon. Not. R. Astron. Soc.* **178**, 75p.
- Mitchell, R. J., J. L. Culhane, P. J. Davison, and J. C. Ives, 1976, *Mon. Not. R. Astron. Soc.* **176**, 29p.
- Mitchell, R. J., R. J. Dickens, S. J. Bell Burnell, and J. L. Culhane, 1979, *Mon. Not. R. Astron. Soc.* **189**, 329.
- Mitchell, R. J., J. C. Ives, and J. L. Culhane, 1977, *Mon. Not. R. Astron. Soc.* **131**, 25p.
- Mitchell, R. J., and R. Mushotzky, 1980, *Astrophys. J.* **236**, 730.
- Morgan, W. W., 1961, *Proc. Natl. Acad. Sci. U. S. A.* **47**, 905.
- Morgan, W. W., S. Kayser, and R. A. White, 1975, *Astrophys. J.* **199**, 545.
- Morgan, W. W., and J. Lesh, 1965, *Astrophys. J.* **142**, 1364.
- Moss, C., and R. J. Dickens, 1977, *Mon. Not. R. Astron. Soc.* **178**, 701.
- Mottmann, J., and G. O. Abell, 1977, *Astrophys. J.* **218**, 53.
- Murray, S. S., W. Forman, C. Jones, and R. Giacconi, 1978, *Astrophys. J. Lett.* **219**, L89.
- Mushotzky, R. F., 1980, in *X-ray Astronomy*, edited by R. Giacconi and G. Setti, p. 171. Dordrecht: Reidel.
- Mushotzky, R. F., 1984, *Phys. Scripta* **T7**, 157.
- Mushotzky, R. F., 1985, in *Proceedings of the Conference on Non-thermal and Very High Temperature Phenomena in X-ray Astronomy*, edited by G. Perola and M. Salvati.
- Mushotzky, R. F., W. A. Baity, and L. E. Peterson, 1977, *Astrophys. J.* **212**, 22.
- Mushotzky, R. F., S. S. Holt, B.W. Smith, E.A. Boldt, and P. J. Serlemitsos, 1981, *Astrophys. J. Lett.* **244**, L47.
- Mushotzky, R. F., P. J. Serlemitsos, B. W. Smith, E. A. Boldt, and S. S. Holt, 1978, *Astrophys. J.* **225**, 21.
- Nepveu, M., 1981a, *Astron. Astrophys.* **98**, 65.
- Nepveu, M., 1981b, *Astron. Astrophys.* **101**, 362.
- Noonan, T. W., 1973, *Astron. J.* **78**, 26.
- Noonan, T. W., 1974, *Astron. J.* **79**, 775.
- Noonan, T. W., 1975, *Astrophys. J.* **202**, 551.
- Noonan, T. W., 1976, *Astrophys. J.* **196**, 683.
- Noonan, T. W., 1980, *Astrophys. J.* **238**, 793.
- Noonan, T. W., 1981, *Astrophys. J. Suppl.* **45**, 613.
- Noordam, J. E., and A. G. de Bruyn, 1982, *Nature* **299**, 597.
- Norman, C., and J. Silk, 1979, *Astrophys. J. Lett.* **233**, L1.
- Nugent, J., K. Jensen, J. Nousek, G. Garmire, K. Mason, F. Walter, C. Bowyer, R. Stern, and G. Riegler, 1983, *Astrophys. J. Suppl.* **51**, 1.
- Nulsen, P. E., 1982, *Mon. Not. R. Astron. Soc.* **198**, 1007.
- Nulsen, P. E., and A. C. Fabian, 1980, *Mon. Not. R. Astron. Soc.* **191**, 887.
- Nulsen, P. E., A. C. Fabian, R. F. Mushotzky, E. A. Boldt, S. S. Holt, F.J. Marshall, and P. J. Serlemitsos, 1979, *Mon. Not. R. Astron. Soc.* **189**, 183.
- Nulsen, P. E., G. C. Stewart, and A. C. Fabian, 1984, *Mon. Not. R. Astron. Soc.* **208**, 185.
- Nulsen, P. E., G. C. Stewart, A. C. Fabian, R. F. Mushotzky, S. S. Holt, W. Ku, and D. Malin, 1982, *Mon. Not. R. Astron. Soc.* **199**, 1089.
- O'Connell, R. W., B. McNamara and C. L. Sarazin, 1987, Preprint.
- O'Dea, C. P., and F. N. Owen, 1986, *Astrophys. J.* **301**, 841.
- Oemler, A. Jr., 1973, *Astrophys. J.* **180**, 11.
- Oemler, A. Jr., 1974, *Astrophys. J.* **194**, 1.
- Oemler, A. Jr., 1976, *Astrophys. J.* **209**, 693.
- Omer, G. C., T. L. Page, and A. G. Wilson, 1965, *Astron. J.* **70**, 440.
- Oort, J. H., 1983, *Ann. Rev. Astron. Astrophys.* **21**, 373.

- Osmer, P. S., 1978, *Phys. Scripta* **17**, 357.
- Osterbrock, D. E., 1974, *Astrophysics of Gaseous Nebulae*. San Francisco: Freeman.
- Ostriker, J. P., 1980, *Comm. Astrophys.* **8**, 177.
- Ostriker, J. P., and M. A. Hausman, 1977, *Astrophys. J. Lett.* **217**, L125.
- Ostriker, J. P., P. J. Peebles, and A. Yahil, 1974, *Astrophys. J. Lett.* **193**, L1.
- Ostriker, J. P., and S. D. Tremaine, 1975, *Astrophys. J. Lett.* **202**, L113.
- Owen, F. N., 1974, *Astrophys. J. Lett.* **189**, 155.
- Owen, F. N., 1975, *Astrophys. J.* **195**, 593.
- Owen, F. N., J. O. Burns, and L. Rudnick, 1978, *Astrophys. J. Lett.* **226**, L119.
- Owen, F. N., J. O. Burns, L. Rudnick, and E. W. Greison, 1979, *Astrophys. J. Lett.* **229**, L59.
- Owen, F. N., C. P. O'Dea, M. Inoue, and J. A. Eilek, 1985, *Astrophys. J. Lett.* **294**, L85.
- Owen, F. N., and L. Rudnick, 1976a, *Astrophys. J.* **203**, 307.
- Owen, F. N., and L. Rudnick, 1976b, *Astrophys. J. Lett.* **205**, L1.
- Owen, F. N., R. A. White, K. C. Hilldrup, and R. J. Hanisch, 1982, *Astron. J.* **87**, 1083.
- Paal, G., 1964, in *Mitt. d. Sternwarted. Ungarischen Akad. d. Wissenschaften*, no. 54.
- Pacholczyk, A. G., and J. S. Scott, 1976, *Astrophys. J.* **203**, 313.
- Pariiskii, Y. N., 1973, *Sov. Astron. AJ* **16**, 1048.
- Peach, J. V., 1969, *Nature* **223**, 1140.
- Peebles, P. J., 1974, *Astron. Astrophys.* **32**, 197.
- Perola, G. C., and M. Reinhardt, 1972, *Astron. Astrophys.* **17**, 432.
- Perrenod, S. C., 1978a, *Astrophys. J.* **224**, 285.
- Perrenod, S. C., 1978b, *Astrophys. J.* **226**, 566.
- Perrenod, S. C., 1980, *Astrophys. J.* **236**, 373.
- Perrenod, S. C., and J. P. Henry, 1981, *Astrophys. J. Lett.* **247**, L1.
- Perrenod, S. C., and C. J. Lada, 1979, *Astrophys. J. Lett.* **234**, L173.
- Perrenod, S. C., and M. P. Lesser, 1980, *Publ. Astron. Soc. Pac.* **92**, 764.
- Peterson, B. M., 1978, *Astrophys. J.* **223**, 740.
- Peterson, B. M., S. E. Strom, and K. M. Strom, 1979, *Astron. J.* **84**, 735.
- Piccinotti, G., R. F. Mushotzky, E. A. Boldt, S. S. Holt, F. E. Marshall, P. J. Serlemitsos, and R. A. Shafer, 1982, *Astrophys. J.* **253**, 485.
- Pravdo, S., E. Boldt, F. Marshall, J. McKee, R. Mushotzky, and B. Smith, 1979, *Astrophys. J.* **234**, 1.
- Press, W. H., 1976, *Astrophys. J.* **203**, 14.
- Press, W. H., and P. Schechter, 1974, *Astrophys. J.* **187**, 425.
- Primini, F., E. Basinska, S. Howe, F. Lang, A. Levine, W. Lewin, R. Rothschild, W. Baity, D. Gruber, F. Knight, J. Matteson, S. Lea, and G. Reichert, 1981, *Astrophys. J. Lett.* **243**, L13.
- Pye, J. P., and B. A. Cooke, 1976, *Mon. Not. R. Astron. Soc.* **177**, 21p.
- Quintana, H., 1979, *Astron. J.* **84**, 15.
- Quintana, H., and D. G. Lawrie, 1982, *Astron. J.* **87**, 1.
- Quintana, H., and J. Melnick, 1982, *Astron. J.* **87**, 972.
- Raymond, J. C., D. P. Cox, and B. W. Smith, 1976, *Astrophys. J.* **204**, 290.
- Raymond, J. C., and B. W. Smith, 1977, *Astrophys. J. Suppl.* **35**, 419.
- Rees, M. J., 1978, *Phys. Scripta* **17**, 193.
- Reichert, G., K. Mason, S. Lea, P. Charles, S. Bowyer, and S. Pravdo, 1981, *Astrophys. J.* **247**, 803.
- Rephaeli, Y., 1977a, *Astrophys. J.* **212**, 608.
- Rephaeli, Y., 1977b, *Astrophys. J.* **218**, 323.
- Rephaeli, Y., 1978, *Astrophys. J.* **225**, 335.
- Rephaeli, Y., 1979, *Astrophys. J.* **227**, 364.
- Rephaeli, Y., 1980, *Astrophys. J.* **241**, 858.
- Rephaeli, Y., 1981, *Astrophys. J.* **245**, 351.
- Rephaeli, Y., and E. E. Salpeter, 1980, *Astrophys. J.* **240**, 20.
- Richstone, D. O., 1975, *Astrophys. J.* **200**, 535.
- Richstone, D. O., 1976, *Astrophys. J.* **204**, 642.
- Richstone, D. O., and E. M. Malumuth, 1983, *Astrophys. J.* **268**, 30.

- Richter, O. E., J. Materne, and W. K. Huchtmeier, 1982, *Astron. Astrophys.* **111**, 193.
- Ricketts, M. J., 1978, *Mon. Not. R. Astron. Soc.* **183**, 51p.
- Riley, J. M., 1975, *Mon. Not. R. Astron. Soc.* **170**, 53.
- Robertson, J. G., 1985, Preprint.
- Roland, J., H. Sol, I. Pauliny-Toth, and A. Witzel, 1981, *Astron. Astrophys.* **100**, 7.
- Roland, J., P. Veron, I. Pauliny-Toth, E. Preuss, and A. Witzel, 1976, *Astron. Astrophys.* **50**, 165.
- Rood, H. J., 1965, Ph.D. thesis University of Michigan.
- Rood, H. J., 1969, *Astrophys. J.* **158**, 657.
- Rood, H. J., 1975, *Astrophys. J.* **201**, 551.
- Rood, H. J., 1976, *Astrophys. J.* **207**, 16.
- Rood, H. J., 1981, *Rep. Prog. Phys.* **44**, 1077.
- Rood, H. J., and G. O. Abell, 1973, *Astrophys. Lett.* **13**, 69.
- Rood, H. J., and A. A. Leir, 1979, *Astrophys. J. Lett.* **231**, L3.
- Rood, H. J., T. L. Page, E. C. Kintner, and I. R. King, 1972, *Astrophys. J.* **175**, 627.
- Rood, H. J., V. C. Rothman, and B.E. Turnrose, 1970, *Astrophys. J.* **162**, 411.
- Rood, H. J., and G. N. Sastry, 1971, *Publ. Astron. Soc. Pac.* **83**, 313.
- Rood, H. J., and G. N. Sastry, 1972, *Astron. J.* **77**, 451.
- Roos, N., and C. A. Norman, 1979, *Astron. Astrophys.* **95**, 349.
- Rose, J. A., 1976, *Astron. Astrophys. Suppl.* **23**, 109.
- Rothenflug, R., L. Vigroux, R. Mushotzky, and S. Holt, 1984, *Astrophys. J.* **279**, 53.
- Rothschild, R. E., W. A. Baity, A. P. Marscher, and W. A. Wheaton, 1981, *Astrophys. J. Lett.* **243**, L9.
- Rowan-Robinson, M., and A. C. Fabian, 1975, *Mon. Not. R. Astron. Soc.* **170**, 199.
- Rubin, V. C., W. K. Ford, C. J. Peterson, and C. R. Lynds, 1978, *Astrophys. J. Suppl.* **37**, 235.
- Rubin, V. C., W. K. Ford, C. J. Peterson, and J. H. Oort, 1977, *Astrophys. J.* **211**, 693.
- Ruderman, M. A., and E. A. Spiegel, 1971, *Astrophys. J.* **165**, 1.
- Rudnick, L., 1978, *Astrophys. J.* **223**, 37.
- Rudnick, L., and F. N. Owen, 1976a, *Astron. J.* **82**, 1.
- Rudnick, L., and F. N. Owen, 1976b, *Astrophys. J. Lett.* **203**, L107.
- Ryle, M., and M. D. Windram, 1968, *Mon. Not. R. Astron. Soc.* **138**, 1.
- Salpeter, E. E., 1955, *Astrophys. J.* **121**, 161.
- Salpeter, E. E., and J. M. Dickey, 1985, *Astrophys. J.* **292**, 426.
- Sandage, A., 1972, *Astrophys. J.* **178**, 1.
- Sandage, A., 1976, *Astrophys. J.* **205**, 6.
- Sandage, A., K. C. Freeman, and N. R. Stokes, 1970, *Astrophys. J.* **160**, 831.
- Sandage, A., J. Kristian, and J. A. Westphal, 1976, *Astrophys. J.* **205**, 688.
- Sandage, A., and G. A. Tammann, 1975, *Astrophys. J.* **197**, 265.
- Sandage, A., and N. Visvanathan, 1978, *Astrophys. J.* **225**, 742.
- Sarazin, C. L., 1979, *Astrophys. Lett.* **20**, 93.
- Sarazin, C. L., 1980, *Astrophys. J.* **236**, 75.
- Sarazin, C. L., 1986a, *Rev. Mod. Phys.* **58**, 1.
- Sarazin, C. L., 1986b, in *Proceedings of the Greenbank Workshop on Gaseous Halos of Galaxies*, edited by J. Bregman and F. Lockman, p. 223.
- Sarazin, C. L., 1987, in *Proceedings of IAU Symposium 117: Dark Matter in the Universe*, edited by G. Knapp and J. Kormendy, p. 183. Dordrecht: Reidel.
- Sarazin, C. L., and J. N. Bahcall, 1977, *Astrophys. J. Suppl.* **34**, 451.
- Sarazin, C. L., and R. W. O'Connell, 1983, *Astrophys. J.* **268**, 552.
- Sarazin, C. L., and H. Quintana, 1987, Preprint.
- Sarazin, C. L., H. J. Rood, and M. F. Struble, 1982, *Astron. Astrophys.* **108**, L7.
- Sargent, W. L., 1973, *Publ. Astron. Soc. Pac.* **85**, 281.
- Sastry, G. N., 1968, *Publ. Astron. Soc. Pac.* **80**, 252.
- Sastry, G. N., and H. J. Rood, 1971, *Astrophys. J. Suppl.* **23**, 371.
- Schallwich, D., and R. Wielebinski, 1978, *Astron. Astrophys.* **71**, L15.
- Schechter, P. L., 1976, *Astrophys. J.* **203**, 297.

- Schechter, P. L., and P. J. E. Peebles, 1976, *Astrophys. J.* **209**, 670.
 Schechter, P. L., and W. H. Press, 1976, *Astrophys. J.* **203**, 557.
 Scheepmaker, A., G. Ricker, K. Brecher, S. Ryckman, J. Ballantine, J. Doty, P. Downey, and W Lewin, 1976, *Astrophys. J. Lett.* **205**, L65.
 Schild, R., and M. Davis, 1979, *Astron. J.* **84**, 311.
 Schipper, L., 1974, *Mon. Not. R. Astron. Soc.* **168**, 21.
 Schipper, L., and I. R. King, 1978, *Astrophys. J.* **220**, 798.
 Schmidt, M., 1978, *Phys. Scripta* **17**, 329.
 Schneider, D. P., and J. E. Gunn, 1982, *Astrophys. J.* **263**, 14.
 Schneider, D. P., J. E. Gunn, and J. G. Hoessel, 1983, *Astrophys. J.* **264**, 337.
 Schnopper, H. W., J. P. Delvaille, A. Epstein, H. Helmken, D. E. Harris, R. G. Strom, G. W. Clark, and J. G. Jernigan, 1977, *Astrophys. J. Lett.* **217**, L15.
 Schreier, E., P. Gorenstein, and E. Feigelson, 1982, *Astrophys. J.* **261**, 42.
 Schwartz, D. A., 1976, *Astrophys. J. Lett.* **206**, L95.
 Schwartz, D. A., 1978, *Astrophys. J.* **220**, 8.
 Schwartz, D. A., M. Davis, R. E. Doxsey, R. E. Griffiths, J. Huchra, M. D. Johnston, R. F. Mushotzky, J. Swank, and J. Tonry, 1980a, *Astrophys. J. Lett.* **238**, L53.
 Schwartz, D. A., J. Schwarz, and W. Tucker, 1980b, *Astrophys. J. Lett.* **238**, L59.
 Schwarz, J., U. Briel, R. Doxsey, G. Fabbiano, R. Griffiths, M. Johnston, D. Schwartz, J. McKee, and R. Mushotzky, 1979, *Astrophys. J. Lett.* **231**, L105.
 Schwarzschild, M., 1954, *Astron. J.* **59**, 273.
 Schwarzschild, M., 1979, *Astrophys. J.* **232**, 236.
 Scott, E. L., 1962, in *Problems in Extra-Galactic Research*, edited by G. C. McVittie, p. 269. Chicago: University of Chicago.
 Scott, J. S., G. D. Holman, J. A. Ionson, and K. Papadopoulos, 1980, *Astrophys. J.* **239**, 769.
 Serlemitsos, P. J., B. W. Smith, E. A. Boldt, S. S. Holt, and J. H. Swank, 1977, *Astrophys. J. Lett.* **211**, L63.
 Sersic, J. L., 1974, *Astrophys. Space Sci.* **28**, 365.
 Shapiro, P. R., and J. N. Bahcall, 1980, *Astrophys. J.* **241**, 1.
 Shaviv, G., and E. E. Salpeter, 1982, *Astron. Astrophys.* **110**, 300.
 Sheetman, S. A., 1982, *Astrophys. J.* **262**, 9.
 Shibasaki, N., R. Hoshi, F. Takahara, and S. Ikeuchi, 1976, *Prog. Theor. Phys.* **56**, 1475.
 Shostak, G., D. Gilra, J. Noordam, H. Nieuwenhuijzen, T. deGraauw, and J. Vermne, 1980, *Astron. Astrophys.* **81**, 223.
 Shu, F. H., 1978, *Astrophys. J.* **225**, 83.
 Shull, J. M., 1981, *Astrophys. J. Suppl.* **46**, 27.
 Shull, J. M., and M. Van Steenberg, 1982, *Astrophys. J. Suppl.* **48**, 95.
 Silk, J., 1976, *Astrophys. J.* **208**, 646.
 Silk, J., 1978, *Astrophys. J.* **220**, 390.
 Silk, J., and S. D. M. White, 1978, *Astrophys. J. Lett.* **226**, L103.
 Simon, A. J., 1978, *Mon. Not. R. Astron. Soc.* **184**, 537.
 Simon, A. J., 1979, *Mon. Not. R. Astron. Soc.* **188**, 637.
 Slingo, A., 1974a, *Mon. Not. R. Astron. Soc.* **166**, 101.
 Slingo, A., 1974b, *Mon. Not. R. Astron. Soc.* **168**, 307.
 Smith, B., R. Mushotzky, and P. Serlemitsos, 1979a, *Astrophys. J.* **227**, 37.
 Smith, H., Jr., 1980, *Astrophys. J.* **241**, 63.
 Smith, H., Jr., P. Hintzen, G. Holman, W. Oegerle, J. Scott, and S. Sofia, 1979b, *Astrophys. J. Lett.* **234**, L97.
 Smith, S., 1936, *Astrophys. J.* **83**, 23.
 Smyth, R. J., and R. S. Stobie, 1980, *Mon. Not. R. Astron. Soc.* **190**, 631.
 Snow, T. P., 1970, *Astron. J.* **75**, 237.
 Sofia, S., 1973, *Astrophys. J. Lett.* **179**, L35.
 Solinger, A. B., and W. H. Tucker, 1972, *Astrophys. J. Lett.* **175**, L107.

- Soltan, A., and J. P. Henry, 1983, *Astrophys. J.* **271**, 442.
- Sparke, L. S., 1983, *Astrophys. Lett.* **23**, 113.
- Spinrad, H., 1985, private communication.
- Spinrad, H., J. Stauffer, and H. Butcher, 1985, *Astrophys. J.* **296**, 784.
- Spitzer, L., Jr., 1956, *Physics of Fully Ionized Gases*. New York; Interscience.
- Spitzer, L., Jr., 1978, *Physical Processes in the Interstellar Medium*. New York; Wiley.
- Spitzer, L., Jr., and W. Baade, 1951, *Astrophys. J.* **113**, 413.
- Spitzer, L., Jr., and J. Greenstein, 1951, *Astrophys. J.* **168**, 283.
- Spitzer, L., Jr., and R. Harm, 1958, *Astrophys. J.* **127**, 544.
- Stauffer, J., 1983, *Astrophys. J.* **264**, 14.
- Stauffer, J., and H. Spinrad, 1978, *Publ. Astron. Soc. Pac.* **90**, 20.
- Stauffer, J., and H. Spinrad, 1979, *Astrophys. J. Lett.* **231**, L51.
- Stauffer, J., and H. Spinrad, 1980, *Astrophys. J.* **235**, 347.
- Stauffer, J., H. Spinrad, and W. Sargent, 1979, *Astrophys. J.* **228**, 379.
- Steiner, J. E., J. E. Grindlay, and T. Maccacaro, 1982, *Astrophys. J.* **259**, 482.
- Stewart, G. C., C. R. Canizares, A. C. Fabian, and P. E. Nulsen, 1984a, *Astrophys. J.* **278**, 536.
- Stewart, G. C., A. C. Fabian, C. Jones, and W. Forman, 1984b, *Astrophys. J.* **285**, 1.
- Stimpel, O., and J. Binney, 1979, *Mon. Not. R. Astron. Soc.* **188**, 883.
- Strom, K. M., and S. E. Strom, 1978a, *Astron. J.* **83**, 73.
- Strom, K. M., and S. E. Strom, 1978b, *Astron. J.* **83**, 1293.
- Strom, S. E., and K. M. Strom, 1978c, *Astron. J.* **83**, 732.
- Strom, S. E., and K. M. Strom, 1978d, *Astrophys. J. Lett.* **225**, L93.
- Strom, S. E., and K. M. Strom, 1979, *Astron. J.* **84**, 1091.
- Struble, M. F., and H. J. Rood, 1981, *Astrophys. J.* **251**, 471.
- Struble, M. F., and H. J. Rood, 1982, *Astron. J.* **87**, 7.
- Struble, M. F., and H. J. Rood, 1985, *Astron. J.* **89**, 1487.
- Sulentic, J. W., 1980, *Astrophys. J.* **241**, 67.
- Sullivan, W. T., G. D. Bothun, B. Bates, and R. A. Schommer, 1981, *Astron. J.* **86**, 919.
- Sullivan, W. T., and P. E. Johnson, 1978, *Astrophys. J.* **225**, 751.
- Sunyaev, R. A., 1981, *Sov. Astron. Lett.* **6**, 213.
- Sunyaev, R. A., and Y. B. Zel'dovich, 1972, *Comm. Astrophys. Space Phys.* **4**, 173.
- Sunyaev, R. A., and Y. B. Zel'dovich, 1980a, *Ann. Rev. Astron. Astrophys.* **18**, 537.
- Sunyaev, R. A., and Y. B. Zel'dovich, 1980b, *Mon. Not. R. Astron. Soc.* **190**, 413.
- Sunyaev, R. A., and Y. B. Zel'dovich, 1981, *Astrophys. Space Sci. Rev.* **1**, 1.
- Takahara, F., and S. Ikeuchi, 1977, *Prog. Theor. Phys.* **58**, 1728.
- Takahara, F., S. Ikeuchi, N. Shibasaki, and R. Hoshi, 1976, *Prog. Theor. Phys.* **56**, 1093.
- Takahara, M., and F. Takahara, 1979, *Prog. Theor. Phys.* **62**, 1253.
- Takahara, M., and F. Takahara, 1981, *Prog. Theor. Phys.* **65**, 1.
- Takeda, H., P. Nulsen, and A. Fabian, 1984, *Mon. Not. R. Astron. Soc.* **208**, 461.
- Tarenghi, M., G. Chincarini, H. Rood, and L. Thompson, 1980, *Astrophys. J.* **235**, 724.
- Tarenghi, M., and J. S. Scott, 1976, *Astrophys. J. Lett.* **207**, L9.
- Tarter, J. C., 1975, Ph. D. thesis University of California, Berkeley.
- Tarter, J. C., 1978, *Astrophys. J.* **220**, 749.
- Thomas, J. C., and D. Batchelor, 1978, *Astron. J.* **83**, 1160.
- Thompson, L. A., 1976, *Astrophys. J.* **209**, 22.
- Thompson, L. A., 1986, *Astrophys. J.* **300**, 639.
- Thompson, L. A., and S. A. Gregory, 1978, *Astrophys. J.* **220**, 809.
- Thompson, L. A., and S. A. Gregory, 1980, *Astrophys. J.* **242**, 1.
- Thuan, T. X., 1980, in *Physical Cosmology: Proceedings of the Les Houches Summer School XXXII*, edited by R. Balian, J. Audouze, and D. N. Schramm, p. 278. Amsterdam: North-Holland.
- Thuan, T. X., and J. Kormendy, 1977, *Publ. Astron. Soc. Pac.* **89**, 466.
- Thuan, T. X., and W. Romanishin, 1981, *Astrophys. J.* **248**, 439.
- Tifft, W. G., 1978, *Astrophys. J.* **222**, 54.

- Tifft, W. G., and S. A. Gregory, 1976, *Astrophys. J.* **205**, 696.
 Tifft, W. G., and M. Tarenghi, 1975, *Astrophys. J. Lett.* **198**, L7.
 Tifft, W. G., and M. Tarenghi, 1977, *Astrophys. J.* **217**, 944.
 Tinsley, B. M., and A. G. Cameron, 1974, *Astrophys. Space Sci.* **31**, 31.
 Toomre, A., and J. Toomre, 1972, *Astrophys. J.* **178**, 623.
 Tonry, J. L., 1984, *Astrophys. J.* **279**, 13.
 Tonry, J. L., 1985a, *Astrophys. J.* **291**, 45.
 Tonry, J. L., 1985b, *Astron. J.* **90**, 2431.
 Toyama, K., and S. Ikeuchi, 1980, *Prog. Theor. Phys.* **64**, 831.
 Tremaine, S., 1981, in *The Structure and Evolution of Normal Galaxies*, edited by M. Fall and D. Lynden-Bell, p.67. Cambridge University Press.
 Tremaine, S. D., and D. O. Richstone, 1977, *Astrophys. J.* **212**, 311.
 Trinchieri, G., and G. Fabbiano, 1985, *Astrophys. J.* **296**, 447.
 Trinchieri, G., G. Fabbiano, and C. R. Canizares, 1986, *Astrophys. J.* **310**, 637.
 Tucker, W. H., and R. Rosner, 1983, *Astrophys. J.* **267**, 547.
 Turner, E. L., and J. R. Gott, 1976a, *Astrophys. J.* **209**, 6.
 Turner, E. L., and J. R. Gott, 1976b, *Astrophys. J. Suppl.* **32**, 409.
 Turner, E. L., and W. L. Sargent, 1974, *Astrophys. J.* **194**, 587.
 Tytler, D., and N. V. Vidal, 1979, *Mon. Not. R. Astron. Soc.* **182**, 33p.
 Ulmer, M. P., and R. G. Cruddace, 1981, *Astrophys. J. Lett.* **246**, L99.
 Ulmer, M. P., and R. G. Cruddace, 1982, *Astrophys. J.* **258**, 434.
 Ulmer, M. P., R. G. Cruddace, and M. P. Kowalski, 1985, *Astrophys. J.* **290**, 551.
 Ulmer, M. P., R. G. Cruddace, K. Wood, J. Meekins, D. Yentis, W. D. Evans, H. Smathers, E. Byram, T. Chubb, and H. Friedman, 1980a, *Astrophys. J.* **236**, 58.
 Ulmer, M. P., and J. G. Jernigan, 1978, *Astrophys. J. Lett.* **222**, L85.
 Ulmer, M. P., R. Kinzer, R. G. Cruddace, K. Wood, W. Evans, E. Byram, T. Chubb, and H. Friedman, 1979, *Astrophys. J. Lett.* **227**, L73.
 Ulmer, M. P., M. Kowalski, R. Cruddace, M. Johnson, J. Meekins, H. Smathers, D. Yentis, K. Wood, D. McNutt, T. Chubb, E. Byram, and H. Friedman, 1981, *Astrophys. J.* **243**, 681.
 Ulmer, M. P., S. Shulman, W. D. Evans, W. N. Johnson, D. McNutt, J. Meekins, G. H. Share, D. Yentis, K. Wood, E. T. Byram, T. A. Chubb, and H. Friedman, 1980b, *Astrophys. J.* **235**, 351.
 Ulrich, M., 1978, *Astrophys. J.* **221**, 422.
 Usen, J. M., and D. T. Wilkinson, 1985, private communication.
 Valentijn, E. A., 1978, *Astron. Astrophys.* **68**, 449.
 Valentijn, E. A., 1979a, *Astron. Astrophys. Suppl.* **38**, 319.
 Valentijn, E. A., 1979b, *Astron. Astrophys.* **78**, 362.
 Valentijn, E. A., 1979c, *Astron. Astrophys.* **78**, 367.
 Valentijn, E. A., 1983, *Astron. Astrophys.* **118**, 123.
 Valentijn, E. A., and W. Bijleveld, 1983, *Astron. Astrophys.* **125**, 223.
 Valentijn, E. A., and R. Giovanelli, 1982, *Astron. Astrophys.* **114**, 208.
 Valentijn, E. A., and G. C. Perola, 1978, *Astron. Astrophys.* **63**, 29.
 Vallee, J. P., 1981, *Astrophys. Lett.* **22**, 193.
 Vallee, J. P., A. H. Bridle, and A. S. Wilson, 1981, *Astrophys. J.* **250**, 66.
 Vallee, J. P., and A. S. Wilson, 1976, *Nature* **259**, 451.
 Vallee, J. P., A. S. Wilson, and H. van der Laan, 1979, *Astron. Astrophys.* **77**, 183.
 Valtonen, M., and G. Byrd, 1979, *Astrophys. J.* **230**, 655.
 van Albada, T. S., 1982, *Mon. Not. R. Astron. Soc.* **201**, 939.
 van Breugel, W. J., 1980, *Astron. Astrophys.* **88**, 248.
 van Breugel, W. J., T. Heckman, and G. Miley, 1984, *Astrophys. J.* **276**, 79.
 van den Bergh, S., 1961a, *Zeit. Astrophys.* **53**, 219.
 van den Bergh, S., 1961b, *Astrophys. J.* **134**, 970.
 van den Bergh, S., 1976, *Astrophys. J.* **206**, 883.
 van den Bergh, S., 1977a, *Publ. Astron. Soc. Pac.* **89**, 746.

- van den Bergh, S., 1977b, *Vistas Astron.* **21**, 71.
 van den Bergh, S., 1983a, *Publ. Astron. Soc. Pac.* **95**, 275.
 van den Bergh, S., 1983b, *Astrophys. J.* **265**, 606.
 van den Bergh, S., and J. de Roux, 1978, *Astrophys. J.* **219**, 352.
 van Gorkom, J. H., and R. D. Ekers, 1983, *Astrophys. J.* **267**, 528.
 Vestrand, W. T., 1982, *Astron. J.* **87**, 1266.
 Vidal, N. V., 1975a, *Astron. Astrophys.* **42**, 145.
 Vidal, N. V., 1975b, *Publ. Astron. Soc. Pac.* **87**, 625.
 Vidal, N. V., 1980, in *IAU Symp. No. 92: Objects of High Redshift*, edited by G. O. Abell and P. J. E. Peebles, p. 69. Dordrecht: Reidel.
 Vidal, N. V., and B. A. Peterson, 1975, *Astrophys. J. Lett.* **196**, L95.
 Vidal, N. V., and D. T. Wickramasinghe, 1977, *Mon. Not. R. Astron. Soc.* **180**, 305.
 Vigroux, L., 1977, *Astron. Astrophys.* **56**, 473.
 Villumsen, J. V., 1982, *Mon. Not. R. Astron. Soc.* **199**, 493.
 Waldthausen, H., C. Haslam, R. Wielebinsky, and P. Kronberg, 1979, *Astron. Astrophys. Suppl.* **36**, 237.
 Weinberg, S., 1972, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. New York: Wiley.
 Wellington, K., G. Miley, and H. van der Laan, 1973, *Nature* **244**, 502.
 West, R. M., 1974, *Euro. South. Obs. Bull.* **10**, 25.
 West, R. M., and S. Frandsen, 1981, *Astron. Astrophys. Suppl.* **44**, 329.
 Westphal, J. A., J. Kristian, and A. R. Sandage, 1975, *Astrophys. J. Lett.* **197**, L95.
 White, R. A., 1978a, *Astrophys. J.* **226**, 591.
 White, R. A., and J. O. Burns, 1980, *Astron. J.* **85**, 117.
 White, R. A., and H. Quintana, 1985, private communication.
 White, R. A., C. L. Sarazin, and H. Quintana, 1987, Preprint.
 White, R. A., C. L. Sarazin, H. Quintana, and W. J. Jaffe, 1981a, *Astrophys. J. Lett.* **245**, L1.
 White, R. E., and Sarazin, C. L., 1987a, *Astrophys. J.* **318**, 612.
 White, R. E., and Sarazin, C. L., 1987b, *Astrophys. J.* **318**, 629.
 White, S. D. M., 1976a, *Mon. Not. R. Astron. Soc.* **174**, 19.
 White, S. D. M., 1976b, *Mon. Not. R. Astron. Soc.* **174**, 467.
 White, S. D. M., 1976c, *Mon. Not. R. Astron. Soc.* **177**, 717.
 White, S. D. M., 1977a, *Comm. Astrophys.* **7**, 95.
 White, S. D. M., 1977b, *Mon. Not. R. Astron. Soc.* **179**, 33.
 White, S. D. M., 1978b, *Mon. Not. R. Astron. Soc.* **184**, 185.
 White, S. D. M., 1979, *Mon. Not. R. Astron. Soc.* **189**, 831.
 White, S. D. M., 1982, in *Morphology and Dynamics of Galaxies*, edited by L. Martinet and M. Mayor, p.289. Geneva: Geneva Observatory.
 White, S. D. M., 1985, private communication.
 White, S. D. M., and M. J. Rees, 1978, *Mon. Not. R. Astron. Soc.* **183**, 341.
 White, S. D. M., and J. Silk, 1980, *Astrophys. J.* **241**, 864.
 White, S. D. M., J. Silk, and J. P. Henry, 1981b, *Astrophys. J. Lett.* **251**, L65.
 Wilkinson, A., and J. B. Oke, 1978, *Astrophys. J.* **220**, 376.
 Willson, M., 1970, *Mon. Not. R. Astron. Soc.* **151**, 1.
 Wilson, A. S., and J. P. Vallee, 1977, *Astron. Astrophys.* **58**, 79.
 Wirth, A., S. J. Kenyon, and D. A. Hunter, 1983, *Astrophys. J.* **269**, 102.
 Wolf, M., 1906, *Astron. Nachr.* **170**, 211.
 Wolf, R. A., and J. N. Bahcall, 1972, *Astrophys. J.* **176**, 559.
 Wolff, R. S., H. Helava, T. Kifune, and M. C. Weisskopf, 1974, *Astrophys. J. Lett.* **193**, L53.
 Wolff, R. S., H. Helava, and M. C. Weisskopf, 1975, *Astrophys. J. Lett.* **197**, L99.
 Wolff, R. S., R. J. Mitchell, P. A. Charles, and J. L. Culhane, 1976, *Astrophys. J.* **208**, 1.
 Wood, K., J. Meekins, D. Yentis, H. Smathers, D. McNutt, R. Bleach, E. Byron, T. Chubb, H. Friedman, and M. Meidav, 1984, *Astrophys. J. Suppl.* **56**, 507.
 Yahil, A., and J. P. Ostriker, 1973, *Astrophys. J.* **185**, 787.

- Yahil, A., and N. V. Vidal, 1977, *Astrophys. J.* **214**, 347.
Young, P. J., 1976, *Astron. J.* **81**, 807.
Zel'dovich, Y. B., and R. A. Sunyaev, 1969, *Astrophys. Space Sci.* **4**, 301.
Zel'dovich, Y. B., and R. A. Sunyaev, 1981, *Sov. Astron. Lett.* **6**, 285.
Zwicky, F., 1933, *Helv. Phys. Acta* **6**, 110.
Zwicky, F., 1957, *Morphological Astronomy*. Berlin: Springer.
Zwicky, F., E. Herzog, P. Wild, M. Karpowicz, and C. T. Kowal, 1961-1968, *Catalogue of Galaxies and Clusters of Galaxies*, Vol. 1-6. Pasadena: Caltech.

Table 1. *Properties of morphological classes of clusters*

Property	Regular	Intermediate	Irregular
Zwick Type	Compact	Medium–Compact	Open
Bautz–Morgan Type	I,I-II,II	II,II-III	II-III,III
Rood–Sastry Type	cD,B,L,C	L,C,F	F,I
Galactic Content E:S0:Sp	Elliptical-rich 3:4:2	Spiral-poor 1:4:2	Spiral-rich 1:2:3
Morgan Type	ii	i-ii	i
Oemler Type	cD,Spiral-poor	Spiral-poor	Spiral-rich
Symmetry	Spherical	Intermediate	Irregular
Central Concentration	High	Moderate	Low
Subclustering	Absent	Moderate	Significant
Richness	Rich $n^* \approx 10^2$	Rich–Moderate $n^* \gtrsim 10^1$	Rich–Poor $n^* \gtrsim 10^0$

Table 2. The results of fits to the X-ray surface brightness profiles of X-ray clusters using the IPC detector on the Einstein X-ray observatory

Cluster	L_x (0.5-3.0 keV) (10^{43} erg/s)	r_x (Mpc)	β	M_g ($10^{14} M_\odot$)
A56 *	19.51±9.17	0.33 [†]	1.00	0.25
A76 *	1.90±0.21	0.75 [†]	1.00	0.22
A85	41.77±0.41	0.19-0.26	0.60-0.65	2.30
A119 *	6.58±0.15	0.71-0.83	1.00	0.42
A133 *	15.24±0.64	0.24-0.28	1.00	0.17
A136 *	3.93±0.95	0.22-1.49	1.00	0.28
A154	3.57±0.20	0.05-0.30	0.40-0.70	1.00
A168	1.94±0.08	0.34-0.88	0.50-1.00	1.50
A194	0.27±0.02	0.09-0.38	0.40-1.00	0.29
A262	3.07±0.06	0.07-0.12	0.50-0.60	0.36
A348 *	8.61±1.48	0.67 [†]	1.00	0.40
A358 *	1.46±0.39	0.51-0.83	1.00	0.16
A376 *	6.28±0.81	0.33-0.47	1.00	0.18
A399	14.79±0.31	0.18-0.25	0.47-0.57	1.20
A400	1.84±0.04	0.13-0.20	0.50-0.65	0.41
A401 *	35.69±0.55	0.53-0.57	1.00	0.64
A407 *	3.22±1.18	0.36-0.48	1.00	0.13
A426	46.10±0.15	0.23-0.34	0.55-0.60	3.00
A496 *	15.33±0.27	0.22-0.23	1.00	0.14
A501 *	6.24±1.40	0.00-0.53	1.00	0.04
A514 *	1.45±0.20	0.19-0.29	1.00	0.04
A539 *	2.70±0.13	0.27-0.29	1.00	0.08
A568 *	2.96 [†]	0.41 [†]	1.00	0.13
A569 *	0.78±0.14	0.01-0.05	1.00	0.00
A576	6.55±0.10	0.09-0.14	0.47-0.52	0.47
A592	3.89±0.13	0.16-0.37	0.63-1.00	1.20
A644 *	43.59±1.42	0.36-0.42	1.00	0.46
A646 *	22.38±2.98	0.14-0.28	1.00	0.14

Table 2. (Continued)

Cluster	L_x (0.5-3.0 keV) (10^{43} erg/s)	r_x (Mpc)	β	M_g ($10^{14} M_\odot$)
A671	3.83 ± 0.26	0.09-0.34	0.55-1.00	1.40
A754 *	18.09 ± 0.45	0.68-0.77	1.00	0.63
A854 *	35.50 ± 2.75	0.23-0.68	1.00	0.44
A882 *	3.77^\dagger	0.46^\dagger	1.00	0.16
A910 *	11.04 ± 1.33	0.86-1.24	1.00	0.80
A957	2.66 ± 0.14	0.05-0.23	0.40-0.63	0.58
A1060	2.37 ± 0.02	0.09-0.11	0.60-0.75	0.40
A1142	1.03 ± 0.17	0.10-0.41	0.43-1.00	0.64
A1185	1.60 ± 0.06	0.07-0.41	0.40-0.80	0.81
A1213 *	1.03^\dagger	0.31^\dagger	1.00	0.05
A1291	2.92 ± 0.16	0.05-0.24	0.50-1.00	1.00
A1314	1.58 ± 0.07	0.21-0.50	0.60-1.00	1.00
A1367	4.49 ± 0.04	0.26-0.60	0.40-0.65	1.30
A1377	1.98 ± 0.11	0.14-0.62	0.43-1.00	1.30
A1569 *	7.41 ± 1.18	0.40-0.97	1.00	0.35
A1631 *	1.03 ± 0.21	0.11-0.23	1.00	0.03
A1656 *	23.55 ± 0.14	0.48-0.52	1.00	0.46
A1677 *	24.91 ± 2.84	0.37-0.57	1.00	0.44
A1767 *	13.59 ± 0.84	0.49-0.58	1.00	0.37
A1775	10.83 ± 0.40	0.12-0.25	0.52-0.80	1.20
A1795	51.26 ± 0.53	0.20-0.40	0.65-0.80	4.20
A1809	6.93 ± 0.47	0.13-0.49	0.50-1.00	2.00
A1890	3.01 ± 0.21	0.17-0.59	0.50-1.00	1.50
A1904 *	1.98 ± 0.48	0.81-1.42	1.00	0.34
A1913	1.73 ± 0.08	0.36-1.10	0.45-1.00	1.70
A1983	2.12 ± 0.44	0.05-0.25	0.50-1.00	0.90
A1991	8.47 ± 0.25	0.03-0.10	0.50-0.63	0.62
A2029	68.18 ± 1.76	0.08-0.30	0.63-0.83	5.30
A2040	1.91 ± 0.08	0.08-0.25	0.50-1.00	0.63
A2052 *	10.88 ± 0.40	0.20-0.24	1.00	0.12
A2063	9.65 ± 0.19	0.15-0.20	0.58-0.67	1.10
A2065 *	20.57 ± 0.47	0.56-0.62	1.00	0.53
A2079 *	2.87 ± 0.54	0.61-1.05	1.00	0.29
A2107	7.08 ± 0.26	0.13-0.19	0.60-1.00	0.89
A2124	5.61 ± 0.26	<0.04 -0.12	0.45-0.53	0.51
A2125 *	6.89^\dagger	0.56^\dagger	1.00	0.29

Table 3. (Continued)

Cluster	L_x (0.5-3.0 keV) (10^{43} erg/s)	r_x (Mpc)	β	M_g ($10^{14} M_\odot$)
A2142 *	69.08±0.90	0.49-0.53	1.00	0.80
A2151 *	1.51±0.10	0.41-0.43	1.00	0.09
A2152 *	2.02±0.20	0.20-0.29	1.00	0.06
A2162 *	0.58±0.32	0.05-0.79	1.00	0.03
A2165 *	3.39±2.26	0.00-0.32	1.00	0.02
A2197 *	0.66±0.11	0.31-0.53	1.00	0.05
A2199	20.68±0.21	0.12-0.16	0.63-0.73	1.50
A2255	13.85±0.24	0.53-0.65	0.70-0.83	3.45
A2256	27.34±0.31	0.43-0.47	0.68-0.77	3.65
A2271	2.70±0.19	0.04-0.23	0.40-0.73	0.70
A2312	6.60±0.35	0.05-0.15	0.50-0.65	0.73
A2319	41.74±0.39	0.36-0.46	0.46-0.63	3.90
A2410	2.42±0.14	0.08-0.63	0.40-0.65	1.60
A2415	9.37±0.57	<0.04-0.13	0.45-0.57	0.72
A2424 *	13.76 [†]	0.42 [†]	1.00	0.28
A2440 *	3.36±2.35	0.22-1.66	1.00	0.12
A2521 *	7.44±1.33	0.27-0.75	1.00	0.21
A2580 *	31.11±3.01	0.20-0.36	1.00	0.26
A2593	4.34±1.41	0.15-0.29	0.50-0.70	0.85
A2626	8.43±0.34	0.08-0.31	0.57-1.00	2.10
A2634	3.67±0.58	0.45-0.85	0.61-1.10	2.20
A2657	7.29±0.11	0.10-0.19	0.50-0.57	0.66
A2670	9.90±0.51	0.05-0.18	0.50-0.67	1.00
SC0107-46	1.83±0.05	0.07-0.52	0.40-1.00	1.40
SC0559-40	5.44±0.01	0.60-0.95	0.65-1.00	0.33

Note. The X-ray luminosities scale as $L_x \propto h_{50}^{-2}$, the X-ray core radii scale as $r_x \propto h_{50}^{-1}$, and the gas masses scale as $M_g \propto h_{50}^{-5/2}$. The data are from Jones and Forman (1984) and Abramopoulos and Ku (1983); the entries from Abramopoulos and Ku are marked with an *. The value of L_x is only for gas within 0.5 Mpc of the cluster center, while the gas mass is for gas within 3.0 Mpc of the center. The values of r_x and β are from fits of the X-ray surface brightness to equation 3.7. In the data from Abramopoulos and Ku, β was assumed to be unity, and was not derived from the data. The data from Abramopoulos and Ku were adjusted for the definitions of L_x and M_g from Jones and Forman, using the fits given by Abramopoulos and Ku.

[†]Very uncertain value.

Table 3. *Morphological classification of X-ray clusters, adapted from Forman and Jones (1982)*

	No X-ray Dominant Galaxy (nXD)	X-ray Dominant Galaxy (XD)
Irregular	Low $L_x \lesssim 10^{44}$ erg/s	Low $L_x \lesssim 10^{44}$ erg/s
	Cool gas $T_g = 1\text{-}4$ keV	Cool gas $T_g = 1\text{-}4$ keV
	X-ray emission around many galaxies	Central galaxy X-ray halo
	Irregular X-ray distribution	Irregular X-ray distribution
	High spiral fraction > 40%	High spiral fraction > 40%
	Low central galaxy density	Low central galaxy density
Regular	Prototype: A1367	Prototype: Virgo/M87
	High $L_x \gtrsim 10^{44}$ erg/s	High $L_x \gtrsim 3 \times 10^{44}$ erg/s
	Hot gas $T_g \gtrsim 6$ keV	Hot gas $T_g \gtrsim 6$ keV
	No cooling flow	Cooling flow onto central galaxy
	Smooth X-ray distribution	Compact, smooth X-ray distribution
	Low spiral fraction $\lesssim 20\%$	Low spiral fraction $\lesssim 20\%$
	High central galaxy density	High central galaxy density
	Prototype: Coma (A1656)	Prototype: Perseus (A426)

Table 4. *Clusters that show evidence for cooling flows*

Cooling rates scale as h_{50}^{-2} . First, Abell clusters are listed, then non-Abell clusters. The references for the X-ray or optical emission line observations are listed below.

Cluster	\dot{M} (M_\odot / yr)	X-ray	References	Optical
A85	120	25		13,4
A262	28	25		14
A400	2	25		
A426 Perseus	300	6,21,1,20		15,4
A496	200	22,20		13,7,4
A576	40	26,23		
A978	500			13
A1060	6	25		
A1126	500			13,14
A1795	400	25,20		13,4
A1983	7	25		
A1991	115	25		
A2029	250	25,20		4
A2052	120			13,14,5
A2063	26	25		
A2107	18	25		
A2142	28	16		4
A2199	110	25,20		4
A2319	75	25,26		4
A2415	15	25		
A2626	10	25		
A2657	36	25		
A2670	78	25		
SC0107-46	4	25		
AWM7	40	3		
3A0335+096	280	24,20		
MKW4	7	18,3		
SC0745-191	1000	9		9
M87/Virgo	3-20	10,11,2,17,20		12
Centaurus	22	19		8
MKW3s	100	3		
AWM4	25	3		
SC1842-63	3	25		
SC2059-247	500	27		

Table 4 (Continued).

References			
1	Canizares (1981)	2	Canizares <i>et al.</i> (1979,1982)
3	Canizares <i>et al.</i> (1983)	4	Cowie <i>et al.</i> (1983)
5	Demoulin-Ulrich <i>et al.</i> (1984)	6	Fabian <i>et al.</i> (1981a)
7	Fabian <i>et al.</i> (1981b)	8	Fabian <i>et al.</i> (1982a)
9	Fabian <i>et al.</i> (1985)	10	Fabricant <i>et al.</i> (1980)
11	Fabricant and Gorenstein (1983)	12	Ford and Butcher (1979)
13	Heckman (1981)	14	Hu <i>et al.</i> (1985)
15	Kent and Sargent (1979)	16	Lea <i>et al.</i> (1981)
17	Lea <i>et al.</i> (1982)	18	Malumuth and Kriss (1986)
19	Matilsky <i>et al.</i> (1985)	20	Mushotzky (1984)
21	Mushotzky <i>et al.</i> (1981)	22	Nulsen <i>et al.</i> (1982)
23	Rothenflug <i>et al.</i> (1984)	24	Schwartz <i>et al.</i> (1980b)
25	Stewart <i>et al.</i> (1984b)	26	White and Silk (1980)
27	White <i>et al.</i> (1981a)		

Figure Captions

Fig. 1. Optical photographs of clusters of galaxies. (a) An optical photograph of the Virgo cluster of galaxies, an irregular cluster that is the nearest cluster to our galaxy. The galaxy M87, on which the X-ray emission is centered, is marked, as are the two bright galaxies M84 and M86. Photograph from the Palomar Observatory Sky Survey (Minkowski and Abell, 1963). (b) The Coma cluster of galaxies (Abell 1656), showing the two dominant D galaxies. Coma is one of the nearest rich, regular clusters. Photograph copyright 1973, AURA, Inc., the National Optical Astronomy Observatories, Kitt Peak. (c) The Perseus clusters of galaxies (Abell 476), showing the line of bright galaxies. NGC1275 is the brightest galaxy, on the east (left) end of the chain. NGC1265 is a head-tail radio galaxy. Photograph from Strom and Strom (1978c). (d) The irregular cluster Abell 1367. Photograph from Strom and Strom (1978c).

Fig. 2. The luminosity function of galaxies in clusters. $N(\leq M_V)$ is the number of galaxies brighter than absolute magnitude M_V . The circles are the composite observed luminosity function derived by Schechter (1976). The solid circles exclude cD galaxies, while the open circles show the changes when they are included. The solid, dashed, and dash-dotted curves are the fitting functions of Schechter, Abell, and Zwicky, as discussed in the text.

Fig. 3. (a) The Rood–Sastry (1971) cluster classification scheme. (b) The revised Rood–Sastry classes from Struble and Rood (1982).

Fig. 4. The dots give the projected galaxy number density observed in 12 regular clusters, from Bahcall (1975). The observed number densities are normalized to the central surface number density σ_o and given as a function of the projected radius b divided by the core radius r_c . These parameters were determined by fitting equation 2.11 (the solid curve) to the observed distributions.

Fig. 5. The projected galaxy distributions in White's (1976c) N-body calculations of the evolution of a collapsing cluster. Each symbol represents a galaxy. In each figure, the bar represents a fixed length scale, one-half of the gravitational radius R_G (equation 2.21). The times t are given in units of the initial collapse time of the cluster. (a) The initial configuration ($t = 0$). (b) An irregular distribution with subclustering at $t = 0.19$. (c) A bimodal distribution in the cluster at $t = 0.97$. (d) The final relaxed configuration at $t = 2.66$.

Fig. 6. The surface photometry of cD galaxies. The surface brightness in magnitudes per square second of arc is plotted against the one-fourth power of the radius. The dots are the observed points, and the straight lines are de Vaucouleurs fits to the inner points. NGC1278 (Oemler, 1976) is a normal giant elliptical galaxy, showing no extended halo. The middle figure shows the cD in the cluster A1413 (Oemler, 1976), with a more extended central de Vaucouleurs profile and a very extended halo. The right panel is the cD in the poor

cluster AWM4 (Thuan and Romanishin, 1981), which has an even more extensive de Vaucouleurs profile, but no apparent halo.

Fig. 7. A radio map of the Perseus cluster of galaxies from Gisler and Miley (1979). Contours of constant radio surface brightness at 610 MHz are shown superimposed on the optical image of the cluster. Note the very strong source associated with the galaxy NGC1275 (the highest contours associated with this source have been removed), and the two head-tail radio sources associated with NGC1265 and IC310. The rings are diffraction features due to NGC1275 and are not real.

Fig. 8. A low resolution radio map at a frequency of 5 GHz of the head-tail radio source associated with the galaxy NGC1265 in the Perseus cluster, from Wellington *et al.* (1973). Contours of constant radio surface brightness are shown superimposed on the optical image of the galaxy.

Fig. 9. A high resolution Very Large Array radio map at a frequency of 4.9 GHz of the head-tail radio galaxy NGC1265, from O'Dea and Owen, 1986. Note the twin radio jets leading from the nucleus of the galaxy out to the radio tails.

Fig. 10. A Very Large Array radio image at a frequency of 1420 MHz of the wide-angle-tail (WAT) radio galaxy 1919+479 associated with a cD galaxy in a Zwicky cluster, from Burns *et al.* (1986).

Fig. 11. The correlation between the gas temperatures derived from X-ray spectra with HEAO-1 A-2 and the cluster X-ray luminosities, from Mushotzky (1984).

Fig. 12. HEAO-1 A-2 low resolution X-ray spectra of clusters, showing the Fe K line at about 7 keV. The plots give the number flux of X-ray photons per $\text{cm}^2\text{-sec-keV}$ versus photon energy in keV, for the (a) Coma (Henriksen and Mushotzky, 1986) and (b) Perseus (Henriksen, 1985) clusters.

Fig. 13. The iron abundance of the gas in clusters as derived from their X-ray spectra, plotted versus their X-ray luminosity, from Mushotzky (1984). The iron abundance is relative to solar, and the X-ray luminosity is in erg/s.

Fig. 14. The moderate resolution X-ray spectrum of the M87/Virgo cluster taken with the Solid State Spectrometer on the *Einstein* satellite (Lea *et al.*, 1982). The spectral lines of Mg, Si, S, and the lower energy (L-shell) lines of Fe are marked.

Fig. 15. The very high resolution X-ray spectrum of the M87/Virgo cluster, showing the O VIII K line, from Canizares *et al.* (1979) using the Focal Plane Crystal Spectrometer on the *Einstein* satellite.

Fig. 16. The X-ray surface brightness of several clusters, as determined by the IPC on the *Einstein* satellite by Jones and Forman (1984). The surface brightness is normalized to its central value and is given as a function of the angular distance from the cluster center. The solid curves give the observed surface brightness, and the dots are the best fit using equation (4.7). The bottom two panels show the A1795 cluster with the

inner eight data points either included in the fit (left) or removed (right). The improvement in the fit in the outer points when the inner regions are removed suggests excess emission in the center, due to a cooling flow.

Fig. 17. The X-ray morphology of several clusters of galaxies from Jones and Forman (1984). Contours of constant X-ray surface brightness are shown superimposed on optical images of the clusters. (Top left), the prototypical irregular nXD cluster A1367. (Top right), the irregular XD cluster A262. (Bottom left), the regular nXD cluster A2256. (Bottom right), the regular XD cluster A85, showing the X-ray emission centered on the cD galaxy.

Fig. 18. The X-ray surface brightness in four double clusters, from Forman *et al.* (1981). Contours of constant X-ray surface brightness are shown superimposed on optical images of the clusters.

Fig. 19. The X-ray surface brightness of the Coma cluster of galaxies from the IPC on *Einstein*, kindly provided by Christine Jones and Bill Forman. Contours of constant X-ray surface brightness are shown superimposed on the optical image of the cluster.

Fig. 20. An optical photograph of the spectacular galaxy NGC1275 in the Perseus cluster from Lynds (1970), copyright 1970 AURA, Inc., the National Optical Astronomy Observatories, Kitt Peak. The photograph was taken with a filter sensitive to the H α emission line, and shows the prominent optical emission line filaments around this galaxy.

Fig. 21. The X-ray surface brightness of the Perseus cluster of galaxies, observed by Branduardi-Raymont *et al.* (1981) with the IPC on the *Einstein* satellite. Contours of constant X-ray surface brightness are shown superimposed on the optical image of the cluster. The center of the galaxy distribution in the cluster is shown as a dashed circle, while the centroid of the extended X-ray emission is the \times . The peak in the X-ray surface brightness is centered on the galaxy NGC1275.

Fig. 22. A higher resolution X-ray image of the Perseus cluster emission around the galaxy NGC1275, from Branduardi-Raymont *et al.* (1981) and Fabian *et al.* (1984), using the HRI on the *Einstein* satellite. Contours of constant X-ray surface brightness are shown superimposed on the optical image of the galaxy NGC1275.

Fig. 23. An optical photograph of the elliptical galaxy M87 in the Virgo cluster, showing the jet in the interior of the galaxy (Arp and Lorre, 1976).

Fig. 24. The X-ray surface brightness of the M87/Virgo cluster of galaxies as observed by Fabricant and Gorenstein (1983) with the IPC on the *Einstein* satellite. The lines are contours of constant X-ray surface brightness. The X-ray emission is centered on M87.

Fig. 25. A higher resolution X-ray image of the M87/Virgo cluster emission centered on the galaxy M87,

from Schreier *et al.* (1982). The lines are contours of constant X-ray surface brightness. (Upper), a lower spatial resolution version, showing the peaking of the emission on the center of M87, and its asymmetrical structure. (Lower), a blowup of the center of the galaxy, showing X-ray emission along the optical jet.

Fig. 26. The X-ray emission from the Virgo cluster galaxies M86 (east or left) and M84 (west or right) from Forman and Jones (1982). Contours of constant X-ray surface brightness are shown superimposed on an optical image of the galaxies. Note the plume of X-ray emission extending to the north of M86, which may indicate that this galaxy is currently being stripped of its gas.

Fig. 27. The X-ray surface brightness of the A1367 cluster of galaxies, observed by Bechtold *et al.* (1983) with the *Einstein* satellite. Contours of constant X-ray surface brightness are shown superimposed on the optical image of the cluster. (a) The lower resolution IPC image, showing the irregular cluster emission. (b) The higher resolution HRI image, showing many discrete sources within the cluster, many of which are associated with individual cluster galaxies.

Fig. 28. The correlation between the X-ray luminosity of clusters observed with HEAO-1 A-2 and their line-of-sight velocity dispersion, from Quintana and Melnick (1982).

Fig. 29. The relationship between the X-ray luminosities of clusters observed with HEAO-1 A-2 and the central galaxy density N_o of Bahcall (1977b), from Mushotzky (1984).

Fig. 30. The correlation between the spiral fraction in clusters and their X-ray luminosity, from Bahcall (1977c). The line shows the relation from equation (4.9).

Fig. 31. The relationship between gas temperatures derived from X-ray spectral observations with HEAO-1 A-2 and the central velocity dispersion ($\Delta V_C = \sigma_r$ at the cluster center), from Mushotzky (1984). The lines show the predictions of polytropic models with various indices (Section 5.5.2). The dots are the data for non-cD clusters; the open circles are cD clusters.

Fig. 32. The correlation between the gas temperatures derived from X-ray spectral measurements with HEAO-1 A-2 and the central galaxy density of Bahcall (1977b), from Mushotzky (1984).

Fig. 33. The X-ray emission from the AWM4 poor cluster of galaxies, from Kriss *et al.* (1983) with the IPC on the *Einstein* satellite. Contours of constant X-ray surface brightness are shown superimposed on the optical image of the cluster. The emission is centered on the cD galaxy.

Fig. 34. The predicted X-ray spectra of intracluster gas at various gas temperatures (shown at the upper right of each panel), from Sarazin and Bahcall (1977). The calculations assume the gas is isothermal, and the intensities are normalized to a sphere of gas with a proton number density of 0.001 cm^{-3} and a radius of 0.5 Mpc. E is the photon energy. The strongest line features are labeled; the lower curves, where present, show the bound-free emission.

Fig. 35. The equivalent width (in eV) of the Fe K line at 7 keV as a function of gas temperature, from Bahcall and Sarazin (1978). For gas temperatures $\gtrsim 2 \times 10^7$ K, this is the strongest X-ray line feature.

Fig. 36. The conduction time as a function of position in an adiabatic model for the intracluster gas. The conduction time is relative to its central value, and the radius is in units of the cluster core radius r_c . The solid (dashed) curve indicates the portion of the cluster where the gas is cooled (heated) by conduction.

Fig. 37. The gas temperature and density in a cluster with a steady-state cooling flow, from the models in Cowie and Binney (1977). The gas density is normalized to its value at the cluster core radius, and the gas temperature is normalized to $\mu m_p \sigma_r^2 / k$, where σ_r is the cluster velocity dispersion and μ is the mean atomic mass. The radius from the cluster center is in units of the cluster core radius.

The Evolution of X-ray Clusters of Galaxies

Piero Rosati ^{1,3}, Stefano Borgani ² and Colin Norman ^{3,4}

¹ ESO – European Southern Observatory, D-85748 Garching bei München, Germany;
 e-mail: prosati@eso.org

² Dipartimento di Astronomia dell'Università di Trieste, via Tiepolo 11, I-34131 Trieste, Italy;
 e-mail: borgani@ts.astro.it

³ Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, MD
 21218, USA; e-mail: norman@stsci.edu

⁴ Space Telescope Science Institute, Baltimore, MD 21218, USA

KEYWORDS: cosmology, intracluster medium, temperature, masses, dark matter

ABSTRACT: Considerable progress has been made over the last decade in the study of the evolutionary trends of the population of galaxy clusters in the Universe. In this review we focus on observations in the X-ray band. X-ray surveys with the *ROSAT* satellite, supplemented by follow-up studies with *ASCA* and *Beppe-SAX*, have allowed an assessment of the evolution of the space density of clusters out to $z \approx 1$, and the evolution of the physical properties of the intra-cluster medium out to $z \approx 0.5$. With the advent of *Chandra* and *Newton-XMM*, and their unprecedented sensitivity and angular resolution, these studies have been extended beyond redshift unity and have revealed the complexity of the thermodynamical structure of clusters. The properties of the intra-cluster gas are significantly affected by non-gravitational processes including star formation and Active Galactic Nucleus (AGN) activity. Convincing evidence has emerged for modest evolution of both the bulk of the X-ray cluster population and their thermodynamical properties since redshift unity. Such an observational scenario is consistent with hierarchical models of structure formation in a flat low density universe with $\Omega_m \simeq 0.3$ and $\sigma_8 \simeq 0.7 - 0.8$ for the normalization of the power spectrum. Basic methodologies for construction of X-ray-selected cluster samples are reviewed and implications of cluster evolution for cosmological models are discussed.

CONTENTS

INTRODUCTION	2
PHYSICAL PROPERTIES OF GALAXY CLUSTERS	3
<i>X-ray properties of clusters</i>	5
<i>Cooling in the Intra Cluster Medium</i>	8
OBSERVATIONAL FRAMEWORK	9
<i>Optically-based Cluster Surveys</i>	9
<i>X-ray Cluster Surveys</i>	11
<i>Strategies and Selection Functions for X-ray Surveys</i>	14
<i>Other methods</i>	17
THE SPACE DENSITY OF X-RAY CLUSTERS	17
<i>Local Cluster Number Density</i>	17
<i>The Cluster Abundance at Higher Redshifts and Its Evolution</i>	18
<i>Distant X-ray Clusters: the Latest View from Chandra</i>	20

COSMOLOGY WITH X-RAY CLUSTERS	22
The cosmological mass function	22
Deriving Ω_m from cluster evolution	24
OUTLOOK AND FUTURE WORK	27

1 INTRODUCTION

Galaxy clusters arise from the gravitational collapse of rare high peaks of primordial density perturbations in the hierarchical clustering scenario for the formation of cosmic structures (e.g. Peebles 1993, Coles & Lucchin 1995, Peacock 1999). They probe the high-density tail of the cosmic density field and their number density is highly sensitive to specific cosmological scenarios (e.g. Press & Schechter 1974, Kofman et al. 1993, Bahcall & Cen 1993, White et al. 1993a). The space density of clusters in the local universe has been used to measure the amplitude of density perturbations on ~ 10 Mpc scales. Its evolution, which is driven by the growth rate of density fluctuations, essentially depends on the value of the matter density parameter Ω_m ¹ (e.g. Oukbir & Blanchard 1992, Eke et al. 1998, Bahcall et al. 1999). Figure 1 shows how structure formation proceeds and the cluster population evolves in two cosmological models, characterized by different values of Ω_m . High and low density universes show largely different evolutionary patterns, which demonstrate how the space density of distant clusters can be used as a powerful cosmological diagnostic. What cosmological models actually predict is the number density of clusters of a given mass at varying redshifts. The cluster mass, however, is never a directly observable quantity, although several methods exist to estimate it from observations.

Determining the evolution of the space density of clusters requires counting the number of clusters of a given mass per unit volume at different redshifts. Therefore, three essential tools are required for its application as a cosmological test: *i*) an efficient method to find clusters over a wide redshift range; *ii*) an observable estimator of the cluster mass and *iii*) a method to compute the selection function or equivalently the survey volume within which clusters are found.

Clusters form *via* the collapse of cosmic matter over a region of several megaparsecs. Cosmic baryons, which represent approximately 10–15% of the mass content of the Universe, follow the dynamically dominant dark matter during the collapse. As a result of adiabatic compression and of shocks generated by supersonic motions during shell crossing and virialization, a thin hot gas permeating the cluster gravitational potential well is formed. For a typical cluster mass of 10^{14} – $10^{15} M_\odot$ this gas reaches temperatures of several 10^7 K, becomes fully ionized and, therefore, emits via thermal bremsstrahlung in the X-ray band.

Observations of clusters in the X-ray band provide an efficient and physically motivated method of identification, which fulfills the three requirements above. The X-ray luminosity, which uniquely specifies the cluster selection, is also a good probe of the depth of the cluster gravitational potential. For these reasons most of the cosmological studies based on clusters have used X-ray selected samples. X-ray studies of galaxy clusters provide: (1) an efficient way of mapping the

¹The matter-density parameter is defined as $\Omega_m = \bar{\rho}/\rho_c$, where $\bar{\rho}$ is the cosmic mean matter density; $\rho_c = 1.88 \cdot 10^{-29} h^2 \text{ g cm}^{-3}$ is the critical density; h and h_{50} denote the Hubble constant H_0 respectively in units of 100 and 50 $\text{km s}^{-1} \text{ Mpc}^{-1}$. Ω_Λ is referred to as the contribution to the total mass-energy density of the Universe associated with the cosmological constant Λ .

overall structure and evolution of the Universe and (2) an invaluable means of understanding their internal structure and the overall history of cosmic baryons.

X-ray cluster studies made substantial progress at the beginning of the 90s with the advent of new X-ray missions. Firstly, the all-sky survey and the deep pointed observations conducted by the *ROSAT* satellite have been a goldmine for the discovery of hundreds of new clusters in the nearby and distant Universe. Follow-up studies with *ASCA* and *Beppo-SAX* satellites revealed hints of the complex physics governing the intra-cluster gas. In addition to gas heating associated with gravitational processes, star formation processes and energy feedback from supernovae and galactic nuclear activity are now understood to play an important role in determining the thermal history of the intra-cluster medium (ICM), its X-ray properties and its chemical composition. Studies utilizing the current new generation of X-ray satellites, *Chandra* and *Newton-XMM*, are radically changing our X-ray view of clusters. The large collecting area of *Newton-XMM*, combined with the superb angular resolution of *Chandra*, have started to unveil the interplay between the complex physics of the hot ICM and detailed processes of star formation associated with cool baryons.

The aim of this article is to provide an up-to-date review on the methodology used to construct X-ray selected cluster samples and to investigate their evolutionary properties. We emphasize the evolution of the space density of clusters and their physical parameters. Additional reviews on galaxy clusters include: Forman & Jones (1982) and Sarazin (1988) for historical reviews on X-ray properties of galaxy clusters; Bahcall (1988) and Borgani & Guzzo (2001) for large-scale structure studies of galaxy clusters; Fabian (1994) for the physics of cooling flows in clusters; Mulchaey (2000) for the X-ray properties of galaxy groups; Birkkinshaw (1999) and Carlstrom et al. (2001) for cluster studies with the Sunyaev-Zeldovich effect; Mellier (1999) for studies of the mass distribution of clusters via gravitational lensing and van Dokkum & Franx (2001) for the study of galaxy populations in clusters.

2 PHYSICAL PROPERTIES OF GALAXY CLUSTERS

Clusters of galaxies were first identified as large concentrations in the projected galaxy distribution (Abell 1958, Zwicky et al. 1966, Abell et al. 1989), containing hundreds to thousands galaxies, over a region of the order of ~ 1 Mpc. The first observations showed that such structures are associated with deep gravitational potential wells, containing galaxies with a typical velocity dispersion along the line-of-sight of $\sigma_v \sim 10^3$ km s $^{-1}$. The crossing time for a cluster of size R can be defined as

$$t_{\text{cr}} = \frac{R}{\sigma_v} \simeq 1 \left(\frac{R}{1 \text{Mpc}} \right) \left(\frac{\sigma_v}{10^3 \text{ km s}^{-1}} \right)^{-1} \text{Gyr}. \quad (1)$$

Therefore, in a Hubble time, $t_H \simeq 10 h^{-1}$ Gyr, such a system has enough time in its internal region, $\lesssim 1 h^{-1}$ Mpc, to dynamically relax – a condition that can not be attained in the surrounding, ~ 10 Mpc, environment. Assuming virial equilibrium, the typical cluster mass is

$$M \simeq \frac{R \sigma_v^2}{G} \simeq \left(\frac{R}{1 h^{-1} \text{Mpc}} \right) \left(\frac{\sigma_v}{10^3 \text{ km s}^{-1}} \right)^2 10^{15} h^{-1} M_\odot. \quad (2)$$

Smith (1936) first noticed in his study of the Virgo cluster that the mass implied by cluster galaxy motions was largely exceeding that associated with the

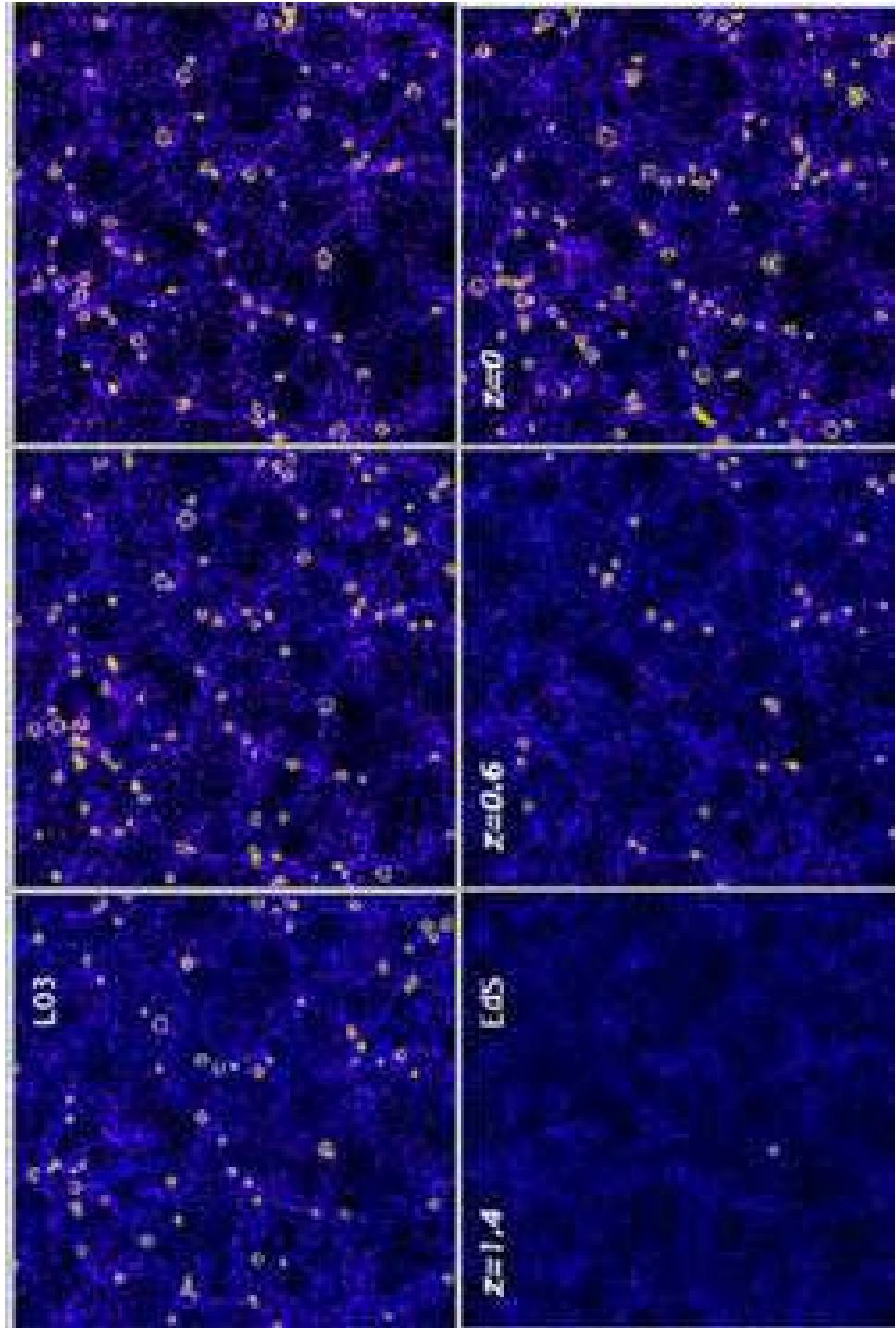


Figure 1: The evolution of the cluster population from N-body simulations in two different cosmologies (from Borgani & Guzzo 2001). Left panels describe a flat, low-density model with $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ (L03); right panels are for an Einstein-de-Sitter model (EdS) with $\Omega_m = 1$. Superimposed on the dark matter distribution, the yellow circles mark the positions of galaxy clusters with virial temperature $T > 3$ keV, the size of the circles is proportional to temperature. Model parameters have been chosen to yield a comparable space density of nearby clusters. Each snapshot is $250h^{-1}$ Mpc across and $75h^{-1}$ Mpc thick (comoving with the cosmic expansion).

optical light component. This was confirmed by Zwicky (1937), and was the first evidence of the presence of dark matter.

2.1 X-ray properties of clusters

Observations of galaxy clusters in the X-ray band have revealed a substantial fraction, $\sim 15\%$, of the cluster mass to be in the form of hot diffuse gas, permeating its potential well. If this gas shares the same dynamics as member galaxies, then it is expected to have a typical temperature

$$k_B T \simeq \mu m_p \sigma_v^2 \simeq 6 \left(\frac{\sigma_v}{10^3 \text{ km s}^{-1}} \right)^2 \text{ keV}, \quad (3)$$

where m_p is the proton mass and μ is the mean molecular weight ($\mu = 0.6$ for a primordial composition with a 76% fraction contributed by hydrogen). Observational data for nearby clusters (e.g. Wu et al. 1999) and for distant clusters (see Figure 2) actually follow this relation, although with some scatter and with a few outliers. This correlation indicates that the idealized picture of clusters as relaxed structures in which both gas and galaxies feel the same dynamics is a reasonable representation. There are some exceptions that reveal the presence of a more complex dynamics.

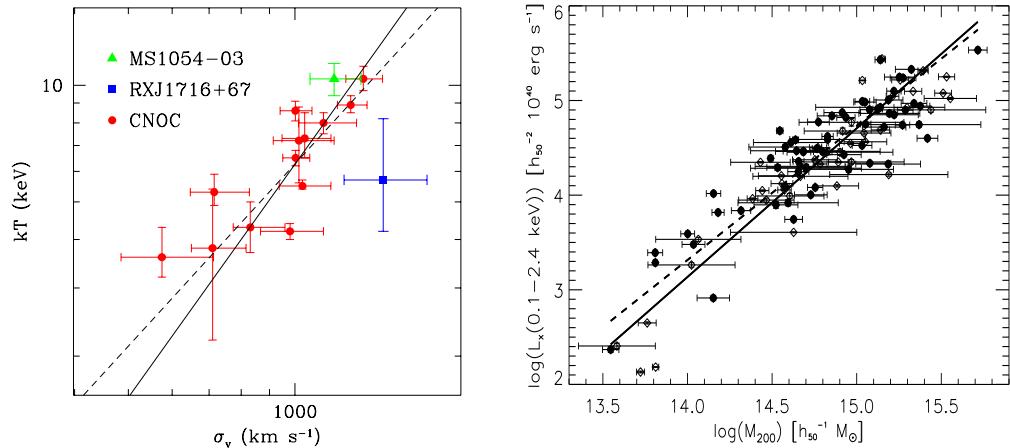


Figure 2: *Left* The relation between galaxy velocity dispersion, σ_v , and ICM temperature, T , for distant ($z > 0.15$) galaxy clusters. Velocity dispersions are taken from Carlberg et al. (1997a) for CNOC clusters and from Girardi & Mezzetti (2001) for MS1054-03 and RXJ1716+67. Temperatures are taken from Lewis et al. (1999) for CNOC clusters, from Jeltema et al. (2001) for MS1054-03 and from Gioia et al. (1999) for RXJ1716+67. The solid line shows the relation $k_B T = \mu m_p \sigma_v^2$, and the dashed line is the best-fit to the low- z $T-\sigma_v$ relation from Wu et al. (1999). *Right* The low- z relation between X-ray luminosity and the mass contained within the radius encompassing an average density $200\rho_c$ (from Reiprich & Böhringer 2002). The two lines are the best log-log linear fit to two different data sets indicated with filled and open circles.

At the high energies implied by Equation 3, the ICM behaves as a fully ionized

plasma, whose emissivity is dominated by thermal bremsstrahlung. The emissivity for this process at frequency ν scales as $\epsilon_\nu \propto n_e n_i g(\nu, T) T^{-1/2} \exp(-h\nu/k_B T)$, where n_e and n_i are the number density of electrons and ions, respectively, and $g(\nu, T) \propto \ln(k_B T/h\nu)$ is the Gaunt factor. Whereas the pure bremsstrahlung emissivity is a good approximation for $T \gtrsim 3$ keV clusters, a further contribution from metal emission lines should be taken into account when considering cooler systems (e.g. Raymond & Smith 1977). By integrating the above equation over the energy range of the X-ray emission and over the gas distribution, one obtains X-ray luminosities $L_X \sim 10^{43}\text{--}10^{45}$ erg s $^{-1}$. These powerful luminosities allow clusters to be identified as extended sources out to large cosmological distances.

Assuming spherical symmetry, the condition of hydrostatic equilibrium connects the local gas pressure p to its density ρ_{gas} according to

$$\frac{dp}{dR} = -\frac{GM(< R)\rho_{\text{gas}}(R)}{R^2}. \quad (4)$$

By inserting the equation of state for a perfect gas, $p = \rho_{\text{gas}}k_B T/\mu m_p$ into Equation (4), one can express, $M(< R)$, the total gravitating mass within R as

$$M(< R) = -\frac{k_B T R}{G \mu m_p} \left(\frac{d \log \rho_{\text{gas}}}{d \log R} + \frac{d \log T}{d \log R} \right). \quad (5)$$

If R is the virial radius, then at redshift z we have $M \propto R^3 \bar{\rho}_0 (1+z)^3 \Delta_{\text{vir}}(z)$, where $\bar{\rho}_0$ is the mean cosmic density at present time and $\Delta_{\text{vir}}(z)$ is the mean overdensity within a virialized region (see also Equation 13, below). For an Einstein-de-Sitter cosmology, Δ_{vir} is constant and therefore, for an isothermal gas distribution, Equation (5) implies $T \propto M^{2/3}(1+z)$.

Such relations show how quantities, such as ρ_{gas} and T , which can be measured from X-ray observations, are directly related to the cluster mass. Thus, in addition to providing an efficient method to detect clusters, X-ray studies of the ICM allow one to measure the total gravitating cluster mass, which is the quantity predicted by theoretical models for cosmic structure formation.

A popular description of the gas density profile is the β -model,

$$\rho_g(r) = \rho_{g,0} \left[1 + \left(\frac{r}{r_c} \right)^2 \right]^{-3\beta/2}, \quad (6)$$

which was introduced by Cavaliere & Fusco-Femiano (1976; see also Sarazin 1988, and references therein) to describe an isothermal gas in hydrostatic equilibrium within the potential well associated with a King dark-matter density profile. The parameter β is the ratio between kinetic dark-matter energy and thermal gas energy (see Equation 3). This model is a useful guideline for interpreting cluster emissivity, although over limited dynamical ranges. Now, with the *Chandra* and *Newton-XMM* satellites, the X-ray emissivity can be mapped with high angular resolution and over larger scales. These new data have shown that Equation 6 with a unique β value cannot always describe the surface brightness profile of clusters (e.g. Allen et al. 2001).

Kaiser (1986) described the thermodynamics of the ICM by assuming it to be entirely determined by gravitational processes, such as adiabatic compression during the collapse and shocks due to supersonic accretion of the surrounding

gas. As long as there are no preferred scales both in the cosmological framework (i.e. $\Omega_m = 1$ and power-law shape for the power spectrum at the cluster scales), and in the physics (i.e. only gravity acting on the gas and pure bremsstrahlung emission), then clusters of different masses are just a scaled version of each other. Because bremsstrahlung emissivity predicts $L_X \propto M \rho_{\text{gas}} T^{1/2}$, $L_X \propto T_X^2 (1+z)^{3/2}$ or, equivalently $L_X \propto M^{4/3} (1+z)^{7/2}$. Furthermore, if we define the gas entropy as $S = T/n^{2/3}$, where n is the gas density assumed fully ionized, we obtain $S \propto T(1+z)^{-2}$.

It was soon recognized that X-ray clusters do not follow these scaling relations. As we discuss in Section 5, below, the observed luminosity–temperature relation for clusters is $L_X \propto T^3$ for $T \gtrsim 2$ keV, and possibly even steeper for $T \lesssim 1$ keV groups. This result is consistent with the finding that $L_X \propto M^\alpha$ with $\alpha \simeq 1.8 \pm 0.1$ for the observed mass–luminosity relation (e.g. Reiprich & Böhringer 2002; see right panel of Figure 2). Furthermore, the low-temperature systems are observed to have shallower central gas-density profiles than the hotter systems, which turns into an excess of entropy in low- T systems with respect to the $S \propto T$ predicted scaling (e.g. Ponman et al. 1999, Lloyd-Davies et al. 2000).

A possible interpretation for the breaking of the scaling relations assumes that the gas has been heated at some earlier epoch by feedback from a non-gravitational astrophysical source (Evrard & Henry 91). This heating would increase the entropy of the ICM, place it on a higher adiabat, prevent it from reaching a high central density during the cluster gravitational collapse and, therefore, decrease the X-ray luminosity (e.g. Balogh et al. 1999, Tozzi & Norman 2001, and references therein). For a fixed amount of extra energy per gas particle, this effect is more prominent for poorer clusters, i.e. for those objects whose virial temperature is comparable with the extra-heating temperature. As a result, the self-similar behavior of the ICM is expected to be preserved in hot systems, whereas it is broken for colder systems. Both semi-analytical works (e.g. Cavaliere et al. 1998, Balogh et al. 1999, Wu et al. 2000; Tozzi et al. 2001) and numerical simulations (e.g. Navarro et al. 1995, Brighenti & Mathews 2001, Bialek et al. 2001, Borgani et al. 2001a) converge to indicate that ~ 1 keV per gas particle of extra energy is required. A visual illustration of the effect of pre-heating is reported in Figure 3, which shows the entropy map for a high-resolution simulation of a system with mass comparable to that of the Virgo cluster, for different heating schemes (Borgani et al. 2001b). The effect of extra energy injection is to decrease the gas density in central cluster regions and to erase the small gas clumps associated with accreting groups.

The gas-temperature distributions in the outer regions of clusters are not affected by gas cooling. These temperature distributions have been studied with the *ASCA* and *Beppe-SAX* satellites. General agreement about the shape of the temperature profiles has still to be reached (e.g. Markevitch et al. 1998, White 2000, Irwin & Bregman 2000). De Grandi & Molendi (2002) analyzed a set of 21 clusters with *Beppe-SAX* data and found the gas to be isothermal out to $\sim 0.2R_{\text{vir}}$, with a significant temperature decline at larger radii. Such results are not consistent with the temperature profiles obtained from cluster hydrodynamical simulations (e.g. Evrard et al. 1996), thus indicating that some physical process is still lacking in current numerical descriptions of the ICM. Deep observations with *Newton-XMM* and *Chandra* will allow the determination of temperature profiles over the whole cluster virialized region.

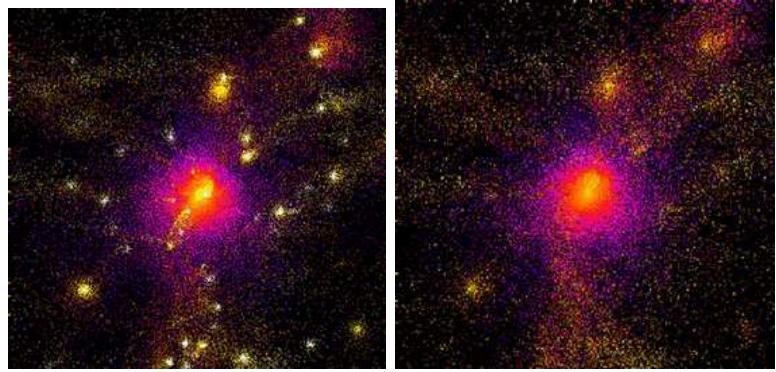


Figure 3: Map of gas entropy from hydrodynamical simulations of a galaxy cluster (from Borgani et al. 2001a). (*Left*): gravitational heating only. (*Right*): entropy floor of $50 \text{ keV}/\text{cm}^2$ imposed at $z = 3$, corresponding to about $1 \text{ keV}/\text{part}$. Light colors correspond to low entropy particles, and dark blue corresponds to high-entropy gas.

X-ray spectroscopy is a powerful means for analyzing the metal content of the ICM. Measurements of over 100 nearby clusters have yielded a mean metallicity $Z \sim 1/3Z_{\odot}$, largely independent of the cluster temperature (e.g. Renzini 1997, and references therein). The spatial distribution of metals has recently been studied in detail with *ASCA* and *Beppo-SAX* data (e.g. White 2000, De Grandi & Molendi 2001). This field will receive a major boost over the next few years particularly with *Newton-XMM*, which, with a ten-fold improvement in collecting area and much better angular resolution, will be able to map the distribution of different metals in the ICM, such as Fe, S, Si, O.

2.2 Cooling in the Intra Cluster Medium

In order to characterize the role of cooling in the ICM, it is useful to define the cooling time-scale, which for an emission process characterized by a cooling function $\Lambda_c(T)$, is defined as $t_{cool} = k_B T / (n\Lambda(T))$, n being the number density of gas particles. For a pure bremsstrahlung emission: $t_{cool} \simeq 8.5 \times 10^{10} \text{ yr} (n/10^{-3} \text{ cm}^{-3})^{-1} (T/10^8 \text{ K})^{1/2}$. (e.g. Sarazin 1988). Therefore, the cooling time in central cluster regions can be shorter than the age of the Universe. A substantial fraction of gas undergoes cooling in these regions, and consequently drops out of the hot diffuse, X-ray emitting phase. Studies with the *ROSAT* and *ASCA* satellites indicate that the decrease of the ICM temperature in central regions has been recognized as a widespread feature among fairly relaxed clusters (see Fabian 1994, and references therein). The canonical picture of cooling flows predicted that, as the high-density gas in the cluster core cools down, the lack of pressure support causes external gas to flow in, thus creating a superpositions of many gas phases, each one characterized by a different temperature. Our understanding of the ICM cooling structure is now undergoing a revolution thanks to the much improved spatial and spectral resolution provided by *Newton-XMM*. Recent observations have shown the absence of metal lines associated with gas at temperature $\lesssim 3 \text{ keV}$ (e.g. Peterson et al. 2001, Tamura et al. 2001), in stark contrast with the standard cooling flow prediction for the presence of low-temperature gas (e.g. Böhringer et al. 2002a, Fabian et al. 2001a).

Radiative cooling has been also suggested as an alternative to extra heating to explain the lack of ICM self-similarity (e.g. Bryan 2000, Voit & Bryan 2002). When the recently shocked gas residing in external cluster regions leaves the hot phase and flows in, it increases the central entropy level of the remaining gas. The decreased amount of hot gas in the central regions causes a suppression of the X-ray emission (Pearce et al. 2000, Muanwong et al. 2001). This solution has a number of problems. Cooling in itself is a runaway process, leading to a quite large fraction of gas leaving the hot diffuse phase inside clusters. Analytical arguments and numerical simulations have shown that this fraction can be as large as $\sim 50\%$, whereas observational data indicates that only $\lesssim 10\%$ of the cluster baryons are locked into stars (e.g. Bower et al. 2001, Balogh et al. 2001). This calls for the presence of a feedback mechanisms, such as supernova explosions (e.g. Menci & Cavaliere 2000, Finoguenov et al. 2000, Pipino et al. 2002; Kravtsov & Yepes 2000) or Active Galactic Nuclei (e.g. Valageas & Silk 1999, Wu et al. 2000, Yamada & Fujita 2001), which, given reasonable efficiencies of coupling to the hot ICM, may be able to provide an adequate amount of extra energy to balance overcooling.

3 OBSERVATIONAL FRAMEWORK

3.1 *Optically-based Cluster Surveys*

Abell (1958) provided the first extensive, statistically complete sample of galaxy clusters. Based on pure visual inspection, clusters were identified as enhancements in the galaxy surface density on Palomar Observatory Sky Survey (POSS) plates, by requiring that at least 50 galaxies were contained within a metric radius $R_A = 3h_{50}^{-1}$ Mpc and a predefined magnitude range. Clusters were characterized by their richness and estimated distance. The Abell catalog has been for decades the prime source for detailed studies of individual clusters and for characterizing the large scale distribution of matter in the nearby Universe. The sample was later extended to the Southern hemisphere by Corwin and Olowin (Abell, Corwin & Olowin, 1989) by using UK Schmidt survey plates. Another comprehensive cluster catalog was compiled by Zwicky and collaborators (Zwicky et al. 1966), who extended the analysis to poorer clusters using criteria less strict than Abell's in defining galaxy overdensities.

Several variations of the Abell criteria defining clusters were used in an automated and objective fashion when digitized optical plates became available. The Edinburgh-Durham Southern Galaxy Catalog, constructed from the COSMOS scans of UK Schmidt plates around the Southern Galactic Pole, was used to compile the first machine-based cluster catalog (Lumsden et al. 1992). In a similar effort, the Automatic Plate Measuring machine galaxy catalog was used to build a sample of ~ 1000 clusters (Maddox et al. 1990, Dalton et al. 1997).

Projection effects in the selection of cluster candidates have been much debated. Filamentary structures and small groups along the line of sight can mimic a moderately rich cluster when projected onto the plane of the sky. In addition, the background galaxy distribution against which two dimensional overdensities are selected, is far from uniform. As a result, the background subtraction process can produce spurious low-richness clusters during searches for clusters in galaxy catalogs. N-body simulations have been extensively used to build mock galaxy catalogs from which the completeness and spurious fraction of Abell-like samples

of clusters can be assessed (e.g. van Haarlem et al. 1997). All-sky, X-ray selected surveys have significantly alleviated these problems and fueled significant progress in this field as discussed below.

Optical plate material deeper than the POSS was successfully employed to search for more distant clusters with purely visual techniques (Kristian et al. 1978, Couch et al. 1991, Gunn et al. 1986). By using red-sensitive plates, Gunn and collaborators were able to find clusters out to $z \simeq 0.9$. These searches became much more effective with the advent of CCD imaging. Postman et al. (1996) were the first to carry out a V&I-band survey over 5 deg^2 (the Palomar Distant Cluster Survey, PDCS) and to compile a sample of 79 cluster candidates using a matched-filter algorithm. This technique enhances the contrast of galaxy overdensities at a given position, utilizing prior knowledge of the luminosity profile typical of galaxy clusters. Olsen et al. (1999) used a similar algorithm to select a sample of 35 distant cluster candidates from the ESO Imaging Survey I-band data. A simple and equally effective counts-in-cell method was used by Lidman & Peterson (1996) to select a sample of 104 distant cluster candidates over 13 deg^2 . All these surveys, by using relatively deep I-band data, are sensitive to rich clusters out to $z \sim 1$. A detailed spectroscopic study of one of the most distant clusters at $z = 0.89$ discovered in this way is reported in Lubin et al. (2000).

Dalcanton (1996) proposed another method of optical selection of clusters, in which drift scan imaging data from relatively small telescopes is used to detect clusters as positive surface brightness fluctuations in the background sky. Gonzales et al. (2001) used this technique to build a sample of ~ 1000 cluster candidates over 130 deg^2 . Spectroscopic follow-up observations will assess the efficiency of this technique.

The advantage of carrying out automated searches based on well-defined selection criteria (e.g. Postman et al. 1996) is that the survey selection function can be computed, thus enabling meaningful statistical studies of the cluster population. For example, one can quantify the probability of detecting a galaxy cluster as a function of redshift for a given set of other parameters, such as galaxy luminosity function, luminosity profile, luminosity and color evolution of cluster galaxies, and field galaxy number counts. A comprehensive report on the performance of different cluster detection algorithms applied to two-dimensional projected distributions can be found in Kim et al. (2002).

The success rate of finding real bound systems in optical surveys is generally relatively high at low redshift ($z < 0.3$, Holden et al. 1999), but it degrades rapidly at higher redshifts, particularly if only one passband is used, as the field galaxy population overwhelms galaxy overdensities associated with clusters. The simplest way to counteract this effect is to observe in the near-infrared bands ($\gtrsim 1\mu\text{m}$). The cores of galaxy clusters are dominated by red, early-type galaxies at least out to $z \simeq 1.3$ for which the dimming effect of the K-correction is particularly severe. In addition, the number counts of the field galaxy population are flatter in the near-IR bands than in the optical. Thus, by moving to z, J, H, K bands, one can progressively compensate the strong K-correction and enhance the contrast of (red) cluster galaxies against the background (blue) galaxy distribution. An even more effective way to enhance the contrast of distant clusters is to use some color information, so that only overdensities of galaxies with peculiar red colors can be selected from the field. With a set of two or three broad band filters, which sample the rest frame UV and optical light at different redshifts, one can separate out early type galaxies which dominate cluster cores from the late type

galaxy population in the field. The position of the cluster red sequence in color-magnitude diagrams, and red clumps in color-color diagrams can also be used to provide an accurate estimate of the cluster redshift, by modeling the relatively simple evolutionary history of early-type galaxies.

The effectiveness of this method was clearly demonstrated by Stanford et al. (1997), who found a significant overdensity of red galaxies with $J - K$ and $I - K$ colors typical of $z > 1$ ellipticals and were able to spectroscopically confirm this system as a cluster at $z = 1.27$ (c.f. see also Dickinson 1997). With a similar color enhancement technique and follow-up spectroscopy, Rosati et al. (1999) confirmed the existence of an X-ray selected cluster at $z = 1.26$. Gladders & Yee (2000) applied the same technique in a systematic fashion to carry out a large area survey in R and z bands (the Red Sequence Survey), which is currently underway and promises to unveil rare, very massive clusters out to $z \sim 1$.

By increasing the number of observed passbands one can further increase the efficiency of cluster selection and the accuracy of their estimated redshifts. In this respect, a significant step forward in mapping clusters in the local Universe will be made with the five-band photometry provided by the Sloan Digital Sky Survey (York et al. 2000). The data will allow clusters to be efficiently selected with photometric redshift techniques, and will ultimately allow hundreds of clusters to be searched directly in redshift space. The next generation of wide field (> 100 deg 2) deep multicolor surveys in the optical and especially the near-infrared will powerfully enhance the search for distant clusters.

3.2 X-ray Cluster Surveys

The *Uhuru* X-ray satellite, which carried out the first X-ray sky survey (Giacconi et al. 1972), revealed a clear association between rich clusters and bright X-ray sources (Gursky et al. 1971, Kellogg et al. 1971). *Uhuru* observations also established that X-ray sources identified as clusters were among the most luminous in the sky (10^{43-45} erg s $^{-1}$), were extended and showed no variability. Felten et al. (1966) first suggested the X-ray originated as thermal emission from diffuse hot intra-cluster gas (Cavaliere et al. 1971). This was later confirmed when the first high quality X-ray spectra of clusters were obtained with the HEAO-1 A2 experiment (e.g. Henriksen and Mushotzsky, 1986). These spectra were best fit by a thermal bremsstrahlung model, with temperatures in the range $2 \times 10^7 - 10^8$ keV, and revealed the 6.8 keV iron K α line, thus showing that the ICM was a highly ionized plasma pre-enriched by stellar processes.

The HEAO-1 X-ray Observatory (Rothschild et al. 1979) performed an all-sky survey with much improved sensitivity compared to *Uhuru* and provided the first flux-limited sample of extragalactic X-ray sources in the 2-10 keV band, with a limiting flux of 3×10^{-11} erg cm $^{-2}$ s $^{-1}$ (Piccinotti et al. 1982). Among the 61 extragalactic sources discovered outside the galactic plane ($|b| > 20^\circ$), 30 were identified as galaxy clusters, mostly in the Abell catalog. This first X-ray flux-limited sample allowed an estimate of the cluster X-ray luminosity function (XLF) in the range $L_X = 10^{43} - 3 \cdot 10^{45}$ erg s $^{-1}$. The derived space density of clusters (all at $z < 0.1$) is fairly close to current values. An earlier determination of the XLF based on optically selected Abell clusters (McKee et al. 1980) and the same HEAO-1 A2 data gave similar results.

The Piccinotti et al. sample was later augmented by Edge et al. (1990), who extended the sample using the *Ariel V* catalog (McHardy et al. 1981) and revised

the identifications of several clusters using follow-up observations by the *Einstein Observatory* and *EXOSAT*. With much improved angular resolution, these new X-ray missions allowed confused sources to be resolved and fluxes to be improved. The resulting sample included 55 clusters with a flux limit a factor of two fainter than in the original Piccinotti catalog.

Confusion effects in the large beam ($\gtrsim 1^\circ$) early surveys, such as *HEAO-1* and *Ariel V*, had been the main limiting factor in cluster identification. With the advent of X-ray imaging with focusing optics in the 80's, particularly with the *Einstein Observatory* (Giacconi et al. 1979), it was soon recognized that X-ray surveys offer an efficient means of constructing samples of galaxy clusters out to cosmologically interesting redshifts.

First, the X-ray selection has the advantage of revealing physically-bound systems, because diffuse emission from a hot ICM is the direct manifestation of the existence of a potential well within which the gas is in dynamical equilibrium with the cool baryonic matter (galaxies) and the dark matter. Second, the X-ray luminosity is well correlated with the cluster mass (see right panel of Figure 2). Third, the X-ray emissivity is proportional to the square of the gas density (Section 2), hence cluster emission is more concentrated than the optical bidimensional galaxy distribution. In combination with the relatively low surface density of X-ray sources, this property makes clusters high contrast objects in the X-ray sky, and alleviates problems due to projection effects that affect optical selection. Finally, an inherent fundamental advantage of X-ray selection is the ability to define flux-limited samples with well-understood selection functions. This leads to a simple evaluation of the survey volume and therefore to a straightforward computation of space densities. Nonetheless, there are some important caveats described below.

Pioneering work in this field was carried out by Gioia et al. (1990a) and Henry et al. (1992) based on the *Einstein Observatory* Extended Medium Sensitivity Survey (EMSS, Gioia et al. 1990b). The EMSS survey covered over 700 square degrees using 1435 imaging proportional counter (IPC) fields. A highly complete spectroscopic identification of 835 serendipitous sources lead to the construction of a flux-limited sample of 93 clusters out to $z = 0.58$. By extending significantly the redshift range probed by previous samples (e.g. Edge et al. 1990), the EMSS allowed the cosmological evolution of clusters to be investigated. Several follow-up studies have been undertaken such as the CNOC survey (e.g. Yee et al. 1996), and gravitational lensing (Gioia & Luppino 1994).

The *ROSAT* satellite, launched in 1990, allowed a significant step forward in X-ray surveys of clusters. The *ROSAT-PSPC* detector, in particular, with its unprecedented sensitivity and spatial resolution, as well as low instrumental background, made clusters high contrast, extended objects in the X-ray sky. The *ROSAT* All-Sky Survey (RASS, Trümper 1993) was the first X-ray imaging mission to cover the entire sky, thus paving the way to large contiguous-area surveys of X-ray selected nearby clusters (e.g. Ebeling et al. 1997, 1998, 2000, 2001; Burns et al. 1996; Crawford et al. 1995; De Grandi et al. 1999; Böhringer et al. 2000, 2001). In the northern hemisphere, the largest compilations with virtually complete optical identification include, the Bright Cluster Sample (BCS, Ebeling et al. 1998), its extension (Ebeling et al. 2000b), and the Northern *ROSAT* All Sky Survey (NORAS, Böhringer et al. 2000). In the southern hemisphere, the *ROSAT-ESO* flux limited X-ray (REFLEX) cluster survey (Böhringer et al. 2001) has completed the identification of 452 clusters, the largest, homogeneous com-

pilation to date. Another on-going study, the Massive Cluster Survey (MACS, Ebeling et al. 2001) is aimed at targeting the most luminous systems at $z > 0.3$ which can be identified in the RASS at the faintest flux levels. The deepest area in the RASS, the North Ecliptic Pole (NEP, Henry et al. 2001) which *ROSAT* scanned repeatedly during its All-Sky survey, was used to carry out a complete optical identification of X-ray sources over a 81 deg^2 region. This study yielded 64 clusters out to redshift $z = 0.81$.

In total, surveys covering more than 10^4 deg^2 have yielded over 1000 clusters, out to redshift $z \simeq 0.5$. A large fraction of these are new discoveries, whereas approximately one third are identified as clusters in the Abell or Zwicky catalogs. For the homogeneity of their selection and the high degree of completeness of their spectroscopic identifications, these samples are now becoming the basis for a large number of follow-up investigations and cosmological studies.

After the completion of the all-sky survey, *ROSAT* conducted thousands of pointed observations, many of which (typically those outside the galactic plane not targeting very bright or extended X-ray sources) can be used for a serendipitous search for distant clusters. It was soon realized that the good angular resolution of the *ROSAT-PSPC* allowed screening of thousands of serendipitous sources and the selection of cluster candidates *solely* on the basis of their flux and spatial extent. In the central 0.2 deg^2 of the *PSPC* field of view the point spread function (PSF) is well approximated by a Gaussian with FWHM= $30 - 45''$. Therefore a cluster with a canonical core radius of $250 h^{-1}\text{kpc}$ (Forman & Jones 1982) should be resolved out to $z \sim 1$, as the corresponding angular distance always exceeds $45''$ for current values of cosmological parameters (important surface brightness biases are discussed below).

ROSAT-PSPC archival pointed observations were intensively used for serendipitous searches of distant clusters. These projects, which are now completed or nearing completion, include: the RIXOS survey (Castander et al. 1995), the *ROSAT* Deep Cluster Survey (RDCS, Rosati et al. 1995, 1998), the Serendipitous High-Redshift Archival *ROSAT* Cluster survey (SHARC, Collins et al. 1997, Burke et al. 1997), the Wide Angle *ROSAT* Pointed X-ray Survey of clusters (WARPS, Scharf et al. 1997, Jones et al. 1998, Perlman et al. 2002), the 160 deg^2 large area survey (Vikhlinin et al. 1998b), the *ROSAT* Optical X-ray Survey (ROXS, Donahue et al. 2001). *ROSAT-HRI* pointed observations, which are characterized by a better angular resolution although with higher instrumental background, have also been used to search for distant clusters in the Brera Multi-scale Wavelet catalog (BMW, Campana et al. 1999).

A principal objective of all these surveys has been the study of the cosmological evolution of the space density of clusters. Results are discussed in Section 4 and 5, below. In Figure 4, we give an overview of the flux limits and surveyed areas of all major cluster surveys carried out over the last two decades. RASS-based surveys have the advantage of covering contiguous regions of the sky so that the clustering properties of clusters (e.g. Collins et al. 2000, Mullis et al. 2001), and the power spectrum of their distribution (Schücker et al. 2001a) can be investigated. They also have the ability to unveil rare, massive systems albeit over a limited redshift and X-ray luminosity range. Serendipitous surveys, or general surveys, which are at least a factor of ten deeper but cover only a few hundreds square degrees, provide complementary information on lower luminosities, more common systems and are well suited for studying cluster evolution on a larger redshift baseline. The deepest pencil-beam surveys, such as the Lockman Hole

with *XMM* (Hasinger et al. 2000) and the Chandra Deep Fields (Giacconi et al. 2002, Bauer et al. 2002), allow the investigation of the faintest end of the XLF (poor clusters and groups) out to $z \sim 1$.

3.3 Strategies and Selection Functions for X-ray Surveys

Ideally, one would like to use selection criteria based on X-ray properties alone to construct a flux-limited sample with a simple selection function. The task of separating clusters from the rest of the X-ray source population is central to this work. At the *ROSAT* flux limit ($\sim 1 \times 10^{-14}$ erg cm $^{-2}$ s $^{-1}$ for clusters) $\sim 10\%$ of extragalactic X-ray sources are galaxy clusters. A program of complete optical identification is very time consuming, as only spectroscopy can establish in many cases whether the X-ray source is associated with a real cluster. The EMSS and NEP samples, for example, were constructed in this way. In some cases, the hardness ratio (a crude estimate of the source's X-ray spectral energy distribution) is used to screen out sources which are incompatible with thermal spectra or to resolve source blends. With the angular resolution provided by *ROSAT*, however, it became possible to select clusters on the basis of their spatial extent. This is particularly feasible with pointed observations, as opposed to all-sky survey data which are characterized by a broader PSF and shallower exposures, so that faint and/or high redshift clusters are not always detected as extended (e.g. Ebeling et al. 1997, Böhringer et al. 2001).

In constructing RASS based samples (shaded circles in Figure 4) most of the authors had to undertake a complete optical identification program of $\sim 10^4$ sources using POSS plates or CCD follow-up imaging in order to build a sample of cluster candidates. Whereas a sizable fraction of these systems can be readily identified in previous cluster catalogs (primarily Abell's), spectroscopy is needed to measure redshifts of newly discovered systems or to resolve ambiguous identifications. We recall that optically selected, X-ray confirmed samples, such as the X-ray Brightest Abell-like Clusters (XBACS, Ebeling et al. 1996), while useful for studying optical–X-ray correlations, lead to incomplete flux-limited samples. Many of the low X-ray luminosity systems (poor clusters or groups) are missed in the optical selection even though they lie above the X-ray flux limit of the RASS.

Most of the *ROSAT* serendipitous surveys (dark circles in Figure 4) have adopted a very similar methodology but somewhat different identification strategies. Cluster candidates are selected from a serendipitous search for extended X-ray sources above a given flux limit in deep *ROSAT*-PSPC pointed observations. Moderately deep CCD imaging in red passbands (or in near-IR for the most distant candidates) is used to reveal galaxy overdensities near the centroid of X-ray emission. Extensive spectroscopic follow-up programs associated with these surveys, have lead to the identification of roughly 200 new clusters or groups, and have increased the number of clusters known at $z > 0.5$ by approximately a factor of ten.

An essential ingredient for the evaluation of the selection function of X-ray surveys is the computation of the sky coverage: the effective area covered by the survey as a function of flux. In general, the exposure time, as well as the background and the PSF are not uniform across the field of view of X-ray telescopes (owing to their inherent optical design), which introduces vignetting and a degradation of the PSF at increasing off-axis angles. As a result, the sensitivity to source detection varies significantly across the survey area so that only bright

sources can be detected over the entire solid angle of the survey, whereas at faint fluxes the effective area decreases. An example of survey sky coverage is given in Figure 5 (left). By integrating the volume element of the Friedmann-Robertson-Walker metric, $dV/d\Omega dz(z, \Omega_m, \Omega_\Lambda)$ (e.g. Carroll et al. 1992), over these curves one can compute the volume that each survey probes above a given redshift z , for a given X-ray luminosity ($L_X = 3 \times 10^{44} \text{ erg s}^{-1} \simeq L_X^*$, the characteristic luminosity, in the figure). The resulting survey volumes are shown in Figure 5 (right). By normalizing this volume to the local space density of clusters (ϕ^* , see below) one obtains the number of L^* volumes accessible in the survey above a given redshift. Assuming no evolution, this yields an estimate of the number of typical bright clusters one expects to discover.

By covering different solid angles at varying fluxes, these surveys probe different volumes at increasing redshift and therefore different ranges in X-ray luminosities at varying redshifts. The EMSS has the greatest sensitivity to the most luminous, yet rare, systems but only a few clusters at high redshift lie above its bright flux limit. Deep *ROSAT* surveys probe instead the intermediate-to-faint end of the XLF. As a result, they have lead to the discovery of many new clusters at $z > 0.4$. The RDCS has pushed this search to the faintest fluxes yet, providing sensitivity to the highest redshift systems with $L_X \lesssim L_X^*$ even beyond $z = 1$. The WARPS, and particularly the 160 deg² survey have covered larger areas at high fluxes thus better studying the bright end of the XLF out to $z \simeq 1$.

Particular emphasis is given in these searches to detection algorithms that are designed to examine a broad range of cluster parameters (X-ray flux, surface brightness, morphology) and to deal with source confusion at faint flux levels. The traditional detection algorithm used in X-ray astronomy for many years, the sliding cell method, is not adequate for this purpose. A box of fixed size is slid across the image, and sources are detected as positive fluctuations that deviate significantly from Poissonian expectations based on a global background map (the latter being constructed from a first scan of the image). Although this method works well for point-like sources, it is less suited to extended, low-surface brightness sources, which can consequently be missed leading to a significant incompleteness in flux-limited cluster samples.

The need for more general detection algorithms, not only geared to the detection of point sources, became important with *ROSAT* observations, which probe a much larger range in surface brightness than previous missions (e.g. *Einstein*). A popular alternative approach to source detection and characterization developed specifically for cluster surveys is based on wavelet techniques (e.g. Rosati et al. 1995, Vikhlinin et al. 1998b, Lazzati et al. 1999, Romer et al. 2000). Wavelet analysis is essentially a multi-scale analysis of the image based on a quasi-orthonormal decomposition of a signal via the wavelet transform which enables significant enhancement of the contrast of sources of different sizes against non-uniform backgrounds. This method, besides being equally efficient at detecting sources of different shapes and surface brightnesses, is well-suited to dealing with confusion effects, and allows source parameters to be measured without knowledge of the background. Another method that has proved to be well-suited for the detection of extended and low surface brightness emission is based on Voronoi Tessellation and Percolation (VTP, Scharf et al. 1997 and references therein).

Besides detection algorithms, which play a central role in avoiding selection ef-

fects, there are additional caveats to be considered when computing the selection function of X-ray cluster surveys. For example, the sky coverage function (Figure 5) depends not only on the source flux but in general on the extent or surface brightness of cluster sources (Rosati et al. 1995, Scharf et al. 1997, Vihklinin et al. 1998). This effect can be tested with extensive simulations, by placing artificial clusters (typically using β -profiles) in the field and measuring the detection probability for different cluster parameters or instrumental parameters.

More generally, as in all flux-limited samples of extended sources (e.g. optical galaxy surveys), one has to make sure that the sample does not become surface brightness (SB) limited at very faint fluxes. As the source flux decreases, clusters with smaller mean SB have a higher chance of being missed, because their signal-to-noise is likely to drop below the detection threshold. SB dimming at high redshifts ($SB \propto (1+z)^{-4}$) can thus create a serious source of incompleteness at the faintest flux levels. This depends critically on the steepness of the SB-profile of distant X-ray clusters, and its evolution. Besides simulations of the detection process, the most meaningful way to test these selection effects is to verify that derived cluster surface or space densities do not show any trend across the survey area (e.g. a decrease in regions with higher background, low exposures, degraded PSF). The task of the observer is to understand what is the fiducial flux limit above which the sample is truly flux-limited and free of SB effects. This fiducial flux limit is typically a factor of 2–3 higher than the minimum detectable flux in a given survey.

An additional source of sample contamination or misidentification may be caused by clusters hosting X-ray bright AGN, or by unrelated point sources projected along the line of sight of diffuse cluster emission. The former case does not seem to be a matter of great concern, because bright AGN have been found near the center of clusters in large compilations (Böhringer et al. 2001) in less than 5% of the cases. The latter effect can be significant in distant and faint *ROSAT* selected clusters, for which high resolution *Chandra* observations (Stanford et al. 2001, 2002) have revealed up to 50% flux contamination in some cases.

Concerning selection biases, a separate issue is whether, using X-ray selection, one might miss systems that, although virialized, have an unusually low X-ray luminosity. These systems would be outliers in the $L_X - M$ or $L_X - T$ relation (Section 5.2). Such hypothetical systems are at odds with our physical understanding of structure formation and would require unusual mechanisms that would (a) lead galaxies to virialize but the gaseous component not to thermalize in the dark matter potential well, (b) allow energy sources to dissipate or remove the gas after collapse, or (c) involve formation scenarios in which only a small fraction of the gas collapses. Similarly, systems claimed to have unusually high mass-to-optical luminosity ratio, M/L , such as MG2016+112 from *ASCA* observations (Hattori et al. 1998) have not held up. MG2016+112 was later confirmed to be an ordinary low mass cluster at $z = 1$ by means of near-infrared imaging (Benítez et al. 1999) and spectroscopic (Soucail et al. 2001) follow-up studies. Chartas et al. (2001) have completely revised the nature of the X-ray emission with *Chandra* observations. Comparing optical and X-ray techniques for clusters' detection, Donahue et al. (2001) carried out an optical/X-ray joint survey in the same sky area (ROXS). They found no need to invoke an X-ray faint population of massive clusters.

3.4 Other methods

X-ray and optical surveys have been by far the most exploited techniques for studying the distribution and evolution of galaxy clusters. It is beyond the scope of this paper to review other cluster-finding methods, which we only summarize here for completeness:

- *Search for galaxy overdensities around high- z radio galaxies or AGN:* searches are conducted in near-IR or narrow-band filters, or by means of follow-up X-ray observations. Although not suited for assessing cluster abundances, this method has provided the only examples of possibly virialized systems at $z > 1.5$ (e.g. Pascarelle et al. 1996; Dickinson 1997; Crawford & Fabian 1996, Hall & Green 1998; Pentericci et al. 2000; Fabian et al. 2001b, Venemans et al. 2002).
- *Sunyaev-Zeldovich effect:* clusters are revealed by measuring the distortion of the CMB spectrum owing to the hot ICM. This method does not depend on redshift and provides reliable estimate of cluster masses. It is possibly one of the most powerful methods to find distant clusters in the years to come. At present, serendipitous surveys with interferometric techniques (e.g. Carlstrom et al. 2001) cannot cover large areas (i.e. more than ~ 1 deg 2) and their sensitivity is limited to the most X-ray luminous clusters.
- *Gravitational lensing:* in principle a powerful method to discover mass concentrations in the universe through the statistical distortion of background galaxy images (see Mellier 1999 for a review).
- *Search for clusters around bent-double radio sources:* radio galaxies with bent lobes are often associated with dense ICM and are therefore good tracers of rich cluster environments (e.g. Blanton et al. 2001).
- *Clustering of absorption line systems:* this method has lead to a few detections of “proto-clusters” at $z \gtrsim 2$ (e.g. Francis et al. 1996). The most serious limitation of this technique is the small sample volume.

4 THE SPACE DENSITY OF X-RAY CLUSTERS

4.1 Local Cluster Number Density

The determination of the local ($z \lesssim 0.3$) cluster abundance plays a crucial role in assessing the evolution of the cluster abundance at higher redshifts. The cluster XLF is commonly modeled with a Schechter function:

$$\phi(L_X)dL_X = \phi^* \left(\frac{L_X}{L_X^*} \right)^{-\alpha} \exp(-L_X/L_X^*) \frac{dL_X}{L_X^*}, \quad (7)$$

where α is the faint-end slope, L_X^* is the characteristic luminosity, and ϕ^* is directly related to the space-density of clusters brighter than L_{min} : $n_0 = \int_{L_{min}}^{\infty} \phi(L)dL$. The cluster XLF in the literature is often written as: $\phi(L_{44}) = K \exp(-L_X/L_X^*) L_{44}^{-\alpha}$, with $L_{44} = L_X/10^{44}$ erg s $^{-1}$. The normalization K , expressed in units of $10^{-7}\text{Mpc}^{-3}(10^{44}\text{ erg s}^{-1})^{\alpha-1}$, is related to ϕ^* by $\phi^* = K (L_X^*/10^{44})^{1-\alpha}$.

Using a flux-limited cluster sample with measured redshifts and luminosities, a binned representation of the XLF can be obtained by adding the contribution

to the space density of each cluster in a given luminosity bin ΔL_X :

$$\phi(L_X) = \left(\frac{1}{\Delta L_X} \right) \sum_{i=1}^n \frac{1}{V_{max}(L_i, f_{lim})}; \quad (8)$$

where V_{max} is the total search volume defined as

$$V_{max} = \int_0^{z_{max}} S[f(L, z)] \left(\frac{d_L(z)}{1+z} \right)^2 \frac{c dz}{H(z)}. \quad (9)$$

Here $S(f)$ is the survey sky coverage, which depends on the flux $f = L/(4\pi d_L^2)$, $d_L(z)$ is the luminosity distance, and $H(z)$ is the Hubble constant at z (e.g. Peebles 1993, pag.312). We define z_{max} as the maximum redshift out to which the object is included in the survey. Equations 8 and 9 can be easily generalized to compute the XLF in different redshift bins.

In Figure 6 we summarize the recent progress made in computing $\phi(L_X)$ using primarily low-redshift *ROSAT* based surveys. This work improved the first determination of the cluster XLF (Piccinotti et al. 1982, see Section 3.2). The BCS and REFLEX cover a large L_X range and have good statistics at the bright end, $L_X \gtrsim L_X^*$ and near the knee of the XLF. Poor clusters and groups ($L_X \lesssim 10^{43} \text{ erg s}^{-1}$) are better studied using deeper surveys, such as the RDCS. The very faint end of the XLF has been investigated using an optically selected, volume-complete sample of galaxy groups detected *a posteriori* in the RASS (Burns et al. 1996).

From Figure 6, we note the very good agreement among all these independent determinations. Best-fit parameters are consistent with each other with typical values: $\alpha \simeq 1.8$ (with 15% variation), $\phi^* \simeq 1 \times 10^{-7} h_{50}^3 \text{ Mpc}^{-3}$ (with 50% variation), and $L_X^* \simeq 4 \times 10^{44} \text{ erg s}^{-1}$ [0.5–2 keV]. Residual differences at the faint end are probably the result of cosmic variance effects, because the lowest luminosity systems are detected at very low redshifts where the search volume becomes small (see Böhringer et al. 2002b). Such an overall agreement is quite remarkable considering that all these surveys used completely different selection techniques and independent datasets. Evidently, systematic effects associated with different selection functions are relatively small in current large cluster surveys. This situation is in contrast with that for the galaxy luminosity function in the nearby Universe, which is far from well established (Blanton et al. 2001). The observational study of cluster evolution has indeed several advantages respect to galaxy evolution, despite its smaller number statistics. First, a robust determination of the local XLF eases the task of measuring cluster evolution. Second, X-ray spectra constitute a single parameter family based on temperature and K-corrections are much easier to compute than in the case of different galaxy types in the optical bands.

4.2 The Cluster Abundance at Higher Redshifts and Its Evolution

A first analysis of the EMSS cluster sample (Gioia et al. 1990a) revealed negative evolution of the XLF – a steepening of the high-end of XLF indicating a dearth of high luminosity clusters at $z > 0.3$. This result was confirmed by Henry et al. (1992) using the complete EMSS sample with an appropriate sky coverage function. Edge et al. (1990) found evidence of a strong negative evolution already at redshifts < 0.2 using a HEAO-1 based cluster sample (see Section 3.2). The

very limited redshift baseline made this result somewhat controversial, until it was later ruled out by the analysis of the first RASS samples (Ebeling et al. 1997). The *ROSAT* deep surveys extended the EMSS study on cluster evolution. Early results (Castander et al. 1995) seemed to confirm and even to reinforce the evidence of negative evolution. This claim, based on a sample of 12 clusters, was later recognized to be the result of sample incompleteness and an overestimate of the solid angle covered at low fluxes and its corresponding search volume (Burke et al. 1997, Rosati et al. 1998, Jones et al. 1998).

If cluster redshifts are not available, X-ray flux-limited samples can be used to trace the surface density of clusters at varying fluxes. In Figure 7, we show several determinations of the cumulative cluster number counts stretching over five decades in flux. This comparison shows a good agreement (at the 2σ level) among independent determinations (see also Gioia et al. 2001). The slope at bright fluxes is very close to the Euclidean value of 1.5 (as expected for an homogeneous distribution of objects over large scales), whereas it flattens to $\simeq 1$ at faint fluxes. The slope of the LogN–LogS is mainly determined by the faint-to-moderate part of the XLF, but it is rather insensitive to the abundance of the most luminous, rare systems. The fact that the observed counts are consistent with no-evolution predictions, obtained by integrating the local XLF, can be interpreted as an indication that a significant fraction of the cluster population does not evolve with redshift (Rosati et al. 1995, 1998, Jones et al. 1998, Vikhlinin et al. 1998a). We have included the recent data from the Chandra Deep Fields North (Bauer et al. 2002) and South (Giacconi et al. 2002), which have extended the number counts by two decades. Note that cosmic variance may be significant because these are only two, albeit deep, pencil beam fields ($\lesssim 0.1 \text{ deg}^2$). Serendipitous surveys with *Chandra* and *XMM* (see Figure 4) will fill the gap between these measurements and the *ROSAT* surveys. The no evolution curves in Figure 7 are computed by integrating the BCS local XLF (Ebeling et al. 1997) according to the evolutionary model in Figure 9.

A much improved picture of the evolution of the cluster abundance emerged when, with the completion of spectroscopic follow-up studies, several cluster samples were used to compute the XLF out to $z \simeq 0.8$. These first measurements are summarized in Figure 8. Although binned representations of the XLF are not straightforward to compare, it is evident that within the error bars there is little, if any, evolution of the cluster space density at $L_X([0.5 - 2]\text{keV}) \lesssim 3 \times 10^{44} \text{ erg s}^{-1} \simeq L_X^*$ out to redshift $z \simeq 0.8$. These results (Burke et al. 1997, Rosati et al. 1998, Jones et al. 1998, Vikhlinin et al. 1998a, Nichols et al. 1999) extended the original study of EMSS to fainter luminosities and larger redshifts, and essentially confirmed the EMSS findings in the overlapping X-ray luminosity range. The ability of all these surveys to adequately study the bright end of the XLF is rather limited, since there is not enough volume to detect rare systems with $L_X > L_X^*$. The 160 deg^2 survey by Vikhlinin et al. (1998a), with its large area, did however confirm the negative evolution at $L_X \gtrsim 4 \times 10^{44} \text{ erg s}^{-1}$. Further analyses of these datasets have confirmed this trend, i.e. an apparent drop of super- L_X^* clusters at $z \gtrsim 0.5$ (Nichol et al. 1999 from the Bright-SHARC survey; Rosati et al. 2000 from the RDCS, Gioia et al. 2001 from the NEP survey). These findings, however, were not confirmed by Ebeling et al. (2000) in an analysis of the WARPS sample.

The evolution of the bright end of the XLF has remained a hotly debated sub-

ject for several years. The crucial issue in this debate is to properly quantify the statistical significance of any claimed evolutionary effect. The binned representation of the XLF in Figure 8 can be misleading and can even lead to biases (Page & Carrera 2000). The full information contained in any flux-limited cluster sample can be more readily recovered by analyzing the unbinned (L_X, z) distribution with a maximum-likelihood approach, which compares the observed cluster distribution on the (L_X, z) plane with that expected from a given XLF model. Rosati et al. (2000) used this method by modeling the cluster XLF as an evolving Schechter function: $\phi(L) = \phi_0(1+z)^A L^{-\alpha} \exp(-L/L^*)$, with $L^* = L_0^*(1+z)^B$; where A and B are two evolutionary parameters for density and luminosity; ϕ_0 and L_0^* the local XLF values (Equation 7). Figure 9 shows an application of this method to the RDCS and EMSS sample, and indicates that the no-evolution case ($A = B = 0$) is excluded at more than 3σ levels in both samples when the most luminous systems are included in the analysis. However, the same analysis confined to clusters with $L_X < 3 \times 10^{44}$ erg s $^{-1}$ yields an XLF consistent with no evolution. In Figure 9 we also report the latest determinations of the XLF out to $z \sim 1$.

In summary, by combining all the results from *ROSAT* surveys one obtains a consistent picture in which the comoving space density of the bulk of the cluster population is approximately constant out to $z \simeq 1$, but the most luminous ($L_X \gtrsim L_X^*$), presumably most massive clusters were likely rarer at high redshifts ($z \gtrsim 0.5$). Significant progress in the study of the evolution of the bright end of the XLF would require a large solid angle and a relatively deep survey with an effective solid angle of $\gg 100$ deg 2 at a limiting flux of 10^{-14} erg cm $^{-2}$ s $^{-1}$.

The convergence of the results from several independent studies illustrates remarkable observational progress in determining the abundance of galaxy clusters out to $z \sim 1$. At the beginning of the *ROSAT* era, until the mid nineties, controversy surrounded the usefulness of X-ray surveys of distant galaxy clusters and many believed that clusters were absent at $z \sim 1$. This prejudice arose from an over-interpretation of the early results of the EMSS survey. Gioia et al. (1990a) did point out that the evolution of the XLF was limited only to the very luminous systems but this important caveat was often overlooked. The original controversy concerning cluster evolution inferred from optical and X-ray data finds an explanation in light of the *ROSAT* results. Optical surveys (Couch et al. 1991, Postman et al. 1996) have shown no dramatic decline in the comoving volume density of rich clusters out to $z \simeq 0.5$. This was considered to be in contrast with the EMSS findings. However, these optical searches covered limited solid angles (much smaller than the EMSS) and therefore did not probe adequately the seemingly evolving high end of the cluster mass function.

4.3 Distant X-ray Clusters: the Latest View from *Chandra*

With its unprecedented angular resolution, the *Chandra* satellite has revolutionized X-ray astronomy, allowing studies with the same level of spatial details as in optical astronomy. *Chandra* imaging of low redshift clusters has revealed a complex thermodynamical structure of the ICM down to kiloparsec scales (e.g. Markevitch et al. 2000, Fabian et al. 2000). At high redshifts, deep *Chandra* images still have the ability to resolve cluster cores and to map ICM morphologies at scales below 100 kpc. Moreover, temperatures of major subclumps can be measured for the first time at $z > 0.6$.

Figure 10 is an illustrative example of the unprecedented view that *Chandra* can offer on distant clusters. We show twelve archival images of clusters at $0.7 < z < 1.7$ all covering 2 Mpc (projected at the cluster redshift) and smoothed at the same physical scale (a Gaussian FWHM of 70 kpc). Point-like sources in each field were removed. The intensity (in false colors) is proportional to the square root of the X-ray emission, so that they roughly map the gas density distribution in each cluster. The images are arranged in three redshift bins ($\sim 0.7, 0.8, > 1$), in each row, with X-ray luminosities increasing from left to right. The upper left image shows one of the highest redshift groups known to date, a system discovered in the megasecond exposure of the Chandra Deep Field South (Giacconi et al. 2002) with a core of a few arcseconds. A close inspection of these images reveal a deviation from spherical symmetry in all systems. Some of them are elongated or have cores clearly displaced with respect to the external diffuse envelope (e.g. Holden et al. 2002).

Three of the most luminous clusters at $z \simeq 0.8$ (RXJ1716: Gioia et al. 1999; RXJ0152: Della Ceca et al. 2000, Ebeling et al. 2000a; MS1054: Jeltema et al. 2001) show a double core structure both in the distribution of the gas and in their member galaxies. It is tempting to interpret these morphologies as the result of on-going mergers, although no dynamical information has been gathered to date to support this scenario. In a hierarchical cold dark matter formation scenario, one does expect the most massive clusters at high redshift to be accreting subclumps of comparable masses, and the level of substructure to increase at high redshifts. With current statistical samples however, it is difficult to draw any robust conclusion on the evolution of ICM substructure, which is also found to be a large fraction of the low- z cluster population (e.g. Schücker et al. 2001b).

The third row in Figure 10 show the most distant clusters observed with *Chandra* to date. The first three systems are also among the most distant X-ray selected clusters discovered in the *ROSAT* Deep Cluster Survey (Stanford et al. 2001, 2002), at the very limit of the *ROSAT* sensitivity. RXJ0848 and RXJ0849 are only 5 arcmin apart on the sky (the Lynx field) and are possibly part of a superstructure at $z = 1.26$, consisting of two collapsed, likely virialized clusters (Rosati et al. 1999). Follow-up *Chandra* observations of the Lynx field (Stanford et al. 2001) have yielded for the first time information on ICM morphologies in $z > 1$ clusters and allowed a measurement of their temperatures (see Figure 14 below), implying masses of $(0.5 - 1) \times 10^{15} h_{50}^{-1} M_\odot$. The discovery and the study of these remote systems have the strongest leverage on testing cosmological models.

In Figure 11, we show color composite optical/near-IR images of two clusters at $z > 1$, with overlaid *Chandra* contours. Already at these large lookback times, the temperature and surface brightness profiles of these systems are similar to those of low redshift clusters. Moreover, the morphology of the gas, as traced by the X-ray emission, is well correlated with the spatial distribution of member galaxies, similar to studies at lower redshifts. This suggests that there are already at $z > 1$ galaxy clusters in an advanced dynamical stage of their formation, in which all the baryons (gas and galaxies) have had enough time to thermalize in the cluster potential well. Another example of a $z > 1$ cluster was reported by Hashimoto et al. (2002) using XMM observations of the Lockman Hole.

At $z > 1.3$, X-ray selection has not yielded any cluster based on *ROSAT* data. Follow-up X-ray observations of distant radio galaxies have been used to search for diffuse hot ICM (e.g. Crawford & Fabian 1996). A relatively short *Chandra* observation of the radio galaxy 3C294 at $z = 1.789$ (bottom right in Figure 10)

(Fabian et al. 2001b) has revealed an extended envelope around the central point source, which is the most distant ICM detected so far. Deeper observations are needed to accurately measure the temperature of this system.

5 COSMOLOGY WITH X-RAY CLUSTERS

5.1 *The cosmological mass function*

The mass distribution of dark matter halos undergoing spherical collapse in the framework of hierarchical clustering is described by the Press-Schechter distribution (PS, Press & Schechter 1974). The number of such halos in the mass range $[M, M + dM]$ can be written as

$$n(M, z) dM = \frac{\bar{\rho}}{M} f(\nu) \frac{d\nu}{dM} dM \quad (10)$$

where $\bar{\rho}$ is the cosmic mean density. The function f depends only on the variable $\nu = \delta_c(z)/\sigma_M$, and is normalized so that $\int f(\nu) d\nu = 1$. $\delta_c(z)$ is the linear-theory overdensity extrapolated to the present time for a uniform spherical fluctuation collapsing at redshift z . This quantity conveys information about the dynamics of fluctuation evolution in a generic Friedmann background. It is convenient to express it as $\delta_c(z) = \delta_0(z) [D(0)/D(z)]$, where $D(z)$ is the linear fluctuation growth factor, which depends on the density parameters contributed by matter, Ω_m and by cosmological constant, Ω_Λ (e.g. Peebles 1993). The quantity $\delta_0(z)$ has a weak dependence on Ω_m and Ω_Λ (e.g. Kitayama & Suto 1997). For a critical-density Universe it is $\delta_0 = 1.686$, independent of z .

The r.m.s. density fluctuation at the mass scale M , σ_M , is connected to the fluctuation power spectrum, $P(k)$, by the relation

$$\sigma_M^2 = \frac{1}{2\pi^2} \int_0^\infty dk k^2 P(k) W^2(kR). \quad (11)$$

The dependence of the power spectrum on the wavenumber k is usually written as $P(k) \propto k^{n_{pr}} T^2(k)$, where $T(k)$ is the transfer function, which depends both on the cosmological parameters of the Friedmann background and on the cosmic matter constituents (e.g. fraction of cold, hot and baryonic matter, number of relativistic species; see Kolb & Turner 1989). For a pure cold dark matter (CDM) model, $T(k)$ depends to a good approximation only on the shape parameter $\Gamma = \Omega_m h$ (e.g. Bardeen et al. 1986), while a correction to this dependence needs to be introduced to account for the presence of the baryonic component (e.g. Eisenstein & Hu 1999). The Harrison-Zel'dovich spectrum is generally assumed with the primordial index, $n_{pr} = 1$, consistent with the most recent analyses of the CMB anisotropies (de Bernardis et al. 2001 and references therein). The amplitude of $P(k)$ is usually expressed in terms of σ_8 , the r.m.s. density fluctuation within a top-hat sphere of $8 h^{-1} \text{Mpc}$ radius. Finally, in Equation 11 $W(x)$ is the Fourier representation of the window function, which describes the shape of the volume from which the collapsing object is accreting matter. The comoving fluctuation size R is connected to the mass scale M as $R = (3M/4\pi\bar{\rho})^{1/3}$ for the top-hat window, i.e. $W(x) = 3(\sin x - x \cos x)/x^3$.

In their original derivation of the cosmological mass function, Press & Schechter (1974) obtained the expression $f(\nu) = (2\pi)^{-1/2} \exp(-\nu^2/2)$ for Gaussian density fluctuations. Despite its subtle simplicity (e.g., Monaco 1998), the PS mass

function has served for more than a decade as a guide to constrain cosmological parameters from the mass function of galaxy clusters. Only with the advent of the last generation of N-body simulations, which are able to span a very large dynamical range, significant deviations of the PS expression from the exact numerical description of gravitational clustering have been noticed (e.g. Gross et al. 1998, Governato et al. 1999, Jenkins et al. 2001, Evrard et al. 2002). Such deviations are interpreted in terms of corrections to the PS approach. For example, incorporating the effects of non-spherical collapse (Sheth et al. 2001) generalizes the above PS expression for $f(\nu)$ to

$$f(\nu) = \sqrt{\frac{2a}{\pi}} C \left(1 + \frac{1}{(a\nu^2)^q} \right) \exp \left(-\frac{a\nu^2}{2} \right), \quad (12)$$

where $a = 0.707$, $C = 0.3222$ and $q = 0.3$ (Sheth & Tormen 1999). The above equation reduces to the PS expression for $a = 1$, $C = 1/2$ and $q = 0$. Fitting formulae for $f(\nu)$, which reproduce N-body results to an accuracy of about 10% (e.g. Evrard et al. 2002) are now currently used to derive cosmological constraints from the evolution of the cluster population.

In practical applications, the observational mass function of clusters is usually determined over about one decade in mass. Therefore, it probes the power spectrum over a relatively narrow dynamical range, and does not provide strong constraints on the shape Γ of the power spectrum. Using only the number density of nearby clusters of a given mass M , one can constrain the amplitude of the density perturbation at the physical scale $R \propto (M/\Omega_m \rho_{crit})^{1/3}$ containing this mass. Since such a scale depends both on M and on Ω_m , the mass function of nearby ($z \lesssim 0.1$) clusters is only able to constrain a relation between σ_8 and Ω_m . In the left panel of Figure 12 we show that, for a fixed value of the observed cluster mass function, the implied value of σ_8 from Equation 12 increases as the density parameter decreases.

Determinations of the cluster mass function in the local Universe using a variety of samples and methods indicate that $\sigma_8 \Omega_m^\alpha = 0.4 - 0.6$, where $\alpha \simeq 0.4 - 0.6$, almost independent of the presence of a cosmological constant term providing spatial flatness (e.g. Bahcall & Cen 1993, Eke et al. 1996, Girardi et al. 1998, Viana & Liddle 1999, Blanchard et al. 2000, Pierpaoli et al. 2001, Reiprich & Böhringer 2002, Seljak 2002, Viana et al. 2002). It is worth pointing out that formal statistical uncertainties in the determination of σ_8 from the different analyses are always far smaller, $\lesssim 5\%$, than the above range of values. This suggests that current discrepancies on σ_8 are likely to be ascribed to systematic effects, such as sample selection and different methods used to infer cluster masses. We comment more on such differences in the following section. Completely independent constraints on a similar combination of σ_8 and Ω_m can be obtained with measurements of the cosmic gravitational lensing shear (e.g. Mellier 1999). The most recent results give $\sigma_8 \Omega_m^{0.6} = 0.45 \pm 0.05$ (van Waerbeke et al. 2001, and references therein).

The growth rate of the density perturbations depends primarily on Ω_m and, to a lesser extent, on Ω_Λ , at least out to $z \sim 1$, where the evolution of the cluster population is currently studied. Therefore, following the evolution of the cluster space density over a large redshift baseline, one can break the degeneracy between σ_8 and Ω_m . This is shown in a pictorial way in Figure 1 and quantified in the right panel of Figure 12: models with different values of Ω_m , which are normalized

to yield the same number density of nearby clusters, predict cumulative mass functions that progressively differ by up to orders of magnitude at increasing redshifts.

5.2 Deriving Ω_m from cluster evolution

An estimate of the cluster mass function is reduced to the measurement of masses for a sample of clusters, stretching over a large redshift range, for which the survey volume is well known.

Velocity dispersions for statistical samples of galaxy clusters have been provided by the ESO Nearby Abell Cluster Survey (ENACS; Mazure et al. 2001) and, more recently, by the 2dF survey (de Propris et al. 2002). Application of this method to a statistically complete sample of distant X-ray selected clusters has been pursued by the CNOC (Canadian Network for Observational Cosmology) collaboration (e.g. Yee et al. 1996). The CNOC sample includes 16 clusters from the EMSS in the redshift range $0.17 \leq z \leq 0.55$. Approximately 100 redshifts of member galaxies were measured for each cluster, thus allowing an accurate analysis of the internal cluster dynamics (Carlberg et al. 1997b). The CNOC sample has been used to constrain Ω_m through the M/L_{opt} method (e.g. Carlberg et al. 1997b), yielding $\Omega_m \simeq 0.2 \pm 0.05$. Attempts to estimate the cluster mass function $n(> M)$ using the cumulative velocity dispersion distribution, $n(> \sigma_v)$, were made (Carlberg et al. 1997b). This method, however, provided only weak constraints on Ω_m owing to the narrow redshift range and the limited number of clusters in the CNOC sample (Borgani et al. 1999, Bahcall et al. 1997). The extension of such methodology to a larger and more distant cluster sample would be extremely demanding from the observational point of view, which explains why it has not been pursued thus far.

A conceptually similar, but observationally quite different method to estimate cluster masses, is based on the measurement of the temperature of the intra-cluster gas (see Section 2). Based on the assumption that gas and dark matter particles share the same dynamics within the cluster potential well, the temperature T and the velocity dispersion σ_v are connected by the relation $k_B T = \beta \mu m_p \sigma_v^2$, where $\beta = 1$ would correspond to the case of a perfectly thermalized gas. If we assume spherical symmetry, hydrostatic equilibrium and isothermality of the gas, the solution of Equation 5 provides the link between the total cluster virial mass, M_{vir} , and the ICM temperature:

$$k_B T = \frac{1.38}{\beta} \left(\frac{M_{vir}}{10^{15} h^{-1} M_\odot} \right)^{2/3} [\Omega_m \Delta_{vir}(z)]^{1/3} (1+z) \text{ keV}. \quad (13)$$

$\Delta_{vir}(z)$ is the ratio between the average density within the virial radius and the mean cosmic density at redshift z ($\Delta_{vir} = 18\pi^2 \simeq 178$ for $\Omega_m = 1$; see Eke et al. 1996 for more general cosmologies). Equation 13 is fairly consistent with hydrodynamical cluster simulations with $0.9 \lesssim \beta \lesssim 1.3$ (e.g. Bryan & Norman 1998, Frenk et al. 2000; see however Voit 2000). Such simulations have also demonstrated that cluster masses can be recovered from gas temperature with a $\sim 20\%$ precision (e.g. Evrard et al. 1996).

Observational data on the $M_{vir}-T$ relation show consistency with the $T \propto M_{vir}^{2/3}$ scaling law, at least for $T \gtrsim 3$ keV clusters (e.g. Allen et al. 2001), but with a $\sim 40\%$ lower normalization. As for lower-temperature systems, Finoguenov et al.

(2001) found some evidence for a steeper slope. Such differences might be due to a lack of physical processes in simulations. For example, energy feedback from supernovae or AGNs and radiative cooling (see Section 2, above) can modify the thermodynamical state of the ICM and the resulting scaling relations.

Measurements of cluster temperatures for flux-limited samples of clusters were made using modified versions of the Piccinotti et al. sample (e.g. Henry & Arnaud 1991). These results have been subsequently refined and extended to larger samples with the advent of *ROSAT*, *Beppo-SAX* and, especially, *ASCA*. With these data one can derive the X-ray Temperature Function (XTF), which is defined analogously to Equation 7. XTFs have been computed for both nearby (e.g. Markevitch 1998, see Pierpaoli et al. 2001, for a recent review) and distant (e.g. Eke et al. 1998, Donahue & Voit 1999, Henry 2000) clusters, and used to constrain cosmological models. The mild evolution of the XTF has been interpreted as a case for a low-density Universe, with $0.2 \lesssim \Omega_m \lesssim 0.6$ (see Figure 13). The starting point in the computation of the XTF is inevitably a flux-limited sample for which $\phi(L_X)$ can be computed. Then the $L_X - T_X$ relation is used to derive a temperature limit from the sample flux limit (e.g. Eke et al. 1998). A limitation of the XTFs presented so far is the limited sample size (with only a few $z \gtrsim 0.5$ measurements), as well as the lack of a homogeneous sample selection for local and distant clusters. By combining samples with different selection criteria one runs the risk of altering the inferred evolutionary pattern of the cluster population. This can give results consistent even with a critical-density Universe (Colafrancesco et al. 1997, Viana & Liddle 1999, Blanchard et al. 2000).

Another method to trace the evolution of the cluster number density is based on the XLF. The advantage of using X-ray luminosity as a tracer of the mass is that L_X is measured for a much larger number of clusters within samples with well-defined selection properties. As discussed in Section 3, the most recent flux-limited cluster samples contain now a large (~ 100) number of objects, which are homogeneously identified over a broad redshift baseline, out to $z \simeq 1.3$. This allows nearby and distant clusters to be compared within the same sample, i.e. with a single selection function. The potential disadvantage of this method is that it relies on the relation between L_X and M_{vir} , which is based on additional physical assumptions and hence is more uncertain than the $M_{\text{vir}} - \sigma_v$ or the $M_{\text{vir}} - T$ relations.

A useful parameterization for the relation between temperature and bolometric luminosity is

$$L_{\text{bol}} = L_6 \left(\frac{T_X}{6 \text{ keV}} \right)^\alpha (1+z)^A \left(\frac{d_L(z)}{d_{L, \text{EdS}}(z)} \right)^2 10^{44} h^{-2} \text{ erg s}^{-1}, \quad (14)$$

with $d_L(z)$ the luminosity-distance at redshift z for a given cosmology. Several independent analyses of nearby clusters with $T_X \gtrsim 2$ keV consistently show that $L_6 \simeq 3$ is a stable result and $\alpha \simeq 2.5-3$ (e.g. White et al. 1997, Wu et al. 1999, and references therein). For cooler groups, $\lesssim 1$ keV, the $L_{\text{bol}} - T_X$ relation steepens, with a slope $\alpha \sim 5$ (e.g. Helsdon & Ponman 2000).

The redshift evolution of the $L_X - T$ relation was first studied by Mushotzky & Scharf (1997) who found that data out to $z \simeq 0.4$ are consistent with no evolution for an Einstein-de-Sitter model (i.e., $A \simeq 0$). This result was extended to higher redshifts using cluster temperatures out to $z \simeq 0.8$ as measured with

ASCA and *Beppo-SAX* data (Donahue et al. 1999, Della Ceca et al. 2000, Henry 2000). The lack of a significant evolution seems to hold beyond $z = 1$ according to recent *Chandra* observations of very distant clusters (Borgani et. al. 2001b, Stanford et al. 2001, Holden et al. 2002), as well as *Newton-XMM* observations in the Lockman Hole (Hashimoto et al. 2002). Figure 14 shows a summary of the observational results on the L_X-T . The high redshift points generally lie around the local relation, thus demonstrating that it is reasonable to assume $A \lesssim 1$ implying at most a mild positive evolution of the $L_{bol}-T_X$ relation. Besides the relevance for the evolution of the mass–luminosity relation, these results also have profound implications for the physics of the ICM (see Section 2).

Kitayama & Suto (1997) and Mathiesen & Evrard (1998) analyzed the number counts from different X-ray flux-limited cluster surveys (Figure 7) and found that resulting constraints on Ω_m are rather sensitive to the evolution of the mass–luminosity relation. Sadat et al. (1998) and Reichart et al. (1999) analyzed the EMSS and found results to be consistent with $\Omega_m = 1$. Borgani et al. (2001b) analyzed the RDCS sample to quantify the systematics in the determination of cosmological parameters induced by the uncertainty in the mass–luminosity relation (Borgani et al. 1998). They found $0.1 \lesssim \Omega_m \lesssim 0.6$ at the 3σ confidence level, by allowing the $M-L_X$ relation to change within both the observational and the theoretical uncertainties. In Figure 15 we show the effect of changing in different ways the parameters defining the $M-L_X$ relation, such as the slope α and the evolution A of the L_X-T relation (see Equation 14), the normalization β of the $M-T$ relation (see Equation 13), and the overall scatter Δ_{M-L_X} . We assume flat geometry here, i.e. $\Omega_m + \Omega_\Lambda = 1$. In general, constraints of cosmological models based on cluster abundance are not very sensitive to Ω_Λ (see Figure 12). To a first approximation, the best fit Ω_m has a slight dependence on Ω_Λ for open geometry: $\Omega_m \simeq \Omega_{m,fl} + 0.1(1 - \Omega_{m,fl} - \Omega_\Lambda)$, where $\Omega_{m,fl}$ is the best fit value for flat geometry.

Constraints on Ω_m from the evolution of the cluster population, like those shown in Figures 13 and 15, are in line with the completely independent constraints derived from the baryon fraction in clusters, f_{bar} , which can be measured with X-ray observations. If the baryon density parameter, Ω_{bar} , is known from independent considerations (e.g. by combining the observed deuterium abundance in high-redshift absorption systems with predictions from primordial nucleosynthesis), then the cosmic density parameter can be estimated as $\Omega_m = \Omega_{\text{bar}}/f_{\text{bar}}$ (e.g. White et al. 1993b). For a value of the Hubble parameter $h \simeq 0.7$, this method yields $f_{\text{bar}} \simeq 0.15$ (e.g. Evrard 1997; Ettori 2001). Values of f_{bar} in this range are consistent with $\Omega_m = 0.3$ for the currently most favored values of the baryon density parameter, $\Omega_{\text{bar}} \simeq 0.02 h^{-2}$, as implied by primordial nucleosynthesis (e.g. Burles & Tytler 1998) and by the spectrum of CMB anisotropies (e.g. de Bernardis et al. 2001, Stompor et al. 2001, Pryke et al. 2002).

Figure 15 demonstrates that firm conclusions about the value of the matter density parameter Ω_m can be drawn from available samples of X-ray clusters. In keeping with most of the analyses in the literature, based on independent methods, a critical density model cannot be reconciled with data. Specifically, $\Omega_m < 0.5$ at 3σ level even within the full range of current uncertainties in the relation between mass and X-ray luminosity.

A more delicate issue is whether one can use the evolution of galaxy clusters for high-precision cosmology, e.g., $\lesssim 10\%$ accuracy. Serendipitous searches of distant clusters from XMM and Chandra data will eventually lead to a significant

increase of the number of high- z clusters with measured temperatures. Thus, the main limitation will lie in systematics involved in comparing the mass inferred from observations with that given by theoretical models. A point of concern, for example, is that constraints on σ_8 from different analyses of the cluster abundance differ by up to 30% from each other. While a number of previous studies found $\sigma_8 \simeq 0.9\text{--}1$ for $\Omega_m = 0.3$ (e.g. Pierpaoli et al. 2001 and references therein), the most recent analyses point toward a low power spectrum normalization, $\sigma_8 \simeq 0.7$ for $\Omega_m = 0.3$ (Borgani et al. 2001b, Reiprich & Böhringer 2002, Seljak 2002, Viana et al. 2002).

A thorough discussion of the reasons for such differences would require an extensive and fairly technical review of the analysis methods applied so far. For instance, a delicate point concerns the different recipes adopted for the mass–temperature and mass–luminosity conversions. The M – T relation, usually measured at some fixed overdensity from observational data, seems to have a lower normalization than that calibrated from hydrodynamical simulations (e.g. Finoguenov et al. 2001, Allen et al. 2001, Ettori et al. 2002). In turn, this provides a lower amplitude for the mass function implied by an observed XTF and, therefore, a smaller σ_8 . Several uncertainties also affect the L_X – T relation. The derived slope depends on the temperature range over which the fit is performed. We are also far from understanding the nature of its scatter, i.e. how much it is due to systematics, and how much it is intrinsic, inherent to complex physical conditions in the gas. For example, the contribution of cooling flows is known to increase the scatter in the L_X – T relation (e.g. Markevitch 1998, Allen & Fabian 1998, Arnaud & Evrard 1999). Adding such a scatter in the mass–luminosity conversion increases the amplitude of the mass–function, especially in the high-mass tail, thus decreasing the required σ_8 .

As an illustrative example, we show in Figure 15 how constraints in the σ_8 – Ω_m plane move as we change the scatter and the amplitude of the M – L_X relation in the analysis of the RDCS. The upper left panel shows the result for the same choice of parameters as in the original analysis by Borgani et al. (2001b), which gives $\sigma_8 \simeq 0.7$ for $\Omega_m = 0.3$. The central lower panel shows the effect of decreasing the scatter of the M – L_X relation by 20%, in keeping with the analysis by Reiprich & Böhringer (2002, see also Ettori et al. 2002). Such a reduced scatter causes σ_8 to increase by about 20%. Finally, if the normalization of the M – T relation is decreased by $\sim 30\%$ with respect to the value suggested by hydrodynamical cluster simulations (lower right panel), σ_8 is again decreased by $\sim 20\%$.

In light of this discussion, a 10% precision in the determination of fundamental cosmological parameters, such as Ω_m and σ_8 lies in the future. With forthcoming datasets the challenge will be in comparing observed clusters with the theoretical clusters predicted by Press-Schechter-like analytical approaches or generated by numerical simulations of cosmic structure formation.

6 OUTLOOK AND FUTURE WORK

Considerable observational progress has been made in tracing the evolution of global physical properties of galaxy clusters as revealed by X-ray observations. The *ROSAT* satellite has significantly contributed to providing the statistical samples necessary to compute the space density of clusters in the local Universe and its evolution. A great deal of optical spectroscopic studies of these samples

has consolidated the evidence that the bulk of the cluster population has not evolved significantly since $z \sim 1$. However, the most X-ray luminous, massive systems do evolve. Similarly, the thermodynamical properties of clusters as indicated by statistical correlations, such as the $L_X - T_X$ relation, do not show any strong evolution. Moreover, the *Chandra* satellite has given us the first view of the gas distribution in clusters at $z > 1$; their X-ray morphologies and temperatures show that they are already in an advanced stage of formation at these large lookback times.

These observations can be understood in the framework of hierarchical formation of cosmic structures, with a low density parameter, $\Omega_m \sim 1/3$, dominated by cold dark matter: structure formation started at early cosmic epochs and a sizable population of massive clusters was in place already at redshifts of unity. In addition, detailed X-ray observation of the intra-cluster gas show that the physics of the ICM needs to be regulated by additional non-gravitational processes.

With *Chandra* and *Newton-XMM*, we now realize that physical processes in the ICM are rather complex. Our physical models and numerical simulations are challenged to explain the new level of spatial details in the density and temperature distribution of the gas, and the interplay between heating and cooling mechanisms. Such complexities need to be well understood physically before we can use clusters as high-precision cosmological tools, particularly at the beginning of an era in which cosmological parameters can be derived rather accurately by combining methods that measure the global geometry of the Universe (the CMB spectrum, type Ia Supernovae (e.g. Leibungut 2001)), and the large-scale distribution of galaxies (e.g. Peacock et al. 2001). It remains remarkable that the evolution of the cluster abundance, the CMB fluctuations, the type Ia Supernovae and large scale structure – all completely independent methods – converge toward $\Omega_m \simeq 0.3$ in a spatially flat Universe ($\Omega_m + \Omega_\Lambda = 1$). Further studies with the current new X-ray facilities will help considerably in addressing the issue of systematics discussed above, although some details of the ICM in $z \gtrsim 1$ clusters, such as temperature profiles or metallicity, will remain out of reach until the next generation of X-ray telescopes. Direct measurements of cluster masses at $z \gtrsim 1$ via gravitational lensing techniques will soon be possible with the *Advanced Camera for Surveys* (Ford et al. 1998) on-board the *Hubble Space Telescope*, which offers an unprecedented combination of sensitivity, angular resolution and field of view.

The fundamental question remains as to the mode and epoch of formation of the ICM. When and how was the gas pre-heated and polluted with metals? What is the epoch when the first X-ray clusters formed, i.e. the epoch when the accreted gas thermalizes to the point at which they would lie on the $L_X - T$ relation (Figure 14)? Are the prominent concentrations of star forming galaxies discovered at redshift $z \sim 3$ (Steidel et al. 1998) the progenitors of the X-ray clusters we observed at $z \lesssim 1$? If so, cluster formation should have occurred in the redshift range 1.5–2.5. Although the redshift boundary for X-ray clusters has receded from $z = 0.8$ to $z = 1.3$ recently, a census of clusters at $z \simeq 1$ has just begun and the search for clusters at $z > 1.3$ remains a serious observational challenge. Using high- z radio galaxies as signposts for proto-clusters has been the only viable method so far to break this redshift barrier. These searches have also lead to the discovery of extended $Ly\alpha$ nebulae around distant radio galaxies (e.g., Venemans et al. 2002), very similar to those discovered by Steidel et al. (2000) in correspondence with large scale structures at $z \simeq 3$. The nature of

such nebulae is still not completely understood, however they could represent the early phase of collapse of cool gas through mergers and cooling flows.

In this review we have not treated the formation and evolution of the galaxies in clusters. This must be linked to the evolution of the ICM and the fact that we are still treating the two aspects as separate points to the difficulty in drawing a comprehensive unified picture of the history of cosmic baryons in their cold and hot phase. Multiwavelength studies are undoubtedly essential to reach such a unified picture. When surveys exploiting the Sunyaev–Zeldovich effect (e.g. Carlstrom et al. 2001) over large solid angles become available, one will be able to observe very large volumes at $z > 1$. In combination with a deep large area X-ray survey (e.g. Wide Field X-ray Telescope, Burrows et al. 1992) and an equivalent deep near-IR survey (e.g. the Primordial Explorer (PRIME), Zheng et al. 2002), this could reveal the evolutionary trends in a number of independent physical parameters, including: the cluster mass, the gas density and temperature, the underlying galactic mass and star formation rates. Advances in instrumentation and observational technique will make this approach possible and will provide vital input for models of structure formation and tight constraints on the underlying cosmological parameters.

ACKNOWLEDGMENTS

We acknowledge useful discussions with Hans Böhringer, Alfonso Cavaliere, Guido Chincarini, Roberto Della Ceca, Stefano Ettori, Gus Evrard, Isabella Gioia, Luigi Guzzo, Brad Holden, Silvano Molendi, Chris Mullis, and Adam Stanford. We thank Paolo Tozzi for his help in producing Figure 10. PR thanks Riccardo Giacconi for continuous encouragement of this work. PR is grateful for the hospitality of the Astronomical Observatory of Trieste. SB and CN acknowledge the hospitality and support of ESO in Garching.

Literature Cited

1. Abell GO. 1958. *Ap. J. Suppl.* 3:211–278.
2. Abell GO, Corwin HG Jr, Olowin RP. 1989. *Ap. J. Suppl.* 70:1–138
3. Allen SW, Fabian AC. 1998. *MNRAS* 297:L57–62
4. Allen SW, Schmidt RW, Fabian AC. 2001. *MNRAS* 328:L37–41
5. Allen SW, Schmidt RW, Fabian AC. 2002. *MNRAS* in press. Preprint astro-ph/0111368
6. Arnaud M, Evrard CE. 1999. *MNRAS* 305:631–640
7. Bahcall NA. 1988. *Annu. Rev. Astron. Astrophys.* 26:631–686
8. Bahcall NA, Cen R. 1993. *Ap. J. Lett.* 407:L49–52
9. Bahcall NA, Fan X, Cen R. 1997. *Ap. J. Lett.* 485:L53–56
10. Bahcall NA, Ostriker JP, Perlmutter S, Steinhardt PJ. 2000. *Science*, 284:1481–88
11. Balogh ML, Babul A, Patton DR. 1999. *MNRAS* 307:463–479
12. Balogh, ML, Pearce FR, Bower RG, Kay ST 2001. *MNRAS* 326:1228–1234
13. Bardeen JM, Bond JR, Kaiser N, Szalay AS. 1986. *Ap. J.* 304:15–61
14. Bauer FE, Alexander DM, Brandt WN, Hornschemeier AE, Miyaji T, Garmire DP. 2002. *Astron. J.* 123:1163–1178
15. Benítez N, Broadhurst T, Rosati P, Courbin F, Squires G, Lidman C, Magain P. 1999. *Ap. J.* 527:31–41
16. de Bernardis P, Ade PAR, Bock JJ, Bond JR, Borrill J. 2001. *Ap. J.* 561:13–21
17. Bialek JJ, Evrard AE, Mohr JJ. 2001. *Ap. J.* 555:597–612
18. Birkinshaw M. 1999. *Phys. Rep.* 310:97–195
19. Blanchard A, Sadat R, Bartlett JG, Le Dour M. 2000. *Astron. Astrophys.* 362:809–824
20. Blanton MR, Dalcanton J, Eisenstein D, Loveday J, Strauss MA, et al. 2001. *Astron. J.* 121:2358–2380

21. Böhringer H, Voges W, Huchra JP, McLean B, Giacconi R, et al. 2000. *Ap. J. Suppl.* 129:435–474
22. Böhringer H, Schuecker P, Guzzo L, Collins CA, Voges W, et al. 2001. *Astron. Astrophys.* 369:826–50
23. Böhringer H, Matsushita K, Churazov E, Ikebe Y, Chen Y. 2002a. *Astron. Astrophys.* 382:804–20
24. Böhringer H, Collins CA, Guzzo L, Schuecker P, Voges W, et al. 2002b. *Ap. J.* 566:93–102
25. Borgani S, Girardi M, Carlberg RG, Yee HKC, Ellingson E. 1999. *Ap. J.* 527: 561–572
26. Borgani S, Guzzo L. 2001. *Nature* 409:39–45
27. Borgani S, Rosati P, Tozzi P, Norman C. 1998. *Ap. J.* 517:40–53
28. Borgani S, Governato F, Wadsley J, Menci N, Tozzi P, et al. 2001a, *Ap. J. Lett.* 559:L71–74
29. Borgani S, Rosati P, Tozzi P, Stanford, Eisenhardt PE, et al. 2001b. *Ap. J.* 561:13–21
30. Bower RG, Benson AJ, Bough CL, Cole S, Frenk CS, Lacey CG. 2001. *MNRAS* 325:497–508
31. Brighenti F, Mathews WG. 2001. *ApJ* 553:103–120
32. Bryan GL 2000. *Ap. J. Lett.* 544:L1–L4
33. Bryan GK, Norman ML. 1998. *Ap. J.* 495:80–99
34. Burns JO, Ledlow MJ, Loken C, Klypin A, Voges W, et al. 1996. *Ap. J. Lett.* 467:L49–52
35. Burke DJ, Collins CA, Sharples RM, Romer AK, Holden, BP, Nichol, RC, et al. 1997. *Ap. J. Lett.* 488:L83–86
36. Burles S, Tytler D. 1998. *Space Sc. Rev.* 84(1/2):65–75
37. Burns JO, Ledlow MJ, Loken C, Klypin A, Voges W., et al. 1996. *Ap. J. Lett.* 467:L49–52
38. Burrows CJ, Burg R, Giacconi R. 1992. *Ap. J.* 392:760–765
39. Campana S, Lazzati D, Panzera MR, Tagliaferri G. 1999. *Ap. J.* 524:423–433
40. Carlberg RG, Yee HKC, Ellingson E. 1997a. *Ap. J.* 478:462–475
41. Carlberg RG, Morris SL, Yee HKC, Ellingson E, 1997b. *Ap. J. Lett.* 479:L19–23
42. Carlstrom JE, Joy M, Grego L, Holder G, Holzapfel WL, et al. 2001. In *Constructing the Universe with Clusters of Galaxies*, eds. F. Durret and G. Gerbal. Preprint astro-ph/0103480
43. Carroll SM, Press WH, Turner EL. 1992. *Annu. Rev. Astron. Astrophys.* 30:499–542
44. Castander FJ, Bower RG, Ellis RS, Aragon-Salamanca A, Mason KO, et al. 1995. *Nature* 377:39–41
45. Cavaliere A, Fusco-Femiano R. 1976. *Astron. Astrophys.* 49:137–144
46. Cavaliere A, Gursky H, Tucker WH. 1971. *Nature* 231:437–438
47. Cavaliere A, Menci N, Tozzi P. 1998. *Ap. J.* 501:493–508
48. Chartas G, Bautz M, Garmire G, Jones C, Schneider DP. 2001. *Ap. J. Lett.* 550:L163–166
49. Colafrancesco S, Mazzotta P, Vittorio N. 1997. *ApJ* 488:566–571
50. Coles P, Lucchin F. 1995. *Cosmology. The origin and evolution of cosmic structure*, Chichester, Wiley
51. Collins CA, Burke DJ, Romer AK, Sharples RM, Nichol RC. 1997. *Ap. J. Lett.* 479:L117–120
52. Collins CA, Guzzo L, Böhringer H, Schücker P, Chincarini G, et al. 2000. *MNRAS* 319:939–948
53. Couch WJ, Ellis RS, MacLaren I, Malin DF. 1991. *MNRAS* 249: 606–628
54. Crawford CS, Edge AC, Fabian AC, Allen SW, Böhringer H, et al. 1995. *MNRAS* 274:75–84
55. Crawford CS, Fabian AC. 1996. *MNRAS* 282:1483–1488
56. Dalcanton JJ. 1996. *Ap. J.* 466:92–103
57. Dalton GB, Maddox SJ, Sutherland WJ, Efstathiou G. 1997. *MNRAS* 289:263–284
58. De Grandi S, Böhringer H, Guzzo L, Molendi S, Chincarini G, et al. 1999. *Ap. J.* 514:148–163
59. De Grandi S, Molendi S. 2001 *Ap. J.* 551:153–159
60. De Grandi S, Molendi S. 2002 *Ap. J.* 567:163–177
61. Della Ceca R, Scaramella R, Gioia IM, Rosati P, Fiore F, Squires G. 2000. *Astron. Astrophys.* 353:498–506
62. de Propris R, the 2dFGRS team 2002. *MNRAS* 329:87–101
63. Dickinson M, 1997, in *The Early Universe with the VLT*, ed. J. Bergeron, (Springer: Berlin) p.274
64. Donahue M, Voit GM, Scharf CA, Gioia IM, Mullis CR, et al. 1999. *Ap. J.* 527:525–534
65. Donahue M, Voit GM. 1999. *Ap. J. Lett.* 523:L137–140
66. Donahue M, Mack J, Scharf C, Lee P, Postman M, et al. 2001. *Ap. J. Lett.* 552:L93–96
67. Ebeling H, Voges W, Böhringer H, Edge AC, Huchra JP, Briel UG. 1996. *MNRAS* 281:799–829
68. Ebeling H, Edge AC, Fabian AC, Allen SW, Crawford CS, Böhringer H. 1997. *Ap. J. Lett.*

- 479:L101–104
69. Ebeling H, Edge AC, Bohringer H, Allen SW, Crawford CS, et al. 1998. *MNRAS* 301:881–914
 70. Ebeling H, Jones LR, Perlman E, Scharf C, Horner D, et al. 2000. *Ap. J.* 534:133–145
 71. Ebeling H, Edge AC, Allen SW, Crawford CS, Fabian AC, Huchra JP. 2000, *MNRAS* 318:333–340
 72. Ebeling A, Edge AC, Henry JP. 2001. *Ap. J.* 553:668–676
 73. Edge AC, Stewart GC, Fabian AC, Arnaud KA, 1990. *MNRAS* 245:559–569
 74. Eke VR, Cole S, Frenk CS. 1996. *MNRAS* 282:263–280
 75. Eke VR, Cole S, Frenk CS, Henry JP. 1998. *MNRAS*, 298:1145–1158
 76. Eisenstein DJ, Hu W. 1999. *Ap. J.* 511:5–15
 77. Ettori S. 2001. *MNRAS* 323:L1–L5
 78. Ettori S, De Grandi S, Molendi S. 2002. *Astron. Astrophys.* in press. Preprint astro-ph/0206120
 79. Evrard AE. 1997. *MNRAS* 292:289–297
 80. Evrard AE, Henry JP. 1991. *Ap. J.* 383:95–103
 81. Evrard, AE, MacFarland TJ, Couchmam HMP, Colberg JM, Yoshida N, et al. 2002. *Ap. J.* 573:7–36
 82. Evrard AE, Metzler CR, Navarro JF. 1996. *Ap. J.* 469:494–507
 83. Fabian AC. 1994. *Annu. Rev. Astron. Astrophys.* 32:277–318
 84. Fabian AC, Sanders JS, Ettori S, Taylor GB, Allen SW, et al. 2000. *MNRAS* 318:L65–68
 85. Fabian AC, Mushotzky RF, Nulsen PEJ, Peterson JR. 2001. *MNRAS* 321:L20–L24
 86. Fabian AC, Crawford CS, Ettori S, Sanders JS. 2001. *MNRAS* 332:L11–L15
 87. Felten JE, Gould RJ, Stein WA, Woolf NJ. 1966. *Ap. J.* 146:955–958
 88. Finoguenov A, David LP, Ponman TJ. 2000. *Ap. J.* 544:188–203
 89. Finoguenov A, Reiprich TH, Böhringer H. 2001. *Astron. Astrophys.* 368:749–759
 90. Ford HC and the ACS Science Team. 1998. In *Space Telescopes and Instruments V*, eds. PY Bely, JB Breckinridge, Proc. SPIE, 3356, 234.
 91. Forman W, Jones C. 1982. *Annu. Rev. Astron. Astrophys.* 20:547–585
 92. Francis PJ, Woodgate BE, Warren SJ, Moller P, Mazzolini M, et al. 1996. *Ap. J.* 457:490–499
 93. Frenk CS, White SDM, Bode P, Bond JR, Bryan GL, et al. 2000. *ApJ* 525:554–582
 94. Giacconi R, Murray S, Gursky H, Kellogg E, Schreier E, Tananbaum H. 1972. *Ap. J.* 178:281–308
 95. Giacconi R, Branduardi G, Briel U, Epstein A, Fabricant D, et al. 1979. *Ap. J.* 230:540–550
 96. Giacconi R, Zirm A, JunXian W, Rosati P, Nonino M, et al. 2002. *Ap. J. Suppl.* 139:369–410
 97. Gioia IM, Henry JP, Maccacaro T, Morris SL, Stocke JT, Wolter A. 1990a. *Ap. J. Lett.* 356:L35–38
 98. Gioia IM, Maccacaro T, Schild RE, Wolter A, Stocke JT, et al. 1990b. *Ap. J. Suppl.* 72:567–619
 99. Gioia IM, Luppino GA. 1994. *Ap. J. Suppl.* 94:583–614
 100. Gioia IM, Henry JP, Mullis CR, Ebeling H, Wolter A. 1999. *Astron. J.* 117:2608–2616
 101. Gioia IM, Henry JP, Mullis CR, Voges W, Briel UG. 2001. *Ap. J. Lett.* 553:L109–112
 102. Girardi M, Borgani S, Giuricin G, Mardirossian F, Mezzetti M. 1998. *Ap. J.* 506:45–52
 103. Girardi M, Mezzetti M. 2001. *Ap. J.* 540:79–96
 104. Gladders MD, Yee HKC. 2000. *Astron. J.* 120:2148–2162
 105. Gonzales AH, Zaritsky D, Dalcanton JJ, Nelson A. 2001. *Ap. J. Suppl.* 137:117–138
 106. Governato F, Babul A, Quinn T, Tozzi P, Baugh CM, et al. 1999. *MNRAS* 307:949–966
 107. Gursky H, Kellogg E, Murray S, Leong C, Tananbaum H, Giacconi, R. 1971. *Ap. J.* 167:L81–84
 108. Gross MAK, Somerville RS, Primack JR, Holtzman J, Klypin, A. 1998. *MNRAS* 301:81–94
 109. Gunn JE, Hoessel JG, Oke JB. 1986. *Ap. J.* 306:30–37
 110. Hashimoto Y, Hasinger G, Aranud M, Rosati P, Miyaji T. 2002. *Astron. Astrophys.* 381:841–847
 111. Hattori M, Matuzawa H, Morikawa K, Kneib J-P, Yamashita K, et al. 1998. *Ap. J.* 503:593–598
 112. Hall PB, Green RF. 1998. *Ap. J.* 558:558–584
 113. Hasinger G, Altieri B, Arnaud M, Barcons X, Bergeron J, et al. 2001. *Astron. Astrophys.* 365:L45–L50
 114. Helsdon SF, Ponman TJ. 2000. *MNRAS* 315:356–370

115. Henriksen MJ, Mushotzky RF. 1986. *Ap. J.* 302:287–295
116. Henry JP. 2000. *Ap. J.* 534:565–580
117. Henry JP, Arnaud KA. 1991. *Ap. J.* 372:410–418
118. Henry JP, Gioia IM, Maccacaro T, Morris SL, Stocke JT, Wolter A. 1992. *Ap. J.* 386:408–419
119. Henry JP, Gioia IM, Mullis CR, Voges W, Briel UG, et al. 2001. *Ap. J. Lett.* 553:L109–112
120. Holden BP, Nichol RC, Romer AK, Metevier A, Postman M, et al. 1999. *Astron. J.* 118:2002–2013
121. Holden B, Stanford SA, Squires GK, Rosati P, Tozzi P, et al. 2002. *Astron. J.*, 124:35–45
122. Kaiser N. 1986. *MNRAS* 222:323–345
123. Kellogg E, Gursky H, Leong C, Schreier E, Tananbaum H, Giacconi R. 1971. *Ap. J.* 165:L49–54
124. Kim RSJ, Kepner JV, Postman M, Strauss MA, Bahcall NA, et al. 2002. *Astron. J.* 123:20–36
125. Kitayama T, Suto Y. 1997. *Ap. J.* 490:557–563
126. Kofman LA, Gnedin NJ, Bahcall NA. 1993. *Ap. J.* 413:1–19
127. Kolb KT, Turner MS. 1989. *The Early Universe* Addison–Wesley Publ.
128. Kravtsov AV, Yepes G. 2000. *MNRAS* 318:227–238
129. Irwin JA, Bregman JN. 2000. *Ap. J.* 538:543–554
130. Jeltema TE, Canizares CR, Bautz MW, Malm MR, Donahue M, Garmire GP. 2001. *Ap. J.* 562:124–132
131. Jenkins A, Frenk CS, White SDM, Colberg J, Cole S, et al. 2001. *MNRAS* 321:372–384
132. Jones LR, Scharf C, Ebeling H, Perlman E, Wegner G, et al. 1998. *Ap. J.*, 495:100–114
133. Jones LR, Ebeling H, Scharf C, Perlman E, Horner D, et al. 2000. in *Constructing the Universe with clusters of Galaxies*, eds Durret F & Gerbal D (CD-rom, website)
134. Kristian J, Sandage A, Westphal JA. 1978. *Ap. J.* 221:383–394
135. Lazzati D, Campana S, Rosati P, Panzera MR, Tagliaferri G. 1999. *Ap. J.* 524:414–422
136. Leibungut B. 2002. *Annu. Rev. Astron. Astrophys.* 39:67–98
137. Lewis AD, Ellingson E, Morris SL, Carlberg RG. 1999. *Ap. J.* 517:587–608
138. Lidman EL, Peterson BA. 1996. *Ap. J.* 112:2454–2470
139. Lloyd-Davies EJ, Ponman TJ, Cannon DB. 2000. *MNRAS* 315:689–702
140. Lubin LM, Brunner R, Metzger MR, Postman M, Oke JB. 2000. *Ap. J.* 531:L5–8
141. Lumsden SL, Nichol RC, Collins CA, Guzzo L. 1992. *MNRAS* 258:1–22
142. Maddox SJ, Efstathiou G, Sutherland WJ, Loveday J. 1990. *MNRAS* 242:43p–47p
143. Markevitch M. 1998. *Ap. J.* 504:27–34
144. Markevith M, Forman WR, Sarazin CL, Vikhlinin A. 1998. *Ap. J.* 503:77–96
145. Markevitch M, Ponman TJ, Nulsen PEJ, Bautz MW, Burke DJ, et al. 2000. *Ap. J.* 541:542–549
146. Mathiesen B, Evrard AE. 1998. *MNRAS* 295:769–780
147. Mazure A, Katgert P, den Hartog R, Biviano A, Dubath P., et al. 2001. *Astron. Astrophys.* 310:31–48
148. McHardy IM, Lawrence A, Pye JP, Pounds KA. 1981. *Ap. J.* 197:893–919
149. McKee JD, Mushotzky RF, Boldt EA, Holt SS, Marshall FE, et al. 1980. *Ap. J.* 242:843–856
150. Mellier Y. 1999. *Annu. Rev. Astron. Astrophys.* 37:127–189
151. Menci N, Cavaliere A. 2000. *MNRAS* 311:50–62
152. Monaco P. 1998. *Fund. Cosm. Phys.* 19:157–317
153. Muanwong O, Thomas PA, Kay ST, Pearce FR, Couchman HMP. 2001. *Ap. J.* 552:L27–30
154. Mulchaey JS. 2000. *Annu. Rev. Astron. Astrophys.* 38:289–335
155. Mullis CR, Henry JP, Gioia IM, Böhringer H, Briel UG, et al. 2001. *Ap. J. Lett.* 553:L115–118
156. Mushotzky RF, Scharf CA. 1997. *Ap. J. Lett.* 482:L13–16
157. Navarro JF, Frenk CS, White SDM. 1995. *MNRAS* 275:720–740
158. Nichol RC, Romer AK, Holden BP, Ulmer MP, Pildis RA, et al. 1999 *Ap. J.* 521:L21–L24
159. Olsen LF, Scoggio M, da Costa LN, Slijkhuis R, Benoist C, et al. 1999. *Astron. Astrophys.* 345:363–368
160. Oukbir J, Blanchard A 1992. *Astron. Astrophys.* 262:L21–24
161. Page MJ, Carrera FJ. 2000. *MNRAS* 311:433–440
162. Pascarelle SM, Windhorst RA, Drivers SP, Ostrander EJ, Keel WC. 1996. *Ap. J. Lett.*

- 456:L21–24
163. Pearce FR, Thomas PA, Couchman HMP, Edge AC. 2000. *MNRAS* 317:1029–1040
 164. Peacock JA. 1999. *Cosmological Physics*, Cambridge University Press, Cambridge
 165. Peacock JA, Cole S, Norberg P, Baugh CM, Bland-Hawthorn J, et al. 2001. *Nature* 4120:169–173
 166. Peebles PJE. 1993. *Physical Cosmology* Princeton, NJ: Princeton Univ. Press
 167. Pentericci L, Kurk JD, Röttgering HJA, Miley GK, van Breugel W, et al. 2000. *Astron. Astrophys.* 361:L25–28
 168. Perlman ES, Horner DJ, Jones LR, Scharf CA, Ebeling H, et al. 2002 *Ap. J. Suppl.* 140:265–301
 169. Peterson JR, Paerels FBS, Kaastra JS, Arnaud M, Reiprich TH, et al. 2001. *Astron. Astrophys.* 365:L104–L109
 170. Piccinotti G, Mushotzky RF, Boldt EA, Holt SS, Marshall FE, et al. 1982. *Ap. J.* 253:485–503
 171. Pierpaoli E, Scott D, White M. 2001. *MNRAS* 325:77–88
 172. Pipino A, Matteucci F, Borgani S, Biviano A. 2002. *New Astron.* 7:227–247
 173. Ponman TJ, Cannon DB, Navarro JF. 1999. *Nature* 397:135–137
 174. Postman M, Lubin LM, Gunn JE, Oke JB, Hoessel JG, Schneider DP, Christensen JA. 1996. *Astron. J.* 111:615–641
 175. Press WH, Schechter P. 1974. *Ap. J.* 187:425–438
 176. Pryke C, Halverson NW, Leitch EM, Kovac J, Carlstrom JE, et al. 2002 *Ap. J.* 568:46–51
 177. Raymond JC, Smith BW. 1977. *Ap. J. Suppl.* 35:419–439
 178. Reichart DE, Nichol RC, Castander FJ, Burke DJ, Romer AK, et al. 1999. *Ap. J.* 518:521–532
 179. Reiprich TH, Böhringer H. 2002 *Ap. J.* 567:716–740
 180. Renzini A. 1997. *Ap. J.* 488:35–43
 181. Romer AK, Nichol RC, Holden BP, Ulmer MP, Pildis RA, et al. 2000. *Ap. J. Suppl.*, 126:209–269
 182. Rosati P, Borgani S, Della Ceca R, Stanford SA, Eisenhardt PR, Lidman C. 2000. In *Large Scale Structure in the X-ray Universe*, ed M Plionis, I Georgantopoulos, p.13. Paris, France: Atlantisciences
 183. Rosati P, Della Ceca R, Burg R, Norman C, Giacconi R. 1995. *Ap. J. Lett.* 445:L11–14
 184. Rosati P, Della Ceca R, Burg R, Norman C, Giacconi R. 1998. *Ap. J. Lett.* 492:L21–24
 185. Rosati P, Stanford SA, Eisenhardt PR, Elston R, Spinrad H, et al. 1999. *Astron. J.* 118:76–85
 186. Rothschild R, et al. 1979 *Space Sci. Instr.* 4:265
 187. Sadat R, Blanchard A, Oukbir J. 1998. *Astron. Astrophys.* 329:21–29
 188. Sarazin C. 1988. *X-Ray Emission from Clusters of Galaxies*. Cambridge: Cambridge University Press
 189. Scharf CA, Jones LR, Ebeling H, Perlman E, Malkan M, et al. 1997. *Ap. J.* 477:79–92
 190. Schücker P, Böhringer H, Guzzo L, Collins CA, Neumann DM, et al. 2001. *Astron. Astrophys.* 368:86–106
 191. Schücker P, Böhringer H, Reiprich TH, Feretti L. 2001. *Astron. Astrophys.* 378:408–427
 192. Seljak U. 2002. *MNRAS* submitted. Preprint astro-ph/0111362
 193. Sheth RK, Tormen G. 1999. *MNRAS* 308:119–126
 194. Sheth RK, Mo HJ, Tormen G. 2001. *MNRAS* 323:1–12
 195. Smith S. 1936. *Ap. J.* 83:23–30
 196. Stanford SA, Elston R, Eisenhardt PR, Spinrad H, Stern D, Dey A. 1997. *Astron. J.* 114:2232–2239
 197. Stanford SA, Holden B, Rosati P, Tozzi P, Borgani S, et al. 2001. *Ap. J.* 552:504–507
 198. Stanford SA, Holden BP, Rosati P, Eisenhardt PR, Stern D, et al. 2002. *Astron. J.* 123:619–626
 199. Steidel CC, Adelberger KL, Dickinson M, Giavalisco M, Pettini M, Kellogg M. 1998. *Ap. J.* 492:428–438
 200. Steidel CC, Adelberger KL, Shapley AE, Pettini M, Dickinson M, Giavalisco M. 2000. *Ap. J.* 532:170–182
 201. Stompor S, Abroe M, Ade P, Balbi A, Barbosa D, et al. 2001. *Ap. J. Lett.* 561:L7–10
 202. Soucail G, Kneib J-P, Jaunsen AO, Hjorth J, Hattori M, Yamada T. 2001. *Astron. Astrophys.* 367:741–747

203. Tamura T, Kaastra JS, Peterson JR, Paerels FBS, Mittaz JPD, et al. 2001. *Astron. Astrophys.* 365:L87–L92
204. Tozzi P, Norman C. 2001. *Ap. J.* 546:63–84
205. Tozzi P, Scharf C, Norman C. 2001. *Ap. J.* 542:106–119
206. Trümper J. 1993. *Science* 260:1769–1771
207. Valageas P., Silk J. 1999. *Astron. Astrophys.* 350:725–742
208. van Dokkum PG, Franx M. 2001 *Ap. J.* 553:90–102
209. van Haarlem MP, Frenk CS, White SDM. 1997. *MNRAS* 287:817–832
210. van Waerbeke L, Mellier Y, Radovich M, Bertin E, Dantel–Fort M, et al. 2001. *Astron. Astrophys.* 374:757–769
211. Venemans BP, Kurk JD, Miley GK, Röttgering HJA, van Breugel W, et al. 2002 *Ap. J.* 569:L11–L14
212. Viana PTP, Liddle AR. 1999. *MNRAS* 303:535–545
213. Viana PTP, Nichol RC, Liddle AR. 2002. *Ap. J.* 569:L75–L78
214. Vikhlinin A, McNamara BR, Forman W, Jones C, Hornstrup A, Quintana H, et al. 1998a. *Ap. J. Lett.* 498:L21–24
215. Vikhlinin A, McNamara BR, Forman W, Jones C, Quintana H, Hornstrup A. 1998b. *Ap. J.*, 502:558–581
216. Voit GM. 2000. *Ap. J.* 543:113–123
217. Voit GM, Bryan GL. 2002. *Nature* 414:425–427
218. White DA. 2000. *MNRAS* 312:663–688
219. White DA, Jones C, Forman W. 1997. *MNRAS* 292:419–467
220. White SDM, Efstathiou G, Frenk CS. 1993. *MNRAS*, 262: 1023–1028
221. White SDM, Navarro JF, Evrard AE, Frenk CS. 1993. *Nature* 366:429–433
222. Wu KKS, Fabian AC, Nulsen PEJ. 2000. *MNRAS* 318:889–912
223. Wu X-P, Xue Y-J, Fang L-Z. 1999. *Ap. J.* 524:22–30
224. Yamada M, Fujita Y. 2001. *Ap. J.* 553:145–148
225. Yee HKC, Ellingson E, Carlberg RG. 1996. *Ap. J. Suppl.* 102:269–287
226. York DG, Adelman J, Anderson JEJr, Anderson SF, Annis J, et al. 2000. *Astron. J.* 120:1579–1587
227. Zheng W, Ford H, Tsvetanov Z, Davidsen A, Szalay A, et al. 2002. In *Lighthouses of the Universe*, in press
228. Zwicky F. 1937. *Ap. J.* 86:217–246
229. Zwicky F, Herzog E, Wild P. 1966. *Catalogue of galaxies and of clusters of galaxies*. Pasadena: California Institute of Technology

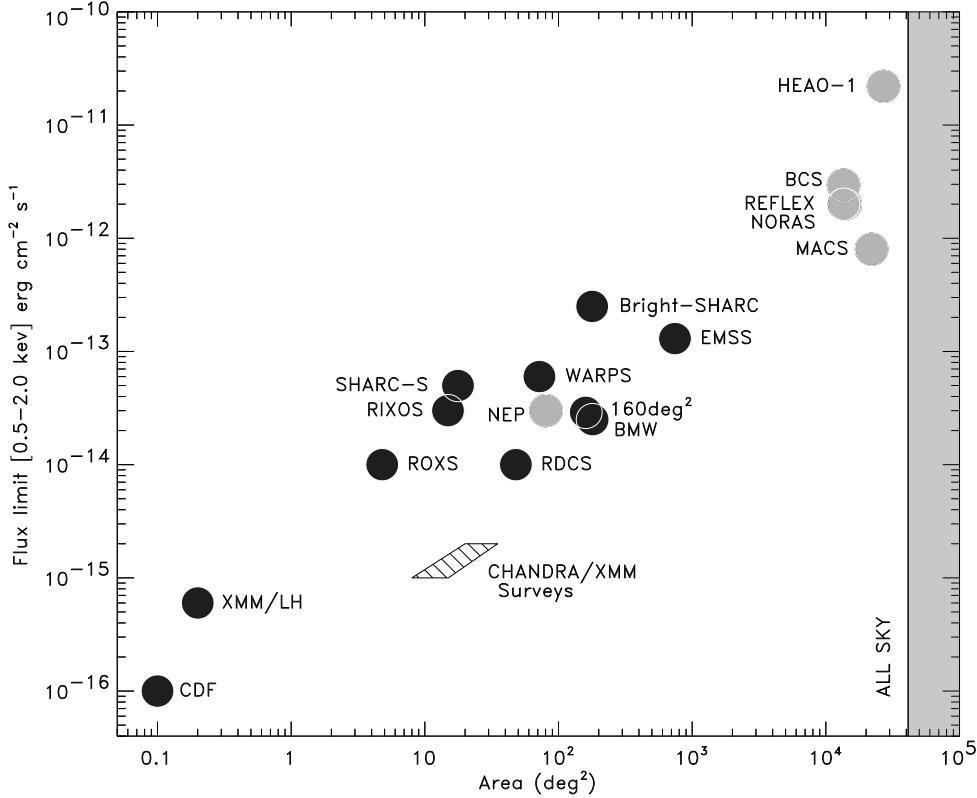


Figure 4: Solid angles and flux limits of X-ray cluster surveys carried out over the last two decades. References are given in the text. Dark filled circles represent serendipitous surveys constructed from a collection of pointed observations. Light shaded circles represent surveys covering contiguous areas. The hatched region is a predicted locus of future serendipitous surveys with *Chandra* and *Newton-XMM*.

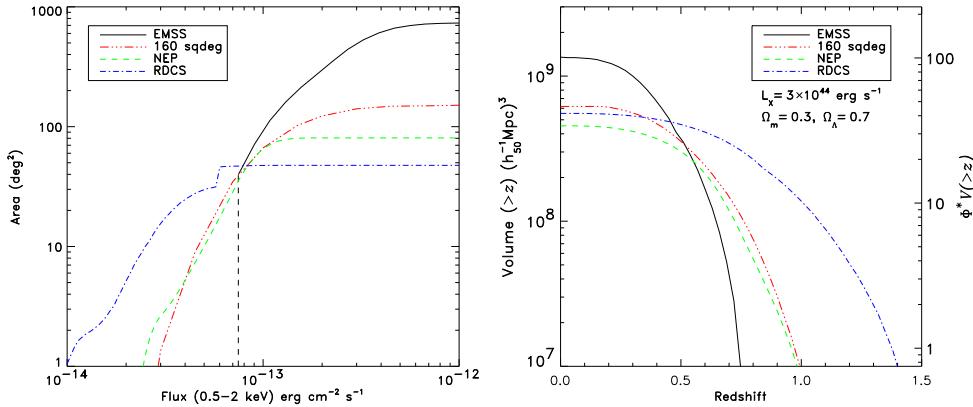


Figure 5: (Left) sky coverage as a function of X-ray flux of several serendipitous surveys; (Right) corresponding search volumes, $V(>z)$, for a cluster of given X-ray luminosity ($L_X = 3 \times 10^{44} \text{ erg s}^{-1} [0.5 - 2 \text{ keV}] \simeq L_X^*$). On the right axis the volume is normalized to the local space density of clusters, ϕ^* .

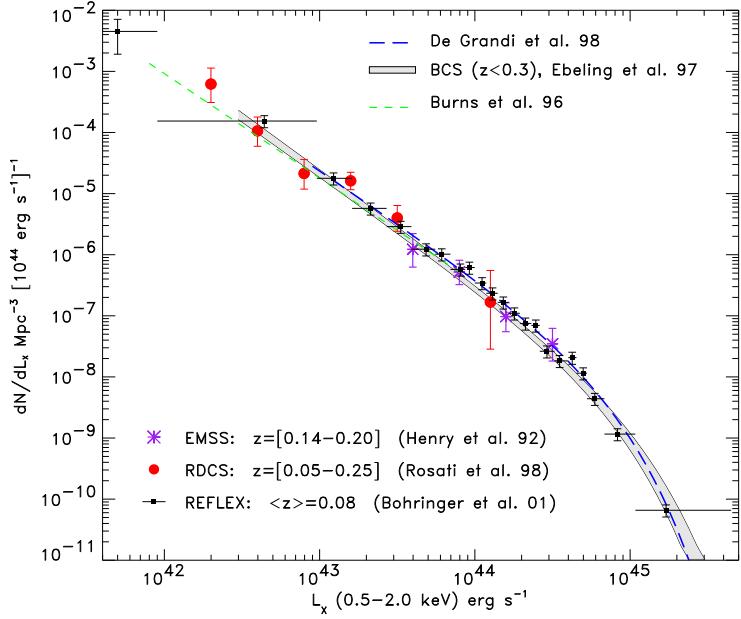


Figure 6: Determinations of the local X-ray Luminosity Function of clusters from different samples (an Einstein-de-Sitter universe with $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ is adopted). For some of these surveys only best fit curves to XLFs are shown.

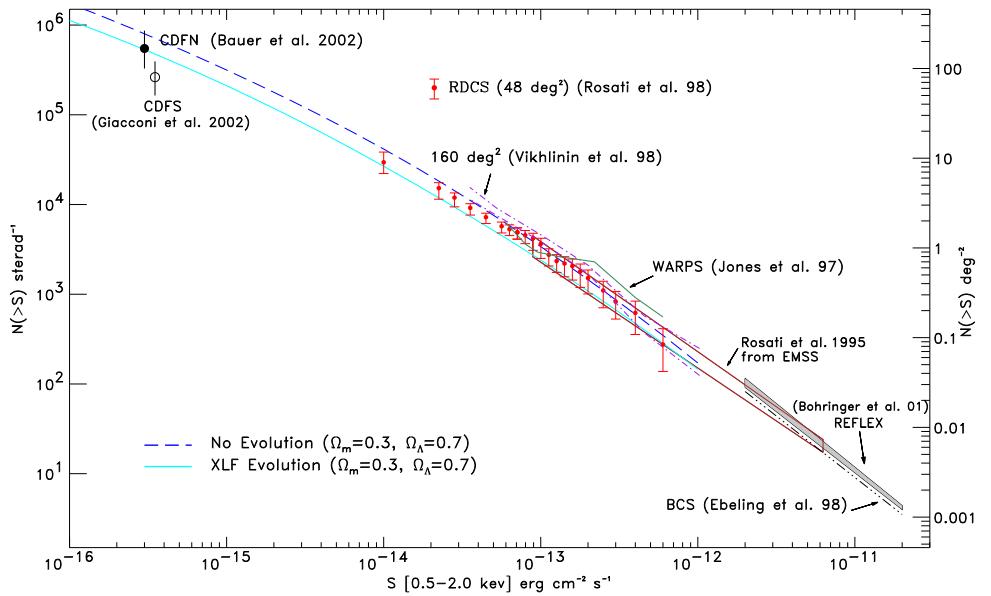


Figure 7: The cluster cumulative number counts as a function of X-ray flux ($\log N - \log S$) as measured from different surveys.

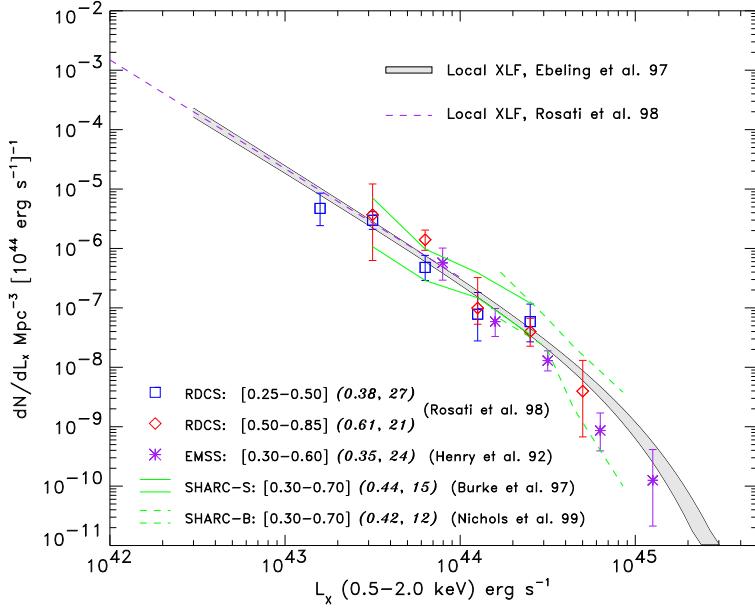


Figure 8: The X-ray Luminosity Function of distant clusters out to $z \simeq 0.8$ compiled from various sources and compared with local XLFs (an Einstein–de-Sitter universe with $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ is adopted). Numbers in parenthesis give the median redshift and number of clusters in each redshift bin.

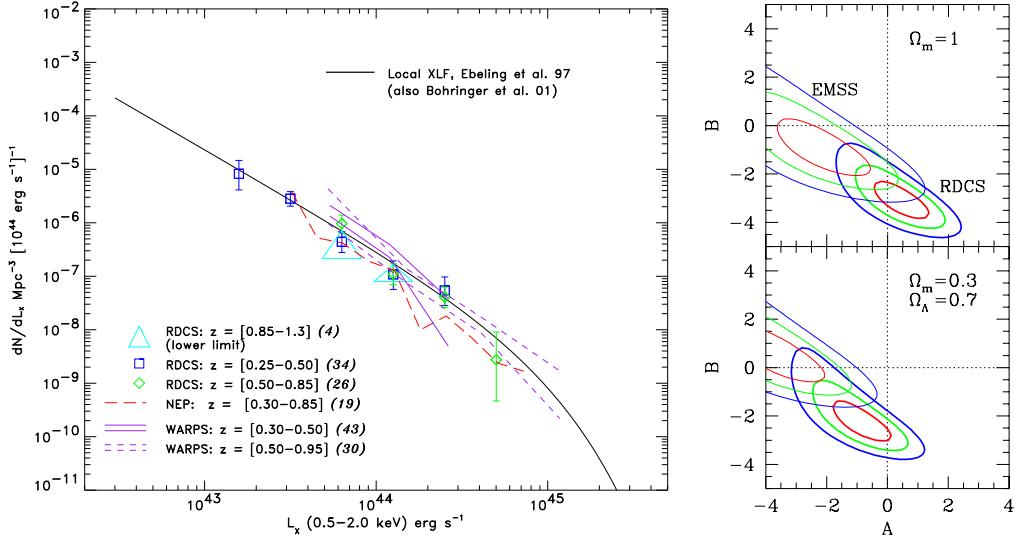


Figure 9: (Left) the latest compilation of distant XLFs (RDCS: Rosati et al. 2000; NEP: Gioia et al. 2001; WARPS: Jones et al. 2000; an Einstein–de-Sitter universe with $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ is adopted). Right panel: Maximum-likelihood contours (1, 2 and 3 σ confidence level) for the parameters A and B defining the XLF evolution for the RDSC and EMSS samples (for two different cosmologies): $\phi^* = \phi_0(1+z)^A$, $L^* = L_0^*(1+z)^B$ (see Equation 7).

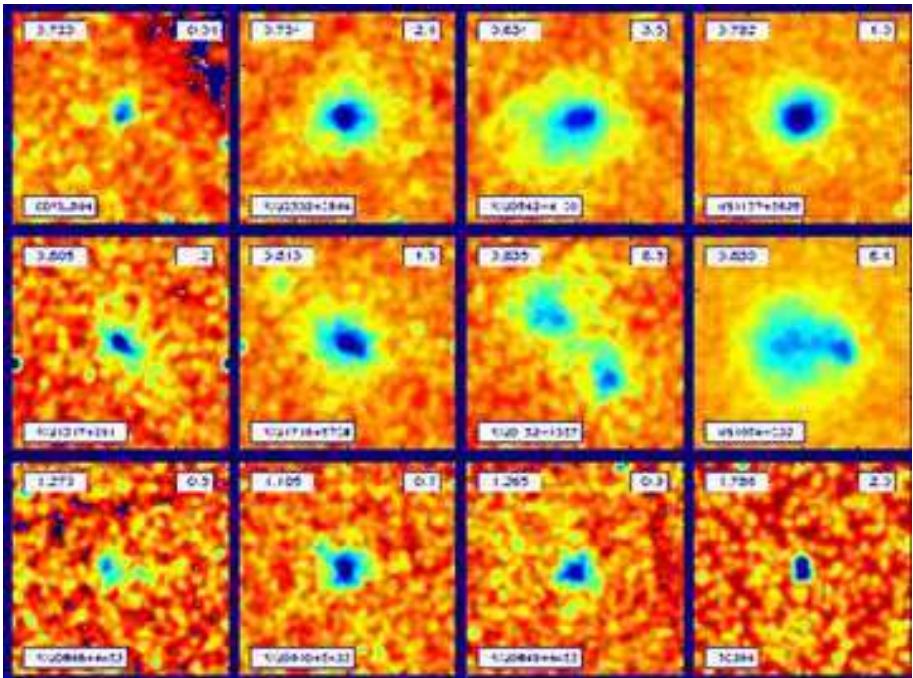


Figure 10: *Chandra* archival images of twelve distant clusters at $0.7 < z < 1.3$. Labels indicate redshifts (upper left) and X-ray luminosities (upper right) in the rest frame [0.5-2] keV band, in units of 10^{44} erg s $^{-1}$. All fields are 2 Mpc across; the X-ray emission has been smoothed at the same physical scale of 70 kpc ($h = 0.7$, $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$).

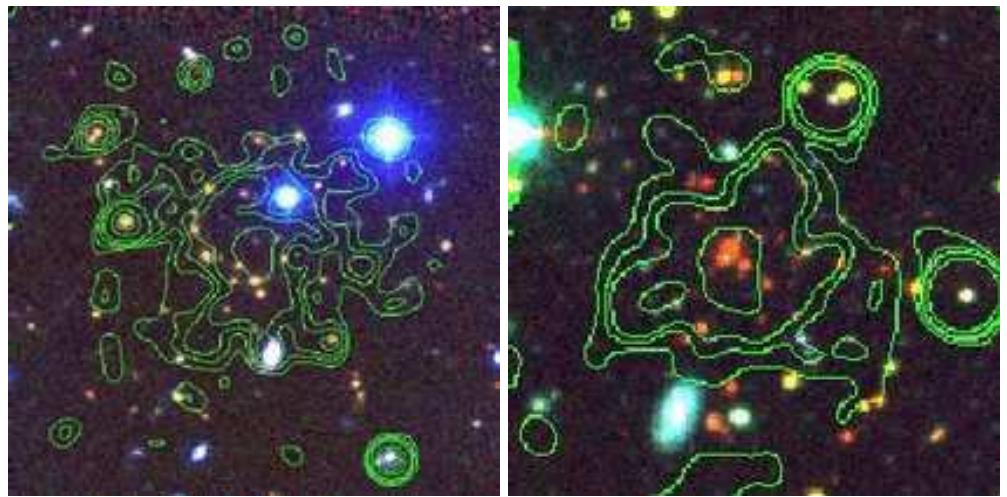


Figure 11: Color composite images combining optical and near-IR imaging of two X-ray selected clusters at $z > 1$. Overlaid contours map the X-ray mission detected by *Chandra/ACIS-I*. (Left) RXJ0910+5422 at $z = 1.11$ (Stanford et al. 2002); (right) RXJ0849+4452 at $z = 1.26$ (Rosati et al. 1999, Stanford et al 2001). The two fields are 1.5 arcmin across ($\simeq 1h_{50}^{-1}$ Mpc at these redshifts).

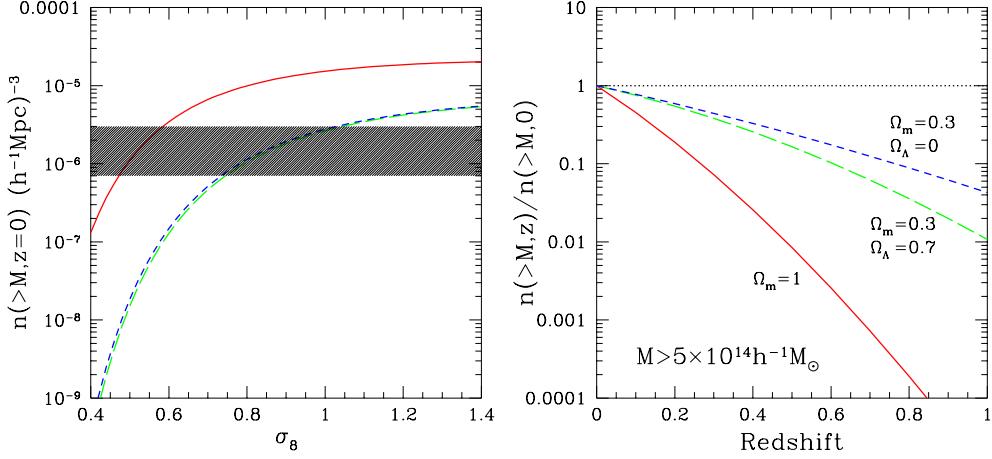


Figure 12: The sensitivity of the cluster mass function to cosmological models. (*Left*) The cumulative mass function at $z = 0$ for $M > 5 \times 10^{14} h^{-1} M_{\odot}$ for three cosmologies, as a function of σ_8 , with shape parameter $\Gamma = 0.2$; solid line: $\Omega_m = 1$; short-dashed line: $\Omega_m = 0.3, \Omega_{\Lambda} = 0.7$; long-dashed line: $\Omega_m = 0.3, \Omega_{\Lambda} = 0$. The shaded area indicates the observational uncertainty in the determination of the local cluster space density. (*Right*) Evolution of $n(>M, z)$ for the same cosmologies and the same mass-limit, with $\sigma_8 = 0.5$ for the $\Omega_m = 1$ case and $\sigma_8 = 0.8$ for the low-density models.

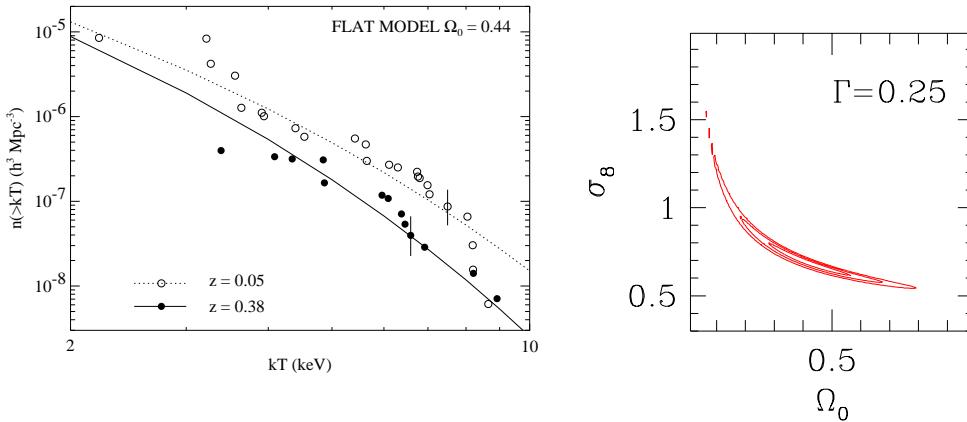


Figure 13: (*Left*) The cumulative X-ray temperature function for the nearby cluster sample by Henry & Arnaud (1991) and for a sample of moderately distant clusters (from Henry 2000). (*Right*) Probability contours in the $\sigma_8 - \Omega_m$ plane from the evolution of the X-ray temperature function (adapted from Eke et al. 1998).

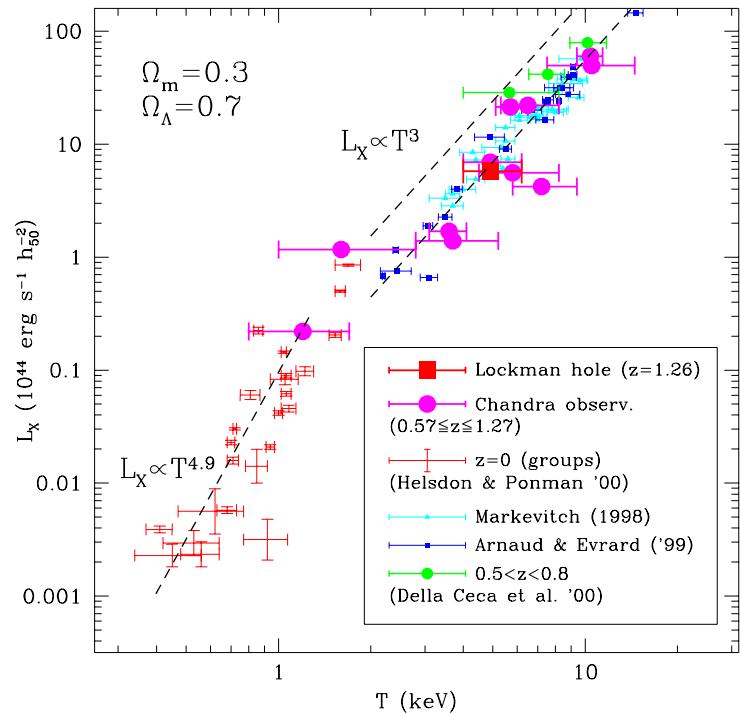


Figure 14: The (bolometric) luminosity–temperature relation for nearby and distant clusters and groups compiled from several sources (see Borgani et al. 2001b, Holden et al. 2002). The two dashed lines at $T > 2$ keV indicate the slope $\alpha = 3$, with normalization corresponding to the local $L_X - T$ relation (lower line) and to the relation of Equation 14 computed at $z = 1$ for $A = 1$. The dashed line at $T < 1$ keV shows the best-fitting relation found for groups by Helsdon & Ponman (2000).

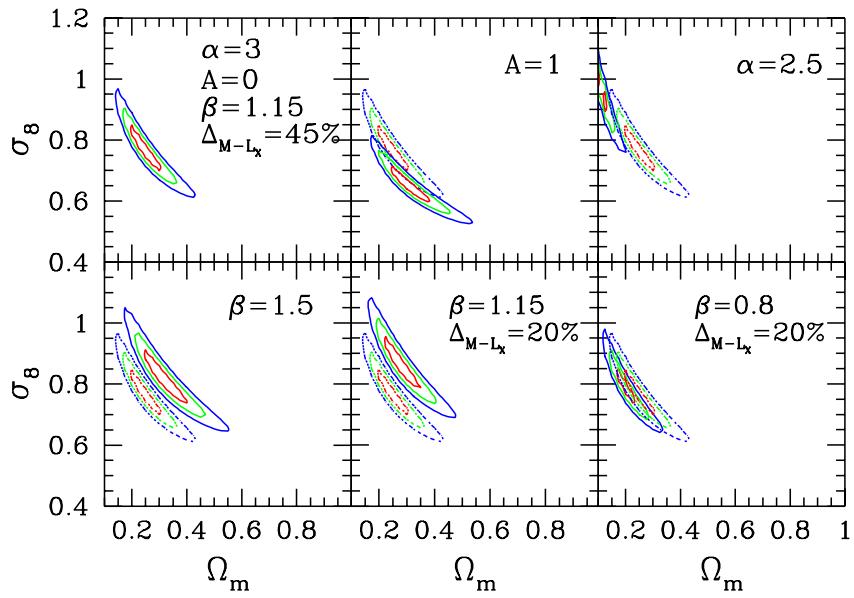


Figure 15: Probability contours in the σ_8 – Ω_m plane from the evolution of the X-ray luminosity distribution of RDCS clusters. The shape of the power spectrum is fixed to $\Gamma = 0.2$. Different panels refer to different ways of changing the relation between cluster virial mass, M , and X-ray luminosity, L_X , within theoretical and observational uncertainties (see also Borgani et al. 2001b). The upper left panel shows the analysis corresponding to the choice of a reference parameter set. In each panel, we indicate the parameters which are varied, with the dotted contours always showing the reference analysis.

Hot plasma in clusters of galaxies, the largest objects in the universe

Craig L. Sarazin

*Department of Astronomy, University of Virginia,
530 McCormick Road, Charlottesville, VA 22903-0818*

(Dated: February 5, 2008)

Clusters of galaxies are the largest organized structures in the Universe. They are important cosmological probes, since they are large enough to contain a fair sample of the materials in the Universe, but small enough to have achieved dynamical equilibrium. Clusters were first discovered as concentrations of hundreds of bright galaxies in a region about 3 megaparsecs (10 million light years) across. However, the dominant observed form of matter in clusters is hot, diffuse intergalactic gas. This intracluster plasma has typical temperatures of $T \sim 7 \times 10^7$ K, and typical electron densities of $n_e \sim 10^{-3}$ cm $^{-3}$. This intracluster plasma mainly emits X-rays, and typical cluster X-ray luminosities are $L_X \sim 10^{43} - 10^{45}$ erg/s. The basic properties of and physical processes in the intracluster plasma will be reviewed. Important observational constraints on plasma processes in these systems will be discussed. Recent X-ray observations of clusters of galaxies with the orbiting Chandra X-ray Observatory will be highlighted.

PACS numbers: 95.30.Qd, 98.65.Cw, 98.65.Hb

I. INTRODUCTION

Clusters of galaxies are the largest relaxed structures in the Universe. They were initially discovered in optical observations, where clusters appear as concentrations containing $\sim 10^2$ bright galaxies and $\sim 10^3$ faint galaxies in a region which is typically ~ 2 Mpc in radius (1 megaparsec = 3.09×10^{24} cm) [1]. For example, Figure (1, left panel) shows the optical image of the central region of the nearby Coma cluster showing many galaxies. In their central regions, clusters are about 10^3 times denser than the average of material in the Universe. Clusters of galaxies are very important cosmological probes [2]. Essentially, they are the only objects in the Universe are both small enough to have achieved dynamical equilibrium during the age of the Universe, and large enough to contain a fair sample of the material in the Universe (e.g., the average ratio of baryonic to dark matter).

Although they were first observed as collections of galaxies, the dominant form of matter which has been observed in clusters of galaxies is hot diffuse plasma [3]. This intracluster medium (ICM) has typical temperatures of $\sim 7 \times 10^7$ K and typical electron number densities of $n_e \sim 10^{-3}$ cm $^{-3}$. At these temperatures, the dominant form of radiation from a plasma is X-ray emission, mainly from thermal bremsstrahlung but also from collisionally excited line emission. As a result, clusters of galaxies are generally very luminous X-ray emitters, with luminosities of $L_X \sim 10^{43} - 10^{45}$ ergs s $^{-1}$. Clusters are second only to quasars as the most luminous X-ray sources in the Universe. For example, Figure (1, right panel) shows the X-ray image of the same central region of the Coma cluster as the left panel. Although the ICM is diffuse, it fills all of the volume between and within the galaxies in clusters, and as a result its mass is large. The total mass of hot plasma in a cluster is typically $M_{\text{gas}} \sim 10^{14} M_\odot$, where $M_\odot = 1.99 \times 10^{33}$ g is the mass of the Sun. In large clusters, the total mass of hot gas exceeds the mass of all

the stars and galaxies by a factor of ~ 5 . Hot intracluster plasma is the dominant form of baryonic matter in clusters. In general, we now believe that most of the baryonic matter in the low redshift Universe is in the form of hot intergalactic plasma.

II. PHYSICAL PROPERTIES OF THE INTRACLUSTER PLASMA

The mean free paths of electrons and ions in a plasma without a magnetic field are determined by Coulomb collisions. The mean free path of electrons (which is nearly the same as that protons) is [4]

$$\lambda_e = \frac{3^{3/2}(kT)^2}{4\pi^{1/2}n_e e^4 \ln \Lambda} \approx 23 \left(\frac{T}{10^8 \text{ K}} \right)^2 \left(\frac{n_e}{10^{-3} \text{ cm}^{-3}} \right)^{-1} \text{ kpc}, \quad (1)$$

where n_e is the electron number density, the Coulomb logarithm $\ln \Lambda \approx 38$, and 1 kpc = 3.09×10^{21} cm. This is about 1% of the radius of a cluster, which suggests that the intracluster plasma can be treated as a fluid. The gyroradii in the intracluster magnetic field are much smaller than this (Sec. III).

The timescale for Coulomb collisions between electrons to bring them into kinetic equilibrium (an isotropic Maxwellian velocity distribution) is about [4]

$$t_{\text{eq}}(e, e) \approx 3.3 \times 10^5 \left(\frac{T}{10^8 \text{ K}} \right)^{3/2} \left(\frac{n_e}{10^{-3} \text{ cm}^{-3}} \right)^{-1} \text{ yr}. \quad (2)$$

The time scale for protons to equilibrate among themselves is $t_{\text{eq}}(p, p) \approx (m_p/m_e)^{1/2} t_{\text{eq}}(e, e)$ or roughly 43 times longer than the value in Equation (2). Similarly, time scale for the electrons and ions to reach equipartition (equal temperatures) is $t_{\text{eq}}(p, e) \approx (m_p/m_e)t_{\text{eq}}(e, e)$, or roughly 1870 times the value in Equation (2). All of these are shorter than the typical ages of clusters of

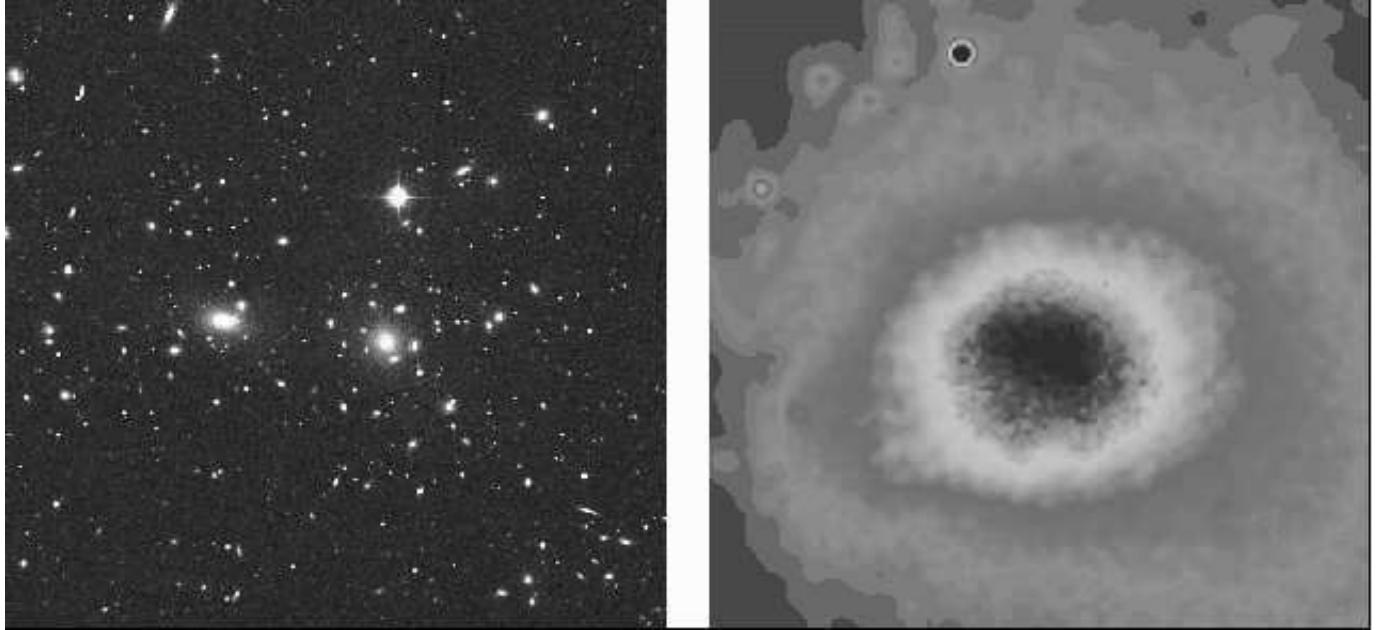


FIG. 1: The left panel shows an optical image of the central region of the nearby Coma cluster of galaxies, showing the galaxies in the cluster. The right panel shows the same region imaged in X-ray emission with the Röntgen Satellite (ROSAT). The X-ray emission comes from hot plasma at about 10^8 K which fills the volume of the cluster.

$\gtrsim 10^9$ yr. Thus, the intracluster plasma can generally be characterized by a Maxwellian distribution at a kinetic temperature T . Because the radiation field is much more diffuse than a blackbody at the kinetic temperature, the gas is far from thermodynamic equilibrium in terms of the populations of excited levels of ions and the ionization state. However, the gas is in “coronal equilibrium”; most bound electrons are in their ground levels, they are in excitation equilibrium, and the gas is in ionization equilibrium.

The sound crossing time for a cluster is

$$t_s \equiv \frac{D}{c_s} \approx 6.6 \times 10^8 \left(\frac{T}{10^8 \text{ K}} \right)^{-1/2} \left(\frac{D}{\text{Mpc}} \right) \text{ yr}, \quad (3)$$

Here, D is the diameter of the cluster, and c_s is the sound speed. This is somewhat smaller than the likely ages of clusters, so unless they are being disturbed (Sec. IV), the gas should be nearly in hydrostatic equilibrium. For a spherical cluster in hydrostatic equilibrium, the gas distribution is given by

$$\frac{1}{\rho_{\text{gas}}} \frac{dP_{\text{gas}}}{dr} = -\frac{GM(r)}{r^2}, \quad (4)$$

where $M(r)$ is the total cluster mass within a radius r . The gas pressure in the ICM is given by the ideal gas law, $P_{\text{gas}} = \rho_{\text{gas}} kT / (\mu m_H)$, where ρ_{gas} is the mass density in the gas, and μm_H gives the mean mass per particle. One often assumes that magnetic forces, pressure from relativistic particles, and other forces are relatively weak in clusters, although it is not certain that this is

correct (Sec. III). Equation (4) has been used to determine the total masses or total density profiles of clusters by solving for $M(r)$. The temperature of the ICM can be determined from observations of the X-ray spectrum of the gas, while the X-ray surface brightness can easily be de-projected to give the gas density. Such measurements indicate that the total masses of large clusters are about $10^{15} M_\odot$, which considerably exceed the total mass of all of the intracluster gas and of all the galaxies combined. As such, clusters provide some of the strongest evidence for the domination of (probably nonbaryonic) dark matter on large scales in the Universe. In a typical large cluster, $\sim 16\%$ of the mass is in hot ICM, $\sim 3\%$ of the mass is in stars and galaxies, and $\sim 81\%$ of the mass appears to be dark matter.

As is true of most materials in the Universe, the ICM consists primarily of ionized hydrogen ($\approx 71\%$ of mass) and helium ($\approx 28\%$ of mass). However, the ICM does contain a significant amount of the common heavier elements (O, Fe, etc.; $\approx 1\%$ of mass). This fraction is only about a factor of 2–3 times smaller than the fraction in the Sun. Many of these heavy elements are detected through X-ray lines observed in the spectra of clusters of galaxies; these lines occur because the heavier elements are not quite completely ionized, even at the high temperatures in clusters. Because most of the baryonic matter in clusters is in the ICM, it turns out that most of the heavy elements are actually located there as well. Hydrogen and helium are formed in the Big Bang, but the only source of the common heavier elements is fusion reactions in the centers of stars. At present, the only significant

populations of stars are located in galaxies. The dispersal of heavy element into the diffuse intracluster gas required that stars in galaxies be disrupted by supernova explosions, and that the enriched gas escape the gravity of the galaxy in which the stars are located. Detailed models for the chemical evolution of the ICM suggest that about 25% of the gas originated in stars in galaxies, and that the remaining 75% came from primordial intergalactic gas. Because there is presently about five times as much mass in the ICM as in stars and galaxies, this requires that the galaxies located in clusters lost significant amounts of their baryonic content. At present, the galactic population in clusters of galaxies consists mainly of elliptical (E) and lenticular (S0) galaxies, which have only low mass stars which do not produce a high rate of supernovae. To explain the large amounts of heavy elements in the ICM, these galaxies must have had much higher rates of star formation and supernovae at earlier times, possibly associated with the formation of the galaxies.

Initially, the very high temperature ($\sim 10^8$ K) of the ICM might seem surprising. However, this is one feature of clusters which is easily understood. Clusters of galaxies contain enormous masses of material, and have very deep gravitational potential wells. Almost any natural process which introduces gas into clusters will cause it to move very rapidly and be shock-heated to roughly the observed temperature. For example, if the ICM fell into clusters, it would be accelerated to roughly the escape speed from clusters, which is ~ 2000 km/s. A portion of the ICM may have been ejected from galaxies; galaxies in clusters move on random orbits with velocities of ~ 1000 km/s. In either case, infalling or ejected gas would encounter other gas moving at similar velocities, and would undergo strong hydrodynamical shocks. Shocks at speeds of ~ 1000 km/s heat gas to temperatures of $\sim 10^8$ K.

III. MAGNETIC FIELDS AND RELATIVISTIC PARTICLES

Although the ICM is dominated in mass and energetically by thermal plasma, it does contain magnetic fields and populations of relativistic, nonthermal particles as well. The most direct way to measure the magnetic field in clusters of galaxies is through the Faraday rotation of the polarization of background or embedded radio sources [5]. Most strong extragalactic radio sources emit synchrotron radiation which is strongly linearly polarized. When this radiation passes through a magnetized plasma, the plane of polarization is rotated through an angle

$$\phi = (RM)\lambda^2, \quad (5)$$

where λ is the wavelength of the radiation. The rotation measure, RM , is given by

$$RM = \frac{e^3}{2\pi m_e^2 c^4} \int n_e B_{\parallel} dl, \quad (6)$$

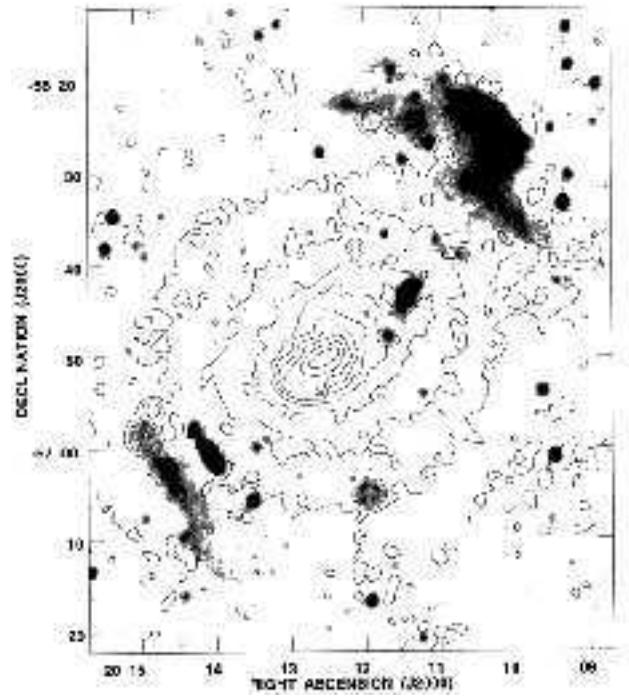


FIG. 2: The grey-scale is the radio image of the cluster Abell 3667 [7], while the contours are an X-ray image. The radio image shows two extended arcs, which are cluster radio relics, located to the southeast (lower left) and to the northwest (upper right) of the cluster center. These radio relics are believed to be associated with cluster merger shocks. (©1997, The Royal Astronomical Society.)

where l is the path length through the medium, and B_{\parallel} is the component of the magnetic field parallel to the direction of propagation of the radiation. The rotation measures seen through the central regions of clusters are $RM \sim 100$ rad m $^{-2}$. Unfortunately, Equation (6) shows that the rotation measure only determines an integral of B_{\parallel} along the line of sight, and this integral is strongly affected by the poorly known topology of the magnetic field. Field reversals along the line of sight will greatly reduce RM . If one assumes that the coherence length of the field is about 10 kpc, the rotation measure observations suggest that the typical magnetic field strength is $B \sim 5 \mu G$ [5]. This would imply that the fields are still significantly subthermal; the ratio of magnetic to gas pressures is only $(P_B/P_{\text{gas}}) \sim 0.05$. In the central regions of clusters with cooling cores (Sec. V), much larger rotations measures are observed ($RM \sim 10^4$ rad m $^{-2}$), which suggest that the magnetic fields in these regions may approach equipartition with the gas pressure.

In addition to the thermal plasma in the intracluster medium, significant population of relativistic electrons are observed in some clusters. The observations are crudely consistent with a power-law distribution for the relativistic particles; that is, $N_e(E) \propto E^{-p}$, where $N_e(E) dE$ gives the number of electrons with energies in

the range $E \rightarrow E + dE$. Similar distributions are seen in other astrophysical plasmas including the Galactic cosmic rays. However, it is likely that the particle distributions are more complex, and also vary spatially. Relativistic electrons interact with the intracluster magnetic field to produce synchrotron radio emission; the electrons which produce the emission typically have energies ~ 10 GeV. Diffuse emission, not associated with any individual galaxy, is seen in ~ 40 clusters of galaxies; when the emission is centrally located, the sources are called “cluster radio halos,” while “cluster radio relics” are peripherally located [6]. For example, Figure (2) shows two radio relic sources in the cluster Abell 3667 [7].

Relativistic electrons in clusters can also produce observable emission through the inverse Compton scattering of low energy photons; the main source of these photons in clusters of galaxies is the Cosmic Microwave Background. Typically, inverse Compton scattering produces emission which is observable either in the extreme-uv/soft X-ray band (near 0.1 keV) or in the hard X-ray band (20–100 keV). Recently, inverse Compton hard X-ray emission has been detected from several clusters [8]. Extreme-uv/soft X-ray emission, which might also be from relativistic electrons, may have also been detected [9]. Detecting both inverse Compton and synchrotron emission from the same population of relativistic electrons is useful, because the synchrotron emission depends on the product of the energy density in relativistic electrons and that in the magnetic field, while inverse Compton emission depends on the product of the energy density in relativistic electrons and that in the Cosmic Microwave Background (which is very well determined). In principle, the combination of these two measurements allows both the total energy in relativistic electrons and the magnetic field strength to be determined. However, one needs to assume that the particles and magnetic field have the same distribution and that both are reasonably uniform, which may not be true. For the very few clusters with such data, the present observations and this simple argument suggest that the magnetic field strengths are $\sim 0.5 \mu\text{G}$, about an order of magnitude smaller than those derived from Faraday rotation. This disagreement may indicate that the assumption of a similar and uniform distribution for the particles and magnetic fields is wrong; for example, the magnetic field in clusters may be very inhomogeneous. Assuming these uncertain values are correct, the energy density and pressure in relativistic electrons may be a few percent of values for the thermal plasma in clusters. However, the total contribution from relativistic particles is also uncertain because the ions have not been detected.

IV. CLUSTER MERGERS: THE MOST ENERGETIC EVENTS SINCE THE BIG BANG

There now is considerable evidence that clusters of galaxies and other large structures in the Universe form

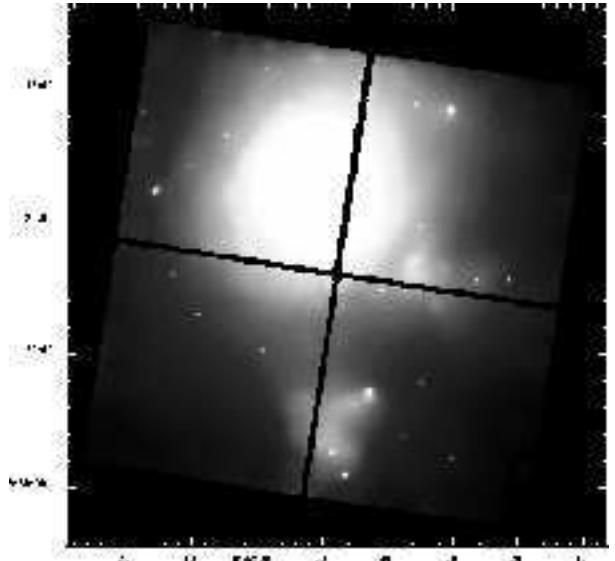


FIG. 3: The Chandra X-ray image of the merger cluster Abell 85 [11]. The grey scale burns out the central cooling flow region to show the outer parts of the cluster. Two subclusters to the south (lower middle) and southwest (lower right) are merging with the main cluster. The southwestern subcluster has a cluster radio relic. The sharp feature at the northwest of the southern subcluster is a “cold front”.

hierarchically; that is, smaller structures form first, and gravity pulls these smaller structures together to make larger structures. Clusters of galaxies form by the merger of smaller subclusters and groups of galaxies.

Major cluster mergers, in which two subclusters with a total mass of $\sim 10^{15} M_\odot$ collide together at velocities of more than 2000 km/s, are the most energetic events which have occurred in the Universe since the Big Bang itself [10]. Cluster mergers release total energies of $\sim 3 \times 10^{64}$ erg. The motions in cluster mergers are transonic, and the mergers drive shocks into the intracluster gas. In major mergers, these merger shocks dissipate total energies of $\sim 3 \times 10^{63}$ erg. Such merger shocks are, in fact, the primary heating source of the intracluster plasma. For example, Figure (3) shows the Chandra image of the merging cluster Abell 85 [11].

Hydrodynamical simulations of cluster formation and evolution have shown the importance of merger shocks [12]. The evolution of the structure of merger shocks is illustrated in Figure 4, which shows an off-center merger between two symmetric subclusters. At early stages in the merger (the first panel and earlier), the shocked region is located between the two subcluster centers and is bounded on either side by two shocks. At this time, the subcluster centers, which may contain cooling cores and central radio sources, are not affected. Later, these shocks sweep over the subcluster centers (between the first and second panels). The main merger shocks pass into the outer parts of the merging system (panel 2), and secondary shocks may appear in the inner regions (panel 3). Eventually, the cluster begins to return to equilibrium

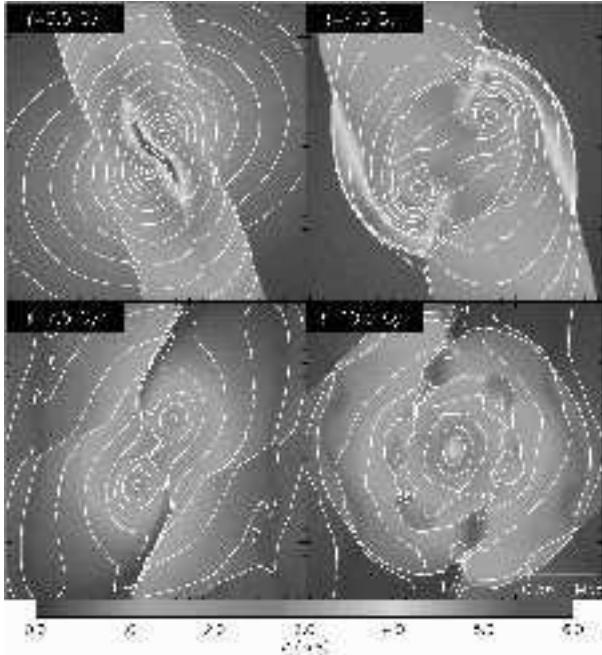


FIG. 4: The results of a hydrodynamical simulation of a symmetric, off-center cluster merger [12]. The grey scale shows the temperature, while the contours are the X-ray surface brightness. Initially, the shocked region is located between the two subcluster centers. Later, the main merger shocks propagate to the outer parts of the cluster, and other weaker shocks also occur. By the end of the simulation, the cluster is beginning to return to equilibrium.

(panel 4).

In addition to their thermal effects, astrophysical shocks at velocities $\gtrsim 1000$ km/s always convert at least a few percent of the shock energy into the acceleration of relativistic electrons and ions [13]. In general, this occurs through a first-order Fermi acceleration process. One would thus expect that merger shocks would produce relativistic electrons, which would be observable through radio synchrotron emission. In fact, cluster radio relic and cluster radio halo sources are seen in many clusters (Sec. III). In every case, these clusters appear to be undergoing a cluster merger; the cluster Abell 3667 in Figure (2) is an example. Recent Chandra X-ray images indicate that the radio relics lie just behind merger shocks; the central radio halos may be due to turbulent particle acceleration after the passage of the merger shock [14].

One exciting discovery made with the Chandra X-ray Observatory is the importance of “cold fronts” in merging clusters. Figure (5) shows the cold front seen in the Chandra X-ray image of the central regions of the merging cluster Abell 3667 [15]; this is the same cluster shown in Figure (2). The cold front is the sharp surface brightness discontinuity to the lower right. When these features were first seen, it was initially assumed that they merger shocks, and that the brighter inner region resulted from shock compression. However, X-ray spectral measure-

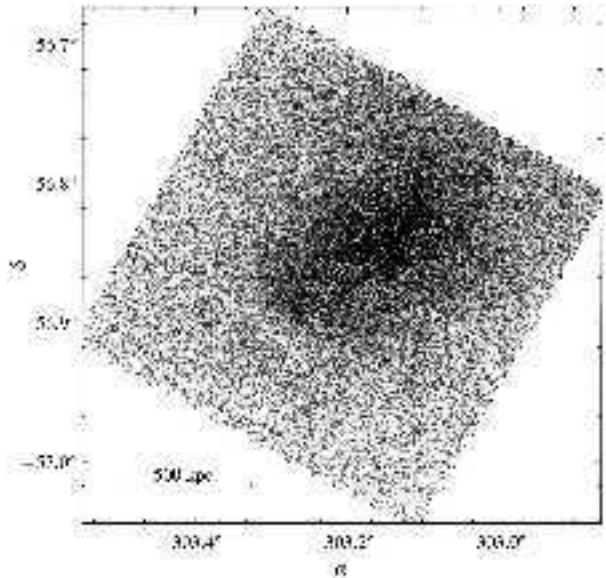


FIG. 5: A “cold front” seen in the Chandra X-ray image of the central region of the merging cluster Abell 3667 [15]. [This is the central region of the same cluster shown in Figure (2).] Note the sharp discontinuity in the X-ray surface brightness. Spectral analysis show that the brighter gas is actually cooler than the surrounding fainter gas. (©2001, The American Astronomical Society.)

ments show that these features are not shocks. The temperature of the denser gas is actually lower than that of the less dense gas in such a way that the pressure is continuous across the surface brightness discontinuity. Thus, the specific entropy is actually lower in the denser region. Shocks are irreversible changes which increase the density, pressure, and entropy. The observed “cold fronts” in clusters are not shocks, but rather contact discontinuities between higher density cool gas and lower density hot gas.

As discussed below (Sec. V), the centers of many clusters often contain relative cool (10^7 K rather than 10^8 K) gas. It is believed that cold fronts occur when clusters containing such cool cores merger. The gas in the core is dense enough to survive for some time after the merger. As the clusters merge, the cool cores move rapidly through the lower density shocked gas, producing the cold fronts.

Observations of cluster merger shocks and cold fronts can be used to derive the kinematics of the merger [15, 16]. Most of these diagnostics give the Mach number of the merger \mathcal{M} , which is the ratio of the merger velocity to the sound speed in the pre-merger gas. The Rankine–Hugoniot jump conditions across a merger shock give the pressure increase across the shock and the shock compression as

$$\frac{P_2}{P_1} = \frac{2\gamma}{\gamma+1} \mathcal{M}^2 - \frac{\gamma-1}{\gamma+1}$$

$$\frac{v_2}{v_1} = \frac{\rho_1}{\rho_2} \equiv \frac{1}{C} = \frac{2}{\gamma+1} \frac{1}{\mathcal{M}^2} + \frac{\gamma-1}{\gamma+1}, \quad (7)$$

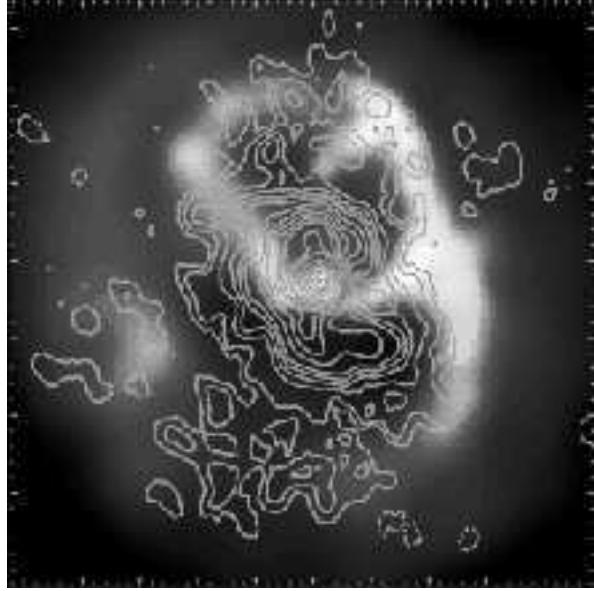


FIG. 6: The grey scale image is the central $\sim 80 \times 80$ kpc of the Chandra X-ray image of the cooling core cluster Abell 2052 [17]. The contours are the radio image of the same region. The hot, X-ray gas is missing from two bubbles to the north (up) and south (down), and bright X-ray shells surround these bubbles. The radio image shows that the bubbles are filled by radio emitting plasma containing relativistic electrons and magnetic fields.

where $\gamma = 5/3$ is the adiabatic index for fully ionized plasma, where $C \equiv \rho_2/\rho_1$ is the shock compression, and the subscripts 1 and 2 denote the pre-shock and post-shock gas. X-ray observations can provide the gas temperature and density on either side of the shock, and the jump conditions yield \mathcal{M} and the merger velocity.

For the case of a cold front, the stagnation condition at the leading edge of the cold front gives [15]

$$\frac{P_{st}}{P_1} = \begin{cases} \left(1 + \frac{\gamma-1}{2}\mathcal{M}^2\right)^{\frac{\gamma}{\gamma-1}}, & \mathcal{M} \leq 1, \\ \mathcal{M}^2 \left(\frac{\gamma+1}{2}\right)^{\frac{\gamma+1}{\gamma-1}} \left(\gamma - \frac{\gamma-1}{2\mathcal{M}^2}\right)^{-\frac{1}{\gamma-1}}, & \mathcal{M} > 1, \end{cases} \quad (8)$$

where P_{st} is the pressure at the stagnation point. If $\mathcal{M} > 1$ there will be a bow shock ahead of the cold front, and one can also apply the shock jump conditions, (Equation 7). The bow shock will be located at some distance (the “stand off” distance d_s) ahead of the cold front, and the ratio of this distance to the radius of curvature of the cold front is a decreasing function of \mathcal{M} . Finally, the opening angle of the Mach cone formed from the merger shock will depend on the Mach number as $\theta_M = \csc^{-1}(\mathcal{M})$.

Applications of these techniques to observed cluster mergers give values for the merger Mach number of $\mathcal{M} \approx 2$ and merger velocities of ≈ 2000 km/s [11, 15].

V. CENTRAL COOLING CORES IN CLUSTERS

At the centers of many clusters of galaxies, the gas temperature is seen to drop significantly from $\sim 10^8$ K further out to $\sim 10^7$ K near the center. At the same time, the gas density rises very rapidly, which gives these regions very large X-ray surface brightnesses. The low gas temperatures and high gas densities are the result of radiative cooling of the gas. Since the X-ray emissivity varies with the square of the density, while the thermal energy density is proportional to the gas density, the X-ray radiation we observe can cool the gas most rapidly in the dense, central regions of clusters. In these regions, the radiative cooling times are $\lesssim 3 \times 10^8$ yr, which is much shorter than the ages of typical clusters of $\sim 10^{10}$ yr. As the central gas cools, the weight of the overlying outer gas compresses the inner gas, resulting in cooler temperatures and very high densities. I will refer to these cool central regions as “cooling cores”.

One mystery with previous X-ray observations was whether the gas continues to cool below $\sim 10^7$ K, and if so, what happens to it. The rates of gas cooling at higher temperature are quite large, and corresponding amounts of much cooling gas are not seen. However, at the center of every cluster with a cooling core, there is a giant cD galaxy. These cD galaxies are the largest galaxies seen in the Universe, and are ~ 10 times larger in mass and radius than other very large galaxies. The cD galaxies at the centers of cooling flows often contain some cool gas and some star formation, although in amounts which are $\lesssim 10\%$ of those expected from the rates of radiative cooling of the X-ray gas.

In nearly every case, these central cD galaxies host radio sources. Recent X-ray observations with the Chandra X-ray Observatory suggest that interactions between the radio source and the X-ray gas in the central regions of clusters strongly affects both components. For example, Figure (6) shows the inner region of the the cluster Abell 2052, which has a cooling core. The central cD galaxy has the strong radio source 3C 317. There are central point sources in X-rays and radio which are coincident with the center of the cD galaxy, which is believed to contain a supermassive black hole. The extended radio emission corresponds with “holes” in the X-ray emission, and the radio source is surrounded by a brightened “shell” of X-ray emission. We refer to these structures as “radio bubbles.” Similar structures are seen in many other cooling core clusters.

The pressures in the X-ray-bright shells are nearly continuous with the pressure of the surrounding gas. There is no clear evidence for strong shocks. Thus, it seems likely that the radio lobes are displacing and compressing the X-ray gas, but are, at the same time, confined by the X-ray gas. The radio bubbles contain relativistic electrons and magnetic fields, and emit radio synchrotron radiation.

The total energy of the the radio plasma is $\sim 10^{59}$ erg. If this energy is eventually dissipated into thermal energy

in the X-ray gas, the energy input would be sufficient to balance cooling for about 10^8 yr, which is about the radiative cooling time of the gas. Thus, energy input from radio sources at the centers of cooling core clusters may partly balance radiative cooling, and may help to explain why only a fraction of the X-ray gas cools to low temperatures.

If there were no physical connection between the radio source and the X-ray emitting plasma in a cluster, then any balance between X-ray cooling and radio source heating would be a coincidence. However, one possibility is that the two are coupled, and form a “feedback loop.” Observations of nearby examples indicate that all large galaxies, such as cluster-central cD galaxies, contain supermassive black holes (SMBHs) with masses of $\sim 10^8 M_\odot$. Radio sources occur when such a SMBH accretes gas from its environment. In radio sources, much of the energy release is converted into the kinetic energy of a pair of oppositely-directed jets of plasma which expand away from the black hole. Consider an inactive SMBH black hole at the center of a cD galaxy in the cooling core of a cluster. If there is no heat source to balance radiative cooling in the X-ray gas, it will cool and flow towards the central supermassive black hole. Part of this gas will be accreted by the supermassive black hole, and the accretion energy (gravitational binding energy) will power radio jets. These will expand into the surrounding X-ray gas, and will inflate two radio bubbles like those in Figure (6). These bubbles will displace and may eventually heat the X-ray gas, balancing its radiative cooling. This will stop the flow of material towards the central SMBH, and eventually turn off the radio source. Heating from the radio source will be unable to prevent cooling of the X-ray gas, and the cycle will start anew. Various arguments suggest that the radio sources in these systems are very strongly active for $\sim 10^7$ yr, and that the cycle

repeats about every $\sim 10^8$ yr.

VI. CONCLUSIONS

Recent observations and theoretical work indicate that the majority of the normal, baryonic matter in the low-redshift Universe is in the form of hot, diffuse plasma. Clusters of galaxies are particularly striking examples of this. Large clusters contain $\sim 10^{14} M_\odot$ of hot plasma at a temperature of $\sim 10^8$ K and typical electron densities of $n_e \sim 10^{-3}$ cm $^{-3}$. The total thermal energy content of this plasma is $\sim 3 \times 10^{63}$ erg. This plasma emits most readily in the X-ray band. Recent X-ray observations with the Chandra X-ray Observatory have confirmed that the clusters are formed by the merger of smaller structures, and that merger shocks heat the gas to high temperatures. Mergers may also accelerate relativistic electrons. In the central regions of clusters, the hot plasma cools radiatively. Recent Chandra X-ray Observatory images indicate that the hot thermal plasma interacts with cluster-central radio sources.

Acknowledgments

I am grateful to Huub Rottgering and Alexey Vikhlinin for their very kind permission to use figures from their publications. Support for this work was provided by the National Aeronautics and Space Administration through Chandra Award Numbers GO1-2123X, GO1-2133X, and GO2-3159X, issued by the Chandra X-ray Observatory Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of NASA under contract NAS8-39073, and by NASA XMM/Newton Grant NAG5-10075.

-
- [1] G. O. Abell, H. G. Corwin, and R. P. Olowin, *Astrophys. J. Suppl.* **70**, 1 (1989).
 - [2] N. A. Bahcall, *Phys. Rep.* **333**, 233 (2000).
 - [3] C. L. Sarazin, *Rev. Mod. Phys.* **58**, 1 (1986).
 - [4] L. J. Spitzer, *Physics of Fully Ionized Gases* (Interscience, New York, 1956).
 - [5] T. E. Clarke, P. P. Kronberg, and H. Böhringer, *Astrophys. J.* **547**, L111 (2001).
 - [6] G. Giovannini and L. Feretti, in *Merging Processes in Galaxy Clusters*, edited by L. Feretti, I. M. Gioia, and G. Giovannini (Kluwer, Dordrecht, 2002), pp. 197–227.
 - [7] H. J. A. Rottgering, M. H. Wieringa, R. W. Hunstead, and R. D. Ekers, *Mon. Not. R. Astron. Soc.* **290**, 577 (1997).
 - [8] R. Fusco-Femiano, D. dal Fiume, L. Feretti, G. Giovannini, P. Grandi, G. Matt, S. Molendi, and A. Santangelo, *Astrophys. J.* **513**, L21 (1999).
 - [9] C. L. Sarazin and R. Lieu, *Astrophys. J.* **494**, L177 (1998).
 - [10] C. L. Sarazin, in *Merging Processes in Galaxy Clusters*, edited by L. Feretti, I. M. Gioia, and G. Giovannini (Kluwer, Dordrecht, 2002), pp. 1–38.
 - [11] J. C. Kempner, C. L. Sarazin, and P. M. Ricker, *Astrophys. J.* **579**, 236 (2002).
 - [12] P. M. Ricker and C. L. Sarazin, *Astrophys. J.* **561**, 621 (2001).
 - [13] R. Blandford and D. Eichler, *Phys. Rep.* **154**, 1 (1987).
 - [14] M. Markevitch and A. Vikhlinin, *Astrophys. J.* **563**, 95 (2001).
 - [15] A. Vikhlinin, M. Markevitch, and S. S. Murray, *Astrophys. J.* **551**, 160 (2001).
 - [16] M. Markevitch, C. L. Sarazin, and A. Vikhlinin, *Astrophys. J.* **521**, 526 (1999).
 - [17] E. L. Blanton, C. L. Sarazin, B. R. McNamara, and M. W. Wise, *Astrophys. J.* **558**, L15 (2001).

X-RAYS FROM CLUSTERS OF GALAXIES

A.C. FABIAN AND S.W. ALLEN

*Institute of Astronomy,
Madingley Road,
Cambridge CB3 0HA,
U.K.*

E-mail: acf,swa@ast.cam.ac.uk

The X-ray emission from clusters of galaxies enables them to be used as good cosmological probes and as an example for massive galaxy formation. The gas mass fraction in clusters should be a universal standard which by means of Chandra observations enables Ω_m to be determined to better than 15 per cent accuracy. Future observations of its apparent variation with redshift will enable Ω_Λ to be measured. The interplay of radiative cooling and heating in cluster cores may reveal the dominant processes acting during the formation of the baryonic part of massive galaxies.

1. Introduction

Clusters of galaxies are luminous X-ray sources, with X-ray luminosities ranging from $10^{43} - 10^{46}$ erg s $^{-1}$. The emission is predominantly thermal bremsstrahlung from hydrogen and helium in the intracluster medium. Line emission, particularly from iron, is also present showing that most of the gas has a mean metallicity of about 0.3 Solar. The total mass of the intracluster medium is about one tenth of the total cluster mass, and about 6 times that of all the stars in the member galaxies. Most of the mass of a cluster is due to dark matter.

Clusters are the most massive bound objects in the Universe and therefore make good cosmological probes. They are the extreme tail of the mass distribution. The number density of clusters in a given mass range is a sensitive measure of the amplitude of the cosmic power spectrum on cluster scales, σ_8 . The gas fraction in clusters, f_{gas} , enables the matter density parameter, Ω_M , to be determined and is a standard measure which should be invariant with redshift. This means that it has a strong potential to be a valuable, independent diagnostic of dark energy, Ω_Λ .

The X-ray emission in the cores of many clusters is sharply peaked.

The radiative cooling time of the gas within 50 kpc of the centre is shorter than the likely age of the cluster. The temperature drops smoothly there by a factor of two to three from that of the outer gas. Although it might seem that a cooling flow should be operating there with gas cooling out of the intracluster medium, spectra from the new generation of X-ray space observatories show that radiative cooling is much reduced and some form of distributed heating is taking place.

This last point is of considerable importance for understanding the gaseous part of galaxy formation, most of which proceeds by radiative cooling of hot gas in dark matter potential wells. The cooling in galaxies predominately occurs in the extreme and far UV and so is not readily observable, but is directly observable in clusters. Whatever is stemming the cooling in clusters may be determining the upper mass cutoff for galaxies.

We review these properties of clusters here with an emphasis on our own recent results.

2. Cosmological parameters from clusters

The equation of hydrostatic equilibrium $dP/dr = -\rho g$, where $n, T, P (= nkT), \rho (= n\mu)m$ and g are the particle density, temperature, pressure, mass density and local gravitational acceleration ($G(M < r)/r^2$), can be rewritten as

$$M_T(< r) = -\frac{kTr^2}{G\mu m} \left(\frac{d \ln n}{dr} + \frac{d \ln T}{dr} \right). \quad (1)$$

The quantities on the right can be measured from X-ray spectral images yielding the gas mass profile $M_{\text{gas}}(r)$, the total mass profile $M_T(r)$ and the gas fraction f_{gas} . Where possible, $M_T(< r)$ can be checked against gravitational lensing data (Fig. 1). Ω_M is then determined using the baryon mass density Ω_b by [50]

$$f_{\text{gas}} = \frac{M_{\text{gas}}}{M_T} = \frac{\Omega_b}{\Omega_M}. \quad (2)$$

A good example is shown by our analysis of the Chandra data on the most luminous cluster known RXJ1347-1145 [6]. The X-ray emission is sharply peaked on the dominant cluster galaxy. Excluding one quadrant which is hotter due to a merging subcluster, we find that the X-ray surface brightness and deprojected temperature profiles are well fit by a hydrostatic model assuming a Navarro-Frenk-White [31] mass profile (Fig. 2). This also agrees with the redshifts of the strong lensing arcs and with weak lensing

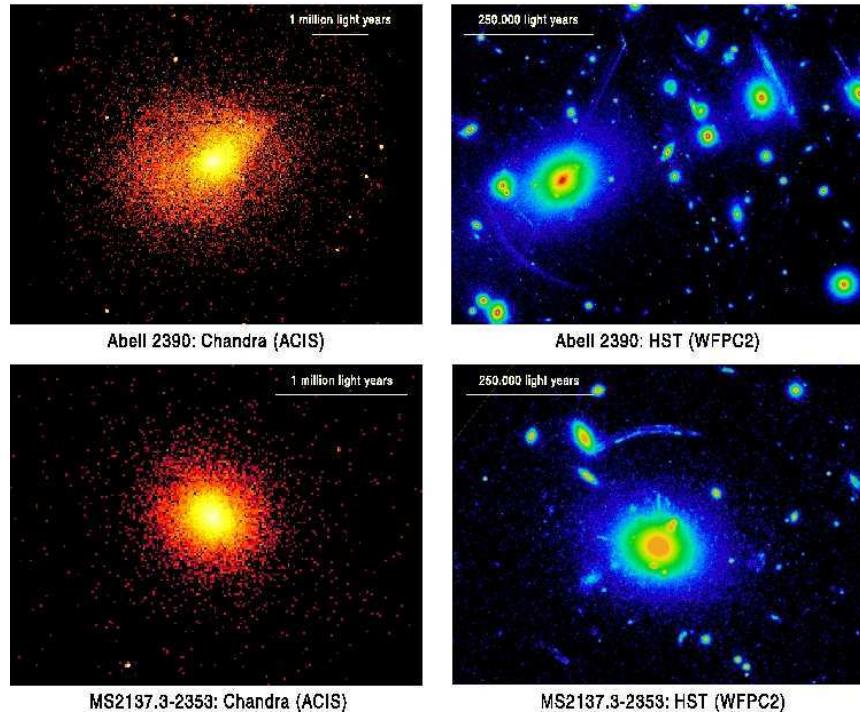


Figure 1. Chandra X-ray (left) and Hubble Space Telescope Wide Field Planetary Camera 2 optical (right) images of two of the dynamically-relaxed, X-ray luminous lensing clusters discussed here. The clusters shown are Abell 2390 ($z = 0.230$) and MS2137.3-2353 ($z = 0.313$). The scale bars indicating distances of 1 million light years correspond to angular sizes of 83 and 67 arcsec for Abell 2390 and MS2137.3-2353, respectively. (A standard Λ CDM cosmology with $h = H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1} = 0.7$, $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ is assumed.) Note the clear gravitational arcs in the HST images.

results. Repeating this for several other massive relaxed clusters shows that the mass profiles are reliably obtained from the X-ray data.

We have measured f_{gas} for 9 massive, relaxed clusters out to the radius where the mean enclosed density is 2500 times the critical mass density of the Universe at the cluster redshift (Fig. 3; [5]). This leads to

$$\Omega_m = \frac{(0.0205 \pm 0.0018)h^{-0.5}}{(0.064 \pm 0.002)(1 + 0.19h^{0.5})} = 0.325 \pm 0.034, \quad (3)$$

where the numerator involves the determination of $\Omega_b h^2$ from deuterium abundance measurements in quasar intergalactic absorption spectra [33]

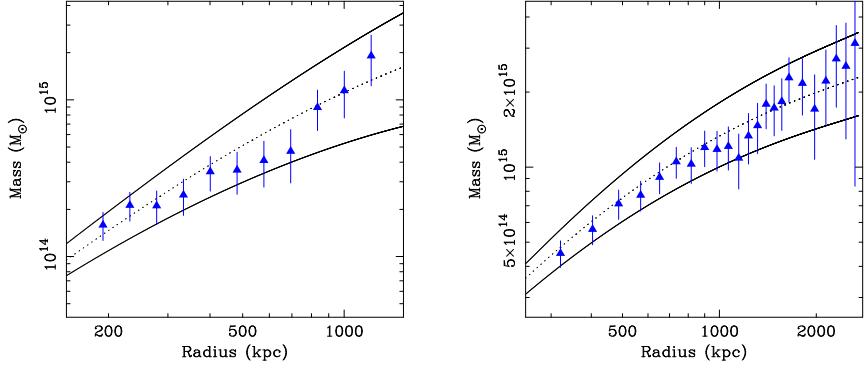


Figure 2. A comparison of the mass measurements obtained from Chandra X-ray observations (solid lines) and wide field weak lensing studies (triangles) of two of the dynamically relaxed clusters in our sample: Abell 2390 (left; [3]) and RXJ1347.5-1145 (right; [6]). Error bars are 68 per cent confidence limits.

and the Hubble constant $h = H_0/100 = 0.72 \pm 0.08$ [24]. The second small factor in the denominator accounts for the baryons in stars.

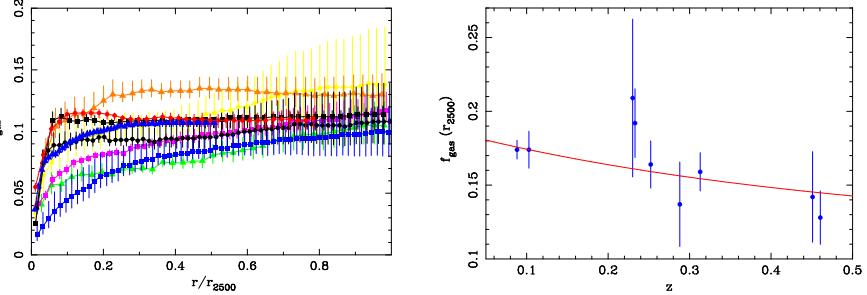


Figure 3. (Left panel) The X-ray gas mass fraction, f_{gas} , as a function of radius (scaled in units of r_{2500}) for the present sample of nine dynamically relaxed clusters observed with Chandra [7]. A Λ CDM cosmology with $h = 0.7$ is assumed. Note how the profiles flatten and converge to universal value within r_{2500} . (Right panel) The apparent redshift variation of the X-ray gas mass fraction measured at r_{2500} (with rms 1σ errors) for a reference SCDM ($h = 0.5$) cosmology. The solid curve shows the predicted $f_{\text{gas}}(z)$ behaviour for the best-fitting, underlying cosmology with $\Omega_m = 0.29$ and $\Omega_\Lambda = 0.68$.

Numerical simulations of clusters show that f_{gas} should be independent of redshift. A small amount of gas may be lost during the assembly of a cluster (remaining fraction 0.93 ± 0.05 ; [8]) but f_{gas} should basically be

a universal ‘standard measure’. We can therefore use this fact to determine the correct cosmological model [42, 34], i.e. adjust the cosmological parameters until we find f_{gas} to be independent of redshift. We find a best fit with $\Omega_m = 0.292^{+0.040}_{-0.036}$, $\Omega_\Lambda = 0.68^{+0.42}_{-0.52}$. The method essentially involves the determination of the angular diameter distance of the cluster D_A ($f_{\text{gas}} \propto D_A^{3/2}$).

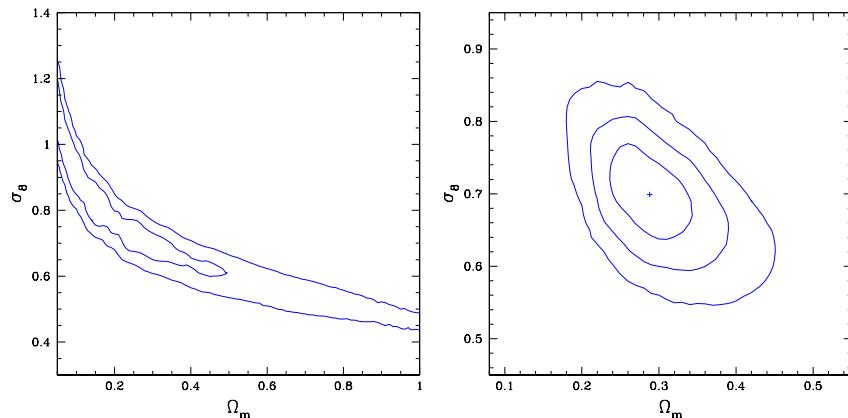


Figure 4. Left: Allowed region in (σ_8, Ω_m) space obtained by applying our mass-temperature relation to the luminosity function of the 120 ROSAT clusters with luminosities above $10^{45} \text{ erg s}^{-1}$. Right: The resultant region when our constraints on Ω_m from f_{gas} are included.

The constraints on Ω_m are competitive with all other methods. Those on Ω_Λ show the potential for now. With another 10 clusters probing the redshift range 0.3–0.8, f_{gas} measurements promise to exceed the precision for Ω_Λ obtained so far by distant supernovae. The importance of clusters is that they complement supernovae and sample the redshift range over which Λ is most effective.

We have also determined the mass–luminosity relation from 17 massive clusters from Chandra observations and weak lensing measurements. In detail the relation is of M_{200} and $L(0.1\text{--}2.4 \text{ keV})$, the latter quantity matching the ROSAT band for which there are now excellent X-ray luminosity functions from the ROSAT All Sky Survey (the eBCS and REFLEX studies). We then use the combined luminosity function for the 120 clusters with $L(0.1\text{--}2.4) > 10^{45} \text{ erg s}^{-1}$ to obtain the cluster mass function. This is then compared with the predicted mass function from the Hubble Volume simu-

lations [18] to yield σ_8 , the rms variation of the present day, linearly evolved, density field, smoothed by a top-hat window function of size $8h^{-1}$ Mpc.

The resulting constraint on σ_8 is a function of Ω_m (Fig. 4). If we now include the $f_{\text{gas}}(z)$ data the degeneracy between σ_8 and Ω_m is broken and we obtain $\sigma_8 = 0.695 \pm 0.042$ [7].

Other important and complementary X-ray cluster work on σ_8 is by Schuecker et al [44] who have also measured the cosmic power spectral shape from the distribution of the REFLEX clusters on the sky, and by Borgani et al [12] who studied the evolution of the cluster luminosity function.

3. Cluster cores

The radiative cooling time within the inner 100 kpc of most cluster cores is less than 10^{10} yr. The gas temperature also drops there by a factor of two to three (Fig. 5). If there is no heating of the gas it should cool out at a rate given by (see [19] for a review)

$$\dot{M} = \frac{2}{5} \frac{L \mu m}{kT} \quad (4)$$

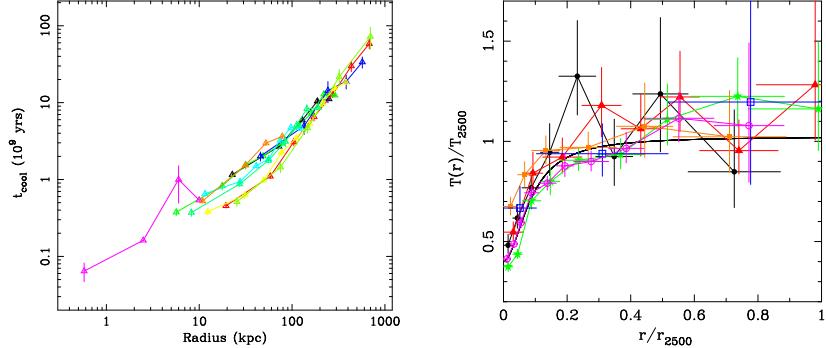


Figure 5. Left: Radiative cooling time versus radius for several cooling flow clusters, from L. Voigt et al (in preparation). Right: Temperature profile for 6 massive clusters [4].

As the gas cools below 1 keV it emits strong Fe L line emission (e.g. FeXV emission at 15 and 19 Å). A major result from the Reflection Grating Spectrometer (RGS) on XMM-Newton was to show that little such emission is seen ([35, 36, 46] Fig. 6). These studies show that the mass cooling rate below about one half to one third of the bulk cluster temperature is less

than one fifth to one tenth of that deduced from the above simple formula. Chandra data are also in agreement with this result.

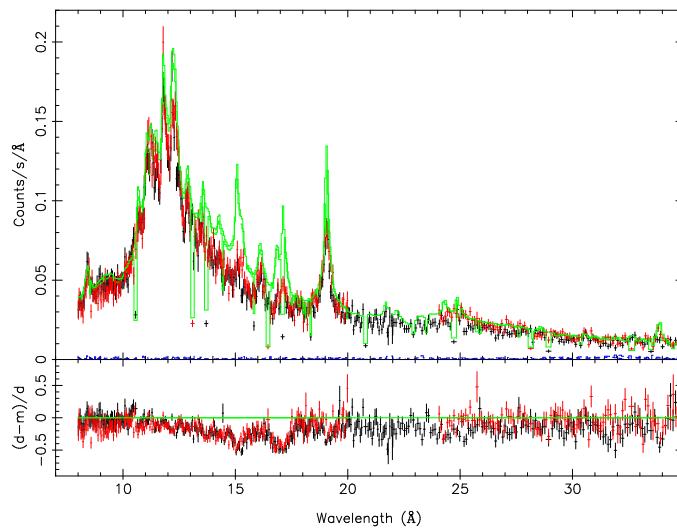


Figure 6. RGS X-ray spectrum of the inner parts of the Virgo cluster around M87 [40]. The faint line shows the predicted emission if gas is cooling to below 10^6 K. The data clearly shows much less emission than predicted by this model in the 13–19 Å region.

This result was present in previous ASCA and ROSAT studies [2] but the lack of emission had been attributed to intrinsic absorption. Although some absorption has not been completely ruled out, the improved new spectra show that it is not dominant.

It is most likely that some heating is taking place, for which there are two plausible candidates. These are heating by a central active nucleus and heating by conduction from the hot outer gas.

All of the relevant *cooling flow* clusters peak on a central galaxy which is expected to host a massive black hole. Many of these galaxies have radio sources, some of which are obviously blowing bubbles of relativistic plasma in the central regions (Perseus, [20], Fig. 7; A2052, [10]). The energy flux from the radio source can be high (10^{43-45} erg s $^{-1}$). The difficulties are however that the X-ray coldest observed gas lies around the bubbles, not all clusters host powerful enough radio sources and that the mechanism for heat transfer from bubbles to the surrounding gas are unclear, despite several impressive computational studies (e.g. [13, 38, 37]).

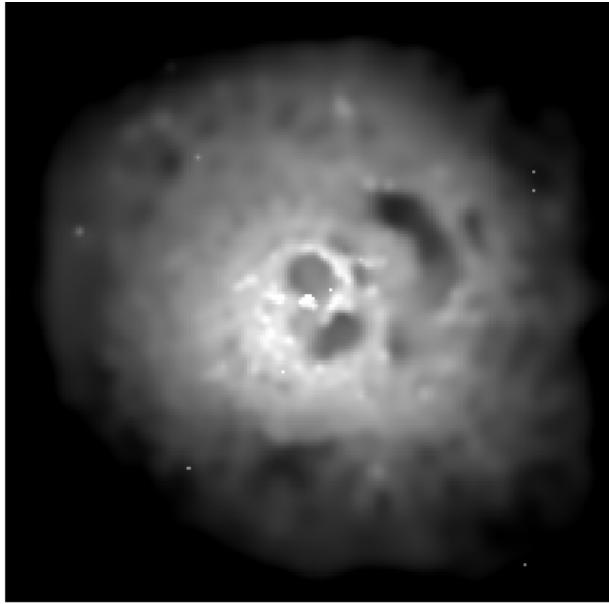


Figure 7. Adaptively smoothed Chandra X-ray image of the core of the Perseus cluster. Note the holes, which coincide with the radio lobes, above and below the nucleus. A buoyant outer bubble is seen to the right. From [20].

The heat must be distributed ([21, 26] Fig. 8) and cannot just heat the innermost, coolest gas (Fig. 8; [26]). Some [9, 45] have argued that radio source activity may be sporadic, which explains why there is little correlation between the present radio source activity and the heating requirement. However that makes very strong demands on the power of the source when it is switched on, particularly in the high luminosity clusters (Fig. 9). There it must exceed $10^{46} \text{ erg s}^{-1}$, which will hardly be contained in simple bubbles.

Note that strong abundance gradients are found in many clusters ([25, 16]), often peaking at radii $\sim 30 \text{ kpc}$ ([41] Fig. 10; [26]), limiting the degree of large scale disturbance which can take place in the central regions.

Thermal conduction has long been considered to be suppressed in clusters because of the observed central temperature drops. Conductive energy flow increases strongly with temperature, unlike radiative cooling which decreases (at constant pressure), so one might assume that it either operates, so making the core isothermal, or is suppressed and radiative cooling dominates. Narayan & Medvedev [30] have however revived the concept

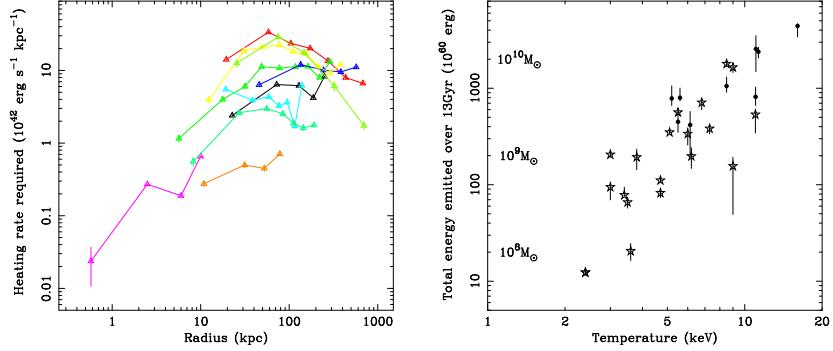


Figure 8. Left: The heating rate per unit radius required to stop radiative cooling in several cooling flow clusters (L. Voigt et al in preparation). The heating needs to be distributed. Right: The total energy radiated over a Hubble time from within the cooling region for a selection of cooling flow clusters. The masses indicate the total mass which must accrete to produce this energy if the efficiency of energy release is 0.1. If the central radio source stems cooling in the hotter clusters then most of the power released must be channeled into heating the intracluster medium [23].

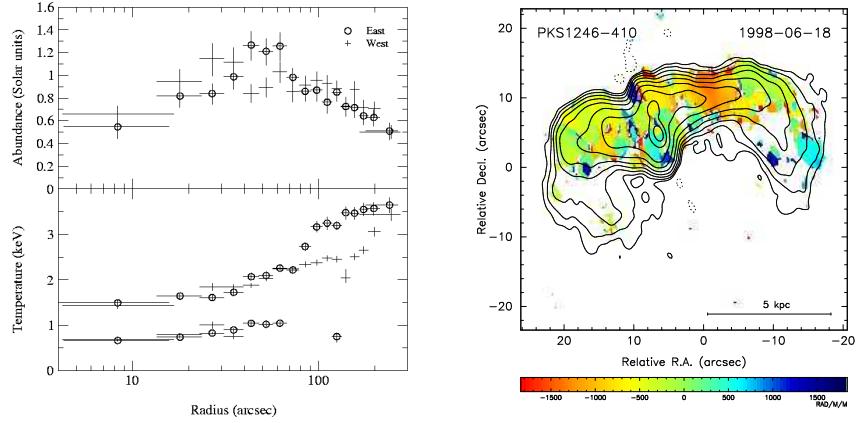


Figure 9. Left: Abundance (top) and temperature (bottom) profiles for the Centaurus cluster [41]. Right: Faraday rotation measure map for the Centaurus cluster [47].

and noted that conduction may account for the observed temperature gradients. This has been explored in more detail by [48, 23, Fig. 11] and [51]. These last authors found some clusters where conduction appears to be in-

sufficient. A major issue here is whether the effective conductivity can be as high as the Spitzer value, or whether magnetic fields suppress it heavily (Narayan & Medvedev [30] argue that it may operate at close to the Spitzer rate). Faraday rotation indicates the presence of magnetic fields in the intracluster medium ([47] Fig. 10).

Ruszkowski & Begelman [39] have incorporated both distributed heating by a radio source with conduction to obtain stable solutions. This still leaves the method by which the heat is distributed as unsolved.

It is unlikely that either radio source heating or conduction can suppress radiative cooling within such a large region completely. Indeed, this is unnecessary, for significant rates of massive star formation [1, 15] and masses of cold gas [17] are found in the central parts. It is probable that

$$\dot{M}_{\text{cool}} = \dot{M}_X/10, \quad (5)$$

where \dot{M}_X is the cooling rate derived simply from the X-ray data on the assumption of no heating.

In summary there are plausible heat sources at the centre and beyond radii of 100 kpc. The main problem is to distribute the energy within 100 kpc without either disrupting the metallicity profiles or exceeding some observational constraint. Beyond any bubbles and plumes, and an occasional cold front (all of which occupy only a small fraction of the volume of the cooler gas) the distribution of surface brightness, temperature, metallicity and entropy of the gas all vary very smoothly.

Further possibilities remain in which cooling dominates but the situation is more complicated so that gas cooling below say 2 keV less observable. This can result if the metals are not uniformly mixed in the hot gas [21, 29] or if the cooler gas mixes with cold gas [22]. The missing soft X-ray luminosity from a simple cooling flow is similar to that in the optical/UV/IR nebulosity at the centre. One reason to continue to consider such models is the detection of strong OVI emission in some clusters [32].

The cooling flow problem, as it has become known, has wider relevance than just to cluster cores. The visible parts of galaxies are due to gas cooling in dark matter potential wells [49, 28] and the cores of clusters are a directly observable example of this process. If it does not operate in cluster cores why does it work in galaxies? It is possible that whatever is stemming cooling in clusters does so in galaxies but operates in such a way as to dominate only in massive systems, so determining the upper mass limit of visible galaxies. A process like conduction, which is more effective in hotter, massive objects, has the right property to allow gas to cool in

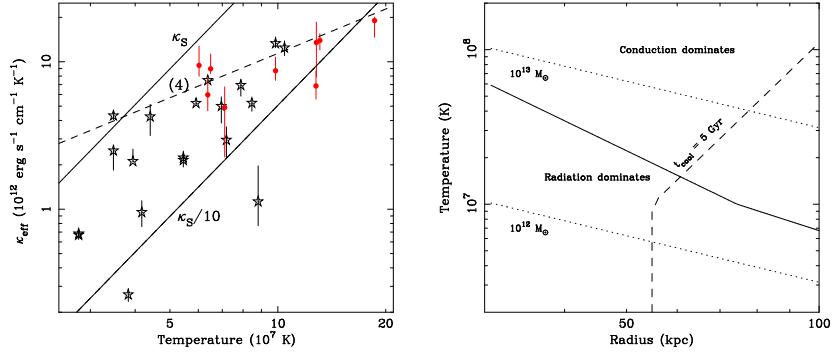


Figure 10. Left: Effective conductivity required to balance radiative cooling in a sample of the brightest clusters [23]. Right: Conduction can only offset radiative cooling in very massive galaxies, and so may determine the upper mass limit of galaxies [23].

normal galaxies but not in more massive systems ([23] Fig. 11).

Although cluster cores are complex regions the superb spatial resolution of Chandra means that it can be separated from the bulk of the cluster and need not affect the cosmological determinations outlined earlier.

4. Acknowledgements

We thank Robert Schmidt and Jeremy Sanders for help and collaboration with these projects.

References

1. Allen S.W., 1995, MNRAS, 276, 947
2. Allen S.W., Fabian A.C., Johnstone R.M., Arnaud K.A., Nulsen P.E.J., 2001, MNRAS, 322, 589
3. Allen, S.W., Ettori S., Fabian, A.C. 2001, MNRAS, 324, 877
4. Allen, S.W., Schmidt, R.W., Fabian, A.C. 2001, MNRAS, 328, L37
5. Allen, S.W., Schmidt, R.W., Fabian, A.C. 2002a, MNRAS, 334, L11
6. Allen, S.W., Schmidt, R.W., Fabian, A.C. 2002b, MNRAS, 335, 256
7. Allen, S.W., Schmidt, R.W., Fabian, A.C., Ebeling H., 2002c, MNRAS, submitted (astro-ph/0208394)
8. Bialek J.J., Evrard A.E., Mohr J.J., 2002, 578, L9
9. Binney J., Tabor G., 1995, MNRAS, 276, 663
10. Blanton E.L., Sarazin C.L., McNamara B.R., Wise M.W., 2001, ApJ,
11. Böhringer H., Matsushita K., Churazov E., Ikebe Y., Chen Y., 2002, A&A, 382, 804
12. Borgani S. et al., 2001, ApJ, 561, 13
13. Brüggen M., Kaiser C.R., 2002, Nature, 418, 301

14. Churazov E., Brüggen M., Kaiser C.R., Böhringer H., Forman W., 2001, ApJ, 554, 261
15. Crawford C.S., Allen S.W., Ebeling H., Edge A.C., Fabian A.C., 1999, MNRAS, 306, 875
16. De Grandi S., Molendi S., 2001, ApJ, 551, 153
17. Edge A. C., 2001, MNRAS, 328, 762
18. Evrard A.E. et al., 2002, ApJ, 573, 7
19. Fabian A.C., 1994, ARAA, 32, 277
20. Fabian A. C. et al., 2000, MNRAS, 318, L65
21. Fabian A.C., Mushotzky R.F., Nulsen P.E.J., Peterson J., 2001, MNRAS, 321, L20
22. Fabian A.C., Allen, S.W., Crawford, C.S., Johnstone R.M., Morris, R.G., Sanders J.S., Schmidt R.W., 2002, MNRAS, 332, L50
23. Fabian A.C., Voigt L., Morris R.G., 2002, MNRAS, 335, L71
24. Freedman W. et al., 2001, ApJ, 553, 47
25. Fukazawa Y. et al., 1994, PASJ, 46, L55 224, 75
26. Johnstone R. M., Allen S.W., Fabian A. C., Sanders J.S., 2002, MNRAS, 336, 299
27. Freedman W. et al., 2001, ApJ, 553, 47
28. Kauffmann G., Colberg J.M., Diaferio A., White S.D.M. 1999, MNRAS, 303, 188
29. Morris R.G., Fabian A.C., 2003, MNRAS, 338, 824
30. Narayan R., Medvedev M. V., 2001, ApJ, 562, L129
31. Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
32. Oegerle W.R., et al 2001, ApJ, 560, 187
33. O'Meara J.M. et al., 2001, ApJ, 552, 718
34. Pen, U. 1997, New Ast., 2, 309
35. Peterson J. R. et al., 2001, A&A, 365, L104
36. Peterson J.A. et al, 2002, astro-ph/0202108
37. Quilis V., Bower R.G., Balogh M.G., 2001, MNRAS, 328, 1091
38. Reynolds C. S., Heinz S., Begelman M. C., 2002, MNRAS, 332, 271
39. Ruszkowski M., Begelman M.C., 2002, ApJ, 581, 223
40. Sakelliou, I., et al 2002, A&A, 391, 903
41. Sanders J. S., Fabian A. C., 2002, MNRAS, 331, 273
42. Sasaki, S. 1996, PASJ, 48, L119
43. Schmidt R.W., Allen S.W., Fabian A.C., 2001, MNRAS, 327, 1057
44. Schuecker P., Böhringer H., Collins C.A., Guzzo L., 2003, A&A, 398, 867
45. Soker N., Blanton E.L., Sarazin C.L., 2002, ApJ, 573, 533
46. Tamura T. et al., 2001, A&A, 365, L87
47. Taylor G.B., Fabian A.C., Allen S.W., 2002, MNRAS, 334, 769
48. Voigt L. M., Schmidt R. W., Fabian A. C., Allen S. W., Johnstone R. M., 2002, MNRAS, 335, L7
49. White S.D.M., Rees M.J., 1978, MNRAS, 183, 341
50. White, S.D.M., Navarro J.F., Evrard A.E., Frenk C.S., 1993, Nature, 366, 429
51. Zakamska N.L., Narayan R., 2003, ApJ, 582, 162

Tracing cosmic evolution with clusters of galaxies

G. Mark Voit

Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824

(Dated: September 10, 2004)

The most successful cosmological models to date envision structure formation as a hierarchical process in which gravity is constantly drawing lumps of matter together to form increasingly larger structures. Clusters of galaxies currently sit atop this hierarchy as the largest objects that have had time to collapse under the influence of their own gravity. Thus, their appearance on the cosmic scene is also relatively recent. Two features of clusters make them uniquely useful tracers of cosmic evolution. First, clusters are the biggest things whose masses we can reliably measure because they are the largest objects to have undergone gravitational relaxation and entered into virial equilibrium. Mass measurements of nearby clusters can therefore be used to determine the amount of structure in the universe on scales of $10^{14}\text{-}10^{15} M_{\odot}$, and comparisons of the present-day cluster mass distribution with the mass distribution at earlier times can be used to measure the rate of structure formation, placing important constraints on cosmological models. Second, clusters are essentially “closed boxes” that retain all their gaseous matter, despite the enormous energy input associated with supernovae and active galactic nuclei, because the gravitational potential wells of clusters are so deep. The baryonic component of clusters therefore contains a wealth of information about the processes associated with galaxy formation, including the efficiency with which baryons are converted into stars and the effects of the resulting feedback processes on galaxy formation. This article reviews our theoretical understanding of both the dark-matter component and the baryonic component of clusters, providing a context for interpreting the flood of new cluster observations that are now arriving from the latest generation of X-ray observatories, large optical surveys, and measurements of cluster-induced distortions in the spectrum of the cosmic microwave background.

Contents

I. Introduction	2	6. Weak Lensing and Mass	27
II. Observable Properties of Clusters	3	7. Baryons and Mass	28
A. Clusters in Optical Light	3	D. Evolution of the Mass Function	28
1. Optical Richness	3	1. Dependence on Cosmology	28
2. Galaxy Velocities	3	2. Evolution of the Observables	29
3. Gravitational Lensing	5	3. Constraints on Dark Matter	30
B. Clusters in X-rays	5	4. Constraints on Dark Energy	33
1. X-ray Surface Brightness	6		
2. Plasma Temperature	6		
3. Measuring Abundances	7		
C. Clusters in Microwaves	9	IV. Evolution of the Baryonic Component	34
1. The S-Z Effect	9	A. Structure Formation and Gravitational Heating	35
2. Comparing S-Z with X-ray	10	1. Intracluster Entropy	35
III. Evolution of the Dark Component	10	2. Entropy Generation by Smooth Accretion	35
A. A Recipe for the Universe	10	3. Entropy Generation by Hierarchical Merging	36
1. Global Dynamics	11	4. Observed Entropy Profiles	37
2. Global Geometry	12	B. Galaxy Formation and Feedback	38
3. Density Perturbations	13	1. Preheating	38
4. Growth of Linear Perturbations	14	2. Radiative Cooling	39
5. The CDM Power Spectrum	15	3. Feedback from Supernovae	40
6. Power Spectrum Normalization	16	4. Feedback from Active Galactic Nuclei	40
7. Summary of Cosmological Parameters	16	5. Transport Processes	41
B. Cluster Formation	18	C. Galaxy Formation and Cluster Observables	41
1. Spherical Collapse	18	1. Role of Cooling	42
2. Cluster Mass Profiles	19	2. Role of Feedback	44
3. Defining Cluster Mass	19	3. Role of Smoothing	44
4. Cluster Mass Function	20	4. Predictions for Evolution	44
5. Cluster Bias	22	D. Constraints on Baryon Condensation	45
C. Measuring the Cluster Mass Function	23	1. Mass and Light in Clusters	46
1. Linking Mass with Observables	23	2. Intergalactic Stars	46
2. Mass-Temperature Relation	24	3. Global S-Z Effect	46
3. Mass-Luminosity Relation	25	4. Cooling Flows in Clusters	47
4. Mass-Richness Relation	26		
5. Velocity Dispersion and Mass	27	V. Concluding Remarks	47
		Acknowledgments	48
		References	48

I. INTRODUCTION

Cosmology has recently reached an important milestone. A wide variety of cosmological observations now support a single model for the overall architecture of the observable universe and the development of galaxies and other structures within it. According to this so-called concordance model, the geometry of the observable universe is indistinguishable from a flat geometry, implying that its total energy density is very close to the critical density needed to close the universe. The two dominant components of the universe appear to be a non-baryonic form of dark matter, whose gravity is responsible for structure formation, and a mysterious form of dark energy, whose pressure is currently causing the expansion of the universe to accelerate. The mean density of baryonic matter is about 15% of the total amount of matter, and we can observe the baryonic matter only because the gravitational attraction of non-baryonic dark matter has drawn the baryonic gas into deep potential wells, where a small fraction of it condenses into stars and galaxies.

This model explains many different features of the observable universe, but it is not entirely satisfying because the nature of the dark matter and the provenance of the dark energy remain unknown. The implications of dark energy for fundamental physics are particularly serious, so we need to be sure that it is absolutely necessary to explain the astronomical observations. In addition, many aspects of galaxy formation remain poorly understood. Dark-matter models successfully account for the spatial distribution of mass in the universe, as traced by the galaxies, but they do not explain all the properties of the galaxies themselves. Dark matter initiates the process of galaxy formation, but once stars begin to form, supernova explosions and disturbances wrought by supermassive black holes can inhibit further star formation by pumping thermal energy into the universe's baryonic gas.

Clusters of galaxies are a particularly rich source of information about the underlying cosmological model, making possible a number of critical tests. According to the concordance model, clusters are the largest and most recent gravitationally-relaxed objects to form because structure grows hierarchically. The universe begins in a state of rapid expansion whose current manifestation is Hubble's Law relating a galaxy's recessional velocity v_r to its distance d through Hubble's constant H_0 : $v_r = H_0 d$. Generalizing this feature of the local universe to all of observable space links an object's cosmological redshift $z = (\lambda_{\text{observed}}/\lambda_{\text{rest}}) - 1$ with a unique time $t(z)$ since the Big Bang, enabling us to probe the evolution of the universe with observations of distant objects.¹ Grav-

ity drives structure formation in this expanding realm because the matter density is nearly equal to the critical density during much of cosmic history. Regions whose density slightly exceeds the mean density are therefore gravitationally bound and eventually decouple from the expansion, collapse upon themselves, and enter a state of virial equilibrium in which the mean speeds of the component particles are approximately half the escape velocity. Because density perturbations in the concordance model have greater amplitudes on smaller length scales, small sub-galactic objects are the first to decouple, collapse, and virialize. These small objects then collect into galaxies, and galaxies later collect into clusters of galaxies, whose masses now top out at roughly 10^{15} times that of the Sun's ($10^{15} M_\odot$). Thus, the growth and development of clusters directly traces the process of structure formation in the universe.

Section II outlines the observable properties of galaxy clusters that enable us to measure their masses. Observables in the optical band include the overall luminosity of a cluster's galaxies, which scales with the overall mass, the velocity dispersion of a cluster's member galaxies, which responds to the depth of the cluster's potential well, and gravitational lensing of background galaxies by the cluster's potential. Observables in the X-ray band include the overall X-ray luminosity of a cluster, coming from the hot gas trapped in the cluster's gravitational potential, the temperature inferred from the X-ray spectrum of that gas, and the abundances of various elements inferred from the emission lines in that spectrum. This hot gas also leaves an imprint on the microwave sky because its electrons Compton scatter the photons of the cosmic microwave background radiation. Microwave observations are therefore an alternative source of information about the hot gas and its temperature.

Once we have measured the masses of a sample of clusters, we can use that sample to study cosmology. Section III explains how the characteristics of the cluster population relate to cosmological models. It begins by summarizing the elements of the concordance model and provides a number of useful analytical approximations to the results of numerical simulations of cluster formation based on this model. Then it covers the dicey middle ground linking those simulations with observations, currently the main source of uncertainty in deriving cosmological parameters from cluster observations. The section concludes with a look at the evolution observed in the cluster population and the constraints that cluster evolution places on cosmological models.

Section IV takes up the subject of the baryonic component of clusters, with two purposes in mind. First, in order to improve the precision of cosmological measurements with clusters, we need to know how the process of galaxy formation affects the relations used to derive cluster masses from observations of a cluster's hot gas and galaxies. Current numerical simulations accurately reproduce the behavior of the dark component, whose interactions are purely gravitational, but fail to reproduce

¹ In this definition, λ_{rest} is the wavelength of a photon emitted by a distant object and $\lambda_{\text{observed}}$ is the wavelength it is observed to have when it reaches Earth.

with similar accuracy the observed behavior of the baryonic component, whose interactions are also hydrodynamical and thermodynamical. These discrepancies between simulations and observations indicate that galaxy formation alters the state of a cluster's hot gas in a way that preserves information about the poorly understood feedback processes that regulated galaxy formation long before the cluster reached its present state. Our second purpose is therefore to try to decipher what the state of the hot gas is saying about the process of galaxy formation, so as to gain insight into those feedback processes. Section V concludes the review with some brief remarks about ongoing and future cluster surveys.

Despite this article's length, it falls somewhat short of being a comprehensive review of cluster physics, which would require more pages than this journal is inclined to provide. Instead, I have tried to assemble a readable introduction to cluster evolution for non-experts, concentrating on the middle ground connecting theory to observations and distilling the key theoretical results into a set of simple analytical tools useful to observers. For more on the subject of clusters and their evolution, readers should consult Sarazin (1988), Borgani *et al.* (2002b), Rosati *et al.* (2002), and Mulchaey *et al.* (2004).

II. OBSERVABLE PROPERTIES OF CLUSTERS

Clusters of galaxies might have been called something different if they had first been discovered in a waveband other than visible light, because all of the stars in all of a cluster's galaxies represent only a small fraction of a cluster's overall mass. Clusters contain substantially more mass in the form of hot gas, observable with X-ray and microwave instruments. This section outlines how clusters are observed in all three of these wavebands and how those observations reveal a cluster's total mass, which turns out to be about seven times the combined baryonic mass in stars and hot gas (Allen *et al.*, 2002; David *et al.*, 1995; Evrard, 1997; White *et al.*, 1993).

A. Clusters in Optical Light

Optical identification of galaxy clusters has been going on for quite a long time. By the end of the eighteenth century Charles Messier (1784) and William Herschel (1785) had already recognized concentrations of galaxies in the constellations Virgo and Coma Berenices. Today these clusters of galaxies are known as the Virgo cluster and the Coma cluster. Optical discoveries of clusters continued to accumulate as observing power grew over the next two centuries (see Biviano, 2000, for a review of the history), culminating with the definitive cluster catalogs of George Abell and collaborators (Abell, 1958; Abell *et al.*, 1989). Abell's catalogs contain most of the known nearby galaxy clusters and are the foundation for much of our modern understanding of clusters.

Abell recognized that projection effects can complicate the identification of clusters in optical galaxy surveys and therefore was careful in defining his clusters. Working from the Palomar Sky Survey plates he estimated the distance of each cluster candidate from the apparent brightness of its tenth brightest member galaxy. He then counted all the galaxies lying within a fixed projected radius and brighter than a magnitude limit two magnitudes fainter than the third brightest member. The bounding radius, which he determined from the distance estimate, is now known to be ~ 2 Mpc and was the same for all clusters.² In order to compensate for projection effects, he subtracted from his galaxy counts a background level equivalent to the mean number of galaxies brighter than the magnitude limit for the cluster in similarly-sized, cluster-free regions of the plate, and retained all cluster candidates with a net excess of 50 galaxies brighter than the limiting magnitude.

Most of the optical cluster identification techniques used today extend and refine Abell's basic approach (e.g., Dalton *et al.*, 1997; Lumsden *et al.*, 1992; Postman *et al.*, 1996), often augmenting it with information about galaxy colors (e.g., Bahcall *et al.*, 2003b; Gladders and Yee, 2000; Nichol, 2004). These improvements are necessary because the contrast of clusters against the background galaxy counts decreases with cluster distance. Galaxy colors can help identify distant clusters because many cluster galaxies are significantly redder than other galaxies at a similar redshift, owing to their lack of ongoing star formation. The colors of their aging stellar populations therefore place these cluster members on a narrow and distinctive locus known as the "red sequence" in a plot of galaxy color versus magnitude (e.g. Gladders and Yee, 2000).

Once suitable cluster candidates are found, their status as true mass concentrations can be checked by measuring the underlying mass. Optical observations offer two complementary ways to perform such measurements, through the orbital velocities of the member galaxies and through the degree to which galaxies lying behind the cluster are lensed by the cluster's gravitational potential. We will discuss both of these methods after a few more words about how galaxy counts relate to the overall optical luminosities of clusters.

1. Optical Richness

To the extent that light traces mass in the universe, the total optical luminosity of a cluster is itself an indicator of a cluster's mass. Measuring the luminosity of every galaxy in a cluster is impractical, especially for distant clusters in which only the brightest galaxies can be

² The Megaparsec is astronomers' favored unit of distance on cluster scales: $1 \text{ Mpc} = 3.09 \times 10^{24} \text{ cm} = 3.26 \times 10^6 \text{ light years}$.

observed. However, because the luminosity distribution function of cluster galaxies is nearly the same from cluster to cluster, observing the high-luminosity tip of that distribution allows one to normalize the overall galaxy luminosity function for the cluster, yielding estimates for both the cluster's total optical luminosity and its mass.

Abell's catalogs encode this information by placing clusters in categories of "richness" corresponding to the net excess of galaxies brighter than the magnitude limit used to define each cluster. The richest clusters (class 5) contain over 300 galaxies brighter than the magnitude limit, while the poorest (class 1) contain only 50-79 such galaxies. Clusters not quite making Abell's cut (30-49 galaxies above the magnitude limit) were assigned to richness class zero. Within this system, the Coma cluster originally ranked as richness class 2.

Invoking assumptions about the shape of the luminosity distribution function helps to link richness more directly to a cluster's total luminosity. Cluster galaxies generally adhere to a luminosity distribution function following the form proposed by Schechter (1976), with the number of galaxies in luminosity range dL about L proportional to $L^{-\alpha} \exp(-L/L_*)$, with $\alpha \approx 1$ (e.g., Balogh *et al.*, 2001a). Assuming this distribution function, Postman *et al.* (1996) define a richness parameter Λ_{cl} equivalent to the number of cluster galaxies brighter than the characteristic luminosity L_* . They find that Λ_{cl} is highly correlated with Abell's richness measure, but the scatter between richness and Λ_{cl} is large.

Another richness parameter in current use is B_{cg} , the amplitude of the correlation function between the cluster center and the member galaxies (Longair and Seldner, 1979; Yee and López-Cruz, 1999). It is derived from the angular correlation function of galaxies measured down to a given magnitude limit, after removing the background counts, and is normalized by dividing out the expected luminosity distribution function of galaxies integrated down to that magnitude limit. This richness parameter also correlates with Abell's richness, but again the scatter is broad. Yee and Ellingson (2003) show that B_{cg} correlates well with other global properties of clusters, suggesting that richness observations may become an inexpensive way to measure cluster masses, but first the mass-richness relation must be calibrated and the scatter in that relation must be quantified.

2. Galaxy Velocities

Once a cluster has been optically identified, obtaining the radial velocities v_r of the cluster galaxies from their redshifts helps in mitigating projection effects and in measuring the cluster's mass. Because the velocity distribution of a relaxed cluster's galaxies is expected to be gaussian in velocity space, galaxies with velocities falling well outside the best-fitting gaussian envelope are unlikely to be cluster members and are generally discarded. Fitting the velocity distribution $\exp[-(v_r - \langle v_r \rangle)^2 / 2\sigma_{1\text{D}}^2]$

to the remaining galaxies then yields a one-dimensional velocity dispersion $\sigma_{1\text{D}}$ for the cluster. If the velocity distribution of a cluster candidate is far from gaussian, then it is probably not a real cluster but rather a chance superposition of smaller structures. Obviously, the accuracy of $\sigma_{1\text{D}}$ depends critically on the number of galaxies with measured velocities and the method for identifying and eliminating non-members.

Zwicky (1933, 1937) was the first to measure a cluster's velocity dispersion, finding $\sigma_{1\text{D}} \sim 700 \text{ km s}^{-1}$ for the Coma cluster. He correctly concluded from this fact and his estimate of the Coma cluster's overall radius that this cluster's mass must be far greater than the observed mass in stars—the first evidence for dark matter in the universe. Shortly thereafter, Smith (1936) showed that the same was true of the Virgo cluster. Zwicky's reasoning involved the virial theorem of classical mechanics, which applies to steady, gravitationally bound systems. Differentiating the system's moment of inertia $I = \sum_i m_i \mathbf{r}_i^2$ twice with respect to time and setting the result to zero produces the virial relation

$$\sum_i m_i \dot{\mathbf{r}}_i^2 = - \sum_i m_i \ddot{\mathbf{r}}_i \cdot \mathbf{r}_i . \quad (1)$$

The left-hand side is twice the total kinetic energy of the cluster's particles, and in a spherically symmetric system of mass M with a gaussian velocity distribution, that kinetic energy is $3M\sigma_{1\text{D}}^2/2$. If the system is isolated, then the right-hand side is equal to the absolute value of the gravitational potential energy, which can be expressed as GM^2/r_G , where

$$\begin{aligned} r_G &\equiv M^2 \left(\sum_i \sum_{i < j} \frac{m_i m_j}{r_{ij}} \right)^{-1} \\ &\approx \frac{\pi}{2} M^2 \left(\sum_i \sum_{i < j} \frac{m_i m_j}{r_{\perp,ij}} \right)^{-1} \end{aligned} \quad (2)$$

and r_{ij} is the separation between particles i and j . The approximation gives r_G for a spherically symmetric system in terms of the projected particle separations $r_{\perp,ij}$ in the plane of the sky (Limber and Mathews, 1960). According to the virial theorem, the mass of a spherical, isolated cluster should therefore be $M = 3\sigma_{1\text{D}}^2 r_G / G$.

Applying this virial analysis to real clusters is not quite so simple because clusters are not isolated systems—there is no clean boundary separating a cluster from the rest of the universe. Segregating the cluster from the outlying regions with an arbitrary bounding surface alters the interpretation of the right-hand side of equation (1). In a steady state, the momentum flux of particles exiting the boundary is equal to that entering, so the bounding surface is formally equivalent to a reflecting wall that adds a pressure correction term, offsetting some of the gravitational potential energy (Carlberg *et al.*, 1997a; The and White, 1986). One must also account for

objects seen in projection, such as infalling galaxies that have not yet entered into virial equilibrium and interlopers that are not true cluster members, problems that have led to the invention of various kinds of projected mass estimators (Bahcall and Tremaine, 1981; Heisler *et al.*, 1985).

Extensive redshift measurements now allow observers to measure much more than just a cluster's velocity dispersion, enabling detailed studies of a cluster's mass profile and dynamical state. Generally a cluster's velocity dispersion declines with projected radius, implying that the relationship between projected radius and the mass enclosed within that radius is somewhat shallower than linear in the cluster's outskirts (Carlberg *et al.*, 1997a; Kent and Gunn, 1982; Rood *et al.*, 1972). Beyond the approximate virial radius of a cluster, the enclosed mass continues to increase and the galaxies move primarily on infalling radial trajectories (Biviano and Girardi, 2003; Diaferio and Geller, 1997; Kaiser, 1987; Regos and Geller, 1989). Even farther out is a thin region where galaxies are nearly stationary with respect to the cluster because there the cluster's gravity has just succeeded in reducing the outward Hubble flow to a standstill (Kaiser, 1987; Rines *et al.*, 2003). Eventually these galaxies will fall back toward the cluster and become cluster members.

Because clusters are dynamical systems that have not quite finished forming and equilibrating, the velocity dispersion and virial theorem by themselves do not yield an exact cluster mass measurement. Detailed information on the spatial distribution of galaxy velocities is of great help in measuring the masses of large, nearby clusters but similar information is very difficult to obtain for the distant clusters so interesting to cosmologists. In lieu of detailed observations, one can use simulations of cluster formation to calibrate the approximate virial relationship between velocity dispersion and cluster mass, but we will postpone discussion of that procedure to the discussion of dark-matter dynamics in Sec. III.

3. Gravitational Lensing

In his remarkable 1937 paper on the Coma cluster, Zwicky also proposed that cluster masses could be measured through gravitational lensing of background galaxies. That technique did not become practical for six more decades but is now one of the primary methods for measuring cluster mass. Lensing is sensitive to the cluster's mass within a given projected radius r_{\perp} because the mass within this radius deflects photons toward our line of sight through the cluster's center. When the deflection angle is small compared to a background galaxy's angular distance from the cluster center, weak lensing shifts each point in the galaxy's image to a slightly larger angular distance from the cluster's center, thereby distorting the image by stretching it tangentially to r_{\perp} . Measuring the weak-lensing distortion of any single galaxy is nearly im-

possible because the exact shape of the unlensed galaxy is generally unknown. Instead, observers must measure the shear distortion of an entire field of background galaxies, under the assumption that any intrinsic deviations of galaxy images from circular symmetry are uncorrelated.

Many excellent articles explain this weak-lensing technique in more detail (Bartelmann and Schneider, 2001; Hoekstra *et al.*, 1998; Kaiser and Squires, 1993; Mellier, 1999; Tyson *et al.*, 1990). Here we wish only to give a flavor of how a cluster's mass can be measured from the lensing it induces. The deflection angle itself depends on the gradient of the gravitational potential in the lensing system, meaning that a mass sheet of constant surface density produces no shear and goes undetected. Additional mass that is distributed symmetrically about the line of sight through a cluster's center bends photon paths by an angle twice that expected from Newtonian physics, $4GM(< r_{\perp})/c^2r_{\perp}$, which can be measured from the shear distortion and redshift distribution of the background galaxies. Obtaining a cluster mass from the mass $M(< r_{\perp})$ along the column bounded by r_{\perp} requires additional assumptions about how mass is distributed within this column. A particularly simple mass configuration would be a singular isothermal sphere, in which σ_{1D} remains constant with radius and $M(r) = 2\sigma_{1D}^2r/G$ (Sec. III.B.2); notice that the boundary pressure term required in this configuration alters the usual virial relation. The deflection angle for this mass distribution is $4\pi\sigma_{1D}^2/c^2$, independent of radius. In general, however, the cluster potential will not be precisely isothermal, nor will the cluster be perfectly spherical.

Simulations of large-scale structure formation suggest that superpositions of other mass concentrations limit the accuracy of weak-lensing masses, at least for clusters defined to be within spherical volumes. Projected mass fluctuations along the line of sight to a distant cluster can be on the order of $\sim 10^{14} M_{\odot}$ (Hoekstra, 2001; Metzler *et al.*, 2001, 1999). On the other hand, weak-lensing masses are expected to correlate quite well with cluster richness, another measure of the mass within a cylindrical region, raising the possibility that at least some of the projected mass can be accounted for by using galaxy colors to separate these mass concentrations from the cluster in redshift space.

B. Clusters in X-rays

Clusters of galaxies are X-ray sources because galaxy formation is inefficient. Only about a tenth of the universe's baryons reside with stars in galaxies, leaving the vast majority adrift in intergalactic space. Most of these intergalactic baryons are extremely difficult to observe, but the deep potential wells of galaxy clusters compress the associated baryonic gas and heat it to X-ray emitting temperatures. The gas temperature inferred from a cluster's X-ray spectrum therefore indicates the depth of a cluster's potential well, and the emission-line strengths

in that spectrum indicate the abundances of elements like iron, oxygen, and silicon in the intracluster medium (ICM). Here we outline the primary characteristics of that X-ray emission. For a more detailed discussion of the physics, see Sarazin (1988).

1. X-ray Surface Brightness

Extended X-ray emission from clusters of galaxies was first observed in the early 1970's (Forman *et al.*, 1972; Gursky *et al.*, 1971; Kellogg *et al.*, 1972), but was correctly attributed to thermal bremsstrahlung several years earlier by Felten *et al.* (1966), who were inspired by a spurious X-ray detection of the Coma cluster. For typical cluster temperatures ($kT \gtrsim 2$ keV) the emissivity of thermal bremsstrahlung dominates that from emission lines, but below ~ 2 keV that situation reverses, given the typical heavy-element abundances relative to hydrogen which are ~ 0.3 times those found in the Sun. The rate at which the ICM radiates energy can be expressed in terms of a cooling function $\Lambda_c(T)$ computed assuming that collisional ionization equilibrium determines the relative abundance of each ion. Many collisional ionization codes have been developed to compute the emissivity and X-ray spectrum of such gas (e.g., Raymond and Smith, 1977). Because these cooling processes all involve electrons colliding with ions, the resulting cooling function is usually defined so that either $n_{\text{enH}}\Lambda_c(T)$ or $n_{\text{enion}}\Lambda_c(T)$ is the luminosity per unit volume. Tozzi and Norman (2001) give a useful fit to the computations of Sutherland and Dopita (1993) for abundances equal to 0.3 times their solar values. For typical ICM temperatures, $\Lambda_c \sim 10^{-23}$ erg cm³ s⁻¹.

In most clusters, the intracluster gas appears to be in approximate hydrostatic equilibrium. Assuming spherical symmetry, the equation of hydrostatic equilibrium can be written

$$\frac{d \ln \rho_g}{d \ln r} + \frac{d \ln T}{d \ln r} = -2 \frac{T_\phi(r)}{T} , \quad (3)$$

where ρ_g is the gas density and $k_B T_\phi(r) = GM(r)\mu m_p/2r$ is the characteristic temperature of a singular isothermal sphere with the same value of $M(r)/r$. Making the additional assumption that the gas is isothermal leads to a classic model for the X-ray surface brightness of clusters known as the beta model (Cavaliere and Fusco-Femiano, 1976). If the velocity distribution of the particles responsible for $M(r)$ is also isothermal with a constant velocity dispersion σ_{1D} , then Poisson's equation implies

$$\frac{d \ln \rho_g}{dr} = -\frac{\mu m_p}{kT} \frac{d\phi}{dr} = \beta \frac{d \ln \rho}{dr} , \quad (4)$$

where the eponymous $\beta \equiv \mu m_p \sigma_{1D}^2 / kT$ (e.g., Sarazin, 1988). Given the approximate isothermal potential of King (1962), $\rho(r) \propto [1 + (r/r_c)^2]^{-3/2}$, in which r_c

is a core radius that keeps the profile from becoming singular at the origin, the gas density profile becomes $\rho_g(r) \propto [1 + (r/r_c)^2]^{-3\beta/2}$. The expected X-ray surface brightness profile for an isothermal gas is then $\propto [1 + (r/r_c)^2]^{-3\beta+1/2}$, and fitting this model to the observations gives the best-fit parameters r_c , β_{fit} , and the normalization of the gas-density distribution.

Beta models generally describe the observed surface-brightness profiles of clusters quite well in the radial range from $\sim r_c$ to $\sim 3r_c$, with $\beta_{\text{fit}} \approx 2/3$ and $r_c \sim 0.1r_G$ giving the best fits for rich clusters (Jones and Forman, 1984) and a possible trend toward lower β values in poorer clusters (Finoguenov *et al.*, 2001b; Helsdon and Ponman, 2000; Horner *et al.*, 1999; Sanderson *et al.*, 2003). The X-ray luminosity integrated over radius converges for $\beta > 0.5$, meaning that most of the observed X-rays come from a relatively small proportion of the ICM. However, beta models often underestimate the central surface brightness (Jones and Forman, 1984) and tend to overestimate the brightness at $r \gg r_c$ (Vikhlinin *et al.*, 1999). These discrepancies arise in part because the intracluster medium is not strictly isothermal (Sec. III.C.2) and because real cluster potentials differ from the King model (Sec. III.B.2).

The centrally concentrated surface-brightness profiles of clusters make X-ray surveys very effective at finding cluster candidates. Because X-ray emission depends on density squared, clusters of galaxies strongly stand out against regions of lesser density, minimizing the complications of projection effects (see Rosati *et al.*, 2002, for a recent review). Surveys of X-ray selected clusters currently extend to $z \approx 1.3$ (e.g. Rosati *et al.*, 2004; Stanford *et al.*, 2001), a limit owing to the decline of surface brightness with redshift (Sec. III.A.2). Unfortunately, X-ray luminosity correlates less well than one would like with the optical properties of clusters. Early studies showed that X-ray luminosity correlates with optical richness but with a large scatter (Bahcall, 1977; Mushotzky, 1984), and that situation has not improved much in the intervening decades (Donahue *et al.*, 2002; Gilbank *et al.*, 2003; Kochanek *et al.*, 2003). The optical properties of very luminous X-ray clusters are well behaved (Lewis *et al.*, 1999), but deep optical surveys have found distant cluster candidates that appear to have velocity dispersions much larger than one would guess from their X-ray luminosity (Lubin *et al.*, 2004). These objects may be superpositions of smaller clusters whose joint velocity distribution seems like that of a larger relaxed cluster.

2. Plasma Temperature

Clusters in hydrostatic equilibrium have a plasma temperature that is closely related to the overall mass. Measuring that temperature requires higher quality data than a simple luminosity measurement, because the photons must be divided among multiple energy bins. Ideally,

one would like enough data to measure both $T(r)$ and $\rho_g(r)$, in which case equation (3) can be solved directly for $M(r)$. Even with the highest-quality data, the derived mass is still slightly model dependent because $T(r)$ and $\rho(r)$ must be determined by deprojecting the surface-brightness information (Fabian *et al.*, 1981; Kriss *et al.*, 1983; Pizzolato *et al.*, 2003; White *et al.*, 1997).

In practice, the quality of the mass measurement depends on what the total number of observed X-ray photons allows. With limited information about the temperature gradient, one can fit a polytropic law³ $T \propto \rho_g^{\gamma_{\text{eff}}-1}$, giving the radial dependence of temperature in terms of an effective adiabatic index γ_{eff} with density as the radial coordinate. However, data on distant clusters often do not allow a temperature gradient to be measured and sometimes are even insufficient to give an accurate temperature. In those cases, one must rely on scaling laws that connect X-ray luminosity with temperature and temperature with mass, calibrated with either high-quality observations or numerical simulations of cluster formation that include all the relevant physics (Sec. III.C).

Limitations in the measurement of cluster temperature systematically affect the mass one infers for the cluster. If only a single temperature can be measured, then the isothermal beta model implies

$$\frac{M(r)}{r} = \frac{3\beta k_B T}{G\mu m_p} \frac{(r/r_c)^2}{1 + (r/r_c)^2}. \quad (5)$$

Note that at large radii this relation approaches the one for isothermal gas in a singular isothermal potential, $M(r)/r = 2k_B T/G\mu m_p$, as long as $\beta = 2/3$. However, single temperatures gleaned from a cluster's overall spectrum need to be treated with caution. Global cluster temperatures quoted in the literature are generally spectral-fit temperatures (T_{sp}) obtained by fitting a single-temperature emission model to an overall cluster spectrum containing multiple temperature components. These spectral-fit temperatures are similar to, but not identical to, the cluster's luminosity-weighted temperature T_{lum} in which each temperature component is weighted by ρ_g^2 . Numerical simulations indicate that both T_{sp} and T_{lum} can differ from the mass-weighted gas temperature T_g and from one another by $\sim 10\text{--}20\%$ (Mathiesen and Evrard, 2001; Mazzotta *et al.*, 2004).

A modest amount of spatially resolved temperature information improves the mass measurement. Allowing for a temperature gradient corresponding to $T \propto \rho_g^{\gamma_{\text{eff}}-1}$ changes the estimated mass to

$$\frac{M(r)}{r} = \frac{3\beta\gamma_{\text{eff}} k_B T(r)}{G\mu m_p} \frac{(r/r_c)^2}{1 + (r/r_c)^2}. \quad (6)$$

Observers are still working toward a consensus on the temperature gradients of clusters (De Grandi and Molendi, 2002; Irwin and Bregman, 2000; Markevitch *et al.*, 1998; Mushotzky, 2004; Pratt and Arnaud, 2002), but measured values of γ_{eff} often range as high as 1.2 (Finoguenov *et al.*, 2001b). Cluster temperatures are extremely difficult to observe in the neighborhood of the virial radius, but extrapolating a $\gamma_{\text{eff}} = 1.2$ gradient to $10r_c$ leads to a gas temperature less than half the core temperature. Including temperature-gradient information can therefore lower the estimated mass for a cluster of temperature T_{lum} by up to $\sim 50\%$.

Despite the potential for systematic uncertainties, the luminosity-weighted temperatures of clusters correlate well with their velocity dispersions. Most of the recent comparisons for low-redshift clusters find that $\sigma_{1D} \propto T_{\text{sp}}^{0.6}$, slightly steeper than expected if both quantities track cluster mass (Lubin and Bahcall, 1993; Xue and Wu, 2000). Those same comparisons find normalizations of this relation for rich clusters in the range $\beta_{\text{sp}} = \mu m_p \sigma_{1D}^2 / kT_{\text{sp}} = 0.9 - 1.0$ (Figure 1). The discrepancy between β_{sp} and β_{fit} is no cause for concern. It arises because the true mass profile is not a King model and because clusters are not in perfect hydrostatic equilibrium (Bahcall and Lubin, 1994; Evrard, 1990). More worrisome are recent observations suggesting that the X-ray temperatures of distant optically-selected clusters with unusually small X-ray luminosities are also considerably cooler than their velocity dispersions would indicate (Lubin *et al.*, 2004). However, more extensive redshift measurements have shown that at least one of these systems is composed of several smaller systems that have not yet merged to form a single large cluster (Gal and Lubin, 2004).

3. Measuring Abundances

Abundances of elements like iron, oxygen, and silicon in the intracluster medium are relatively easy to measure from their emission line fluxes, as long as the temperature of the line-emitting gas is well defined. Because of the low density of intracluster gas, collisional deexcitation is negligible, so every collisional excitation produces a photon that leaves the cluster. Thus, one can fit the optically-thin spectrum of a collisionally-ionized, single-temperature plasma to the observed spectrum, adjusting the abundances in the model to produce the best fit. The high spectral resolution of today's X-ray observatories, *Chandra* and *XMM-Newton*, allows abundance determinations for individual elements if enough photons can be gathered. Otherwise, the solar pattern of abundance ratios is assumed for elements other than H and He and the normalization of the overall pattern is fit to the observations. Because the most abundant elements are almost completely ionized in the hottest clusters, these abundance determinations depend heavily on the strength of

³ Note that this is *not* an actual equation of state for the gas but only a fitting formula for $T(r)$ as a function of $\rho_g(r)$.

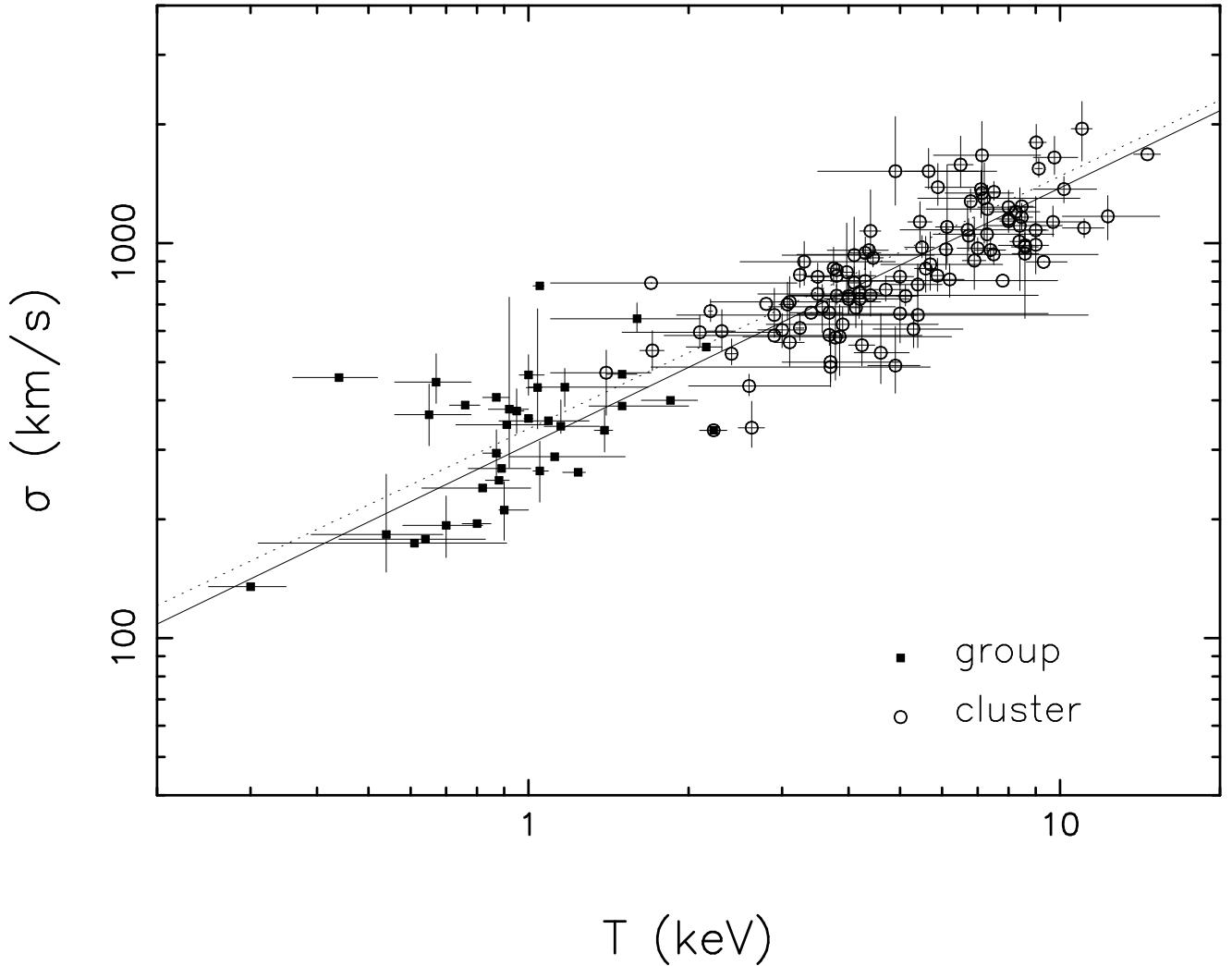


FIG. 1 Relation between velocity dispersion and temperature for a heterogeneous sample drawn from the literature. Solid squares illustrate data on galaxy groups and open circles give the cluster data. The dotted and solid lines show the best power-law fits for groups and clusters, respectively. The best-fitting relation to the combined sample is $\sigma_{1D} = 10^{2.51 \pm 0.01} \text{ km s}^{-1} (T/1 \text{ keV})^{0.61 \pm 0.01}$, corresponding to $\beta_{sp} = 0.97$ at 6 keV. (Figure from Xue and Wu, 2000)

the K-shell emission lines of iron, sometimes the only lines that are measurable.

On average, the overall abundances of heavy elements with respect to hydrogen in clusters are about 0.3 times the solar ratios. Just as with temperature, this determination is weighted toward the cluster core because of the ρ_g^2 emissivity. Spatially resolved observations of Fe K -line emission show that iron abundances, at least, can be higher at the cluster's center, particularly when a giant, central-dominant galaxy is there. This iron excess is consistent with being supernova debris from the giant galaxy's stars (De Grandi *et al.*, 2003). Farther out in clusters, these Fe gradients appear to flatten at ~ 0.3 times the solar level, extending to about $\sim 5r_c$, beyond which point the X-ray surface brightness is too low for accurate abundance and temperature measurements. This abundance level does not seem to have sub-

stantially changed from redshift $z \sim 1$ to the present (Donahue *et al.*, 1998, 1999; Tozzi *et al.*, 2003).

The total amount of iron implied by extrapolating this ratio over an entire cluster is quite impressive, exceeding the total amount of iron contained within all the stars in the cluster's galaxies (Renzini, 1997). Explaining how all that iron got into the intracluster medium is challenging. It is comparable to the total amount of iron produced by all the supernovae thought to have exploded during the history of the cluster, and according to some estimates, it requires a disproportionately large number of massive stars to have formed in order to produce enough supernovae (David *et al.*, 1991; Gibson and Matteucci, 1997; Loewenstein, 2001; Loewenstein and Mushotzky, 1996; Matteucci and Gibson, 1995; Portinari *et al.*, 2003).

Presumably all these supernovae could have driven strong gaseous outflows known as galactic winds that ex-

elled the heavy elements into the intracluster medium (Heckman *et al.*, 1990; Larson and Dinerstein, 1975). However, such powerful galactic winds are hard to produce in numerical simulations of galaxies because much of the energy released by massive-star (Type II) supernovae is transferred to cool gas within the galaxy, where it is radiated away before it manages to drive a powerful wind (Mac Low and Ferrara, 1999). Alternatively, some of this iron may come from exploding white dwarfs (Type Ia supernovae), whose iron yields are higher than those of Type II supernovae. In either case, the total amount of kinetic energy released by the supernovae that created these elements is enormous, corresponding to $\sim 0.3 - 1$ keV per particle in the intracluster medium (Finoguenov *et al.*, 2001a; Pipino *et al.*, 2002). Yet, the efficiency of energy transfer from supernovae to the ICM remains an open question (Kravtsov and Yepes, 2000).

In principle, one can probe the origins of elements in the ICM and assess whether massive stars were disproportionately common earlier in time by comparing the abundances of massive-star products like oxygen to that of iron, which may come largely from Type Ia supernovae. No clear answer has yet emerged from such studies, which depend heavily on a proper understanding of the gas temperature distribution to get the correct elemental abundances (e.g., Buote *et al.*, 2003). Some studies have concluded that the relative abundance patterns in the intracluster medium are near solar, implying that the stellar populations producing those supernovae were similar to those in our own galaxy (Renzini, 2004). Other studies find an excess of oxygen and other elements of similar atomic number, suggesting that the cluster's galaxies produced an unusually large number of massive stars early in the cluster's history (e.g., Finoguenov *et al.*, 2003).

C. Clusters in Microwaves

Hot gas in clusters can also be observed through its effects on the cosmic microwave background. The background itself has a virtually perfect blackbody spectrum (Mather *et al.*, 1990). Soon after the discovery of this background radiation, Weymann (1965, 1966) computed how Compton scattering would distort its spectrum, slightly shifting some of the microwave photons to higher energies as they passed through hot intergalactic gas. Sunyaev and Zeldovich (1970, 1972) then predicted that hot gas in clusters of galaxies would indeed produce such a distortion, now known as the Sunyaev-Zeldovich (S-Z) effect.

1. The S-Z Effect

Two decades after this prediction there were only a few marginal detections (Birkinshaw, 1991), but many clusters were detected at high significance in the ensuing

decade (Birkinshaw, 1999; Carlstrom *et al.*, 2000). With multiple new and highly capable S-Z instruments coming on line in the next few years, another quantum leap in this area is poised to happen, enabling wide-field cosmological studies of clusters to extend through much of the observable universe (Carlstrom *et al.*, 2002). A number of recent reviews elucidate the details of the S-Z effect (e.g., Birkinshaw, 1999; Carlstrom *et al.*, 2002). Here we summarize only a few fundamentals.

To lowest order, the shape of the distorted spectrum depends on a single parameter proportional to the product of the probability that a photon passing through the cluster will Compton scatter and the typical amount of energy a scattered photon gains:

$$y = \int \frac{k_B T}{m_e c^2} n_e \sigma_T dl , \quad (7)$$

where σ_T is the Thomson cross-section and the integral is over a line of sight through the cluster. Because the optical depth of the cluster is small, the change in microwave intensity at any frequency is linearly proportional to $y \ll 1$, with reduced intensity at long wavelengths and enhanced intensity at short wavelengths. Relativistic corrections in hot clusters add a slight frequency dependence to the magnitude of the effect, making cluster temperatures measurable with precise observations of the microwave distortion at several frequencies (see Carlstrom *et al.*, 2002, for a discussion). A cluster's motion with respect to the microwave background produces additional distortion, known as the kinetic S-Z effect, but here we will concern ourselves only with the thermal S-Z effect.

Cosmological applications of the thermal S-Z effect in clusters benefit greatly from the fact that the effect is independent of distance, unlike optical and X-ray surface brightness. Thus, a dedicated S-Z cluster survey efficiently finds clusters out to arbitrarily high redshifts. Because not all these clusters will be well resolved, the surveys will be measuring an integrated version of the distortion parameter:

$$Y = \int y dA \propto \int n_e T dV ; \quad (8)$$

where the first integral is over a cluster's projected surface area and the second is over its volume. The Y parameter therefore tells us the total thermal energy of the electrons, from which one easily derives the total gas mass times its mass-weighted temperature within a given region of space. If these regions can be chosen so that the gas mass is always proportional to the cluster's total mass, then the observable Y can be used a measure of cluster mass, once the relationship between Y and mass has been calibrated.

The impressive power of the S-Z effect for finding distant clusters also has a significant drawback, namely sky confusion owing to projection effects. Along any line of sight through the entire observable universe, the probability of passing within the virial radius of a cluster or

group of galaxies is of order unity (e.g., Voit *et al.*, 2001). Because a cluster's S-Z distortion does not diminish with distance, many of the objects in a highly sensitive S-Z survey will therefore significantly overlap. Information on galaxy colors will help to separate nearby objects from more distant ones, but the implications of sky confusion for making accurate mass measurements are still a matter to be reckoned with (e.g., White *et al.*, 2002). One way to avoid the problem of sky confusion will be to measure the statistical S-Z properties of clusters in the angular power spectrum of the microwave sky instead of analyzing the clusters themselves (da Silva *et al.*, 2001; Holder and Carlstrom, 2001; Seljak *et al.*, 2001). In fact, this statistical signal may already have been detected (Kuo *et al.*, 2004; Pearson *et al.*, 2003)

2. Comparing S-Z with X-ray

Comparisons between a cluster's X-ray properties and S-Z properties are useful in several different ways. X-ray observations are nicely complementary to S-Z observations of clusters because they give the integral of ρ_g^2 along lines of sight through a cluster in addition to a gas temperature. Assuming that clusters are spherical objects with smooth gas distributions, one can divide the product of temperature and the line-of-sight integral of ρ_g^2 by the observed y value to obtain a cluster's gas density profile. Combining the data in this way can be particularly useful in studying the outskirts of clusters, where the X-ray surface brightness is difficult to observe but the S-Z signal remains substantial. With this density profile in hand, one can then derive the line-of-sight thickness of the cluster from either the X-ray or S-Z observations. This type of information could help to solve the S-Z projection problem in fields where there are high-quality X-ray and S-Z data.

If a cluster is indeed spherical, then a comparison of its cluster's physical thickness with its apparent angular size directly gives the cluster's distance, which can be used to determine the scale and geometry of the universe (Birkinshaw *et al.*, 1991). Deriving the scale of the universe in this way is subject to numerous systematic effects. For example, clusters are not all perfectly spherical. Many appear slightly ellipsoidal in X-ray images, calling for a sample of clusters with random orientations to beat down this systematic effect, although three-dimensional reconstructions are possible with the addition of gravitational-lensing data (e.g. Zaroubi *et al.*, 2001). Note also that comparisons of X-ray images to S-Z images would produce nonsensical distances if the intracluster medium were highly clumpy, owing to the ρ_g^2 X-ray emissivity. The fact that cluster distances found in this way are consistent with the standard calibrations of Hubble's Law indicates that the X-ray emitting gas is well-behaved and that most clusters are in approximate hydrostatic equilibrium.

III. EVOLUTION OF THE DARK COMPONENT

Cluster masses measured with the techniques outlined in the previous section range from around $10^{14} M_\odot$ to more than $10^{15} M_\odot$, the vast majority of which appears to be dark matter that emits no detectable radiation. Even using alternative theories of gravity, it is difficult to explain the cluster observations without dark matter dominating the overall mass (Sanders, 2003). In contrast, explaining the characteristics of clusters and their evolution with redshift is much easier with models in which non-baryonic cold dark matter dominates the mass density of the universe.

This section explains how the evolution of the dark component of the universe, including both dark matter and dark energy, is thought to be reflected in the evolution of cluster properties. It begins with a summary of the concordance model for cosmology and some closely related alternatives, all of which are predicated on the existence of non-baryonic cold dark matter. It then explains how dark matter drives cluster formation in such models, providing some simple analytical approximations to the extensive numerical work that has been done on the subject. These models do a good job of accounting for the basic properties of observed clusters, allowing astronomers to measure several of the parameters in the concordance model using cluster observations, most notably the overall mass density of the universe and the amplitude of the initial spectrum of density perturbations that eventually produces all the structure we observe.

The accuracy of those parameter measurements is currently limited by uncertainties in the relationships between cluster masses and the observable properties that trace those masses. Numerical simulations of cluster formation do not yet provide precise calibrations of these relations because they do not yet account for all of the thermodynamical processes associated with galaxy formation. The third part of this section surveys the mass-observable relations and how the uncertainties in those relations affect cosmological parameters derived from them. The fourth part of this section examines how the properties of clusters evolve and how fitting that evolution with cosmological models improves the accuracy of the derived cosmological parameters. Even though current surveys of distant clusters contain relatively few objects, they already place strong constraints on the overall matter density. Larger cluster surveys in both the microwave and X-ray bands have the potential to place much stronger constraints on the overall cosmological model, measuring both dark matter and dark energy parameters to 5% statistical accuracy, independently of other cosmological observations.

A. A Recipe for the Universe

Our current understanding of cluster evolution is an outgrowth of the overall cosmological model, whose pri-

many features depend on just a handful of parameters. One set of parameters specifies the global cosmological model, which describes the overall geometry of the universe, the mean density of its contents, and how its scale changes with time. The other important set of parameters specifies the initial spectrum of density perturbations that grew into the galaxies and clusters of galaxies we see today. Here we define both sets of parameters and their roles in the context of the overall model. More extensive and detailed discussions of this recipe for the universe can be found in some of the excellent books on cosmology (e.g., Peacock, 1999; Peebles, 1993).

1. Global Dynamics

The expansion of the universe can be characterized by a time-dependent scale factor $a(t)$ proportional to the mean distance between the universe's galaxies. Hubble's Law relating the distance d between two galaxies and the speed v at which they appear to move apart can then be written as $v = H(t)d$, where $H(t) = \dot{a}/a$ is the Hubble parameter. Many independent measurements indicate that the value of this parameter at the current time t_0 is $H(t_0) = H_0 = 71 \pm 7 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Freedman *et al.*, 2001). The value of H_0 , known as Hubble's constant, is often further distilled in the literature into the dimensionless quantity $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$. Sometimes this review will use the more suitable alternative $h_{70} = H_0/(70 \text{ km s}^{-1} \text{ Mpc}^{-1})$ when characterizing observable cluster properties.

On very large scales, the universe appears homogenous and isotropic. Astronomers therefore assume that the time-dependent behavior of $H(t)$ obeys the Friedmann-Lemaître model of the universe, in which

$$\frac{\ddot{a}}{a} = -\frac{4}{3}\pi G \left(\rho + \frac{3p}{c^2} \right), \quad (9)$$

where $\rho(t)c^2$ is the mean density of mass-energy and $p(t)$ is the pressure owing to that energy density. Local energy conservation requires that

$$\dot{\rho}c^2 = -3\frac{\dot{a}}{a}(\rho c^2 + p), \quad (10)$$

and we can use this expression to integrate the dynamical equation as long as we know the equation of state linking ρ and p . If the equation of state has the form $p = w\rho c^2$, then density changes with the expansion as $\rho \propto a^{-3(1+w)}$. For a single mass-energy component with a constant value of w we therefore obtain

$$\dot{a}^2 = \frac{8\pi G}{3}\rho_0 a^{-(1+3w)} + \text{const.}, \quad (11)$$

where ρ_0 is the value of the energy density when $a = 1$ and the constant of integration is related to the global curvature of the universe.

It is most convenient to normalize the scale factor so that it equals unity at the current time. Then the cosmological redshift z of radiation from distant objects is simply related to the scale factor of the universe when that radiation was emitted: $a = (1+z)^{-1}$. This definition allows us to link the constant of integration to more familiar parameters, obtaining

$$\left(\frac{\dot{a}}{a} \right)^2 = H_0^2[\Omega_0(1+z)^{3(1+w)} + (1-\Omega_0)(1+z)^2], \quad (12)$$

where Ω_0 is the current energy density ρ_0 in units of the current critical density $\rho_{\text{cr}0} = 3H_0^2/8\pi G$.

Several different components of the universe, each with a different equation of state, can influence the overall expansion history. Non-relativistic particles with a mass density ρ_M contribute negligible pressure, corresponding to $w = 0$. The energy density $\rho_R c^2$ in photons and other relativistic particles exerts a pressure with $w = 1/3$. Einstein's cosmological constant acts like an energy density $\rho_\Lambda c^2$ that remains constant while the universe expands and therefore exerts a pressure corresponding to $w = -1$. Including each of these components yields the dynamical equation

$$H^2(z) = \left(\frac{\dot{a}}{a} \right)^2 = H_0^2[\Omega_M(1+z)^3 + \Omega_R(1+z)^4 + \Omega_\Lambda + (1-\Omega_0)(1+z)^2] \quad (13)$$

where Ω_x is the current mass-energy density in component x in units of $\rho_{\text{cr}0}$ and $\Omega_0 = \Omega_M + \Omega_R + \Omega_\Lambda$. The value of Ω_x at an arbitrary redshift is given by $\Omega_x(z) = \Omega_x(1+z)^{3(1+w)}[H(z)/H_0]^{-2}$.

Each of these energy-density parameters can be further articulated. The matter density parameter Ω_M consists of a contribution Ω_b from baryons and a contribution Ω_{CDM} from non-baryonic cold dark matter. The radiation density parameter includes contributions from the photons of the microwave background, Ω_{CMB} , and from relict neutrinos produced in the Big Bang, Ω_ν , as long as they remain relativistic particles. Finally, because the physical origin of the Ω_Λ term remains mysterious, it may not be correct to assume that the energy density responsible for it stays constant with time. In order to check this possibility observationally, one can replace the Ω_Λ term with a generalized dark-energy term $\Omega_\Lambda(1+z)^{3(1+w)}$ and attempt to measure the value of w (Turner and White, 1997; Wang and Steinhardt, 1998).

Recent observations, including the cluster studies we will discuss later, have provided approximate values for many of these energy-density parameters, allowing us to estimate when each of the various energy components dominated the dynamics (Figure 2). Dark energy with $\Omega_\Lambda \approx 0.7$ seems to be most important at the current epoch, and because of the scaling of other terms with redshift, it will grow increasingly dominant as time progresses. Non-relativistic matter appears to have a density corresponding to $\Omega_M \approx 0.3$, implying that matter dominated the dynamics at $z \gtrsim 1$. The radiation term was

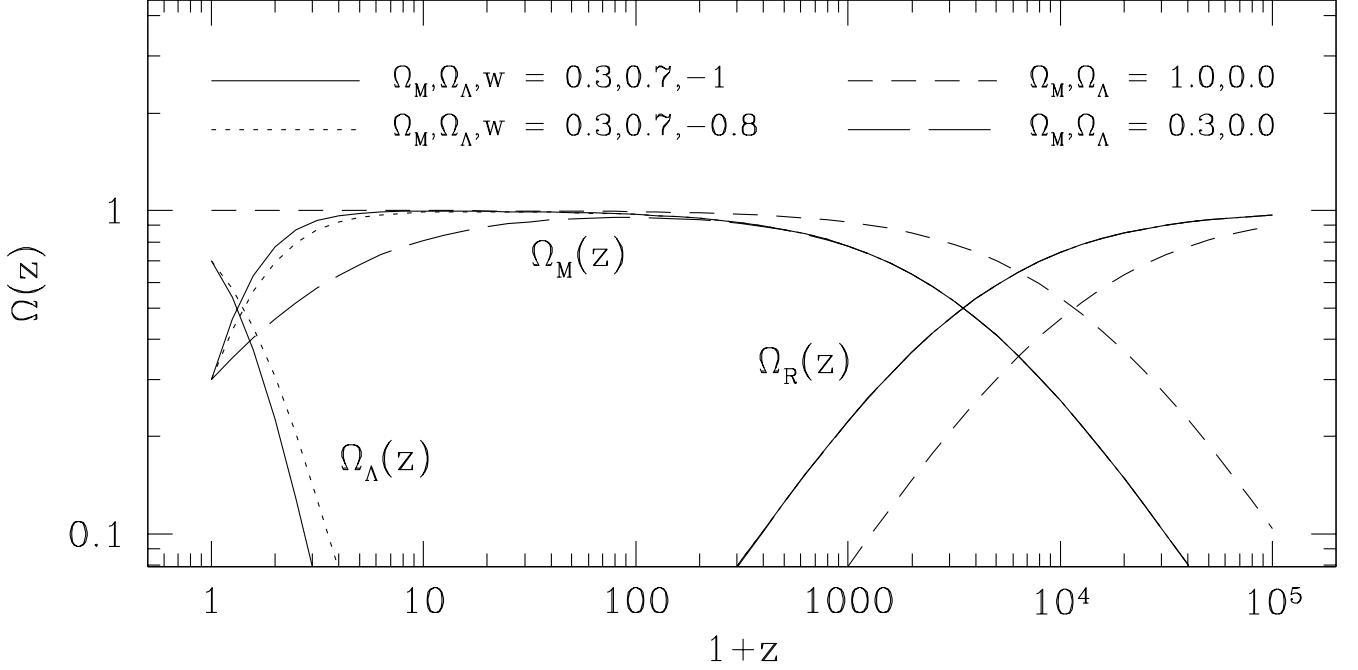


FIG. 2 Evolution of energy densities with redshift. The various lines show the dependence of $\Omega_M(z)$, $\Omega_\Lambda(z)$, and $\Omega_R(z)$ on redshift for various sets of present-day cosmological parameters. Structure in the universe grows most rapidly while $\Omega_M(z) \approx 1$, because positive density perturbations then exceed the critical density. This period of time occurs between the redshift z_{eq} when $\Omega_M(z_{\text{eq}}) = \Omega_R(z_{\text{eq}})$ and the redshift at which Ω_M begins to drop. Notice that the redshift z_{eq} is earlier for larger present-day values of Ω_M and that the redshift at which $\Omega_M(z)$ begins to decline depends on the characteristics of dark energy. Observations of clusters and their evolution provide opportunities to constrain the values of Ω_M , Ω_Λ , and w because the timing of both of these epochs influences the properties of the cluster population.

most important in the distant past, prior to the redshift $z_{\text{eq}} = \Omega_M/\Omega_R - 1$ of matter-radiation equality. Neutrinos with masses less than a few eV will be relativistic particles at this epoch, leading to

$$z_{\text{eq}} = 2.37 \times 10^4 \Omega_M h^2 \quad (14)$$

for $T_{\text{CMB}} = 2.73$ K at $z = 0$ and three families of neutrinos.

2. Global Geometry

Geometry in a universe that is homogenous and isotropic has the same radius of curvature everywhere, but its overall architecture can be either positively curved, flat, or negatively curved, depending on the value of Ω_0 . Because the scale of the universe is changing with time, the most sensible coordinate system to use when describing its geometry is one that expands along with the universe. In such a comoving coordinate system, a radial interval in spherical coordinates has length $a(t)dr$, and the interval corresponding to a small transverse angle $d\psi = \sqrt{d\theta^2 + \sin^2 \theta \cdot d\phi^2}$ depends on the radius of curvature $a(t)R_\kappa$. For positive curvature, analogous to the surface of a sphere, the transverse interval

is $a(t)R_\kappa \sin(r/R_\kappa)d\psi$, and for negative curvature it is $a(t)R_\kappa \sinh(r/R_\kappa)d\psi$.

We can therefore write the Robertson-Walker metric that describes such a universe as

$$c^2 d\tau^2 = c^2 dt^2 - a^2(t) [dr^2 + R_\kappa^2 S_\kappa^2(r/R_\kappa) d\psi^2] , \quad (15)$$

where $S_\kappa(x) = \sin x$ for positive curvature ($\kappa = 1$), $S_\kappa(x) = \sinh x$ for negative curvature ($\kappa = -1$), and a flat universe ($\kappa = 0$) corresponds to $R_\kappa \rightarrow \infty$. The metric can be written in the more familiar form

$$c^2 d\tau^2 = c^2 dt^2 - a^2(t) \left[\frac{dr_\kappa^2}{1 - \kappa r_\kappa^2/R_\kappa^2} + r_\kappa^2 d\psi^2 \right] \quad (16)$$

with the definition $r_\kappa \equiv R_\kappa S_\kappa(r/R_\kappa)$. Plugging this metric into Einstein's field equations leads to

$$H^2(z) = \left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G\rho}{3} - \frac{\kappa c^2}{a^2 R_\kappa^2} , \quad (17)$$

which relates the radius of curvature to other cosmological parameters:

$$R_\kappa = \frac{c}{H_0} \sqrt{\frac{\kappa}{\Omega_0 - 1}} . \quad (18)$$

Notice that the universe at early times is effectively flat as long as $\Omega_M + \Omega_R > 0$ because the horizon size of

the observable patch is $\sim c/H(z) \ll (1+z)^{-1}R_\kappa$ for observers at times corresponding to large values of the redshift z .

The low-redshift universe may also be effectively flat, but that is not guaranteed. Consequently, both the expansion of the universe and its curvature need to be taken into account when we observe highly redshifted objects like distant clusters of galaxies. Because the metric relates the comoving radial coordinate r to redshift through $dr/dz = -c/H(z)$, the coordinate distance to an object with an observed redshift z is

$$r(z) = c \int_0^z \frac{dz}{H(z)} . \quad (19)$$

Relations involving the divergence of light paths can then be compactly written in terms of $r_\kappa(z) = R_\kappa S_\kappa[r(z)/R_\kappa]$, which reduces to $r(z)$ in a flat universe. For example, the angle subtended at coordinate distance $r(z)$ by the transverse length l becomes

$$\psi = \frac{(1+z)l}{r_\kappa(z)} . \quad (20)$$

In a flat, static universe, an object of physical size l would subtend this same angle if it were at the distance $d_A(z) = r_\kappa(z)/(1+z)$, sometimes called the angular-size distance. Likewise, the comoving volume within a solid angle $d\Omega$ and a redshift interval dz is given by

$$\frac{d^2V_{co}}{d\Omega dz} = \frac{cr_\kappa^2(z)}{H(z)} . \quad (21)$$

These formulae are useful to cluster cosmology because they allow us to constrain $H(z)$ and the cosmological parameters that go into it if we know either the transverse sizes of high-redshift clusters or their number density within a given comoving volume. Figure 3 shows how the comoving volume of the universe depends on redshift for several different sets of cosmological parameters.

When surveying the universe for clusters, we also need to know how the geometry and expansion of the universe affect the apparent brightness of a cluster and the galaxies within it. The expansion alone reduces the energy flux received from a distant object by two factors of $1+z$, with one factor coming from the time dilation of the photon flux owing to expansion and the other from the redshift of the photons themselves. The observed energy flux from an object of luminosity L is therefore

$$F = \frac{L}{4\pi(1+z)^2r_\kappa^2(z)} . \quad (22)$$

We would measure the same flux from an equivalent object in a flat, static universe if it were at a distance $d_L(z) = (1+z)r_\kappa(z)$, sometimes called the luminosity distance. The consequences for surface brightness, equal to flux per unit solid angle, are even more dramatic. In a flat, static universe, an object's surface brightness remains constant, but its surface brightness in an expanding universe is reduced by a factor $d_A^2/d_L^2 \propto (1+z)^{-4}$,

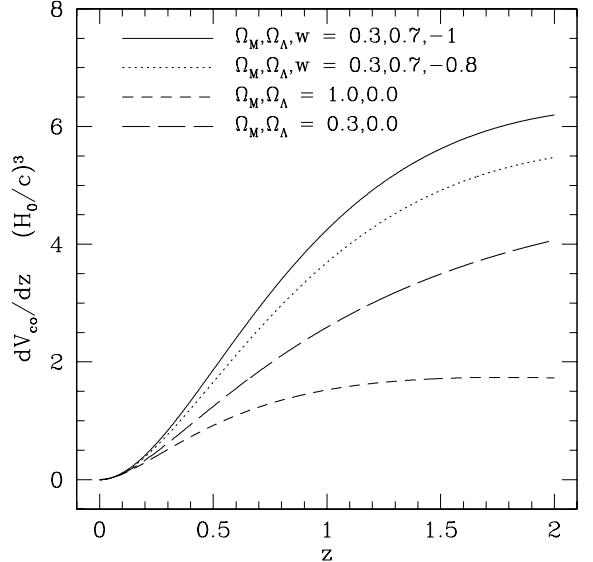


FIG. 3 Redshift dependence of comoving volume in various cosmologies. The quantity dV_{co}/dz is the comoving volume of the entire sky between redshift z and $z + dz$, divided by the redshift interval dz . If clusters were a non-evolving population of objects, one could distinguish between these cosmologies simply by counting the number of clusters on the sky in each redshift interval.

meaning that extended objects like clusters are far less bright at high redshifts.

It would be nice if $r_\kappa(z)$ could be expressed analytically for a general cosmology, but in most cases it cannot. However, a useful analytical expression for the divergence factor does exist for the case in which Ω_R and Ω_Λ are negligible (Mattig 1958):

$$r_\kappa(z) = \frac{2c}{H_0} \frac{\Omega_M z + (\Omega_M - 2)(\sqrt{1 + \Omega_M z} - 1)}{\Omega_M^2(1+z)} . \quad (23)$$

Usually one needs to integrate equation (19) numerically and then insert the results into the S_κ function to obtain the rest of the relations.

3. Density Perturbations

The very existence of galaxy clusters and the human beings who observe them demonstrates that the universe is not perfectly homogeneous. Therefore, the matter density in the early universe must have been slightly lumpy. At some early time these perturbations away from the mean density $\langle \rho_M \rangle$ correspond to an overdensity field

$$\delta(\mathbf{x}) = \frac{\rho_M(\mathbf{x}) - \langle \rho_M \rangle}{\langle \rho_M \rangle} \quad (24)$$

with Fourier components

$$\delta_{\mathbf{k}}(k) = \int \delta(\mathbf{x}) e^{i\mathbf{k} \cdot \mathbf{x}} d^3x . \quad (25)$$

In the plausible case that $\delta(\mathbf{x})$ is isotropic, it can be characterized by an isotropic power spectrum

$$P(k) \equiv \langle |\delta_k|^2 \rangle . \quad (26)$$

If $\delta(\mathbf{x})$ is also a Gaussian random field, then $P(k)$ is a complete statistical description of the initial perturbation spectrum.

The physical meaning of $P(k)$ becomes clearer if we assume it has a power-law form, with $P(k) \propto k^n$, and consider the variance in mass within identical volume elements corresponding to the length scale k^{-1} . For example, let $W(r)$ be a spherical window function that goes quickly to zero outside some characteristic radius r_W and whose integral over all of space is unity. The mass perturbation smoothed over the window is

$$\frac{\delta M}{M}(\mathbf{r}) = \int \delta(\mathbf{x}) W(|\mathbf{x} - \mathbf{r}|) d^3x . \quad (27)$$

Using the convolution theorem, we can then write down the variance $\sigma^2 \equiv \langle |\delta M/M|^2 \rangle$ on this mass scale in terms of W_k , the Fourier transform of $W(r)$:

$$\sigma^2 = \frac{1}{(2\pi)^3} \int P(k) |W_k|^2 d^3k . \quad (28)$$

The variance in mass on scale k for a power-law perturbation spectrum is therefore $\sigma^2 \propto k^{n+3}$, because the windowing averages out modes with $k \gg r_W^{-1}$. Thus, the typical mass fluctuation on mass scale $M \propto k^{-3}$ is

$$\frac{\delta M}{M} \propto M^{-\frac{n+3}{6}} . \quad (29)$$

Notice that large-scale homogeneity of the universe requires $n > -3$.

It is also illuminating to consider how $P(k)$ relates to fluctuations in the gravitational potential, $\delta\Phi \propto k\delta M$. The potential fluctuations owing to a power-law perturbation spectrum scale as $\delta\Phi \propto k^{(n-1)/2}$. The magnitude of these fluctuations therefore diverges on either the high-mass end or the low-mass end, except in the case of $n = 1$. This special property of the $P(k) \propto k$ power spectrum was noted independently by Harrison (1970), Peebles and Yu (1970), and Zeldovich (1972). Not only is this the most natural power-law spectrum, it also appears to be a good approximation to the true power spectrum of density fluctuations in the early universe. Inflationary models for the seeding of structure in the universe produce a Gaussian density field with a power-law index close to $n = 1$ (Guth and Pi, 1982), which is consistent with the observed fluctuations in the cosmic microwave background (e.g., Spergel *et al.*, 2003).

4. Growth of Linear Perturbations

Once the universe has been seeded with density perturbations they begin to grow because the gravity of slightly

overdense regions attracts matter away from neighboring, slightly underdense regions. A complete treatment of perturbation growth is beyond the scope of this review, but some key features can be clarified with a simple toy model consisting of a uniform-density sphere that is slightly denser than its surroundings. The equation of motion for the radius R of an expanding homogeneous sphere is analogous to the one governing the universe as whole. Integrating equation (9) with $a = R/R_0$, where R_0 is an arbitrary fiducial radius at which $\rho = \rho_0$, gives

$$\frac{\dot{R}^2}{2} - \frac{4\pi G\rho_0 R_0^{3+3w}}{3} R^{-(1+3w)} = \epsilon . \quad (30)$$

The constant of integration ϵ in this equation is again related to spatial curvature but can also be interpreted as the net specific energy of the sphere.

Now consider the behavior of two nearly identical spheres that both begin expanding from $R = 0$ at $t = 0$ but have specific energies that differ by a small amount $\delta\epsilon \ll \dot{R}^2/2$. As these two spheres evolve, their radii will become slightly different by an amount $R_2 - R_1 = \delta R$, which satisfies the equation

$$\int_0^R \frac{dR_1}{R_1} = \int_0^{R+\delta R} \frac{dR_2}{R_2} . \quad (31)$$

In the linear regime, we can make the substitution $\dot{R}_2^{-1} = (1 - \dot{R}_1^{-2}\delta\epsilon)\dot{R}_1^{-1}$. If we then take the sphere of radius R_1 to be representative of the universe at large, we obtain

$$\frac{\delta R}{R} = \frac{\delta\epsilon}{R_0^2} \frac{\dot{a}}{a} \int_0^a \frac{da}{\dot{a}^3} . \quad (32)$$

Because $\delta\rho/\rho = -3(1+w)\delta R/R$, this model leads to the following growth function for linear perturbations:

$$D(a) \propto \frac{\delta\rho}{\rho} \propto \frac{\dot{a}}{a} \int_0^a \frac{da}{\dot{a}^3} , \quad (33)$$

which is conventionally normalized so that $D(a) = 1$ at $z = 0$. Notice that the rate of perturbation growth implied by $D(a)$ does not depend on the scale of the perturbation, implying that density perturbations on all scales grow in unison.

This expression for the growth function is identical to those obtained through more rigorous arguments (e.g., Heath, 1977; Peebles, 1993). In a matter-dominated universe, perturbation amplitudes grow in proportion to the scale factor a . In a radiation-dominated universe, they grow $\propto a^2$. Handy numerical algorithms for computing $D(a)$ can be found in Hamilton (2001). A good approximation for the general case with a constant dark-energy density is

$$D(z) = \frac{5\Omega_M(z)}{2(1+z)} \left\{ \Omega_M(z)^{4/7} - \Omega_\Lambda(z) + \left[1 + \frac{\Omega_M(z)}{2} \right] \left[1 + \frac{\Omega_\Lambda(z)}{70} \right] \right\}^{-1} \quad (34)$$

(see Carroll *et al.*, 1992; Lahav *et al.*, 1991).

If the dark-energy density is homogeneous but not constant in time, then the dark-energy density in the perturbed sphere of radius R_2 does not depend on its radius. In that case, one must solve a differential equation to determine the evolution of $\delta \equiv \delta\rho/\rho$ in the linear regime. Differentiating $R_2 = R_1(1 - \delta/3)$ twice with respect to time and keeping only the lowest order terms leads to

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = 4\pi G\rho_M(z)\delta \quad (35)$$

in a universe with negligible radiation density. Wang and Steinhardt (1998) derive a useful approximation to the growth function by defining α_w such that

$$\frac{d\ln\delta}{d\ln a} = [\Omega_M(z)]^{\alpha_w}. \quad (36)$$

For a slowly varying equation of state ($|dw/d\Omega_M(z)| \ll [1 - \Omega_M(z)]^{-1}$), they find that

$$\alpha_w = \frac{3}{5 - w/(1 - w)} + \frac{3}{125} \frac{(1 - w)(1 - 3w/2)}{(1 - 6w/5)^3} [1 - \Omega_M(z)] \quad (37)$$

to lowest order in $1 - \Omega_M(z)$. Using this expression for α_w in the integral

$$D(a) \approx a \exp \left(\int_a^1 \{1 - [\Omega_M(z)]^{\alpha_w}\} \frac{da}{a} \right) \quad (38)$$

reproduces the growth function obtained from numerical integration of equation (33) to better than 1% for $\Omega_M(z) > 0.2$.

These growth functions are valid only as long as pressure gradients do not alter the dynamics of the perturbation. Pressure effects are not an issue when the scale of a perturbation is larger than the Hubble length cH^{-1} . In that regime the growth functions found by solving equations (33) and (35) remain valid. Yet, as the universe ages, it encompasses perturbations of increasingly larger scale and additional physical effects enter the picture.

The bad news is that a variety of processes alter the scale-free nature of the original perturbation spectrum. The good news is that the imprint of these processes on $P(k)$ can tell us a great deal about the contents and dynamics of the universe. During the radiation-dominated era of the universe ($z > z_{\text{eq}}$), pressure effects begin to alter the growth of a given mode when its wavelength is finally contained within the horizon length $\sim cH^{-1}$. Then radiation pressure can effectively resist gravitational compression, inhibiting further growth of modes at that wavelength. Instead, these modes in the coupled photon-baryon fluid begin to oscillate as acoustic waves, and eventually damp owing to photon diffusion out of higher-density, higher-temperature regions. Perturbation growth in the dark-matter component therefore stalls near the amplitude at which the perturbations were first contained within the horizon because the gravitationally dominant photon component no longer spurs

mode growth. These perturbations then resume growing at z_{eq} , when matter begins to dominate the dynamics. The transition from radiation domination to matter domination therefore imprints a bend in $P(k)$ on a length scale corresponding to the horizon scale at z_{eq} .

Perturbation growth is scale-independent during the matter-dominated era only insofar as the matter can be considered cold on the scale of the perturbation. If the characteristic velocities of the matter particles are not small compared to the escape velocity from the perturbation, then both pressure forces and particles streaming out of denser regions can damp small-scale perturbations. Each effect of this type imprints its own characteristic feature on $P(k)$.

All of these scale-imprinting effects that alter $P(k)$ from the time the primordial power spectrum is created until the present day are typically subsumed into a single quantity known as the transfer function, defined to be

$$T(k) \equiv \frac{\delta_k(z=0)}{\delta_k(z)D(z)}, \quad (39)$$

where the symbol k refers to comoving modes with wavenumber $(1+z)k$ in physical space, a convention implicit throughout this review. The redshift z in this definition is assumed to be large enough that $\delta_k(z)$ reflects the original power spectrum imprinted by inflation or some other process. The transfer function therefore represents all the alterations of the original power spectrum that subsequently occur, except for those involving mode growth in the non-linear regime. If the primordial spectrum is a power law of index $n_p \approx 1$, then the power spectrum of linear perturbations at $z = 0$ is $P(k) \propto k^{n_p} T^2(k)$.

5. The CDM Power Spectrum

The most successful models for the formation of large-scale structures like clusters of galaxies assume that cold dark matter (CDM) is responsible. Particles that interact only through gravity exert negligible pressure, and if their random velocities are small then they will not be able to escape from incipient potential wells on the scales of interest. That is, they will be too “cold” to damp the relevant perturbations by freely streaming out of them. Thus, the transfer function for a universe containing only radiation and cold dark matter has just one feature, corresponding to the wavenumber of the mode that enters the horizon at the matter-radiation equality redshift z_{eq} , with a comoving size $l_{\text{eq}} \sim cH_0^{-1}(\Omega_M z_{\text{eq}})^{-1/2} \sim 20(\Omega_M h^2)^{-1}$ Mpc.

Growth of modes with smaller comoving wavelengths temporarily stalls from the redshift at which they enter the horizon until z_{eq} . Because radiation dominates the universe during this time interval, the comoving size of the horizon scales as a while the growth function scales as a^2 . Short-wavelength perturbations therefore miss out on a growth factor $\sim (kl_{\text{eq}})^2$, corresponding to the square

of the change in scale factor from the time a perturbation enters the horizon to the time of matter-radiation equality. Growth of long-wavelength modes, on the other hand, does not stall at all. The behavior of the CDM transfer function in the two extremes is $T(k) \approx 1$ for $k \ll l_{\text{eq}}^{-1}$ and $T(k) \approx (kl_{\text{eq}})^{-2}$ for $k \gg l_{\text{eq}}^{-1}$. For $n_p = 1$, these scalings translate to $\delta M/M \sim M^{-2/3}$ on large scales and $\delta M/M \sim \text{const.}$ on small scales, meaning that structure formation in a CDM universe is hierarchical, with small-scale perturbations reaching the non-linear regime before larger-scale ones.

Numerical computations are needed to derive the exact CDM transfer function, but many authors have provided useful analytical fits to those numerical results. One such expression is

$$T(k) = \frac{\ln(1 + 2.34q)}{2.34q} [1 + 3.89q + (16.1q)^2 + (5.46q)^3 + (6.71q)^4]^{-1/4}, \quad (40)$$

with $q = k(\Omega_M h^2)^{-1}$ Mpc (Bardeen *et al.*, 1986). Allowing for trace populations of baryons and massive neutrinos alters the CDM power spectrum in minor but interesting ways. For example, a small proportion of baryons lowers the apparent dark-matter density parameter, causing a shape-preserving shift in the CDM transfer function (Peacock and Dodds, 1994). This shift can be reproduced by setting $q = k(\Gamma h)^{-1}$ Mpc, so that it includes a shape parameter $\Gamma = \Omega_M h \exp[-\Omega_b(1 + \sqrt{2h}/\Omega_M)]$ (Sugiyama, 1995). Fitting formulae accommodating additional modifications owing to baryons and massive neutrinos can be found in Eisenstein and Hu (1998, 1999).

6. Power Spectrum Normalization

The preceding sections give the theoretical expectations for the shape and growth rate of the density perturbation spectrum but do not specify its normalization. Because inflationary theories do not make firm predictions about the amplitude of the primordial power spectrum, the normalization of $P(k)$ must be determined observationally. For example, measurements of the present-day mass distribution of the universe indicate that $\delta M/M \approx 1$ within comoving spheres of radius $8 h^{-1}$ Mpc (Sec. III.C), as suggested by early galaxy surveys showing that the variance in galaxy counts was of order unity on this length scale (Davis and Peebles, 1983).

This feature of the universe is the motivation for expressing the power-spectrum normalization in terms of the quantity σ_8 , where

$$\sigma_8^2 = \frac{1}{(2\pi)^3} \int P(k) |W_k|^2 d^3k \quad (41)$$

is the variance defined with respect to a top-hat window function $W(r)$ having a constant value inside a comoving

radius of $8 h^{-1}$ Mpc and vanishing outside this radius. When using this formula, one must keep in mind that $P(k)$ refers to the power spectrum of linear perturbations evolved to $z = 0$ according to the growth function $D(z)$, which is valid only for small perturbations. There are other ways of characterizing the power-spectrum normalization, but σ_8 is the most widely-used parameter.

7. Summary of Cosmological Parameters

At the beginning of this recipe, we promised to encapsulate the overall cosmological model in two small sets of parameters. The set governing the global behavior of the universe consists of H_0 , Ω_M , Ω_b , Ω_R , Ω_Λ , and w . The set governing the initial density perturbation spectrum consists of σ_8 and n_p . The shape parameter Γ is not a free parameter in standard cold dark matter models but is sometimes treated as a free parameter in order to test variants of the standard model.

In the concordance model, also known as the Λ CDM model, to denote cold dark matter with a cosmological constant, these parameters are all assigned values close to the most likely values implied by observations:

- **Hubble's constant.** The consensus value of this parameter, measured primarily from the expansion rate of the local universe is $H_0 = 71 \pm 7 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Freedman *et al.*, 2001).
- **Matter density.** Several different methods involving clusters indicate that $\Omega_M \approx 0.3$ (Sec. III.C, Sec. III.D.3). Combining the results of distant supernova observations and observations of temperature patterns in the microwave background gives a similar value for this parameter. Figure 4 shows one example of these mutual constraints in the Ω_M - Ω_Λ plane.
- **Baryon density.** The abundances of light elements formed during primordial nucleosynthesis indicate that $\Omega_b = 0.02h^{-2}$, equal to $\Omega_b = 0.04$ for the value of Hubble's constant given above (e.g., Burles *et al.*, 2001). This value is consistent with the baryon density inferred from the fluctuations in the cosmic microwave background (e.g., Spergel *et al.*, 2003).
- **Radiation density.** The energy density Ω_R in electromagnetic radiation is simply calculated from the microwave background temperature $T_{\text{CMB}} = 2.728 \pm 0.004$ (Fixsen *et al.*, 1996) and Hubble's constant. Neutrinos may also contribute to the energy density in relativistic matter, if their masses are sufficiently small, but this contribution is currently too small to affect the global dynamics.
- **Dark energy density.** Observations of distant supernovae imply that the expansion of the universe is accelerating at a rate consistent with a constant

dark-energy density corresponding to $\Omega_\Lambda \approx 0.7$ (Perlmutter *et al.*, 1999; Riess *et al.*, 1998, 2004). Combining the matter density inferred from clusters with the flat geometry inferred from temperature patterns in the microwave background corroborates this result (e.g., Bahcall *et al.*, 1999, see also Figure 4).

- *Dark energy equation of state.* Observations of microwave background patterns, when combined with observations of large-scale structure are consistent with Einstein’s cosmological constant ($w = -1.0$) but not with $w \gtrsim -0.8$ (Spergel *et al.*, 2003). Alternatively, combining cluster surveys with observations of distant supernovae leads to similar constraints, $w = -0.95^{+0.30}_{-0.35}$ (Schuecker *et al.*, 2003). However, theoretical arguments suggest that the parameter w may be redshift dependent (Peebles and Ratra, 2003).
- *Normalization of density perturbations.* The cluster observations discussed in Sec. III.C and Sec. III.D.3 indicate that the power-spectrum normalization falls into the range $\sigma_8 \approx 0.7 - 1.0$. This range is consistent with both structures in the cosmic microwave background and other observations of large-scale structure.
- *Slope of primordial perturbation spectrum.* All available information indicates that $n_p \approx 1$. Constraints from observations of the microwave background, when combined with optical observations of large-scale structure, give $n_p = 0.97 \pm 0.03$ (Spergel *et al.*, 2003).
- *Shape parameter of perturbation spectrum.* The concordance values of Ω_M , Ω_b , and H_0 given above imply $\Gamma \approx 0.2$, which agrees with the value of Γ derived from observations of large-scale structure (e.g., Peacock and Dodds, 1994; Schuecker *et al.*, 2001; Szalay *et al.*, 2003), an important element of self-consistency in the concordance model.

Several other closely related models have been pursued during the last two decades, but none of them have proven as successful in explaining such a large number of observations. Here are a few variants some of which will be discussed later in connection with the parameter constraints derived from cluster surveys:

- *Standard cold dark matter (SCDM).* In this model, the universe is assumed to be flat, with no dark energy, so $\Omega_M = 1$ and $\Omega_\Lambda = 0$. For this value of the matter density, measurements of large-scale structure, including clusters, imply that $\sigma_8 = 0.4 - 0.5$ (Sec. III.C). However, the shape parameter implied for this model, given the observed Hubble constant, is $\Gamma \approx 0.7$, which conflicts with observations of large-scale structure (e.g., Szalay *et al.*,

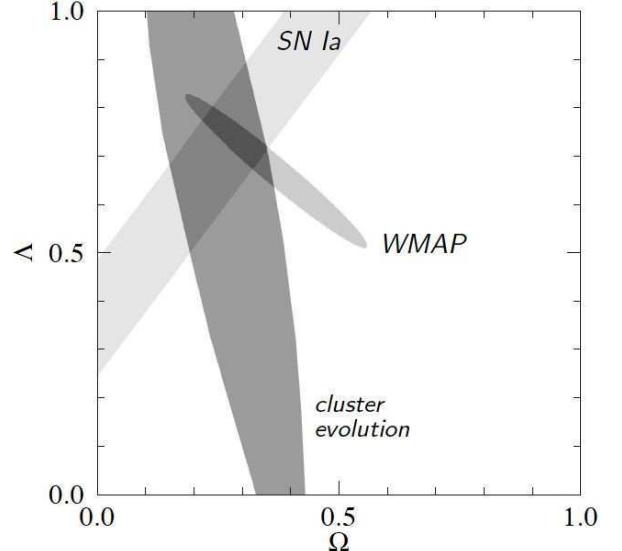


FIG. 4 Cosmological constraints from cluster evolution (Vikhlinin *et al.*, 2003), supernovae (Perlmutter *et al.*, 1999; Riess *et al.*, 1998), and WMAP observations of the cosmic microwave background (Spergel *et al.*, 2003). The horizontal axis labeled Ω gives the value of Ω_M ; the vertical axis labeled Λ gives the value of Ω_Λ . These particular constraints from cluster evolution are based on the baryonic mass function of clusters (Sec. III.D.3), but other measures of cluster evolution give similar results. The complementarity of these constraints is evident from the figure, and the common region of overlap near $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$ is reassuring evidence of consistency in the overall picture. (Figure courtesy of Alexey Vikhlinin.)

2003). Also, this value of Ω_M leads to a baryon-to-dark-matter ratio $f_b = \Omega_b/\Omega_M$ that is inconsistent with cluster observations (Sec. III.C.7).

- *Tilted cold dark matter.* One way to make models with $\Omega_M = 1$ more consistent with large-scale structure observations is to assume that the primordial perturbation spectrum is “tilted” so that it is significantly shallower than $n_p = 1$ (e.g., Cen *et al.*, 1992). However, such models conflict with the strong constraints on n_p inferred from microwave background observations.
- *Ad hoc power spectrum (τ CDM).* Another option for making $\Omega_M = 1$ models more consistent with observations is to arbitrarily adjust the shape of the perturbation spectrum to fit the observations. One such realization is the τ CDM model (Jenkins *et al.*, 1998), which sets $\Omega_M = 1$, $\sigma_8 = 0.5$ and $\Gamma = 0.2$, even though there is little physical justification for having such a low value of Γ in a cosmology with such a large matter density (but see White *et al.*, 1995).
- *Open cold dark matter (OCDM).* There is also the option of accepting the evidence that $\Omega_M \approx 0.3$ but

dispensing with dark energy ($\Omega_\Lambda = 0$) so that the universe has an open geometry. In this case, the perturbation spectrum can be identical to the one in the Λ CDM concordance model, but the growth rate of those perturbations differs because of the changed expansion rate at late times.

This review does not consider models involving forms of dark matter other than cold dark matter, but it does consider models with generalized forms of dark energy having $w \neq -1$ (w CDM). Current cluster surveys are not yet large enough to place strong constraints on this equation-of-state parameter but they might provide much stronger constraints in the coming decade (Sec. III.D.4).

B. Cluster Formation

Cluster formation from perturbations in the density distribution of cold dark matter is a hierarchical process. Small subclumps of matter are the first pieces of the cluster to deviate from the Hubble flow and undergo gravitational relaxation because the density perturbations have larger amplitudes on smaller mass scales. These small pieces then merge and coalesce to form progressively larger structures as perturbations on larger mass scales reach the non-linear regime. A full understanding of the details of this hierarchical merging process requires numerical simulations, but simplified, spherically symmetric models of cluster formation illustrate many of the important concepts. This part of the review shows how a cluster would grow from a spherically symmetric mass perturbation and then refines the details of that simplified approach, based on what we have learned from numerical simulations.

1. Spherical Collapse

The most basic features of cluster formation can be understood in terms of a spherically symmetric collapse model (e.g., Bertschinger, 1985; Fillmore and Goldreich, 1984; Gunn and Gott, 1972). In such a model, the matter that goes on to form a cluster begins as a low-amplitude density perturbation that initially expands along with the rest of the universe. The perturbation's gravitational pull gradually slows the expansion of that matter, eventually halting and reversing the expansion. A cluster of matter then forms at the center of the perturbation, and the rate at which additional matter accretes onto the cluster depends on the distribution of density with radius in the initial perturbation.

In a geometry that is perfectly spherically symmetric, the behavior of an individual mass shell in the presence of a homogeneous generalized dark-energy field follows the equation of motion

$$\ddot{r}_{\text{sh}} = -\frac{GM_{\text{sh}}}{r_{\text{sh}}^2} - \frac{1+3w}{2}\Omega_\Lambda H_0^2(1+z)^{3(1+w)}r_{\text{sh}} , \quad (42)$$

where r_{sh} is the shell radius and M_{sh} is the mass enclosed within r_{sh} . Throughout the early evolution of a spherical perturbation, the value of M_{sh} within a given mass shell remains constant. Thus, if the dark-energy term is negligible, the radius of a mass shell obeys the parametric solution $r_{\text{sh}} = r_{\text{ta}}[(1-\cos\theta_M)/2]$, $t = t_c[(\theta_M - \sin\theta_M)/2\pi]$, with a turnaround radius $r_{\text{ta}} = [(2GM_{\text{sh}}t_c^2)/\pi^2]^{1/3}$ for a shell that collapses to the origin at time t_c . The solution for $\Omega_\Lambda \approx 0.7$ and $w \approx -1$ is not much different because the dark-energy term remains $\lesssim 15\%$ of the matter term during the trajectory of all shells that collapse to the origin by the present time. If greater accuracy is needed, a shell's trajectory can be computed numerically from equation (42).⁴

Once a shell collapses, the mass within it no longer remains constant. Because the dark matter within a collapsing shell is collisionless, shells on different trajectories can easily interpenetrate. The radii of collapsed shells in this idealized geometry therefore oscillate symmetrically about the origin, and the amplitudes of these oscillations modestly decrease with time as mass associated with other collapsed shells accumulates within the oscillations' turning points (Gunn, 1977).

The accretion process in real clusters is not so symmetric. Instead, gravitational forces between infalling clumps of matter produce a time-varying gravitational potential that randomizes the velocities of the infalling particles, yielding a Maxwellian velocity distribution in which temperature is proportional to the particle mass. This process, known as “violent relaxation” (Lynden-Bell, 1967), leads to a state of virial equilibrium in which the total kinetic energy E_K is related to the total gravitational potential energy E_G through the equation

$$E_G + 2E_K = 4\pi P_b r_b^3 \quad (43)$$

where P_b is the effective pressure owing to infalling matter at the boundary r_b of the collapsed system (Sec. II.A.2). Setting P_b to zero yields the usual form of the virial theorem for gravitationally bound systems.

A common toy model for estimating the location of a cluster's outer boundary is the spherical top-hat model, which assumes that the perturbation leading to a cluster is a spherical region of constant density. All of the mass shells in such a perturbation move in unison and collapse to the origin simultaneously. The virial theorem therefore suggests that the bounding radius of the cluster after it collapses and relaxes should be in the neighborhood of half the turnaround radius. Numerical simulations indeed show that particle velocities within this radius are

⁴ Here we are making the standard assumption that the collapsing dark matter has no effect on the local dark-energy density (e.g., Wang and Steinhardt, 1998; Weinberg and Kamionkowski, 2003). If in fact the dark-matter collapse alters the local properties of dark energy, the dynamics could be somewhat altered (Mota and van de Bruck, 2004).

generally isotropic and those outside this radius are generally infalling, but the boundary between the isotropic and infalling regions is not particularly distinct (Evrard, 1990; Navarro and White, 1993).

The spherical top-hat model has actually led to several different definitions for the virial radius of a cluster. If one assumes that all the mass in the original top-hat perturbation ends up within $r_{\text{ta}}/2$, then the mass density in that region is $6M/\pi r_{\text{ta}}^3$. In a matter-dominated universe with zero dark energy, this density is equal to $\Delta_v = 8\pi^2/(Ht)^2$ times the critical density $\rho_{\text{cr}} \equiv 3H^2/8\pi G$. Thus, for a flat, matter-dominated universe in which $Ht = 2/3$, the mean density of a perturbation that has just collapsed is taken to be $18\pi^2 \approx 178$ times the critical density. A useful approximation for Δ_v in a flat universe with a non-zero cosmological constant ($w = -1$) is

$$\Delta_v = 18\pi^2 + 82[\Omega_M(z) - 1] - 39[\Omega_M(z) - 1]^2 \quad (44)$$

(Bryan and Norman, 1998). Because the outer radius of a real cluster is not distinct, one pragmatic definition of the virial radius is then the radius r_v within which the mean matter density is $\Delta_v \rho_{\text{cr}}$ (Eke *et al.*, 1996). However, the numerical value of Δ_v in a flat, matter-dominated universe has inspired other definitions. A common alternative is the scale radius r_{200} , within which the mean matter density is $200\rho_{\text{cr}}$. Another frequently used scale radius is r_{180m} , within which the mean matter density is 180 times the mean background density $\Omega_M(z)\rho_{\text{cr}}$. As long as $\Omega_M(z) \approx 1$, both of these scale radii are nearly identical to r_v , but because $\Omega_M \approx 0.3$ at the present time, these radii are now somewhat different, with $r_{200} < r_v < r_{180m}$. This multiplicity of definitions for the radius of a cluster is a potential source of confusion, but as we will see below, each of these scale radii can be particularly well suited to certain applications.

2. Cluster Mass Profiles

Observations of galaxy clusters have long indicated that the velocity dispersion of a cluster's galaxies remains relatively constant with distance from the cluster's center, implying an underlying mass-density profile $\rho_M(r) \propto r^{-2}$. The simplest analytical cluster model consistent with such a density profile is the singular isothermal sphere, in which the velocity dispersion σ_v is constant and isotropic at every point and $\rho_M(r) = \sigma_v^2/2\pi Gr^2$ (e.g., Binney and Tremaine, 1987). This model is useful for making analytical estimates of cluster properties, but it is incomplete because the total mass diverges linearly with radius.

Numerical simulations of cluster formation produce dark-matter halos whose density profiles are shallower than isothermal at small radii and steeper than isothermal at large radii. A generic form for representing these profiles is

$$\rho_M(r) \propto r^{-p}(r + r_s)^{p-q}, \quad (45)$$

where the parameters p and q describe the inner and outer power-law slopes and the radius r_s specifies where the profile steepens. Groups that have fit such profiles to simulated clusters disagree about the best values of p and q but typically find $1 \lesssim p \lesssim 1.5$ and $2.5 \lesssim q \lesssim 3$. Specific examples include the NFW profile, with $p = 1$ and $q = 3$ (Navarro *et al.*, 1997), the Moore profile, with $p = 1.5$ and $q = 3$, and the Rasia *et al.* (2003) profile, with $p = 1$ and $q = 2.5$. Both optical and X-ray observations indicate that density profiles of this sort are good representations of the underlying mass profiles of clusters, at least outside of the innermost regions (Carlberg *et al.*, 1997b; Lewis *et al.*, 2003; Pratt and Arnaud, 2002). Observing the asymptotic inner slope p is currently a matter of great observational interest, as the cuspiness of dark-matter density profiles at $r = 0$ is one of the acid tests of the CDM paradigm for structure formation (see Navarro *et al.*, 2003, and references therein). However, we will not discuss that issue here because the global properties of clusters depend little on the value of p . In this review, we will use the NFW profile when necessary because it remains the most widely used fitting formula for representing the results of cosmological simulations.

The transition of the density profile from shallow to steep can also be expressed in terms of a concentration parameter $c = r_b/r_s$, which expresses the bounding radius of the cluster in units of r_s . Because the concentration parameter depends on r_b , numerical values of c depend somewhat on whether the bounding radius is taken to be r_v , r_{200} , r_{180m} , or something similar. However, these radii are not vastly different because they are generally several times larger than r_s , meaning that the enclosed mass is not rapidly diverging in the neighborhood of the virial radius. Typical concentration parameters for simulated clusters are in the range $c \sim 4 - 10$, with a scatter in $\ln c$ of 0.2-0.35 (Jing, 2000). Also, lower-mass objects tend to have higher halo concentrations because they formed earlier in time, when the overall density of the universe was greater (Bullock *et al.*, 2001; Eke *et al.*, 2001; Navarro *et al.*, 1997).

3. Defining Cluster Mass

Even with these more sophisticated forms for the density profile, mass still diverges with radius. Thus, a cluster's mass and all the relations linking that mass to other observable quantities depend on how one chooses to define a cluster's outer boundary. One would like to define that boundary so as to maximize the simplicity of the relationships between cluster mass and other observables, but no single definition is best for all applications.

The easiest way to link observations to theoretical models is through definitions taking the mass of a cluster to be M_Δ , the amount of matter contained in a spherical region of radius r_Δ whose mean density is $\Delta \cdot \rho_{\text{cr}}$. It is also common for cluster mass to be defined with respect to the background mass density, so that the mean density of

matter within the virial radius is $\Delta \cdot \Omega_M(z) \rho_{\text{cr}}$, but applying this definition to observations requires prior knowledge of Ω_M . Spherical top-hat collapse suggests that Δ_v is a good choice for the density threshold. However, observers often prefer to raise that threshold to $\Delta = 200$ or even $\Delta = 500$ for two reasons. The properties of a cluster are easier to observe in regions where the density contrast is higher, and simulations show that the region within r_{500} is considerably more relaxed than the region within r_v .

As an example of such definitions in action, consider the relation between velocity dispersion and the virial mass M_v obtained by truncating a singular isothermal sphere at the virial radius r_v :

$$M_v = f_\sigma \frac{4\sigma_v^3}{GH\Delta_v^{1/2}} , \quad (46)$$

where the factor f_σ is a parameter that can be adjusted to account for the fact that clusters are not perfect isothermal spheres (Bryan and Norman, 1998; Eke *et al.*, 1996; Evrard, 1989). The presence of this parameter is a reminder that the derivation of this relation should not be taken too literally. Truncation of the mass distribution at r_Δ formally implies a non-zero boundary pressure that shifts the virial relation for this configuration so that $E_K = -3E_G/4$, which is inconsistent with the definition of r_v (Voit, 2000). This functional form for the M_v - σ_v is useful primarily as a fitting formula that accounts for most of the cosmology-dependent changes in the normalization of the relation. However, because the density profiles of dark matter halos defined in this way depend on both mass and redshift, the correction factor f_σ is not a universal constant.

Recent work has shown that defining a cluster's mass using the threshold $\Delta = 200$ leads to an M_{200} - σ_v relation that is remarkably independent of cosmology. Evrard (2004) finds that the relation

$$M_{200} = \frac{10^{15} h^{-1} M_\odot}{H/H_0} \left(\frac{\sigma_v}{1080 \text{ km s}^{-1}} \right)^3 \quad (47)$$

is an excellent fit to a wide range of simulated clusters sampled over a wide range of redshifts. This relation is equivalent to setting $f_\sigma = 1.2$ and $\Delta_v = 200$ in equation (46).

Conversions between M_v and M_Δ defined with respect to an arbitrary Δ are straightforward as long as a cluster's concentration parameter is known. From the definitions of these masses, we have

$$\frac{M_\Delta}{M_v} = \frac{\Delta}{\Delta_v} \left(\frac{r_\Delta}{r_v} \right)^3 , \quad (48)$$

and the halo concentration gives the relationship between r_Δ and r_v . Hu and Kravtsov (2003) have provided a useful approximation for this relation in the case of an NFW profile. Recasting their formulae in slightly different notation, one can write the halo concentration c_Δ defined

with respect to r_Δ in terms of the concentration c_v defined with respect to r_v :

$$c_\Delta = \frac{1}{\left[a_1 f_c^{2p_c} + \left(\frac{3}{4} \right)^2 \right]^{-1/2} + 2f_c} , \quad (49)$$

with $f_c = (\Delta/c_v^3 \Delta_v)[\ln(1+c_v) - c_v/(1+c_v)]$, $p_c = a_2 + a_3 \ln f_c + a_4 (\ln f_c)^2$ and $(a_1, \dots, a_4) = (0.5116, -0.4283, -3.13 \times 10^{-3}, -3.52 \times 10^{-5})$. Plugging the ratio $r_\Delta/r_v = c_\Delta/c_v$ given by this approximation into equation (48) converts cluster masses with an accuracy $\sim 0.3\%$ for the halo concentrations typical of clusters.

Some other definitions of cluster mass are useful in certain contexts but are more difficult to relate to the top-hat collapse model. For example, observers who measure cluster mass using gravitational lensing or the total optical luminosity are essentially measuring cluster mass within a cylinder along the line of sight rather than a sphere. In principle, these observables can be linked with cluster masses defined with respect to a cylindrical boundary, but the relationships between those cylindrical masses and models of structure formation are not as well understood as their spherical counterparts. On the theoretical side, the masses of clusters identified in numerical simulations are sometimes defined using a “friends-of-friends” algorithm that links neighboring mass particles (e.g., Davis *et al.*, 1985). However, clusters defined in this way often have irregular boundaries (White, 2001), making this sort of definition difficult to apply to observations. Masses defined within spheres also have their shortcomings, particularly in cases where two clusters are just beginning to merge, but in general provide the most direct link between cosmological models and observations.

4. Cluster Mass Function

Some of the most powerful constraints on current cosmological models come from observations of how clusters evolve with time. Because cosmological time scales are so long, we cannot observe how individual clusters evolve but rather observe how the demographics of the entire cluster population changes with redshift. A important conceptual tool in this effort is the cluster mass function, $n_M(M)$ which gives the number density of clusters with mass greater than M in a comoving volume element. Notice that the cluster mass function inevitably depends on how one defines cluster mass.

Combining spherical top-hat collapse with the growth function for linear perturbations has led to a widely used semi-analytical method for expressing the cluster mass function in terms of cosmological parameters. Press and Schechter (1974) pioneered the basic approach, which was refined and extended by Bond *et al.* (1991), Bower (1991), and Lacey and Cole (1993). This class of models simplifies the problem of structure formation by assuming that all density perturbations continue to grow

according to the linear growth rate $D(z)$ even when their amplitudes become non-linear. When perturbations are treated in this way, their variance on mass scale M as a function of redshift is

$$\sigma^2(M, z) = \frac{D^2(z)}{(2\pi)^3} \int P(k) |W_k(M)|^2 d^3k , \quad (50)$$

where $W_k(M) = 3(\sin kr_M - kr_M \cos kr_M)/(kr_M)^3$ with $r_M = (3M/4\pi\Omega_M\rho_{\text{cr}0})^{1/3}$ is the Fourier-space representation of a top-hat window function that encloses mass M . The normalization of $P(k)$ is set so that $\sigma(M_8, 0) = \sigma_8$ for $M_8 \equiv (8h^{-1}\text{Mpc})^3 H_0^2 \Omega_M / 2G = 6.0 \times 10^{14} \Omega_M h^{-1} M_\odot$. These perturbations are then assumed to collapse and virialize when their density contrast $\delta = \delta\rho/\rho$ exceeds a critical threshold δ_c .

Suppose the initial density perturbations are gaussian with a variance $\sigma^2(M, z)$ that declines monotonically with mass. Then according to the Press-Schechter approach, the probability that a region of mass M exceeds the collapse threshold at redshift z is $\text{erfc}[\delta_c/\sqrt{2}\sigma(M, z)]$, where $\text{erfc}(x)$ is the complementary error function. Implicit in this expression is the notion that all the mass in the universe belongs to collapsed, virialized objects when viewed on sufficiently small mass scales. It then follows that the cluster mass function on scale M at redshift z is

$$n_M(M, z) = \frac{\Omega_M \rho_{\text{cr}0}}{M} \text{erfc} \left[\frac{\delta_c}{\sqrt{2}\sigma(M, z)} \right] . \quad (51)$$

This expression implies that the shape of the mass function depends only on $\sigma(M, z)$ and remains invariant with respect to the characteristic collapsing mass scale $M_*(z)$ at which $\sigma(M_*, z) = \delta_c$. Observers often work with the mass function in a differential form, such as $dn_M/d\ln M$, but theorists prefer expressing the differential form in terms of the shape-governing function $\sigma(M, z)$. Then the differential mass function takes the form

$$\frac{dn_M}{d\ln \sigma^{-1}} = \sqrt{\frac{2}{\pi}} \frac{\Omega_M \rho_{\text{cr}0}}{M} \frac{\delta_c}{\sigma} \exp \left[-\frac{\delta_c^2}{2\sigma^2} \right] . \quad (52)$$

Both of these forms for the mass function can be straightforwardly extended to cases in which the perturbations are non-gaussian (e.g., Robinson *et al.*, 2000).

The value of the critical threshold δ_c was originally inferred from spherical top-hat collapse. Expanding the parametric solution for spherical collapse in powers of θ_M leads to the following relation at early times:

$$r_{\text{sh}} \approx \left(\frac{9GMt^2}{2} \right)^{1/3} \left[1 - \frac{1}{20} \left(\frac{12\pi t}{t_c} \right)^{2/3} \right] . \quad (53)$$

The leading term in this expression characterizes the behavior of a critical-density sphere and the second term describes how the evolution of a slightly overdense sphere deviates from that of a critical-density sphere. Assuming that this deviation grows according to equation (53) until the moment of collapse and virialization ($t = t_c$) gives

the value of the critical threshold in a flat universe with $\Omega_M = 1$: $\delta_c = 3(12\pi)^{2/3}/20 \approx 1.686$. Generalizing this treatment to cases where $\Omega_M \neq 1$ produces only minor differences in δ_c for interesting values of the cosmological parameters (Eke *et al.*, 1996; Lacey and Cole, 1993).

The preceding derivation of the cluster mass function is not terribly rigorous, but it is useful because adopting $\delta_c = 1.686$ leads to mass functions that agree reasonably well with those derived from numerical simulations. Treating perturbation collapse as ellipsoidal rather than spherical improves that agreement (Sheth *et al.*, 2001). Sheth and Tormen (1999) have shown that the expression

$$\begin{aligned} \frac{dn_M}{d\ln \sigma^{-1}} &= A_s \sqrt{\frac{2a_s}{\pi}} \left[1 + \left(\frac{\sigma^2}{a_s \delta_c^2} \right)^{p_s} \right] \\ &\times \frac{\Omega_M \rho_{\text{cr}0}}{M} \frac{\delta_c}{\sigma} \exp \left[-\frac{a_s \delta_c^2}{2\sigma^2} \right] , \end{aligned} \quad (54)$$

with $A_s = 0.3222$, $a_s = 0.707$, and $p_s = 0.3$ is quite an accurate representation of the mass functions from several different numerical simulations. However, because semi-analytical mass functions like these are not rigorously derived, they are essentially just fitting formulae that conveniently express the simulation results and should be treated cautiously outside the cosmological models against which they have been tested.

A particularly well-tested fitting formula for cluster mass functions has been provided by Jenkins *et al.* (2001). Combining results for simulated clusters spanning a mass range from $< 10^{12} M_\odot$ to $> 10^{15} M_\odot$ and sampled at a number of different redshifts, they found that the form of $dn_M/d\ln \sigma^{-1}$ was nearly invariant if they defined cluster mass to be M_{180m} , the mass within a sphere of mean density $180 \Omega_M(z)\rho_{\text{cr}}$. When this definition of cluster mass is used, the formula

$$\frac{dn_M}{d\ln \sigma^{-1}} = A_J \frac{\Omega_M \rho_{\text{cr}0}}{M} \exp[-|\ln \sigma^{-1} + B_J|^{\epsilon_J}] \quad (55)$$

with $A_J = 0.301$, $B_J = 0.64$, and $\epsilon_J = 3.82$, reproduces the cluster mass function to $\sim 20\%$ accuracy for all the cosmologies tested, including ΛCDM , τCDM , and OCDM . In this expression, A_J governs the fraction of the total mass in collapsed objects, e^{B_J} functions as a collapse threshold analogous to δ_c , and ϵ_J stretches the mass function to fit the simulations. The Sheth-Tormen mass function of equation (54) fits these same numerical simulations nearly as well.

The exponential sensitivity to mass and redshift evident in these expressions for the cluster mass function is both a blessing and a curse. On the one hand, it makes cluster counts and their evolution with redshift a very powerful probe of cosmological parameters. Figure 5, showing the cluster mass function and its evolution with time for five different cosmologies, illustrates how sensitive mass-function evolution is to the matter density. On the other hand, any systematic errors in the measurement of cluster mass, including inconsistencies in the definition

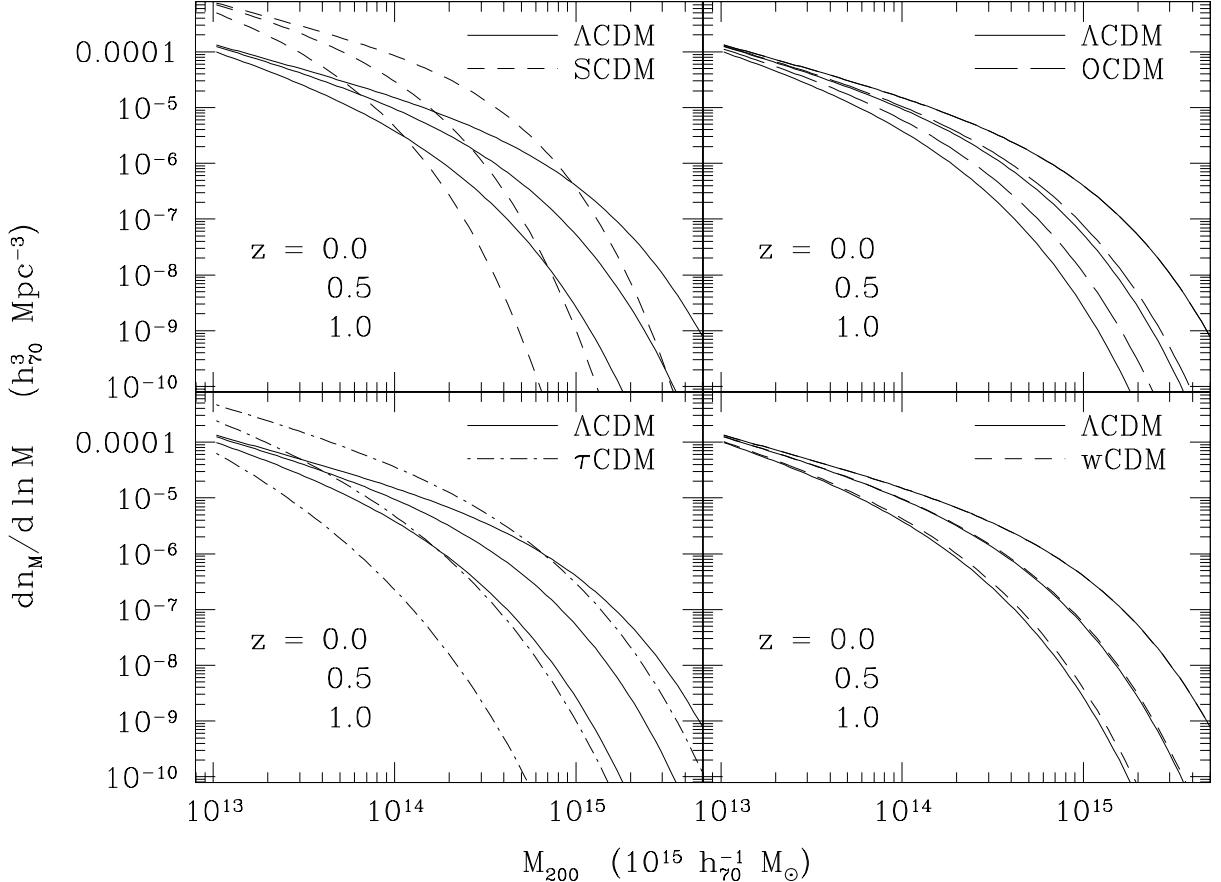


FIG. 5 Mass-function evolution in five different cosmologies. The fiducial model in all cases is the Λ CDM model with $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, $w = -1$, and $\sigma_8 = 0.9$. The upper left panel compares cluster evolution in the Λ CDM case with a standard cold-dark-matter model (SCDM) having $\Omega_M = 1.0$, $\Omega_\Lambda = 0.0$, and $\sigma_8 = 0.5$. Evolution in the SCDM case is much more dramatic, and the steeper slope of the mass function strongly disagrees with observations of local clusters (e.g., Reiprich and Böhringer, 2002). Retaining $\Omega_M = 1.0$ and $\Omega_\Lambda = 0.0$ while adjusting the power spectrum so that $\Gamma = 0.21$ gives a τ CDM model (lower left) in which the slope of the low-redshift mass function is more acceptable, but the evolution is still very strong. Dispensing with dark energy while keeping the matter density low gives an OCDM model ($\Omega_M = 0.3$, $\Omega_\Lambda = 0$, $\sigma_8 = 0.9$; upper right) with less evolution than the Λ CDM case because structure formation starts to ramp down earlier in time (see Figure 2). Dark energy in a w CDM model identical to the Λ CDM model except with $w = -0.8$ (lower right) also slows cluster evolution relative to the Λ CDM case.

of cluster mass, are also exponentially amplified by the steepness of the mass function.

5. Cluster Bias

Another observable feature of the cluster population, closely related to the mass function, is the tendency of galaxy clusters to cluster with one another. Fluctuations in the number density of clusters on large scales are observed to be more pronounced than the fluctuations of the underlying matter density (e.g., Bahcall *et al.*, 2003a; Bahcall and Soneira, 1983; Collins *et al.*, 2000; Klypin and Kopylov, 1983; Postman *et al.*, 1992). In other words, the fractional deviation of $dn_M/d\ln\sigma^{-1}$ from its mean value within a given volume of the universe is observed to be larger than $\delta\rho/\rho$ in that same

volume. The ratio $b(M)$ between the perturbation in the number density of clusters of mass M and the perturbation amplitude of the matter density is known as the bias parameter, and it is taken to be independent of length scale, as long as that length scale is much larger than a cluster.

Cluster bias can be interpreted as a modulation of the collapse threshold by long-wavelength density modes (Cole and Kaiser, 1989; Kaiser, 1984; White *et al.*, 1987). The idea here is that a long-wavelength density enhancement of amplitude $\delta\rho/\rho = \epsilon$ lowers the effective collapse threshold for smaller-scale structures to $\delta_c - \epsilon$, thereby inducing an offset in $dn_M/d\ln\sigma^{-1}$ from its mean value on mass scale M . This contribution adds to the perturbation ϵ in cluster number density owing to the amplitude of the large-scale mode. Dividing the sum of these two offsets by ϵ leads to an expression relating

the bias parameter to the mass function (Mo and White, 1996; Sheth and Tormen, 1999):

$$b(M) = 1 - \frac{d}{d\delta_c} \left[\ln \left(\frac{dn_M}{d\ln \sigma^{-1}} \right) \right] . \quad (56)$$

Plugging in the Sheth-Tormen mass function of equation (54) produces

$$b(M) = 1 + \frac{1}{\delta_c} \left[\frac{a_s \delta_c^2}{\sigma^2} - 1 + \frac{2p_s}{1 + (a_s \delta_c^2 / \sigma^2)^{p_s}} \right] . \quad (57)$$

Hu and Kravtsov (2003) show that the parameter values $a_s = 0.75$ and $p_s = 0.3$ accurately reproduce the bias of cluster-sized halos seen in large-scale numerical simulations, when cluster mass is taken to be M_{180m} . Notice that small values of $\sigma(M)$ lead to large values of $b(M)$, meaning that rare, high-mass objects are much more likely to be found in regions of the universe where the surrounding matter density is higher than average.

C. Measuring the Cluster Mass Function

Equations (52), (54), and (55) illustrate why cosmologists are so enthusiastic about the cluster mass function. Dividing an accurate measurement of the mass function by $\Omega_M \rho_{\text{cr0}}$ directly leads to an accurate measurement of the primordial power spectrum $\sigma(M)$ on mass scales $\sim 10^{14} - 10^{15} M_\odot$. Furthermore, any uncertainty in ρ_{cr0} scales out of the power spectrum's normalization, because measured values of cluster number density scale as h^3 , making the quantity $M_8 \rho_{\text{cr0}}^{-1} (dn_M / d\ln \sigma^{-1})$ independent of h . One is left only with a degeneracy between σ_8 and Ω_M . Taking the logarithmic derivative of (52) with respect to σ at constant M shows that the mass function is roughly $\propto \sigma^2$ in the region where $\sigma \approx 1$. Hence, the measured level of that normalization in the local universe reflects the parameter combination $\sigma_8 \Omega_M^\alpha$, with $\alpha \approx 0.5$.

This degeneracy can be broken in three ways. First, one can simply measure Ω_M or σ_8 in some other way. Second, one can measure the cluster mass function over a range of masses and rely on a precise measurement on the mass function's shape to break the degeneracy, assuming that the CDM power spectrum (Sec. III.A.5) is valid. Or, third, one can measure the evolution of the cluster mass function, which is highly sensitive to Ω_M . We will explore that option in more depth in Sec. III.D, but first we need to examine some of the obstacles to accurate mass-function measurements.

1. Linking Mass with Observables

In order to measure the mass function using a large sample of clusters, we need to relate cluster mass to an easily observable quantity. Doing this properly requires a consistent definition of mass (Sec. III.B.3) and a well-calibrated relation linking that mass to some observable, but which mass definition works best? Expressions

like the Jenkins mass function (equation 55) appear to be cosmologically invariant when cluster masses are defined with respect to the background mass density (e.g., M_{180m} , see Sec. III.B.4). On the other hand, the structure of a cluster, as reflected in its dark-matter velocity dispersion, seems to be cosmologically invariant when cluster masses are defined with respect to the critical mass density (e.g., M_{200} , see Sec. III.B.3). To paraphrase Evrard (2004), Nature appears to do *accounting* relative to the mean mass density and *dynamics* relative to the critical density.

Because simulations suggest that dynamical quantities like the galaxy velocity dispersion and the X-ray determined gas temperature should be more tightly correlated with M_{200} than with M_{180m} , we will take M_{200} to be the primary definition of cluster mass. Methods like those outlined in Sec. III.B.3 can then be used to convert a mass function in M_{200} to one in M_{180m} . Alternatively, one can fit the results of large-scale simulations to determine a cosmology-dependent correction to the Jenkins mass function for use with the mass definition M_{200} . Evrard *et al.* (2002) have done that, finding that substituting $A_J = 0.27 - 0.07(1 - \Omega_M)$, $B_J = 0.65 + 0.11(1 - \Omega_M)$, and $\epsilon_J = 3.8$ into equation (55) reproduces the M_{200} mass function in simulations at $z = 0$. Despite the tight relationship between M_{200} and the dark-matter velocity dispersion in simulations, the link between M_{200} and observable quantities is still a potentially large source of systematic error. Even if the galaxy velocity dispersion were identical to that of the dark matter, accurately measuring that dispersion within a *sphere* of radius r_{200} requires an enormous amount of data to minimize projection effects (e.g., Rines *et al.*, 2003).

To see how systematic errors corrupt the mass-function measurement, consider the general case for a generic observable X . Suppose a cluster survey determines the comoving number density distribution $dn_M / d\ln X$ within logarithmic bins of the observable. Converting this distribution to a mass function $dn_M / d\ln \sigma^{-1}$ via the chain rule requires, at minimum, knowledge of the normalization and effective power-law index $\alpha_X \equiv d\ln X / d\ln M$ of the $M_{200}-X$ relation over the observed range in X , as well as the effective power-law index $\alpha_M \equiv d\ln \sigma^{-1} / d\ln M$ of the mass fluctuations. Fitting a semi-analytic expression for the mass function like equation (55) to the observations for a fixed value of Ω_M then determines $\sigma_{\text{fit}} = \sigma(M_{\text{fit}})$ on a particular mass scale M_{fit} , and consequently determines $\sigma_8 \approx (M_{\text{fit}}/M_8)^{\alpha_M} \sigma_{\text{fit}}(M_{\text{fit}})$.

Any systematic offset $\Delta M/M$ in the normalization of the $M_{200}-X$ relation produces a corresponding offset in the measured power-spectrum normalization:

$$\begin{aligned} \frac{\Delta \sigma_8}{\sigma_8} &= \left(\alpha_M + \frac{d\ln \sigma_{\text{fit}}}{d\ln M_{\text{fit}}} \right) \frac{\Delta M}{M} \\ &= \alpha_X \left[\alpha_M + \frac{1}{\epsilon_J (\ln \sigma^{-1} + B_J)^{\epsilon_J - 1}} \right] \frac{\Delta X}{X} \end{aligned} \quad (58)$$

(e.g., Evrard *et al.*, 2002; Seljak, 2002; Voit, 2000). The second line of this equation assumes that σ_{fit} has been

determined using the Jenkins mass function of equation (55). On the mass scale of rich clusters ($\gtrsim 10^{14.5} M_\odot$), Evrard *et al.* (2002) find that the factor in parentheses is ≈ 0.4 , implying that a systematic 25% error in mass would lead to a 10% error in the measurement of σ_8 . Below this mass scale the factor in parentheses increases, leading to an even larger error in the power-spectrum normalization for a given mass offset (Huterer and White, 2002).

Dispersion in the value of the mass-tracing observable for clusters of fixed mass is another important source of uncertainty that must be dealt with carefully because of the exponential slope of the mass function (e.g., Pierpaoli *et al.*, 2003). Significant scatter boosts the normalization of $dn_M/d\ln X$ over the expectation for the no-scatter case, as the overall number of lower-mass clusters scattering to higher values of X far exceeds the number of higher-mass clusters scattering in the opposite direction. Underestimating this scatter leads to an overestimate of σ_8 that can be particularly severe if the scatter has a long non-gaussian tail to large values of X . Situations in which such a tail could arise include merger shocks that substantially boost the temperature and luminosity in a significant subset of X-ray selected clusters (e.g., Randall *et al.*, 2002; Ricker and Sarazin, 2001) and superpositions of galaxies that boost the apparent richness and velocity dispersion in an optically-selected sample (van Haarlem *et al.*, 1997).

Surveys that probe deep into the universe for clusters must also cope with redshift evolution in the mass-observable relation. That is, if $M_{200} \propto X^{\alpha_X} (1+z)^{b_X}$, then one needs to know the value of b_X . This source of uncertainty affects both the mapping of X onto mass for individual clusters and the number density one infers for clusters of a given mass from a survey based on the observable X .

A sufficiently large cluster survey can circumvent many of these systematic problems through self-calibration (Hu, 2003; Levine *et al.*, 2002; Majumdar and Mohr, 2003, 2004). This procedure treats all parameters describing the systematic uncertainties in such things as the scatter, normalization, and evolution of the mass-observable relation as free parameters in the overall cosmological model. Fitting a large number of clusters spanning a wide range in redshift to this overall model, one can then determine not only the global cosmological parameters but also the most likely values of the free parameters in the mass-observable relation. However, treating the systematic uncertainties in this way has a cost. Each free parameter added weakens the statistical constraints on the cosmological measurements (Majumdar and Mohr, 2003).

2. Mass-Temperature Relation

Among the observables that trace cluster mass, X-ray temperature has received considerable recent attention

because it is closely related to the depth of a cluster's potential well and can be readily observed to $z \sim 1$ with current X-ray telescopes (Sec. II.B.2). Henry and Arnaud (1991) pioneered the technique of measuring the cluster mass function with X-ray temperatures, using cluster temperatures determined at $z \approx 0$ with the *Einstein*, *Exosat*, and *HEAO/OSO* satellites. Cluster temperatures measured with the *ASCA* satellite improved the precision of this measurement (Ikebe *et al.*, 2002), and temperatures measured with the *Chandra* and *XMM-Newton* telescopes should improve that precision even more. Because the data are now of such high quality, systematic uncertainty in the link between mass and temperature is the main factor limiting this technique.

Mass and temperature ought to be simply related for a cluster in hydrostatic equilibrium. The gas temperature of a singular isothermal sphere with mass M_{200} inside radius r_{200} is

$$\begin{aligned} k_B T_{200} &= \frac{GM_{200}\mu m_p}{2r_{200}} = \frac{\mu m_p}{2} [10 GM_{200} H(z)]^{2/3} \\ &= (8.2 \text{ keV}) \left(\frac{M_{200}}{10^{15} h^{-1} M_\odot} \right)^{2/3} \left[\frac{H(z)}{H_0} \right]^{2/3} \end{aligned} \quad (59)$$

Realistic departures from hydrostatic equilibrium can be assessed with simulations of structure formation that include hydrodynamics, but they do not have a large effect on the mass-temperature relation. These models indeed find that gas temperature scales with $M_{200}^{2/3}$, so that $T \approx T_{200}$, with a scatter of only 10-15% (e.g., Evrard *et al.*, 1996; Frenk *et al.*, 1999). However, to calibrate the mass-temperature relation more precisely, we need a more specific definition of temperature.

Clusters are not perfectly isothermal, so any single number specifying a cluster's gas temperature is some sort of weighted mean. The luminosity-weighted mean temperature T_{lum} obtained by weighting each gas parcel's temperature by $\propto \rho_g^2$ is a popular choice for comparing theory with observations because each temperature component contributes in proportion to its photon flux in the cluster's overall spectrum (Sec. II.B.2). The spectral-fit temperature T_{sp} has not yet received much attention in theoretical work because it depends somewhat on the procedure used to fit the spectrum, but Mazzotta *et al.* (2004) have recently developed a temperature weighting scheme for theoretical models that appears to track T_{sp} quite closely.

A bewildering variety of parameters has been used in the literature to express the mass-temperature relation's normalization. Here we will express the normalization of the M_{200} - T_{lum} relation in terms of T_{lum}/T_{200} . This choice has two advantages: It does not link the normalization to any particular mass or temperature scale, and $T_{\text{lum}}/T_{200} = 1$ for both a singular isothermal sphere and an isothermal beta model with $\beta = 2/3$.

Uncertainty in this normalization factor is currently the single most important issue afflicting cluster mass-function measurements with X-ray observations. Table I

TABLE I Normalization of the Mass-Temperature Relation

Models without Radiative Cooling	T_{lum}/T_{200}^a
Navarro <i>et al.</i> (1995)	0.99
Evrard <i>et al.</i> (1996) ^b	0.91
Bryan and Norman (1998)	0.80
Thomas <i>et al.</i> (2001)	0.98
Muanwong <i>et al.</i> (2002) ^c	0.57
Muanwong <i>et al.</i> (2002) ^d	0.90
Observations	T_{lum}/T_{200}^a
Horner <i>et al.</i> (1999) ^e	$(1.08 \pm 0.04)T_6^{-0.19}$
Horner <i>et al.</i> (1999) ^f	$(1.40 \pm 0.16)T_6^{-0.02}$
Nevalainen <i>et al.</i> (2000) ^g	$(1.20 \pm 0.12)T_6^{-0.20}$
Finoguenov <i>et al.</i> (2001b) ^{b,h}	$(1.18 \pm 0.10)T_6^{-0.33}$
Finoguenov <i>et al.</i> (2001b) ^{b,i}	$(1.26 \pm 0.11)T_6^{-0.19}$
Finoguenov <i>et al.</i> (2001b) ^{b,j}	$(1.33 \pm 0.18)T_6^{-0.02}$
Models with Radiative Cooling	T_{lum}/T_{200}^a
Muanwong <i>et al.</i> (2002)	$0.79 T_6^{-0.31}$
Muanwong <i>et al.</i> (2002) ^k	$0.88 T_6^{-0.09}$
Borgani <i>et al.</i> (2003) ^b	$(1.03 \pm 0.03)T_6^{-0.06}$
Borgani <i>et al.</i> (2003) ^{b,l}	$(1.24 \pm 0.03)T_6^{-0.06}$

^a $T_6 \equiv k_B T_{\text{lum}}/6 \text{ keV}$ ^bConversion from M_{500} assumes $M_{200} = 1.4M_{500}$.^c T_{lum} computation includes gas with cooling time $< 6 \text{ Gyr}$ ^d T_{lum} computation excludes gas with cooling time $< 6 \text{ Gyr}$ ^eMasses estimated using isothermal beta model^fMasses estimated using polytropic beta model^gConversion from M_{1000} assumes $M_{200} = 2.0M_{1000}$.^hFull sample, masses from isothermal beta modelⁱFull sample, masses from polytropic beta model^jSubset with $k_B T_{\text{lum}}$, polytropic beta model^k T_{lum} computed with cooling cores removed^lMasses inferred from polytropic beta-model fits.

provides some recent observational and theoretical calibrations of this relation, in three different groups. The first group gives calibrations from hydrodynamical simulations that do not account for galaxy formation, which generally fall into the range $T_{\text{lum}}/T_{200} = 0.8 - 1.0$. In some cases, an M_{500} - T_{lum} relation has been converted to M_{200} assuming $M_{200} = 1.4M_{500}$, appropriate for halo concentration $c = 5$. The second group gives calibrations inferred from observations, which fall into the range $T_{\text{lum}}/T_{200} = 1.1 - 1.4$. In other words, clusters of a given temperature seem to be 30% to 60% less massive than one would expect from the simulations. Apparently galaxy formation changes the normalization of the M_{200} - T_{lum} relation, for reasons discussed in detail in Sec. IV, although some of this discrepancy may also stem from systematic offsets in the observational interpretation. The third group of normalization factors, which tend to lie in between the first two groups, come from simulations that attempt to account for the effects of galaxy formation. Given this uncertainty in the mass-temperature normalization, the systematic uncertainty in σ_8 values derived from cluster temperatures is about 25%, because $\sigma_8 \propto (T_{\text{lum}}/T_{200})^{3/5}$ according to equation (59) for rich clusters (Evrard *et al.*, 2002).

Efforts are underway to reconcile the observed nor-

malization of the M_{200} - T_{lum} relation with theoretical expectations. Some of the discrepancy probably arises from systematic errors in the masses derived from X-ray observations. Hydrostatic equilibrium is usually assumed, but the turbulent velocities in simulated clusters can sometimes be $\sim 20\text{-}30\%$ of the sound speed, in which case the hydrostatic assumption would lead to masses underestimated by 10-15% (Rasia *et al.*, 2003; Ricker and Sarazin, 2001). In addition, the beta model formalism often used to derive cluster mass may have systematic problems. Applying this model to simulated clusters tends to underestimate their masses (Borgani *et al.*, 2003; Muanwong *et al.*, 2002, see the last line of Table I), and recent *XMM-Newton* observations suggest that the correction for temperature gradients may be excessive (Mushotzky, 2004; Pratt and Arnaud, 2002, 2003).

Alternative mass measurements would be very helpful in solving these problems. Calibration of the mass-temperature relation with lensing observations has met with mixed success. Weak-lensing measurements of massive, relaxed clusters agree with the masses derived from X-ray data under the assumption of hydrostatic equilibrium (Allen *et al.*, 2001). However, that agreement seems to be poorer for less relaxed clusters (Smith *et al.*, 2003). Measurements of cluster mass from the galaxy velocity field in and around a few very well observed clusters also tend to support the X-ray derived masses (Rines *et al.*, 2003). Because the calibration may depend systematically on how clusters are selected, self-calibration of a large cluster survey may ultimately be the best way of calibrating the M_{200} - T_{lum} relation. A thorough understanding of how galaxy formation affects that relation would help reduce the number of free parameters that need to be calibrated, thereby reducing the statistical uncertainties achievable with self-calibration.

3. Mass-Luminosity Relation

X-ray luminosity also correlates well with cluster mass and is easier to measure than X-ray temperature, allowing for mass-function measurements using much larger cluster samples. However, the correlation between mass and luminosity is not as tight as that between mass and temperature, having a scatter $\sim 50\%$ (Reiprich and Böhringer, 2002). Additionally, the normalization and slope of the relation depend heavily on the physics of galaxy formation (Sec. IV.C). Because our understanding of the connection between galaxy formation and a cluster's X-ray luminosity is not yet mature enough to calibrate the mass-luminosity relation with simulations, we need to rely solely on observational calibrations.

One common way to calibrate the mass-luminosity relation is to combine the mass-temperature relation with the observed luminosity-temperature relation (e.g., Borgani *et al.*, 1999). On cluster scales, the relation between the total (bolometric) X-ray luminosity and

TABLE II Luminosity-Temperature Relation at $z \approx 0$

Source	L_6^a	α_{LT}
Edge and Stewart (1991)	6.3 ± 1.3	2.62 ± 0.10
David <i>et al.</i> (1993)	5.6 ± 0.9	3.37 ± 0.05
Markevitch (1998) ^b	6.4 ± 0.6	2.64 ± 0.27
Allen and Fabian (1998) ^c	5.7 ± 3.4	2.92 ± 0.45
Allen and Fabian (1998) ^d	14.6 ± 7.3	3.08 ± 0.58
Arnaud and Evrard (1999) ^e	5.9 ± 0.4	2.88 ± 0.15
Xue and Wu (2000)	7.6 ± 1.2	2.79 ± 0.08
Novicki <i>et al.</i> (2002)	6.0 ± 4.2	2.82 ± 0.43
Ettori <i>et al.</i> (2002)	7.3 ± 1.8	2.54 ± 0.42

^aBolometric X-ray luminosity is $L_X = L_6(T_{\text{lum}}/6 \text{ keV})^{\alpha_{LT}}$ with L_6 in units of $10^{44} h_{70}^{-2} \text{ erg s}^{-1}$.

^bCores of clusters excised to avoid cool cores.

^cClusters without cool cores.

^dClusters with cool cores.

^eSample avoids clusters with cool cores.

T_{lum} is approximately a power law. Normalizing the relation at 6 keV, in the heart of the temperature range for rich clusters, leads to the expression $L_X = L_6(T_{\text{lum}}/6 \text{ keV})^{\alpha_{LT}}$, and Table II gives some representative values of L_6 and α_{LT} . Excising the central regions of clusters, out to about 100 kpc in radius, reduces the scatter in the relation because cooling and non-gravitational heating processes affect the temperature and luminosity of these regions differently from cluster to cluster (Allen and Fabian, 1998; Fabian *et al.*, 1994; Markevitch, 1998; Voit *et al.*, 2002).

The power-law index of the L_X-T_{lum} relation clearly indicates that galaxy formation has affected the L_X-T_{lum} relation. If the density distribution of intracluster gas within r_{200} were self-similar, independent of cluster mass, then one would expect bremsstrahlung emission to give $L_X \propto \rho_g M_{200} T_{\text{lum}}^{1/2} \propto T_{\text{lum}}^2$ (Kaiser, 1986). The steepness of the observed power-law index indicates that non-gravitational processes have raised the entropy of the intracluster gas, making it harder to compress, particularly in the shallower potential wells of cool clusters. This excess entropy therefore lowers the luminosities of all clusters by lowering the mean gas density and steepens the L_X-T_{lum} relation because the impact of excess entropy decreases as cluster temperature rises (Evrard and Henry, 1991; Kaiser, 1991, Sec. IV.C).

Calibrating the mass-luminosity relation by coupling the mass-temperature relation with the observed luminosity-temperature relation leads, not surprisingly, to values of σ_8 that are similar to those derived from the mass-temperature relation alone and are subject to the same systematic uncertainties that plague the mass-temperature calibration. There is, however, a route to the mass-luminosity calibration that circumvents the middle step involving the mass-temperature relation.

The mass-luminosity relation can be calibrated more directly with high-quality X-ray imaging and temperature data on a complete sample of clusters. Reiprich and Böhringer (2002) have done this with

ROSAT imaging data and *ASCA* temperatures, finding

$$L_X = 10^{45.0 \pm 0.3} h_{70}^{-2} \text{ erg s}^{-1} \left(\frac{M_{200}}{10^{15} h_{70}^{-1} M_\odot} \right)^{1.8}. \quad (60)$$

Their mass calibration assumes that the cluster gas is in hydrostatic equilibrium and obeys an *isothermal* beta model. The masses they derive are therefore higher than those one would find after correcting for a possible negative temperature gradient at large radii but do not account for any turbulent pressure support. With this mass-luminosity relation, they find a cluster mass function whose normalization corresponds to $\sigma_8 = 0.68(\Omega_M/0.3)^{-0.38}$.

Furthermore, because their observed cluster sample extends over two decades in mass, Reiprich and Böhringer (2002) attempted to break the $\sigma_8-\Omega_M$ degeneracy by fitting the mass-function's shape with a CDM power spectrum, finding a best fit of $\Omega_M = 0.12^{+0.06}_{-0.04}$ and $\sigma_8 = 0.96^{+0.15}_{-0.12}$, with $\Omega_M < 0.31$ at the 3σ level. The unusually low best-fit value of Ω_M arises because their derived mass function is shallower than that expected for $\Omega_M = 0.3$. However, Pierpaoli *et al.* (2003) have applied that same L_X-M_{200} relation to the larger *REFLEX* cluster sample (Böhringer *et al.*, 2002), finding $\sigma_8 = 0.86^{+0.12}_{-0.16}$ and $\Omega_M = 0.23^{+0.10}_{-0.06}$. Results similar to these latter values are also found from the mass-luminosity relation when cluster evolution is used to break the $\sigma_8-\Omega_M$ degeneracy (Sec. III.D).

4. Mass-Richness Relation

Optical telescopes have gathered much larger cluster samples than have X-ray telescopes, but deriving a mass function from these samples is not so straightforward. Projection effects complicate both the measurement of cluster mass and the computation of the sample volume associated with a given mass. Clusters in optical surveys are selected on the basis of richness, which depends on the number of galaxies observed within a certain projected radius from the center of the cluster (Sec. II.A.1). Thus, even if optical luminosity traces mass exactly, galaxy concentrations lying outside r_{200} but projected along the same line of sight can boost the apparent mass, introducing non-gaussian uncertainties in the mass-richness relation. Likewise, the effective volume associated with a given cluster mass in an richness-selected survey is harder to quantify than in a survey with a definite X-ray flux cutoff. Nevertheless, when richness is rigorously defined, it correlates well with a cluster's X-ray properties (Yee and Ellingson, 2003). However, the scatter between optical richness and X-ray luminosity is still large compared with the accuracy to which one would like to derive cosmological parameters (Donahue *et al.*, 2001, 2002; Gilbank *et al.*, 2003; Kochanek *et al.*, 2003).

Measuring cluster masses purely on the basis of galaxy-count richness necessitates a different approach to defin-

ing a cluster's radius and therefore its mass. Because traditional measures of richness depend on the radius within which galaxies are counted, they are defined with respect to a fixed physical radius, independent of mass, at each redshift. For this reason, observations of cluster richness are sometimes compared with simulations on the basis of cluster masses measured within a constant physical or comoving radius (Bode *et al.*, 2001). Deriving a mass-richness relation from simulations of galaxy formation also involves an observational calibration of the cluster mass-to-light ratio, which according to equation (59) introduces a systematic uncertainty in σ_8 that is $\sim 40\%$ of the uncertainty in the mass-to-light conversion.

Making such a comparison with the early clusters from the Sloan Digital Sky Survey, Bahcall *et al.* (2003b) find a mass-function normalization implying $\sigma_8 = 0.69 \pm 0.07(\Omega_M/0.3)^{-0.6}$. Adding mass-function shape information to break the degeneracy leads to $\Omega_M = 0.19^{+0.08}_{-0.07}$ and $\sigma_8 = 0.9^{+0.3}_{-0.2}$, in reasonably good agreement with the X-ray derived values. Unfortunately, because there is as yet no simple parametric form, analogous to the Jenkins mass function, giving the mass function defined with respect to a fixed radius, it is not clear how to self-calibrate the associated mass-richness relation to high accuracy with a large survey.

5. Velocity Dispersion and Mass

Velocity dispersion is the optical analog to X-ray temperature. Thus, one would expect a mass function defined on the basis of velocity dispersion to coincide with those defined with respect to cluster temperature. Measurements of rich clusters indicate that $\sigma_{1D}^2 = (1.0 \pm 0.1)k_B T_{lum}/\mu m_p$ (e.g., Xue and Wu, 2000, Sec. II.B.2), which reassuringly suggests that both quantities accurately trace mass. On the other hand, Evrard *et al.* (2002) have pointed out that combining this relation with the observational calibration of the M_{200} - T_{lum} relation ($T_{lum} \approx 1.2 T_{200}$) leads to a puzzle. While it might be possible for non-gravitational effects associated with galaxy formation to boost T_{lum} , it is more difficult to imagine why non-gravitational effects would boost the galaxy velocity dispersion by a similar factor.

Mass functions derived from velocity dispersion measurements also suggest that the masses derived from those measurements are larger than those derived from X-ray data. Using the virial theorem with a pressure correction term (Sec. II.A.2), Girardi *et al.* (1998) derive a cluster mass function from velocity dispersions whose normalization indicates $\sigma_8 = (1.01 \pm 0.07)(\Omega_M/0.3)^{-0.43}$, implying an overall number density at a given cluster mass about two times larger than the X-ray measurements. A discrepancy in σ_8 as large as 30% could arise if the optically determined masses were over 50% larger, but the actual mass discrepancies appear not to be quite so large. Reiprich and Böhringer (2002) find that in the 42 clusters they have in common with Girardi *et al.*

(1998) the virial masses are 25% larger, on average, than the X-ray masses. Another factor that could contribute to this discrepancy is scatter in the M_{200} - σ_{1D} relation. An underestimate of the scatter would drive up the inferred mass-function amplitude, raising the best-fitting value of σ_8 .

Part of the discrepancy between the optical and X-ray determined masses may stem from how velocity dispersions are observed. Because σ_{1D} declines with projected radius, its observed value will depend on the cutoff radius. Also, any foreground or background interlopers projected onto the cluster can contaminate the velocity-dispersion measurement. Ideally, one would like to cut off the measurement at a spherical boundary with radius r_{200} , inside of which the relation between dark-matter velocity dispersion and mass is well-calibrated, but the large number of galaxy velocities needed to accurately measure the mass profile near the virial radius make this approach impractical for large cluster samples. In a small sample of eight rich clusters, Rines *et al.* (2003) have used an average of almost 200 galaxy velocities per cluster, extending to well beyond r_{200} , to measure the mass M_{200} within r_{200} . The masses they find are consistent with both the X-ray determined masses and with the virial theorem including a surface-pressure correction.

6. Weak Lensing and Mass

Weak lensing is a very promising method for measuring cluster masses that is independent of a cluster's baryon content, dynamical state, and mass-to-light ratio. The main systematic problem in weak-lensing mass measurements comes from the lensing done by excess mass outside the virial radius but along the line of sight through the cluster (Hoekstra, 2001; Metzler *et al.*, 2001, 1999). So far, weak-lensing's main contribution to cluster studies has been to assist in the calibration of other mass estimators (e.g., Allen *et al.*, 2001).

Techniques for compiling cluster samples selected on the basis of weak lensing are still in their infancy. Only a few clusters with confirmed spectroscopic redshifts have been detected in weak lensing surveys (e.g., Dahle *et al.*, 2003; Schirmer *et al.*, 2003; Wittman *et al.*, 2003, 2001). However, deep optical surveys covering wide patches of the sky should turn up many more such clusters in the coming decade. In the meantime, smaller weak-lensing surveys sensitive to large-scale structure are complementing the cluster work because they provide values of σ_8 that are independent of the cluster measurements. Numbers currently in the literature span approximately the same range as those derived from clusters, going from $\sigma_8 = (0.72 \pm 0.08)(\Omega_M/0.3)^{-0.57}$ (Jarvis *et al.*, 2003) on the low end to $\sigma_8 = (0.97 \pm 0.14)(\Omega_M/0.3)^{-0.44}$ (Bacon *et al.*, 2003) on the high end.

7. Baryons and Mass

Yet another technique for measuring the cluster mass function relies on the constancy of the ratio of baryons to dark matter in massive clusters. X-ray observations from *Chandra* indicate that the ratio of hot baryonic gas to total gravitating matter within a given radius asymptotically approaches $(0.113 \pm 0.005)h_{70}^{-3/2}$ in relaxed, high-mass clusters (Allen *et al.*, 2002). Correcting for the baryons in stars, whose mass is approximately $0.16h_{70}^{1/2}$ times that of the hot gas (Sec. IV.D, raises the overall ratio of baryons to dark matter in clusters to $f_b = 0.13$ for $h_{70} = 1.0$.

This ratio is itself one of the best tools for measuring Ω_M (Allen *et al.*, 2002; David *et al.*, 1995; Evrard, 1997; White *et al.*, 1993). No known hydrodynamic process can drive a large proportion of a rich cluster's baryons out of the cluster's deep potential well. Thus, the ratio of baryons to dark matter in a cluster is expected to be similar to the global ratio in the universe. Dividing the mean baryon density $\Omega_b = 0.045 h_{70}^{-2}$ consistent with both the abundances of light elements (Burles *et al.*, 2001) and microwave background fluctuations (Spergel *et al.*, 2003) by the baryon fraction f_b implies $\Omega_M \approx 0.3$. Allen *et al.* (2002) find $\Omega_M = 0.30^{+0.04}_{-0.03}$ after marginalizing over the uncertainties in Ω_b and Hubble's constant.

One can also use ratio of baryons to dark matter to constrain dark energy (Pen, 1997; Sasaki, 1996). Measurements of this ratio in clusters depend on the relationship between transverse size and redshift, which depends on both Ω_M and Ω_Λ (Sec. III.A.2). If the actual ratio remains constant with redshift, then the measured ratios will be independent of redshift only if the correct values of Ω_M and Ω_Λ are used in the measurement. Allen *et al.* (2004) have recently shown that the measured baryon to dark matter ratio in a sample of 26 clusters ranging up to $z \approx 0.9$ is consistent with the low-redshift ratio for $\Omega_\Lambda = 0.94^{+0.21}_{-0.23}$ (see also Allen *et al.*, 2002; Ettori *et al.*, 2003). However, the degree to which the actual ratio is redshift-independent is not yet known.

If the ratio of baryons to dark matter were completely independent of cluster mass and radius, then measurements of the baryon mass inside a radius containing a mean baryon density of $200f_b\rho_{\text{cr}}$ would directly give M_{200} . The cluster mass function could then be determined by measuring the baryon masses within a given scale radius (Vikhlinin *et al.*, 2003; Voevodkin and Vikhlinin, 2004). In fact, the baryon fraction is not quite constant in clusters, probably owing to the same galaxy-formation effects that shift the M - T and L - T relations (Sec. IV.C). For example, Mohr *et al.* (1999) find that the ratio of gas mass to dark matter is $\propto T_{\text{lum}}^{0.36 \pm 0.22}$ in clusters cooler than about 6 keV and is statistically inconsistent with a constant value at the 99% level. Other studies concur that the proportion of hot gas in low-mass clusters is smaller than that in high-mass clusters (Neumann and Arnaud, 2001; Sanderson *et al.*, 2003).

After correcting for this effect, Voevodkin and Vikhlinin (2004) infer a cluster mass function from the baryon mass function signifying $\sigma_8 = 0.72 \pm 0.04$ for the assumed cosmology ($\Omega = 0.3$, $\Omega_\Lambda = 0.7$, and $h = 0.71$). Furthermore, the shape parameter $\Gamma = 0.13 \pm 0.07$ of the mass function is consistent with the CDM power spectrum given the assumed values of Ω_M and h . Notice that this value of σ_8 agrees with those derived from the observationally calibrated M_{200} - T_{lum} and M_{200} - L_X relations, even though it does not explicitly rely on those calibrations.

D. Evolution of the Mass Function

Measurements of evolution in the cluster mass function can considerably tighten all these constraints on cosmological parameters. What we actually observe, of course, is the dependence on redshift of the observables that trace the cluster mass function. For a given cluster sample we can measure the number of clusters dN within a given solid angle $d\Omega$ and redshift interval $[z, z + dz]$ that fall into the range $[X, X + dX]$ of the observable X . With full knowledge of the mass-observable relation $M(X, z)$ and its scatter as a function of redshift, we could then derive the redshift distribution

$$\frac{d^3 N}{dM d\Omega dz}(M, z) = \frac{dn_M}{dM}(M, z) \cdot \frac{d^2 V_{\text{co}}}{dz d\Omega}(z) \quad (61)$$

for clusters of mass M directly from the observations. This distribution of clusters with redshift would then provide strong constraints on cosmological models through both the mass-function evolution factor dn_M/dM and the comoving volume factor $d^2 V_{\text{co}}/d\Omega dz$ from equation (21).

As the reader probably suspects by now, our ability to constrain cosmological parameters through the redshift distribution of clusters is currently limited by our understanding of evolution in the mass-observable relations. However, this problem is not as severe as one might expect because the evolution in the mass function itself is so dramatic, especially for $\Omega_M \approx 1$. This part of the review discusses what we have learned about structure formation and cosmological parameters by observing cluster evolution. It begins with a description of how mass-function evolution depends on cosmological parameters and then considers the complications arising from evolution of the observables themselves. It concludes with a summary of current constraints on Ω_M from cluster evolution and a look at the prospects for measuring Ω_Λ and w with large cluster surveys.

1. Dependence on Cosmology

Evolution of the mass function is highly sensitive to cosmology because the matter density controls the rate at which structure grows. When the mass function can be expressed in terms of formulae like (52),

(54), or (55), its evolution is controlled entirely by the growth function $D(z)$, which is a well defined function of Ω_M , Ω_Λ , and w (Sec. III.B.4). Small-amplitude density perturbations grow as $D(z) = (1+z)^{-1}$ when $\Omega_M(z) \approx 1$, but perturbation growth stalls when $\Omega_M(z) \ll 1$. This effect manifests itself most strongly in high-mass clusters because they are the latest objects to form in a hierarchical cosmology with a CDM-like power spectrum (Eke *et al.*, 1996; Evrard, 1989; Oukbir and Blanchard, 1992; Peebles *et al.*, 1989; Viana and Liddle, 1996). The exponential dependence of the mass function on $\sigma(M, z) = D(z)\sigma(M, 0)$ makes the effect quite dramatic for objects sufficiently massive that $\sigma(M, 0) < 1$.

Dependence of the mass function on Ω_Λ and w is a little more subtle. These parameters affect mass-function evolution by altering the redshift at which $\Omega_M(z)$ departs significantly from unity for a given value of Ω_M at $z = 0$ (Haiman *et al.*, 2001). The time at which dark energy begins to dominate the dynamics of the universe is later for both larger values of Ω_Λ and smaller (more negative) values of w (see Figure 2), leading to greater evolution of the mass function between $z \sim 1$ and the present (Battye and Weller, 2003; Wang and Steinhardt, 1998; Weller *et al.*, 2002).

Measurements of how the mass function changes with redshift can provide additional information about Ω_Λ and w through the expansion rate of the universe. If the mass function of clusters is precisely known, then number counts of clusters exceeding a given mass in each redshift interval dz reveal the volume associated with that redshift interval and can be used to determine the dynamics of the universe's expansion. The number of clusters with mass $> M$ on the celestial sphere in the redshift interval dz is given by

$$\frac{dN}{dz}(M) = \frac{4\pi r_\kappa^2(z)c}{H(z)} n_M(M, z) . \quad (62)$$

Figure 6 shows this number-redshift distribution for several different cosmological models. Notice that the statistical power of cluster surveys is ultimately limited by the total number of massive clusters in the observable universe, which is of order 10^5 .

2. Evolution of the Observables

All of the mass-observable relations discussed in Sec. III.C evolve with redshift, partly because the definition of M_{200} is pinned to the critical density and partly because of galaxy-formation physics. Clusters of a given mass are hotter earlier in time because their matter density is larger; both T_{200} and the square of the dark-matter velocity dispersion for a fixed value of M_{200} vary with redshift as $H^{2/3}(z)$ (Sec. III.C.2). One therefore expects T_{lum} and the square of the galaxy velocity dispersion to depend on redshift in the same way, but it is possible that the physics of galaxy formation adds additional redshift

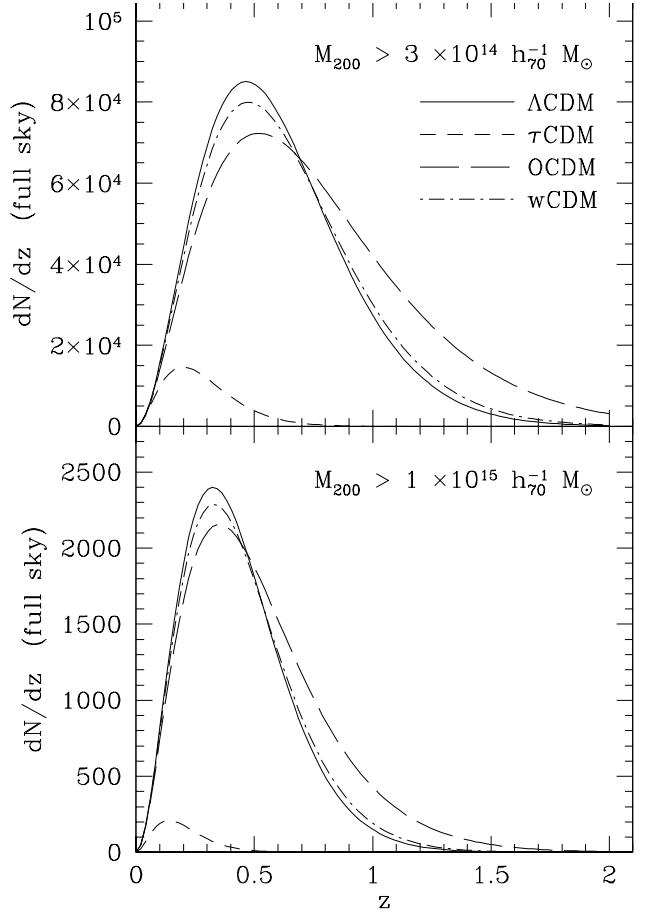


FIG. 6 Predicted number of clusters on the sky as a function of redshift in different cosmologies. The upper panel shows the number of clusters per unit redshift with $M_{200} > 3 \times 10^{14} h_{70}^{-1} M_\odot$ over the entire sky. Notice that there are a few tens of thousands of such clusters on the sky in models with $\Omega_M = 0.3$, most of them at $z < 1$. There are many fewer massive clusters at $z > 0.5$ in the τ CDM model with $\Omega_M = 1$ because cluster evolution is so rapid in that case. The lower panel shows the numbers of clusters with $M_{200} > 1 \times 10^{15} h_{70}^{-1} M_\odot$. Differences between models with $\Omega_M \approx 0.3$ but differing values of Ω_Λ and w should be detectable in large cluster surveys containing $\sim 10^4$ clusters and extending to $z \sim 1$.

evolution that must be accounted for in precise cosmological measurements. Galaxy formation plays a more explicit role in the mass-richness and M_{200} - L_X relations, because the optical luminosities of galaxies evolve with time and because the physics of galaxy formation alters the L_X - T_X relation (Sec. IV.C). Scatter in the mass-observable relation might also be larger at higher redshifts, given that the proportion of relaxed clusters may well be smaller earlier in time.

As an example of how mass-observable evolution affects observations of mass-function evolution, consider its effects on X-ray surveys. The upper left of Figure 7 shows how the cluster mass function evolves for two dif-

ferent cosmologies, a standard Λ CDM model ($\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, $w = -1$, $\sigma_8 = 0.9$) and a τ CDM model ($\Omega_M = 1.0$, $\Omega_\Lambda = 0.0$, $\sigma_8 = 0.5$, $\Gamma = 0.21$) whose power spectrum has been adjusted by hand so that its shape is similar to that of the Λ CDM model, as required by observations of large-scale structure (Sec. III.C). Mass-function evolution is quite pronounced in both models but is far stronger in the τ CDM model.

Evolution in the mass-temperature relation weakens the observed amount of cluster evolution when cluster number density is plotted as a function of temperature. The upper right of Figure 7 shows the result of using a M_{200} - T_{lum} relation with $T_{\text{lum}}/T_{200} = 1$ and zero dispersion. Because clusters of a given mass are hotter at earlier times, the higher redshift curves have translated to higher temperatures, compared with their positions in the upper-left panel. Additional mass-temperature evolution exceeding that predicted by the virial theorem and corresponding to values of T_{lum}/T_{200} that increase with redshift, would further reduce the evolution, but there is currently no evidence for such evolution.

Redshift-dependent changes in the luminosity-temperature relation can have additional evolution-softening effects. The $L_X \propto T_{\text{lum}}^3$ power-law form of the relation appears to remain the same with redshift, but the amount of evolution in the normalization is uncertain. Early assessments suggested no evolution in the normalization (Borgani *et al.*, 1999; Donahue *et al.*, 1999; Mushotzky and Scharf, 1997). The lower right of Figure 7 shows the evolution of cluster number density plotted against luminosity for a non-evolving normalization and $L_X = 6 \times 10^{44} h_{70}^{-2} \text{ erg s}^{-1} (T_{\text{lum}}/6 \text{ keV})^3$, again with no dispersion. These curves differ from the temperature-function curves only in the labeling of the horizontal axis. More recent results indicate that higher-redshift clusters of a given temperature are more luminous, with an evolving relation $L_X \propto T_{\text{lum}}(1+z)^{b_{LT}}$ where $0.5 \lesssim b_{LT} \lesssim 1.5$ (Ettori *et al.*, 2003; Lumb *et al.*, 2003; Vikhlinin *et al.*, 2002, see Sec. IV.C.4). Figure 7 shows the same distribution functions for $b_{LT} = 1.5$, at the high end of the suggested range. The extra redshift dependence in this case slides the high-redshift curves even further to the right, roughly compensating for all of the evolution in the underlying mass function.

These examples underscore the importance of constraining evolution in the mass-observable relations, even if the observables could be perfectly measured. In addition, one must bear in mind that the observations themselves can introduce spurious redshift dependences in the mass-observable relations, largely because distant clusters are more difficult to observe than nearby ones. Optical projection effects become progressively harder to deal with at high redshift, complicating observations of richness and velocity dispersion, observations of weak lensing have fewer background galaxies to measure, and the decline in X-ray surface brightness makes cluster temperature measurements more difficult. In many ways, the Sunyaev-Zeldovich effect is the most promising observ-

able for characterizing high-redshift clusters because its magnitude does not depend on redshift (Sec. II.C).

There are three basic ways to deal with evolution in the normalization and perhaps the scatter of a mass-observable relation:

- Assume a model for the evolution of the relation. Numerical simulations can be very helpful in providing a model for evolution of the normalization and scatter of mass-observable relations but give accurate results only if they include all the relevant physics.
- Assume a parametric form for the mass-observable relation inspired by theoretical models and try to calibrate it directly with observations. The normalization of the relation is usually assumed to be proportional to either $(1+z)$ or $H(z)$ raised to a power determined by a fit to observations. In practice, however, the mass-observable relations for distant clusters are not directly calibrated. What we have instead are relations that link one easily observed quantity, such as X-ray luminosity, to another that is more closely related to mass, like X-ray temperature or the weak-lensing distortion.
- Assume a parametric form for the mass-observable relation and apply self-calibration techniques to a large cluster survey to find the most likely parameters describing mass-observable evolution (Hu, 2003; Levine *et al.*, 2002; Majumdar and Mohr, 2003, 2004). Parameters involving redshift-dependent scatter in the relation can also be included in such an analysis. This technique is very promising but requires large surveys of distant clusters which are not yet in hand. Its accuracy is limited by the number of free parameters needed to describe the mass-observable relations—the fewer, the better. Having a realistic physical model for mass-observable evolution helps boost the accuracy achievable with self-calibration by reducing the number of unknown parameters.

A decade from now, when much larger cluster samples will be available, self-calibration will probably be the best way to calibrate the mass-observable relations. In the meantime, it would be wise to spend some effort on direct observational calibrations through cross-comparisons of multiple mass-tracing observables.

3. Constraints on Dark Matter

Surveys of distant clusters find modest evolution in their comoving number density fully consistent with cosmological models in which $\Omega_M \approx 0.3$. Because the rate of mass-function evolution at moderate redshifts ($z \lesssim 0.5$) is governed primarily by the overall matter density, this conclusion does not depend strongly on the value of Ω_Λ .

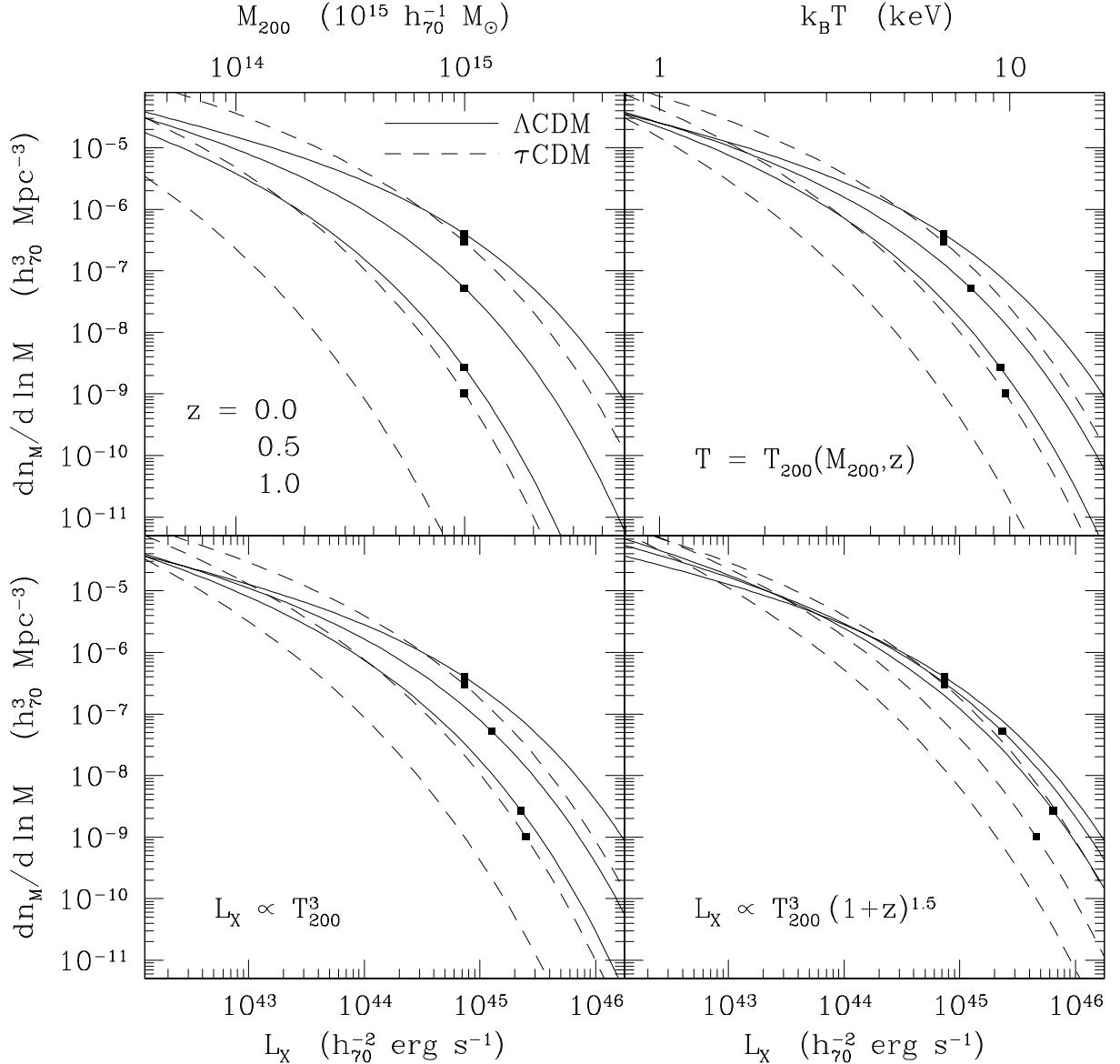


FIG. 7 Evolution of the cluster mass function and its manifestations in temperature and luminosity. The two models in this comparison are Λ CDM ($\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, $w = -1$, $\sigma_8 = 0.9$) and τ CDM ($\Omega_M = 1.0$, $\Gamma = 0.21$, $\sigma_8 = 0.5$). Evolution of the mass function, shown at the upper left, is far more pronounced in τ CDM (dashed lines) than in Λ CDM (solid lines) because it is so sensitive to the current matter density. Each set of three lines shows the differential mass function $dn_M/d \ln M$ at $z = 0$, 0.5 , and 1.0 , from top to bottom, and black squares show the value of the mass function at a fiducial mass of $10^{15} h_{70}^{-1} M_\odot$. The upper right panel shows the same mass functions plotted against temperature, assuming $T = T_{200}(M_{200}, z)$. Notice that the higher-redshift curves have shifted to the right, weakening the evolution in temperature space, because clusters of a given mass have higher temperatures at higher redshifts. In order to convert these curves to temperature functions, one would need to convolve them with the scatter in the mass-temperature relation and multiply by $d \ln M / d \ln T \approx 1.5$. The lower two panels show these same curves as a function of luminosity, assuming $L_X = (6 \times 10^{44} h_{70}^{-2} \text{ erg s}^{-1}) T_{200}^3$ at $z = 0$ and two different redshift dependences of the L_X-T relation. In the case without L_X-T evolution on the left-hand side, the curves are just relabeled versions of the ones in the upper-right panel. However, the strong L_X-T evolution in the right-hand panel ($L_X \propto T^3(1+z)^{1.5}$) shifts the three curves in the Λ CDM case nearly on top of one another at $L_X \approx 10^{44} h_{70}^{-1} \text{ erg s}^{-1}$. Convolving these curves with the dispersion in the mass-luminosity relation and multiplying by $d \ln M / d \ln L \approx 0.5$ converts them to luminosity functions.

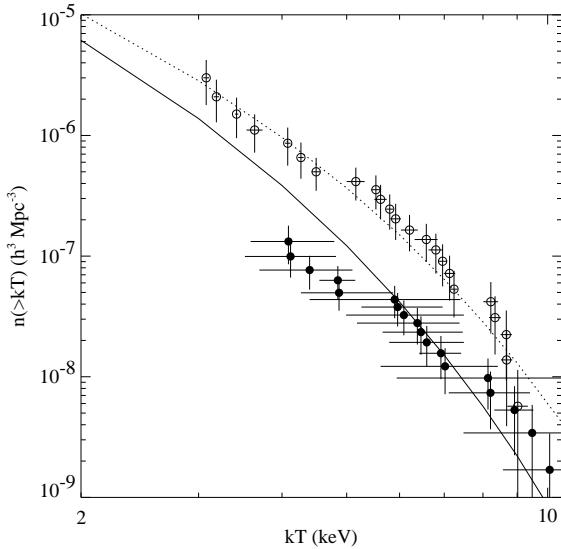


FIG. 8 Observed evolution in the integrated cluster temperature function $n(>kT)$ giving the comoving number density of clusters with temperatures exceeding kT . Open circles give the low-redshift temperature function from a sample of clusters with mean redshift $z = 0.051$. Filled circles give the observed temperature function of clusters with a mean redshift of $z = 0.429$. The data points in each case are correlated because $n(>kT)$ at a given temperature is a cumulative function depending on all data points at higher temperatures. Lines give the predicted temperature functions at $z = 0.051$ (dotted line) and $z = 0.429$ (solid line) for the best-fitting model: $\Omega_M = 0.28$, $\Omega_\Lambda = 0.98$, $\sigma_8 = 0.68$. (Figure courtesy of Pat Henry.)

Here we focus on the constraints on Ω_M derived from X-ray surveys, whose observables— L_X , T_X , and baryonic mass—are related to the spherical mass M_{200} through simple parametric relations.

Evolution in the X-ray temperature function was first observed by Henry (1997), who showed that the comoving number density of ~ 5 keV clusters at $z \sim 0.35$ was only slightly smaller than it is today. Assuming standard evolution of the mass-temperature relation, Eke *et al.* (1998) derived matter-density constraints $\Omega_M = 0.38 \pm 0.2$ for $\Omega_\Lambda = 1 - \Omega_M$ and $\Omega_M = 0.44 \pm 0.2$ for $\Omega_\Lambda = 0$ from these data using a maximum likelihood analysis to take full advantage of the sparse temperature data. More conservative analyses that simply counted clusters hotter than a given temperature found weaker constraints (Viana and Liddle, 1999). Henry (2000) provides a complete discussion of the cluster temperature data and the maximum-likelihood analysis technique.

Temperature measurements of a handful of hot clusters at higher redshifts have shown that the rate of cluster temperature evolution remains modest at higher redshifts (Donahue, 1996; Donahue *et al.*, 1998, 1999; Henry, 2000). The comoving number density of $\gtrsim 8$ keV clusters at $z \sim 0.5 - 0.8$ is no less than about one tenth of its current value, in strong disagreement with

the standard expectation in an $\Omega_M = 1$ universe (see Figure 7). Including these hot, distant clusters in the analysis further strengthens the constraints on the matter density, ruling out $\Omega_M = 1$ at the 3σ level for standard mass-temperature evolution (Bahcall and Fan, 1998; Donahue and Voit, 1999; Donahue *et al.*, 1998; Evrard *et al.*, 2002). In order for such hot clusters to exist in a flat, matter-dominated universe, the mass-temperature relation would have to evolve in a non-standard way, with an increase in either the scatter or the normalization at $z \gtrsim 0.5$. Evrard *et al.* (2002) have shown that the τ CDM mass function of Figure 7 is consistent with the temperature function observations only if the mass-temperature normalization factor T_{lum}/T_{200} is ~ 1.5 times higher at $z \sim 0.5$ than at present. Such a big change seems unlikely in light of alternative observations of these hot, high-redshift clusters that agree with the large masses inferred from the standard normalization (Donahue *et al.*, 1998; Luppino and Gioia, 1995; Tran *et al.*, 1999).

Observations of evolution in the X-ray luminosity function have greater statistical power because many more clusters have known luminosities than have known temperatures, but uncertainties in luminosity-temperature evolution dilute the constraints they place on Ω_M . Many X-ray surveys have shown that the comoving number density of clusters at a given luminosity changes very little from redshift $z \sim 0.8$ to the present for $L_X \lesssim 10^{44} \text{ erg s}^{-1}$; significant evolution is seen only for clusters with $L_X \gtrsim 10^{45} \text{ erg s}^{-1}$ (Mullis *et al.*, 2004; Rosati *et al.*, 2002). Evolution this mild is generally expected in cosmological models with $\Omega_M \approx 0.3$. Strong evolution in the luminosity-temperature relation must occur in models with $\Omega_M = 1$ for the observed evolution in the luminosity function to be so weak (see Figure 7). An extensive analysis by Borgani *et al.* (2001) of luminosity-function evolution in the *ROSAT* Deep Cluster Survey, which extends to $z \sim 1$, indicates that $\Omega_M = 0.35^{+0.13}_{-0.10}$, where the error bars signify the 1σ confidence interval. Models with $\Omega_M = 1$ fall outside the 3σ confidence interval, even when the normalization of the luminosity-temperature relation is allowed to vary with redshift as $L_X \propto T_{\text{lum}}(1+z)$.

The evolution of the baryon mass function observed with X-ray telescopes agrees with the conclusions drawn from the luminosity and temperature functions. Vikhlinin *et al.* (2003) have measured the baryon mass function in a sample of clusters at $z \sim 0.5$, finding that the comoving number density of massive clusters at that redshift is roughly one tenth of the current value. This result implies $\Omega_M = 0.25 \pm 0.1$ (1σ confidence interval) for $\Omega_\Lambda = 1 - \Omega_M$.

Optical studies concur that cluster evolution has been relatively modest since $z \sim 0.5$, buttressing the conclusion that $\Omega_M < 1$. In fact, the evolution of optically selected clusters appears even milder than the evolution in X-ray selected clusters (e.g., Postman *et al.*, 2002), which would imply an even smaller value of Ω_M . However, it is not yet clear how much of this discrepancy arises from

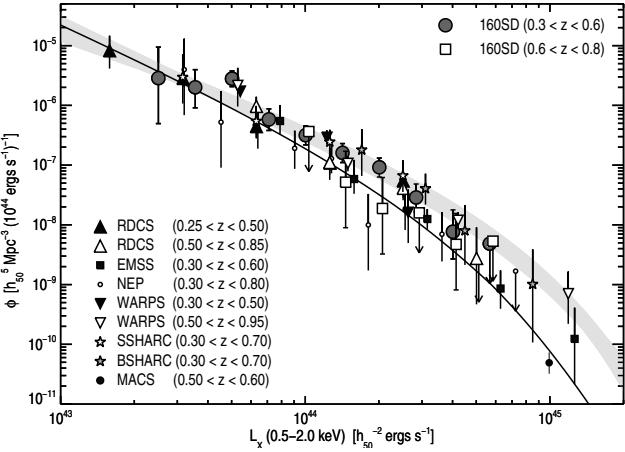


FIG. 9 Observed evolution in the cluster luminosity function. Many different cluster surveys spanning the range $0 \lesssim z \lesssim 1$ are shown on this figure. The vertical axis gives the luminosity functions derived from these surveys in terms of $\phi \equiv dn/dL_X$, and the shaded region shows the luminosity function at $z \approx 0$. Significant evolution is seen only at $L_X \gtrsim 10^{45} \text{ erg s}^{-1}$, consistent with ΛCDM models with a moderate amount of evolution in the L_X-T relation (see Figure 7.) (Figure from Mullis *et al.* (2004).)

differences between the projected masses measured by optical surveys and the spherical masses measured by X-ray surveys.

4. Constraints on Dark Energy

Existing cluster surveys, taken by themselves, do not yet place strong constraints on dark energy, but that situation is likely to change in the coming decade, with the advent of large, deep cluster surveys in the optical, X-ray, and microwave bands. Currently, the most interesting information that clusters provide about dark energy comes from combining the results of cluster surveys with other information. If the overall geometry of the universe is indeed flat, as seems quite evident from the temperature patterns in the cosmic microwave background (e.g., Spergel *et al.*, 2003), then the matter density inferred from clusters implies $\Omega_\Lambda = 1 - \Omega_M = 0.7 \pm 0.1$, in agreement with measurements of Ω_Λ from the supernova magnitude-redshift relation (Perlmutter *et al.*, 1999; Riess *et al.*, 1998). Geometrical arguments involving clusters provide weaker support for this conclusion. If the baryon fraction of clusters at a given temperature is assumed to remain constant with time, then the transverse sizes of clusters as a function of redshift can be used to constrain the geometry of the universe. Studies using such methods disfavor $\Omega_\Lambda = 0$ (Arnaud *et al.*, 2002; Mohr *et al.*, 2000).

Large cluster surveys extending to $z \sim 1$ have the potential to place much stronger constraints on the dark-energy parameters Ω_Λ and w , independent of other information,

as long as these surveys are large enough to permit self-calibration of the mass-observable relationships (Holder *et al.*, 2001; Levine *et al.*, 2002; Weller *et al.*, 2002). The accuracy achievable with self-calibration depends critically on the nature of cluster evolution, because the self-calibration procedure requires evolution of the relevant mass-observable relation to be expressed in a parametric form. Constraints on the cosmological parameters are considerably weaker if the actual evolution does not follow the assumed parametric form. However, cross-calibration of mass-observable evolution through intensive supplementary observations of a small subset of the large survey restores much of the potential inherent in self-calibration (Majumdar and Mohr, 2003).

Including information on cluster bias inherent in a large cluster survey further tightens the constraints on dark energy. Because the tendency of clusters to cluster with one another depends in a simple way on the cosmological model (Sec. III.B.5), folding this information into the self-calibration procedure improves the accuracy with which cosmological parameters can be measured (Majumdar and Mohr, 2004). Figure 10 shows how the estimated constraints on Ω_M and w tighten when information about cluster bias is added. It assumes that the universe is flat ($\Omega_\Lambda = 1 - \Omega_M$) and considers three different planned cluster surveys: two large Sunyaev-Zeldovich surveys (SPT and *Planck*) and a large X-ray survey (DUET), each of which will find 20,000 to 30,000 cluster to $z \gtrsim 1$ (see Majumdar and Mohr, 2004, for details). In the most optimistic cases, the parameters Ω_M , Ω_Λ , and w will be measured with $\sim 5\%$ accuracy.

A large survey also minimizes the sample variance that arises from cluster bias (Evrard *et al.*, 2002; Hu and Kravtsov, 2003). Because clusters tend to be clustered, the variance in the number of clusters within small sample volumes is larger than the gaussian expectation, adding systematic uncertainty to the measured mass function. This effect is generally not large for current cluster surveys but should be taken into account if one is designing a cluster survey for making high-precision cosmological measurements.

In summary, observations of cluster evolution already constrain the density of gravitating matter to be $\Omega_M \approx 0.3 \pm 0.1$, meaning that $\Omega_\Lambda \approx 0.7 \pm 0.1$ if the universe is flat. Using this value of Ω_M to break the $\Omega_M-\sigma_8$ degeneracy leads to $\sigma_8 \approx 0.7 - 1.0$, depending on the mass-temperature calibration. The major source of uncertainty in all these cosmological parameters comes not from the statistics of the survey but rather from uncertainties in the normalization and rate of evolution in the mass-observable relations. In order to better understand these relations and how they evolve, we need to know how galaxy formation affects the evolution of the stuff we can observe—the baryons in clusters. That is where we turn our attention next.

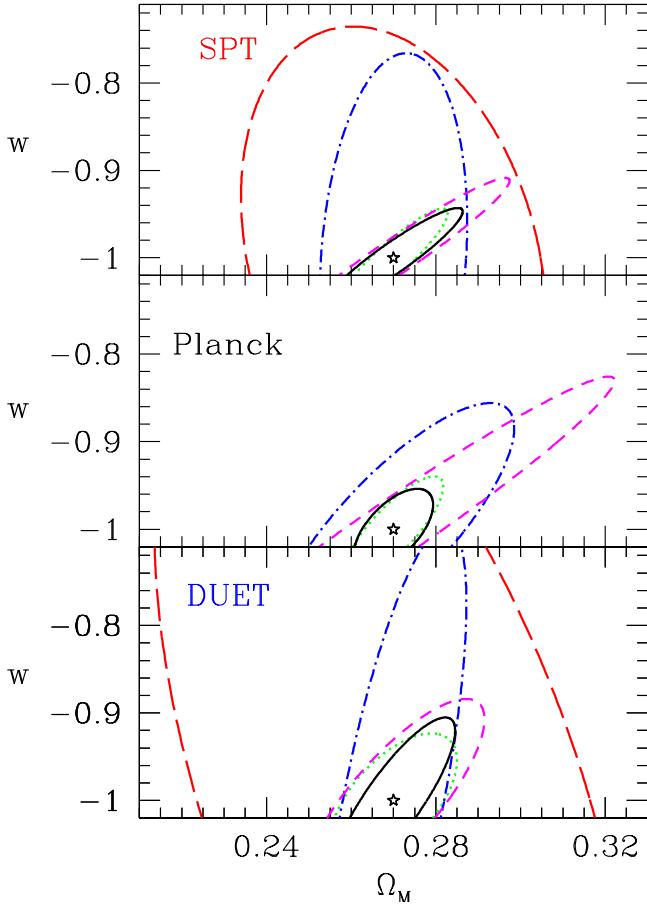


FIG. 10 Expected constraints on cosmological parameters from self-calibrated surveys. The SPT and *Planck* surveys will find 20,000 to 30,000 clusters through the S-Z effect. DUET is a proposed X-ray survey designed to find $\sim 20,000$ clusters. Dotted contours show the expected constraints on w and Ω_M from self-calibration if redshift evolution of the cluster observables behaves exactly according to the standard scaling relations, allowing the redshift dependences of those scaling relations to be fixed. Long-dashed contours show how the constraints loosen when redshift evolution is determined as part of the self-calibration. Dot-dashed contours show how the constraints begin to tighten when information about cluster bias is included in the calibration. Solid contours show the best-case scenario in which the self-calibration includes both information about cluster bias and supplementary follow-up calibrations of a small subset of the survey. (Figure from Majumdar and Mohr (2004).)

IV. EVOLUTION OF THE BARYONIC COMPONENT

Those who are cosmologists at heart are interested in how galaxy formation affects the intracluster medium primarily because they would like to know how to measure cluster masses more accurately. Those who are astronomers at heart are interested in the intracluster medium as well, but for them its main attraction is that the hot gas contains valuable information about the physical processes that govern galaxy formation. Clarifying

the connections between galaxy formation and the mass-observable relations is therefore important to both of these lines of research.

One of the nagging mysteries in our current picture of the universe is why so few of the universe's baryons have turned into stars (Cole, 1991; White and Frenk, 1991; White and Rees, 1978). Numerical simulations of cosmological structure formation that include baryonic hydrodynamics and the radiative cooling processes that lead to galaxy formation predict that $\gtrsim 20\%$ of the baryons should have condensed into galaxies, but $\lesssim 10\%$ have been found in the form of stars (e.g., Balogh *et al.*, 2001b). Some form of feedback, involving supernovae and perhaps outflows from active galactic nuclei, seems to have stymied condensation, but we are still largely ignorant about how this feedback works.

Galactic winds like those observed from nearby starburst galaxies, in which multiple clustered supernovae are driving the powerful outflows, are likely to be important in regulating early star formation, but observational constraints on the mass and energy flux in such winds are sketchy at best (Heckman, 2002; Martin, 1999), particularly at high redshift (Adelberger *et al.*, 2003; Pettini *et al.*, 2001, 2000). These galactic winds presumably had a dramatic impact on the intergalactic medium and subsequent galaxy formation, with effects that may have persisted until the present day (e.g., Benson and Madau, 2003; Oh and Benson, 2003). Likewise, quasars and other forms of activity driven by black-hole growth in the nuclei of young galaxies may also have produced powerful outflows with lasting consequences for the intergalactic gas, but the energy input from these objects is still highly uncertain (Inoue and Sasaki, 2001; Nath and Roychowdhury, 2002; Scannapieco and Oh, 2004; Voit, 1994, 1996).

Unfortunately, the low-redshift intergalactic medium, where most of the universe's baryons are thought to reside, is notoriously hard to observe. Because the majority of this gaseous matter remains undetected, it is sometimes referred to as the "missing baryons" (e.g. Cen and Ostriker, 1999). A handful of quasars are bright enough beacons for probing the missing baryons via absorption-line studies with the ultraviolet spectrographs on the Hubble Space Telescope (e.g., Penton *et al.*, 2002; Shull *et al.*, 1996), and that number will increase if the Cosmic Origins Spectrograph is installed on Hubble. However, the inferences drawn from such studies depend critically on the uncertain heavy-element abundance and ionization state of these intergalactic clouds (e.g., Shull *et al.*, 2003; Tripp *et al.*, 2000).

Clusters of galaxies are still the only places in the universe where we have anything approaching a complete accounting of intergalactic baryons, their thermal state, and their heavy-element enrichment. Thus, observations of the intracluster medium can provide unique insights into the cooling and feedback processes that govern galaxy formation. In order to interpret the signatures of galaxy formation in the intracluster medium we need

to understand how the thermodynamic properties of today’s clusters are linked to the physics of the intergalactic baryons at $z \gtrsim 2$, the epoch of galaxy formation.

This section of the review discusses the current understanding of the interactions between galaxy formation and the intracluster medium, focusing in particular on how those interactions affect the mass-observable relations so crucial to cosmology. It begins by outlining the properties that clusters would have if radiative cooling of the universe’s baryons and subsequent galaxy formation were suppressed. Because these properties do not agree with observations, radiative cooling and galaxy formation must somehow have altered the structure of the intracluster medium, with important consequences for the mass-observable relations. The middle of this section summarizes some of the recent progress that has been made in understanding the role of galaxy formation and its impact on the observable properties of clusters. It then concludes with a brief discussion of the existing constraints on baryon condensation in clusters.

A. Structure Formation and Gravitational Heating

People who study clusters of galaxies are sometimes asked how the X-ray emitting gas gets so hot. The answer to that question is simple. If radiative cooling is negligible, then gravitationally driven processes will heat diffuse gas to the virial temperature of the potential well that confines it. A tougher question would be to ask why the intracluster medium has the density that it does. In order to answer that question, one needs to know what produces the entropy of the X-ray emitting gas. Without galaxy formation in the picture, shocks driven by hierarchical structure formation are the only source of entropy for the intracluster medium, and this mode of entropy production leads to clusters whose density and temperature structures are nearly self-similar.

1. Intracluster Entropy

Entropy is of fundamental importance for two reasons: it determines the structure of the intracluster medium and it records the thermodynamic history of the cluster’s gas. Entropy determines structure because high-entropy gas floats and low-entropy gas sinks. A cluster’s intergalactic gas therefore convects until its isentropic surfaces coincide with the equipotential surfaces of the dark-matter potential. Thus, the entropy distribution of a cluster’s gas and the shape of the dark-matter potential well in which that gas sits completely determine the large-scale X-ray properties of a relaxed cluster of galaxies. The gas density profile $\rho_g(r)$ and temperature profile $T(r)$ of the intracluster medium in this state of convective and hydrostatic equilibrium are just manifestations of its entropy distribution.

This review adopts the approach of other work in this

field and defines “entropy” to be

$$K \equiv \frac{k_B T}{\mu m_p \rho_g^{2/3}} . \quad (63)$$

The quantity K is the constant of proportionality in the equation of state $P = K \rho_g^{5/3}$ for an adiabatic monatomic gas, and is directly related to the standard thermodynamic entropy per particle, $s = k_B \ln K^{3/2} + s_0$, where s_0 is a constant that depends only on fundamental constants and the mixture of particle masses. Another quantity frequently called “entropy” in the cluster literature is $S = k_B T n_e^{-2/3}$. In order to avoid confusion with the classical definition of entropy, we will call this quantity K_e . For the typical elemental abundances in the intracluster medium, one can convert between these definitions using the relation

$$\begin{aligned} K_e &= k_B T n_e^{-2/3} \\ &= 960 \text{ keV cm}^2 \left(\frac{K}{10^{34} \text{ erg cm}^2 \text{ g}^{-5/3}} \right) . \end{aligned} \quad (64)$$

A cluster achieves convective equilibrium when $dK/dr \geq 0$ everywhere, and the entropy distribution that determines the gas configuration in this state can be expressed as $K(M_g)$, where the inverse relation $M_g(K)$ is the mass of gas with entropy $< K$.

Comparisons between the entropy distributions of clusters that differ in mass can be simplified by casting those distributions into dimensionless form (e.g., Voit *et al.*, 2002). Combining the mean density of dark matter within the scale radius r_{200} , the global baryon fraction $f_b = \Omega_b/\Omega_M$, and the characteristic halo temperature T_{200} gives the characteristic entropy scale

$$\begin{aligned} K_{200} &= \frac{k_B T_{200}}{\mu m_p (200 f_b \rho_{\text{cr}})^{2/3}} \\ &= \frac{1}{2} \left[\frac{2\pi G^2 M_{200}}{15 f_b H(z)} \right]^{2/3} . \end{aligned} \quad (65)$$

For $f_b = 0.022 h^{-2}$, this entropy scale reduces to

$$\begin{aligned} K_{e,200} &= 362 k_B T_{\text{lum}} \text{ cm}^2 \left(\frac{T_{200}}{T_{\text{lum}}} \right) \\ &\times \left[\frac{H(z)}{H_0} \right]^{-4/3} \left(\frac{\Omega_M}{0.3} \right)^{-4/3} . \end{aligned} \quad (66)$$

Writing the entropy scale in this way makes explicit the fact that the observed temperature of a cluster is not necessarily a reliable guide to the characteristic entropy K_{200} of its halo. If the intracluster medium of a real cluster is either hotter or cooler than T_{200} , then one must apply the correction factor T_{200}/T_{lum} when computing the cluster’s value of K_{200} .

2. Entropy Generation by Smooth Accretion

One way to approach the problem of gravitationally driven entropy generation is through spherically

symmetric models of smooth accretion, in which gas passes through an accretion shock as it enters the cluster (e.g., Knight and Ponman, 1997; Tozzi and Norman, 2001; Voit *et al.*, 2003). If the incoming gas is cold, then the accretion shock is the sole source of intracluster entropy. If instead the incoming gas has been heated before passing through the accretion shock, then the Mach number of the shock is smaller and the intracluster entropy level reflects both the amount of preheating and the production of entropy at the accretion shock.

Let us first consider the case of cold accreting gas, in which the pressure and entropy of the incoming gas are negligible. Suppose that mass accretes in a series of concentric shells, each with baryon fraction f_b , that initially comove with the Hubble flow as in the spherical collapse model of Sec. III.B.1. In this simple model, a shell that initially encloses total mass M reaches zero velocity at the turnaround radius r_{ta} and falls back through an accretion shock at a radius r_{ac} in the neighborhood of the virial radius $r_{\text{ta}}/2$.

Because the cold accreting gas is effectively pressureless, the equations that determine the postshock entropy are

$$\dot{M}_g = 4\pi r_{\text{ac}}^2 \rho_1 v_{\text{ac}} \quad (67)$$

$$v_{\text{ac}}^2 = \frac{2GM}{r_{\text{ta}}} \quad (68)$$

$$k_B T_2 = \frac{1}{3} \mu m_p v_{\text{ac}}^2 \quad (69)$$

$$\rho_2 = 4\rho_1 , \quad (70)$$

where $\dot{M}_g = f_b \dot{M}$ is the gas accretion rate, ρ_1 is the preshock gas density, T_2 and ρ_2 are postshock quantities, and the accretion radius has been set to $r_{\text{ac}} = r_{\text{ta}}/2$. Equations (69) and (70) are restatements of the jump conditions for strong shocks, assuming that the postshock velocity is negligible in the cluster rest frame (e.g., Cavaliere *et al.*, 1997; Landau and Lifshitz, 1959), and equation (68) is exact only for cosmologies with $\Omega_\Lambda = 0$.

The postshock entropy produced by smooth accretion of cold gas at time t is therefore

$$\begin{aligned} K_{\text{sm}} &= \frac{v_{\text{ac}}^2}{3(4\rho_1)^{2/3}} \quad (71) \\ &= \frac{1}{3} \left(\frac{\pi G^2}{f_b} \right)^{2/3} \left[\frac{d \ln M}{d \ln t} \right]^{-2/3} (Mt)^{2/3} . \end{aligned}$$

Because the entropy generated at the shock front increases monotonically with time, such an idealized cluster never convects but rather accretes shells of baryons as if they were onion skins. The resulting entropy distribution in dimensionless form is

$$\begin{aligned} \frac{K_{\text{sm}}(M_g)}{K_{200}} &= \frac{2}{3} \left(\frac{15}{2} \right)^{2/3} (H_0 t_0)^{2/3} \\ &\times \left[\frac{d \ln \eta}{d \ln t} \right]^{-2/3} \left[\frac{\eta t(\eta)}{t_0} \right]^{2/3} , \quad (72) \end{aligned}$$

where $\eta \equiv M_g(t)/f_b M_{200}(t_0)$ is effectively a radial coordinate corresponding to the amount of gas accreted by time t divided by the amount accreted by the present time t_0 . Given these assumptions, the entropy profile arising from smooth accretion of cold gas depends entirely on the mass accretion history $M(t)$, and the profiles of objects with similar accretion histories should be self-similar with respect to K_{200} .

This simple model yields entropy distributions whose overall shape agrees with cluster observations (Voit *et al.*, 2003). The rate at which a cluster accretes matter through hierarchical structure formation depends on the growth function $D(t)$ and the power-law slope $\alpha_M \equiv d \ln \sigma^{-1} / d \ln M$ of the perturbation spectrum on the mass scale of the cluster: $M(t) \propto [D(t)]^{1/2\alpha_M}$ (Lacey and Cole, 1993; Voit and Donahue, 1998). Clusters ranging in mass from $10^{14} M_\odot$ to $10^{15} M_\odot$ grow roughly as $M(t) \propto t$ to $M(t) \propto t^2$ in the concordance model (Tozzi and Norman, 2001; Voit *et al.*, 2003). Plugging these growth rates into equation (72) leads to entropy distributions between $K \propto M_g$ and $K \propto M_g^{4/3}$. Throughout much of a cluster, the gas mass encompassed within a given radius rises approximately linearly with radius (Sec. II.B.1), meaning that the $K(r)$ relation should be slightly steeper than linear. Numerical models of smooth accretion by Tozzi and Norman (2001) find $K(r) \propto r^{1.1}$. The entropy profile observed outside the core regions of clusters also obey $K(r) \propto r^{1.1}$ (Pratt and Arnaud, 2002, 2003), but the extent to which this agreement is coincidental is not clear.

If the accreting gas is not cold, then the intracluster entropy profile produced by smooth accretion has an isentropic core with an entropy level similar to the preshock entropy (Balogh *et al.*, 1999; Tozzi and Norman, 2001). A non-zero initial entropy level K_1 changes the cold-accretion model outlined above by altering the jump conditions represented by equations (69) and (70). When K_1 is no larger than the entropy generated at the accretion shock, then the entropy profile created by smooth accretion of warm gas can be closely approximated by adding $0.84K_1$ to the entropy profile $K_{\text{sm}}(M_g)$ from the cold-accretion case (Dos Santos and Doré, 2002; Voit *et al.*, 2003). If K_1 is large compared with K_{sm} , then the accretion shock is weak or non-existent, and accretion is nearly adiabatic, leading to an isentropic entropy profile with the constant value K_1 .

3. Entropy Generation by Hierarchical Merging

In real clusters the accreting gas is lumpy, not smooth, which transforms the nature of entropy generation. Incoming gas associated with accreting sublumps of matter enters the cluster with a wide range of densities. There is no well-defined accretion shock but rather a complex network of shocks as different lumps of infalling gas mix with the intracluster medium of the main halo. Numerical simulations of this process beginning with cosmolog-

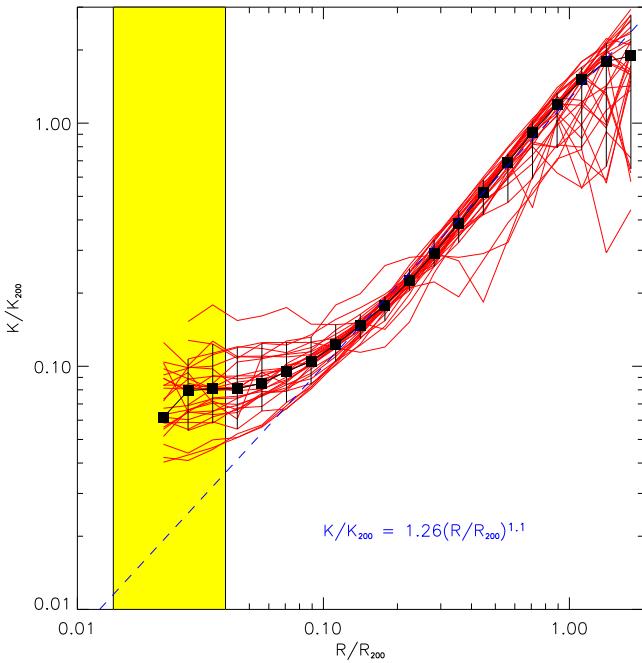


FIG. 11 Dimensionless entropy K/K_{200} as a function of scale radius r/r_{200} for 30 clusters simulated without radiative cooling or feedback. Black squares show the median profile, and the dashed line illustrates the power-law relation $K/K_{200} = 1.26(r/r_{200})^{1.1}$. Most of the entropy profiles shown lie close to this relation in the radial range $0.1 \lesssim r/r_{200} \lesssim 1.0$. At smaller radii, the entropy profiles generally flatten, and their dispersion increases. This flattening is likely to be a real effect, as it sets in well outside the shaded box showing the gravitational softening length of the simulation.

ical initial conditions produce clusters that have nearly self-similar entropy structure (e.g., Navarro *et al.*, 1995), as expected from the scaling properties of hierarchical structure formation (Kaiser, 1986).

Figure 11 shows entropy profiles of 30 clusters generated with a numerical simulation of a Λ CDM cosmology including hydrodynamics but not radiative cooling (Kay, 2004; Voit *et al.*, 2004). The masses of these clusters span more than a factor of 10, but when their entropy profiles are divided by the appropriate value of K_{200} , they lie nearly on top of one another, at least outside the approximate core radius $0.1r_{200}$. This result is not unique to the simulation method. All codes with sufficiently high resolution find that non-radiative clusters have approximately self-similar entropy structure, and consequently self-similar density and temperature structure (Frenk *et al.*, 1999; Voit *et al.*, 2004).

The self-similarity of the entropy profile in non-radiative clusters is a very useful point of comparison for sleuthing the effects of galaxy formation. Deviations from this baseline profile are likely to be due to a combination of radiative cooling and the feedback processes that ensue. Voit *et al.* (2004) find that a good representation of the baseline entropy profile produced outside the cores of clusters by hierarchical structure

formation is given by the power law $K_{\text{PL}} = (1.35 \pm 0.2)K_{200}(r/r_{200})^{1.1}$. Specifying the baseline entropy profile within the cluster core ($< 0.1r_{200}$) is more difficult both because there is more dispersion in that region among simulated clusters and because the results there depend somewhat on the hydrodynamical method used in the simulations.

Another notable aspect of self-similarity in non-radiative clusters is that the gas density profile and the dark-matter density profile outside $0.1r_{200}$ have virtually identical shapes (Frenk *et al.*, 1999; Navarro *et al.*, 1995). This feature leads to another useful approximation to the entropy profiles of non-radiative clusters (Bryan, 2000). One can specify the gas density profile by assuming that it obeys an NFW density profile (Sec. III.B.2) with the same concentration as the dark matter and a total baryon mass $f_b M_{200}$ within r_{200} and then compute the temperature profile that would keep the gas in hydrostatic equilibrium. The temperature and density profiles in this kind of model approximately obey the polytropic relation $T(r) \propto [\rho(r)]^{\gamma_{\text{eff}} - 1}$, with $\gamma_{\text{eff}} \approx 1.1 - 1.2$ (Komatsu and Seljak, 2001; Voit *et al.*, 2002). Combining them produces an alternative baseline entropy profile that depends on the concentration parameter c_{200} of the dark-matter halo and that this review will denote as $K_{\text{NFW}}(r)$.

Despite the complexity of the shock structure in hierarchical accretion, the numerically simulated entropy profiles are similar in shape to those created by smooth accretion models (Borgani *et al.*, 2002a, 2001). However, these profiles have lower overall entropy levels than the smooth-accretion profiles (Voit *et al.*, 2003). Figure 12 demonstrates this point by comparing the two approximations, K_{PL} and K_{NFW} , of simulated non-radiative clusters with two entropy profiles drawn from smooth accretion models, one from the numerical computations of Tozzi and Norman (2001) and the other from equation (72) assuming $M \propto t^{3/2}$ and $H_0 t_0 = 1$, which are reasonable assumptions for Λ CDM models.

The likely reason for this discrepancy is that smooth accretion maximizes entropy production because it minimizes the mean mass-weighted density of accreting gas (Ponman *et al.*, 2003; Voit *et al.*, 2003). Smoothing the accreting gas does not change the accretion velocity but does reduce the mean density of accreting gas lumps. Because postshock entropy scales as $v_{\text{ac}}^2 \rho_1^{-2/3}$, the mean entropy of lumpy accreted gas is therefore less than in the smooth-accretion case. This effect of smoothing might not be entirely academic, because the observed entropy profiles of low-temperature clusters show a similar entropy boost relative to the baseline profile (Voit and Ponman, 2003).

4. Observed Entropy Profiles

Astronomers have known for more than a decade that the structure of the intracluster medium in

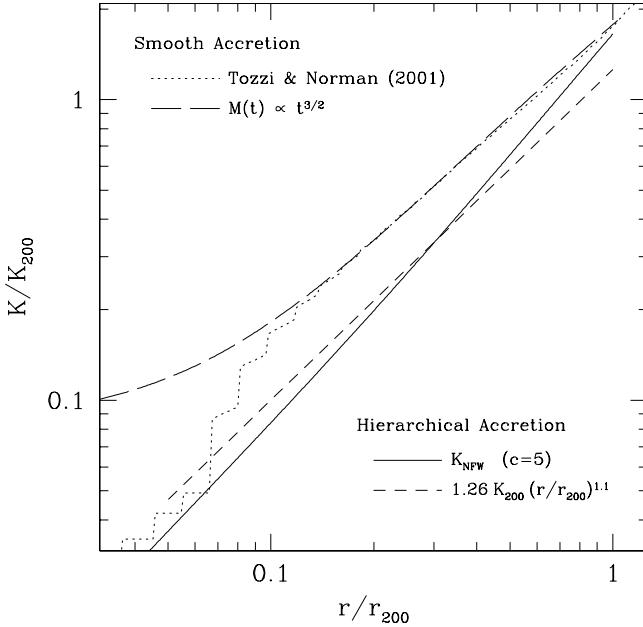


FIG. 12 Entropy profiles from smooth accretion and hierarchical accretion. Smoothing of the gas accreting onto a cluster boosts entropy production while maintaining the characteristic $K(r) \propto r^{1.1}$ entropy profile. The two lower lines illustrate approximate entropy profiles produced by hierarchical accretion, including the power-law expression from Figure 11 and the K_{NFW} model described in the text, with $c_{200} = 5$. The two upper lines illustrate entropy profiles resulting from smooth accretion models, including profile computed by Tozzi and Norman (2001) for a $10^{15} h^{-1} M_\odot$ cluster with 300 keV cm^2 of preheating and radiative cooling implemented (dotted line) and a profile computed from equation (72) and preheating amounting to $0.1K_{200}$. The two smooth models run roughly parallel to the hierarchical accretion models but their normalizations are ~ 1.5 times higher.

real clusters cannot be self-similar because the luminosity-temperature relation of clusters does not agree with self-similar scaling (Edge and Stewart, 1991; Evrard and Henry, 1991; Kaiser, 1991, Sec. III.C.3). Only within the last couple of years has the nature of that deviation from self-similarity become clear. High-quality cluster observations with the *XMM-Newton* satellite are showing that intracluster entropy profiles have the $K(r) \propto r^{1.1}$ shape characteristic of gravitational structure formation outside of the core, but the overall normalization of these profiles scales as $T_{\text{lum}}^{2/3}$ instead of T_{lum} , as in the baseline profiles (Pratt and Arnaud, 2003). Analyses of much larger cluster samples observed with earlier X-ray telescopes have arrived at the same conclusion. Instead of self-similarity with $K(r/r_{200}) \propto T_{\text{lum}}$, Ponman *et al.* (2003) find altered similarity with $K(r/r_{200}) \propto T_{\text{lum}}^{2/3}$ at both the core radius $0.1r_{200}$ and farther out in clusters, at the scale radius $r_{500} \approx 0.66r_{200}$. The question to be answered is therefore how galaxy for-

mation and feedback manage to produce such a shift in the overall normalization of cluster entropy profiles without substantially changing their shape.

B. Galaxy Formation and Feedback

In the decade since astronomers became aware of similarity breaking in clusters there have been many numerical simulations devoted to understanding it. Our understanding of this problem remains incomplete because including galaxy formation in cosmological models of cluster formation is a formidable computational challenge, requiring codes that simulate three-dimensional hydrodynamics spanning an enormous dynamical range in length scales and that track a large number of physical processes. The volume required to model a cosmologically significant sample of clusters is of order 10^{26} cm in linear scale, individual galaxies have sizes $\sim 10^{23} \text{ cm}$, star-forming regions within those galaxies can be as small as 10^{19} cm , and the stars themselves are only $\sim 10^{11} \text{ cm}$ in size. Sophisticated hydrodynamical techniques are now able to model the formation of the first stars from cosmological initial conditions (Abel, Bryan, & Norman 2001), but are far from being able to track in detail the formation of an entire galaxy's worth of stars, let alone all the feedback processes that can occur.

For the time being, the difficulty of solving this problem from first principles means that modelers have to be selective about the physical processes and conditions that merit modeling. Important clues to what the essential processes are can be gleaned from the observed characteristics of clusters. This part of the review sifts through some of those clues, showing that radiative cooling is likely to be the process that sets the entropy scale of similarity breaking but that radiative cooling cannot act alone. Otherwise, too much baryonic matter would condense into stars and cold gas clouds.

1. Preheating

Early approaches to the problem of similarity breaking in clusters postulated that some sort of heating process imposed a universal minimum entropy—an “entropy floor”—on the intergalactic gas before it collected into clusters (Evrard and Henry, 1991; Kaiser, 1991). Imposing a global entropy floor helps to bring the theoretical $L_{\text{X}}-T_{\text{lum}}$ relation into better agreement with observations because this extra entropy makes the gas harder to compress in cluster cores, where entropy is smallest, particularly in the shallower potential wells of low-temperature clusters. This resistance to compression breaks cluster similarity by lowering the core density and therefore the X-ray emissivity in low- T clusters more than in high- T clusters, thereby steepening the $L_{\text{X}}-T_{\text{lum}}$ relation.

According to this preheating picture, the core entropy level and scaling relations of clusters should reflect the

global entropy floor produced at early times. Initial measurements of entropy at the core radius $r_{0.1}$ demonstrated that low-temperature clusters had greater amounts of entropy than expected from self-similarity and suggested that the level of the entropy floor was $\sim 135 \text{ keV cm}^2$ (Lloyd-Davies *et al.*, 2000; Ponman *et al.*, 1999). This result matched well with numerical simulations of cluster formation with preheating levels of $50 - 100 \text{ keV cm}^2$ that produced clusters with approximately the right $L_X - T_{\text{lum}}$ relation (Bialek *et al.*, 2001).

However, simple preheating now appears to be too crude an explanation for similarity breaking. In the preheating picture, low-temperature clusters should have large isentropic cores (Balogh *et al.*, 1999; Tozzi and Norman, 2001), but this prediction disagrees with the observations showing that the shapes of cluster entropy profiles do not depend significantly on temperature (Sec. IV.A.4). In addition, the abundant evidence for intergalactic gas at $\lesssim 10^5 \text{ K}$ from quasar absorption line studies clearly shows that preheating cannot be global at $z \gtrsim 2$, and the preheating models themselves do not explain why the level of the entropy floor should be $\sim 135 \text{ keV cm}^2$.

2. Radiative Cooling

In contrast, the observed entropy scale of similarity breaking emerges naturally from the process of radiative cooling. Intergalactic gas both inside and outside of clusters radiates thermal energy at a rate given by the cooling function $\Lambda_c(T)$, described in more detail in Sec. II.B.1. Cooling that radiates an energy Δq per particle reduces the entropy by $\Delta \ln K^{3/2} = \Delta q / k_B T$. Thus, the equation expressing these radiative losses can be written

$$\frac{dK^{3/2}}{dt} = -\frac{3}{2} \frac{K_c^{3/2}(T)}{t_0}, \quad (73)$$

where

$$K_c(T) = \left[\frac{2}{3} \left(\frac{n_e n_p}{\rho^2} \right) \frac{(k_B T)^{1/2} \Lambda_c(T)}{(\mu m_p)^{1/2}} \right]^{2/3} t_0^{2/3} \quad (74)$$

is the entropy level at which constant-density gas at temperature T radiates an energy equivalent to its thermal energy in the time t_0 . The latter formula reduces to

$$K_c(T) \approx 81 \text{ keV cm}^2 \left(\frac{t_0}{14 \text{ Gyr}} \right)^{2/3} \left(\frac{T}{1 \text{ keV}} \right)^{2/3} \quad (75)$$

when pure bremsstrahlung cooling is assumed.

The fact that the entropy threshold below which gas cools within the universe's lifetime is close to the entropy floor inferred from clusters with $\sim 2 \text{ keV}$ temperatures suggests that radiative cooling sets the entropy scale for similarity breaking (Voit and Bryan, 2001). Voit and Ponman (2003) further quantify this point. Figure 13 shows how entropy measurements at $0.1r_{200}$ in a

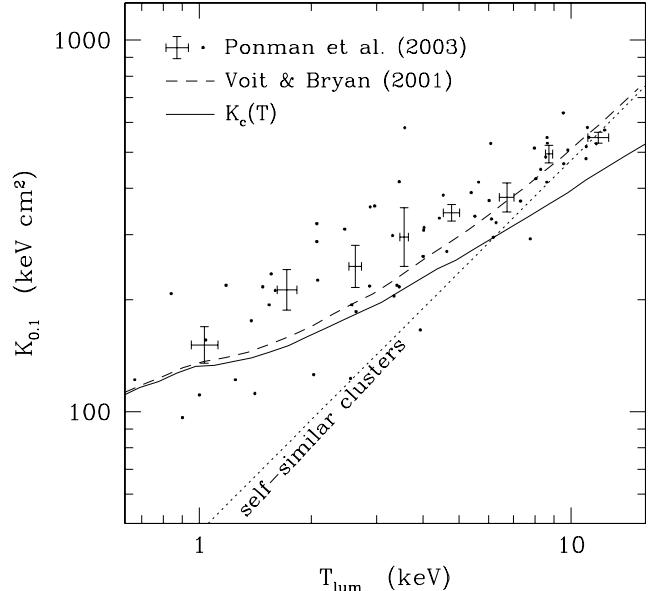


FIG. 13 Comparison between entropy measured at $0.1r_{200}$ and the cooling threshold in a large sample of clusters. Small points show the entropy $K_{0.1}$ measured at $0.1r_{200}$ in a sample of 64 clusters, and points with error bars show the mean entropy measurement in temperature bins of eight clusters each. The dotted line gives the mean entropy predicted by simulations of clusters without radiative cooling or feedback. The solid line shows the value of the cooling threshold $K_c(T)$ computed for heavy-element abundances 0.3 times their solar values and $t_0 = 14 \text{ Gyr}$. The dashed line shows the entropy predicted at $0.1r_{200}$ by the simple analytical model of Voit and Bryan (2001).

large sample of clusters (Ponman *et al.*, 2003) compare with the cooling threshold $K_c(T)$ for gas with heavy-element abundances equal to 30% of their solar values relative to hydrogen. Both the measured core entropies and the entropy threshold for cooling scale as $T^{2/3}$, and they are approximately equal, although the scatter in the data is quite significant.

Section IV.C.1 below shows that radiative cooling also accounts well for the scaling relations of global X-ray properties like L_X and T_{lum} with mass. However, casting equation (75) in dimensionless form illustrates why at least some feedback must compensate for cooling:

$$\frac{K_c(T)}{K_{200}} \approx 0.2 (Ht)^{2/3} \left[\frac{H(z)}{H_0} \right]^{2/3} \left(\frac{T}{1 \text{ keV}} \right)^{-1/3}. \quad (76)$$

The cooling threshold in low-temperature clusters at the present time is $\sim 20\%$ of the characteristic entropy K_{200} and greater than that if emission-line cooling from heavy elements is included. At earlier times, the dimensionless cooling threshold is even higher, meaning that a large proportion of the baryons belonging to the progenitor objects that ultimately assembled into present-day clusters should have condensed into stars or cold gas clouds. This is one of the manifestations of the

classic overcooling problem of hierarchical galaxy formation (Blanchard *et al.*, 1992; Cole, 1991; White and Rees, 1978). Because the observed mass ratio of stars to hot gas in clusters is only about 10% (Sec. IV.D), wholesale baryon condensation doesn't seem to have happened.

Recognition of this overcooling problem led Voit and Bryan (2001) to propose a way for radiative cooling to determine the entropy scale of similarity breaking without acting alone. The basic idea is that gas with entropy less than $K_c(T)$ cannot persist indefinitely. It must either cool and condense or be heated until its entropy exceeds $K_c(T)$. At any given time, feedback is triggered by condensing gas parcels with entropy less than the cooling threshold and acts until those parcels are eliminated by either cooling, heating, or some combination of the two. Thus, the joint action of cooling and feedback imprint an entropy scale corresponding to the cooling threshold, regardless of how strong the feedback is. This kind of effect has now been seen in a number of numerical simulations that include cooling and differing forms of feedback (Borgani *et al.*, 2002a; Borgani *et al.*, 2003; Davé *et al.*, 2002; Kay *et al.*, 2003; Valdarnini, 2003).

The fact that similarity breaking is not very sensitive to the efficiency of feedback is good news for cosmologists but bad news for astrophysicists. It offers hope that we can understand the mass-observable relations of clusters without solving all the messy astrophysical problems of feedback. Yet, it also implies that the mass-observable relations alone do not tell us much about the nature of that feedback. Instead, we must look to the spatially resolved entropy profiles of clusters (Kay, 2004; Voit and Ponman, 2003) and the ratio of condensed baryons to hot gas in clusters (Balogh *et al.*, 2001b; Borgani *et al.*, 2002a; Borgani *et al.*, 2003; Kay *et al.*, 2003).

3. Feedback from Supernovae

Supernovae are the most obvious candidate for supplying the feedback that suppresses condensation, but it is not clear that supernova heating and the galactic winds it drives can provide enough entropy to keep the fraction of condensed baryons below about 10%. Heavy-element abundances in clusters imply that the total amount of supernova energy released during a cluster's history amounts to $\sim 0.3 - 1$ keV per gas particle in the intracluster medium (Finoguenov *et al.*, 2001a; Pipino *et al.*, 2002). The amount of energy input needed to explain the mass-observable relations while avoiding overcooling is ~ 1 keV (Tornatore *et al.*, 2003; Voit *et al.*, 2002; Wu *et al.*, 2001), at the upper end of the range inferred from heavy elements, but the transfer of supernova energy to the intracluster medium must be highly efficient, which seems unlikely (Kraftsov and Yepes, 2000). Supernova energy would have to be converted to almost entirely to thermal energy with very little radiated away.

In order to avoid radiative losses, supernova heating

must raise the entropy of the gas it heats to at least 100 keV cm^2 . An evenly distributed thermal energy input of order 1 keV would therefore have to go into gas significantly less dense than 10^{-3} cm^{-3} to avoid such losses. Gas near the centers of present-day clusters, not to mention the galaxies where supernovae occur, is denser than that, particularly at earlier times when most of the star formation happened. Simulations that spread supernova feedback evenly therefore produce too many condensed baryons in clusters (Borgani *et al.*, 2002a). Artificial algorithms that target supernova feedback at gas parcels that would otherwise cool are more successful at preventing overcooling (Kay *et al.*, 2003). However, efforts to implement a more realistic version of targeted feedback in the form of galactic winds are still not entirely successful at preventing overcooling (Borgani *et al.*, 2003).

It remains to be seen whether supernova feedback alone can account for the observed entropy profiles of clusters. Voit *et al.* (2003) and Ponman *et al.* (2003) have proposed that entropy input from galactic winds preceding the accretion of gas onto clusters could lead to a form of entropy amplification that would explain the observations. If galactic winds are strong enough to significantly smooth out the lumpiness of the local intergalactic gas, then the mode of accretion of this gas onto clusters will be closer to smooth accretion than to hierarchical accretion, thereby boosting the entropy generated through accretion shocks without changing the profile's characteristic shape. This effect is a plausible explanation for the altered similarity of the observed entropy profiles, but it has not yet been thoroughly tested in simulations. Intriguing results by Kay (2004) show that an extremely targeted feedback model, in which supernovae heat the local gas to 1000 keV cm^2 , successfully reproduces both the normalization and shape of the observed entropy profiles.

4. Feedback from Active Galactic Nuclei

If supernovae cannot prevent overcooling, then perhaps supermassive black holes in the nuclei of galaxies are what stop it (Cavaliere *et al.*, 2002; Valageas and Silk, 1999; Wu *et al.*, 2001). The omnipresence of supermassive black holes at the centers of galaxies (Magorrian *et al.*, 1998) and the excellent correlation of their masses with the bulge and halo properties of the host galaxy (Ferrarese and Merritt, 2000; Gebhardt *et al.*, 2000) strongly suggest that the growth of black holes in the nuclei of galaxies goes hand-in-hand with galaxy formation. Furthermore, the centers of many clusters with low-entropy gas whose cooling time is less than the age of the universe also contain active galactic nuclei that are ejecting streams of relativistic plasma into the intracluster medium (Burns, 1990). It is therefore plausible that supermassive black holes at the centers of clusters provide feedback that suppresses further cooling whenever condensing intracluster gas accretes onto the

central black hole.

Such a feedback loop is attractive and consistent with the circumstantial evidence, but the precise mechanism of heating remains unclear. The bubbles of relativistic plasma being inflated by the active galactic nuclei in clusters appear not to be expanding fast enough to shock heat the intracluster medium because the rims of the bubbles are no hotter than their surroundings (Fabian *et al.*, 2000; McNamara *et al.*, 2000). Also, if active galactic nuclei simply injected heat energy into the center of a cluster, then one would expect to see a flat or reversed entropy gradient in clusters with strong nuclear activity, indicating that convection is carrying heat outward. Instead, the entropy gradients in these cluster cores increase monotonically outward (David *et al.*, 2001; Horner *et al.*, 2004). One possibility is that heating is episodic (Kaiser and Binney, 2003) and that we have not yet found a cluster in the midst of an intense heating episode. Another is that heating is somehow spread evenly throughout the cluster core in a way that maintains the entropy gradient (Brüggen and Kaiser, 2002; Ruszkowski and Begelman, 2002). Yet another possibility is that bursts of relativistic plasma drive sound waves into the intracluster medium that eventually dissipate into heat (Fabian *et al.*, 2003).

Unfortunately, the none of these heating mechanisms have yet been tested in the context of cosmological structure formation, so we do not know their overall impact on either baryon condensation or the global entropy profiles of clusters. Also, many aspects of the relationship between cosmology and nuclear activity in galaxies remain highly uncertain. A major role for quasar feedback is plausible. However, the connection between the growth of central black holes in galaxies and galaxy formation itself is not well understood, and the efficiency with which black holes convert accretion energy into outflows is unknown.

5. Transport Processes

Heat transport processes like thermal conduction and turbulent mixing may also mitigate radiative cooling because gas that condenses sets up a temperature gradient along which heat energy can flow. In gas without magnetic fields, electrons conduct heat along temperature gradients giving a heat flux $\kappa_s \nabla T$, with $\kappa_s \approx 6 \times 10^{-7} T^{5/2} \text{ erg cm}^{-1} \text{ s}^{-1} \text{ K}^{-7/2}$ (Spitzer, 1962), the so-called Spitzer rate, valid when the scale length of the temperature gradient is longer than the electron mean free path. Clusters with central cooling times less than H_0^{-1} indeed tend to have positive temperature gradients within the central ~ 100 kpc, raising the possibility that heat conduction at least partially balances radiative losses. Many models for conduction in cluster cores have been developed (e.g., Bertschinger and Meiksin, 1986; Bregman and David, 1988; Rosner and Tucker, 1989; Sparks, 1992; Tucker and Rosner, 1983), but con-

duction does not satisfactorily balance radiative cooling. Temperature-gradient observations are inconsistent with steady-state balance between cooling and conduction in a number of cluster cores (Horner *et al.*, 2004; Voigt and Fabian, 2004). However, mixing of hot gas with cooler gas facilitated by intracluster turbulence (Kim and Narayan, 2003) or AGN activity (Brüggen and Kaiser, 2002) could enhance the effectiveness of heat conduction.

It is possible that cooling, conduction, feedback, and perhaps mixing as well are all needed for a complete solution that explains the observed core temperature gradients without overcooling. Conduction that balances cooling in a steady state has often been dismissed on the grounds that it is not stable enough to preserve the observed temperature and density gradients for periods of order $\gtrsim 1$ Gyr (Cowie and Binney, 1977; Fabian, 1994). Because of conduction's extreme sensitivity to temperature, it is difficult for radiative cooling and conduction to achieve precise thermal balance with a globally stable temperature gradient (Bregman and David, 1988; Soker, 2003). On the other hand, conduction would have to be suppressed by at least two orders of magnitude for radiative cooling to produce the observed gradients (Binney and Cowie, 1981; Fabian *et al.*, 1981). Recent theoretical analyses of conduction have concluded that this level of suppression is unrealistically high (Malyshkin, 2001; Malyshkin and Kulsrud, 2001; Narayan and Medvedev, 2001). Combining cooling, conduction, and feedback offers a way out of this dilemma. Hybrid models in which conduction compensates for cooling in the outer parts of the core while feedback from an active galactic nucleus compensates for it in the inner parts have had some success in reproducing the observations (Brighenti and Mathews, 2003; Ruszkowski and Begelman, 2002).

C. Galaxy Formation and Cluster Observables

Earlier we saw that the X-ray properties of the self-similar clusters produced by purely gravitational structure formation do not agree with observations (Sec. IV.A). Observed clusters of a given mass appear to be hotter than their theoretical counterparts and also less luminous, especially at the cool end of the cluster temperature range. Such disagreements have been worrisome to cosmologists who would like to understand what governs the cluster observables used to measure mass, but these problems are on their way to being solved. Both analytical work and hydrodynamical simulations performed during the last several years are showing that the observed L_X - T_{lum} , M_{200} - T_{lum} , and L_X - M_{200} relations are natural outcomes of galaxy formation. Significant uncertainties remain, but the theoretical foundation for the mass-observable relations essential for probing cosmology with clusters is growing firmer.

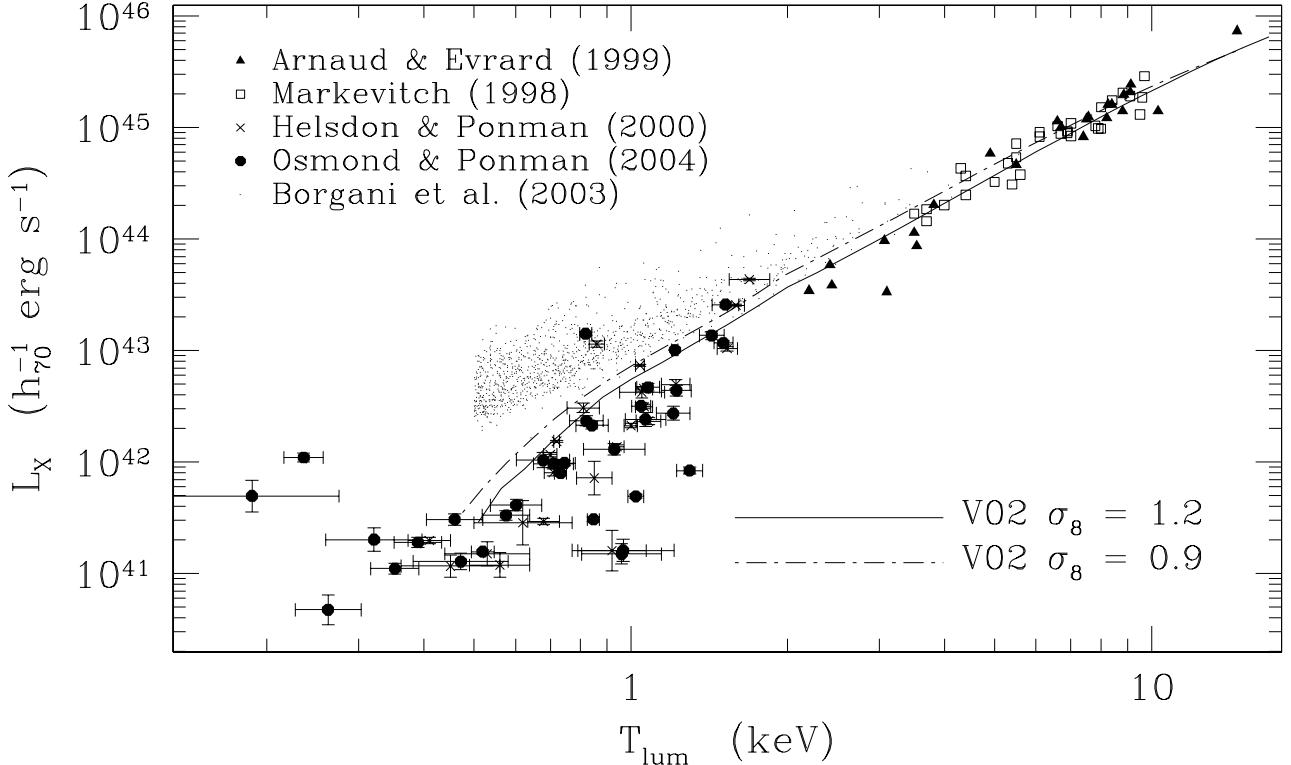


FIG. 14 Luminosity-temperature relation. Points show cluster data from Arnaud and Evrard (1999) (solid triangles) who avoided clusters with cool cores, cluster data from Markevitch (1998) (open squares) with cool cores excised, two sets of group data from Helsdon and Ponman (2000) (crosses) and Osmond and Ponman (2004) (solid octagons) that were not corrected for cool cores, and simulated clusters from Borgani *et al.* (2003) (small points). These simulations implement radiative cooling and supernova feedback in the form of galactic winds. Lines show modified-entropy models from Voit *et al.* (2002) with entropy truncated at the cooling threshold. There is a slight dependence on σ_8 in these models because higher values of σ_8 lead to dark-matter halos with more concentrated cores. Both the analytical and numerical models agree well with the data at $k_B T_{\text{lum}} \gtrsim 2 \text{ keV}$. Agreement is less good at lower temperatures, but the reasons for the disagreements are unclear. More feedback may be needed in the numerical models to suppress the luminosities, and the large scatter in the observations at $\lesssim 1 \text{ keV}$ may reflect a wide range in the effectiveness of feedback.

1. Role of Cooling

Radiative cooling turns out to be the most important process to include. While it might seem paradoxical, allowing the intracluster medium to radiate thermal energy actually causes its luminosity-weighted temperature to rise. The reason for this behavior is that cooling selectively removes low-entropy gas from the intracluster medium, raising the mean entropy of what remains (Bryan, 2000; Knight and Ponman, 1997; Pearce *et al.*, 2000). In non-radiative cluster simulations, the entropy of gas in the vicinity of the cluster core is below the cooling threshold K_c . This aspect of non-radiative models is unphysical, because gas with entropy less than K_c would radiate an amount of energy greater than its total thermal energy content over the course of the simulations. When cooling is allowed to occur, this low-entropy core gas condenses out of the intracluster medium and is replaced by higher entropy core gas having a higher temperature, a lower density, and therefore a lower luminosity.

A simple analytical model illustrates the effect of the cooling threshold on the L_X - T_{lum} and M_{200} - T_{lum} relations (Voit and Bryan, 2001; Voit *et al.*, 2002; Wu and Xue, 2002). The model assumes that the intracluster entropy distribution in the absence of galaxy formation would be the $K_{\text{NFW}}(M_g)$ distribution derived from the density profile of the dark matter. Because condensation and feedback both act to eliminate gas below the cooling threshold, the model simply truncates the entropy distribution at $K_c(T_{200})$ and discards all the gas with lower entropy. One can interpret this gas removal either as condensation or as extreme feedback that heats the sub-threshold gas to a much higher entropy level. This cooling and feedback need not occur at the center of the cluster. In a hierarchical cosmology, much of the low-entropy gas cools, condenses into galaxies, and produces feedback long before the cluster is finally assembled.

Computing the hydrostatic configuration of the modified entropy distribution in the original dark-matter potential gives L_X and T_{lum} as a function of the mass M_{200}

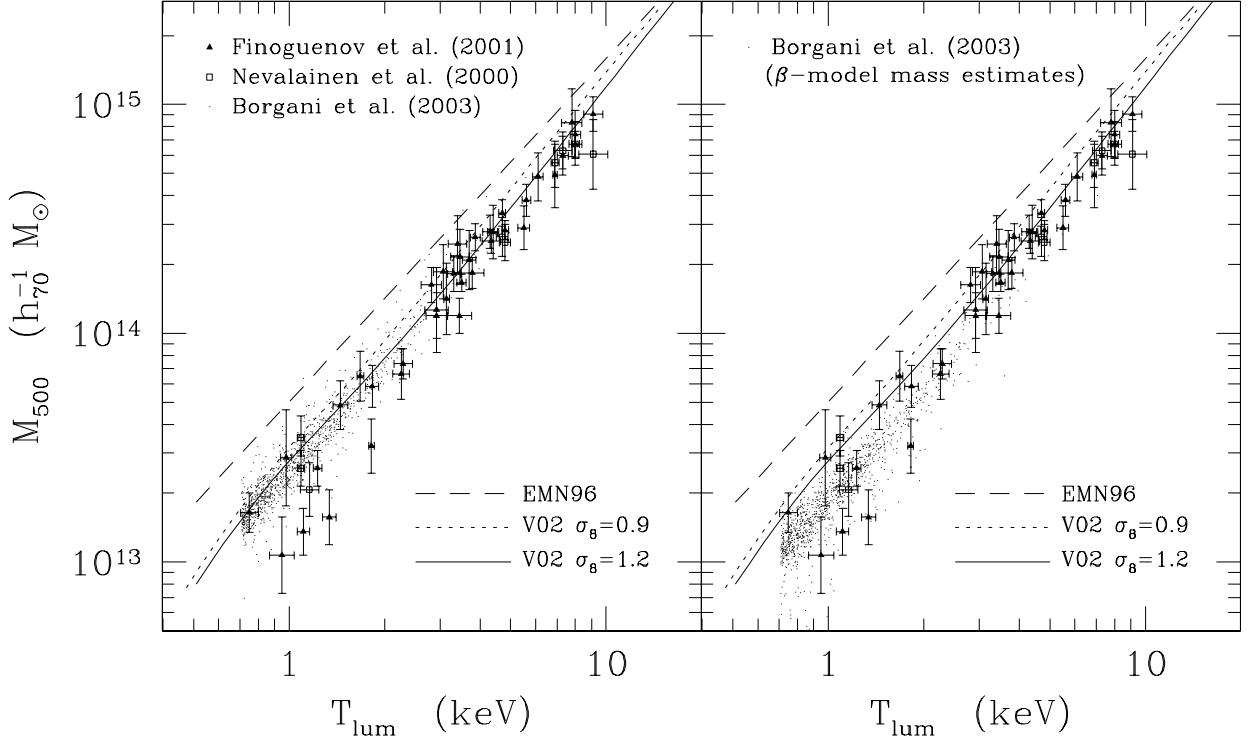


FIG. 15 Mass-temperature relation. Large points show cluster data from Finoguenov *et al.* (2001b) (solid triangles) and Nevalainen *et al.* (2000) (open squares), in which cluster masses were inferred from fitting polytropic beta models (see Sec. sec:tx). Dashed lines illustrate the M_{500} - T_{lum} relation measured in clustered simulated without cooling and feedback by Evrard *et al.* (1996), which clearly disagree with the data points. The other lines show the M_{500} - T_{lum} relations predicted by the analytical models of Voit *et al.* (2002), which agree much better with the data. There is a slight difference between models with $\sigma_8 = 0.9$ (dotted lines) and $\sigma_8 = 1.2$ because higher values of σ_8 lead to clusters with higher halo concentrations that produce slightly higher temperatures. Tiny points show data for clusters simulated by Borgani *et al.* (2003) with radiative cooling and feedback in the form of supernova-driven galactic winds. The left-hand panel uses the actual values of M_{500} , which agree with the analytical models. The right-hand panel uses values of M_{500} inferred from fitting polytropic beta models to the observations, which underestimate true cluster masses, especially at low temperature, suggesting there may be a systematic observational bias in this method of mass measurement.

and concentration c_{200} of the dark-matter halo. Figures 14 and 15 show that the resulting L_X - T_{lum} and M_{200} - T_{lum} relations generally agree well with observations but may slightly overpredict L_X for objects cooler than ~ 2 keV and do not account for the large scatter at low temperatures. There are no free parameters in this model, other than the cosmological parameters, because the M_{200} - c_{200} relation and the age of the universe used to compute K_c depend only on cosmology, and the heavy-element abundance used to compute the cooling threshold is taken from observations.

Numerical simulations in which feedback is either weak or non-existent produce clusters whose properties are quite similar to the ones in this simple analytical model. Early numerical investigations of cooling in individual clusters gave inconclusive results (Lewis *et al.*, 2000; Sugino and Ostriker, 1998), but simulations by Muanwong *et al.* (2001) showed that adding cooling to a large-scale cluster simulation could give an L_X - T_{lum} like the observed one. Subsequent numerical work

has confirmed that result (e.g., Borgani *et al.*, 2002a; Davé *et al.*, 2002; Kay *et al.*, 2003; Valdarnini, 2003). Adding radiative cooling to the cosmological model produces good agreement with observations at all cluster temperatures $\gtrsim 2$ keV.

Even when the simulations implement strong feedback, the X-ray scaling relations change remarkably little from the cooling-only case (Borgani *et al.*, 2002a; Kay *et al.*, 2003). The main effect on the L_X - T_{lum} relation of adding strong feedback to simulations that already include cooling is to slightly reduce the luminosity of cool ($\lesssim 2$ keV) clusters, bringing them into better agreement with observations. This insensitivity to the efficiency of feedback is another strong indication that the cooling threshold governs the entropy scale for similarity breaking.

One point of disagreement between the analytical models, the simulations, and the observations concerns the central temperature gradient. Many observed clusters have a relatively small amount of gas in their cores whose cooling time is less than the age of the universe, and in

those clusters the core temperature gradient is generally positive ($dT/dr > 0$). In the simple analytical models outlined above, no gas is allowed to be below the cooling threshold, resulting in a core that is nearly isentropic and thus has a negative temperature gradient ($dT/dr < 0$). Likewise, simulations with cooling and feedback also tend to have flat or negative temperature gradients in the neighborhood of the core radius (~ 100 kpc).

This problem deserves attention because elevated core temperatures in models with cooling are what bring the theoretical M_{200} - T_{lum} relation into agreement with observations. Making the analytical model slightly more realistic brings the predicted temperature gradient into better agreement with observations. The discontinuous cooling threshold applied by the simplest models is overly crude because it completely removes gas just below the threshold while gas just above the threshold does not cool at all. Instead, cooling acts upon the entropy distribution as described by equation (73). Voit *et al.* (2002) show that modifying the baseline profile K_{NFW} using this equation with $T = T_{200}$ for a time t_0 leads to an entropy distribution that reproduces the observed temperature gradients.

Simulations involving pure cooling do not agree with this result. The temperature-gradient discrepancy between analytical models and simulations in the pure-cooling case is still not understood, but may have something to do with the implicit stability of the cooling process in the analytical model. In that model the present-day intracluster medium is spherically symmetric with a positive entropy gradient, by definition, whereas thermal instabilities in the simulations that lead to a more heterogeneous entropy pattern at each radius, may be at the root of the negative temperature gradient. Perhaps the observations are telling us that a stabilizing influence like conduction erases small-scale thermal instabilities without shutting off global cooling.

2. Role of Feedback

The primary role of feedback is to regulate how many baryons condense into stars and cold gas clouds. As mentioned in the discussion of cooling, strong feedback does not have a large effect on the $L_{\text{X}}\text{-}T_{\text{lum}}$ relation, aside from a slight decrease in the luminosity of low-temperature clusters, as long as it is strong enough to shut off cooling in the gas parcels that it affects. However, moderate feedback that heats gas to $\lesssim 100$ keV cm 2 can boost L_{X} because it does not allow the core gas to cool but rather maintains it in an entropy state that allows it to radiate considerable thermal energy Kay *et al.* (2003).

Some preheating and feedback models adequately explain the scaling relations without explicitly including cooling (e.g., Babul *et al.*, 2002; Balogh *et al.*, 1999; Bialek *et al.*, 2001; Tozzi and Norman, 2001). In these models, the minimum entropy level introduced by heating is typically a free parameter that is adjusted to give

the best-fitting $L_{\text{X}}\text{-}T_{\text{lum}}$ relation. The value of this best-fitting entropy level turns out to be 100 – 400 keV cm 2 , approximately corresponding to the level of the cooling threshold. This correspondence is consistent with the idea that the amount of heating needed to explain the mass-observable relation is determined by the need to shut cooling, in which case cooling still sets the entropy scale of similarity breaking, even when it is not explicitly included in the model (Voit *et al.*, 2002).

From the standpoint of the mass-observable relations, the most important effect of feedback itself has to do with cluster richness. In both the simulations and the analytical models, pure cooling leads to a larger fraction of condensed baryons in cool clusters (Borgani *et al.*, 2002a; Borgani *et al.*, 2003; Davé *et al.*, 2002; Muanwong *et al.*, 2001; Voit *et al.*, 2002), implying that these objects might have a higher star-to-baryon ratio and therefore a lower mass-to-light ratio. There are some observational indications that the ratio of stellar luminosity to mass in clusters is a function of mass (Lin *et al.*, 2003), but not all such studies agree. This issue will need to be settled in order for optical richness measurements to deliver high-precision mass functions (Sec. IV.D).

3. Role of Smoothing

A full understanding of the $L_{\text{X}}\text{-}T_{\text{lum}}$ relation may involve feedback indirectly, through its smoothing effects on the intergalactic medium (Sec. IV.A.3). If the observed preservation of $K(r) \propto r^{1.1}$ entropy profiles is indeed due to smoothing of the intergalactic medium followed by accretion onto clusters, then the present-day entropy profiles of clusters are evidence that galactic winds were widespread prior to the accretion of gas into today's clusters. Rather than just affecting the core entropy of clusters, a modest amount of entropy produced by early winds may have been amplified by smooth accretion, boosting the entire entropy profile by a common factor determined by the cooling threshold (Voit and Ponman, 2003). If that is indeed what happens, then it would explain the observed alteration of cluster similarity such that $K(r/r_{200}) \propto T_{\text{lum}}^{2/3}$ (Ponman *et al.*, 2003; Pratt and Arnaud, 2003), which leads directly to the relation $L_{\text{X}} \propto T_{\text{lum}}^3 (T_{200}/T_{\text{lum}})^{1.5}$ for pure bremsstrahlung emission, in agreement with the observations.

4. Predictions for Evolution

Preheating, the cooling threshold, and the altered similarity indicative of smoothing affect the time-dependent behavior of the $L_{\text{X}}\text{-}T_{\text{lum}}$ relation differently, offering a way to gather further information about their relative influence on cluster structure. Defining

$$\hat{L} = \int_0^{r/r_{200}} \left(\frac{\rho_g}{200 f_b \rho_{\text{cr}}} \right)^2 \hat{r}^2 d\hat{r}, \quad (77)$$

one can express the scaling of a cluster's integrated X-ray luminosity as

$$L_X \propto \Lambda_c(T_{\text{lum}}) M_{200} \rho_{\text{cr}} \hat{L} \quad (78)$$

$$\propto T_{\text{lum}}^2 \left(\frac{T_{200}}{T_{\text{lum}}} \right)^{3/2} H(z) \hat{L}, \quad (79)$$

where the first line assumes the cluster is approximately isothermal and the second line is an approximation that assumes pure bremsstrahlung emission. The self-similar case,

$$L_X \propto T_{\text{lum}}^2 H(z) \quad (80)$$

is well known to be a poor description of the data because its power-law slope at $z \approx 0$ is too shallow.

The modified-entropy models of Voit *et al.* (2002) show that enforcing a minimum core entropy level K_{\min} breaks self-similarity in such a way that $\hat{L} \propto K_{\min}^{-3/2} T_{200}^{3/2} H^{-2}$, if K_{\min} is a significant fraction of the cluster's characteristic entropy K_{200} . In the pure preheating case, K_{\min} is assumed to be independent of both cluster mass and of redshift, leading to

$$L_X \propto T_{\text{lum}}^{3.5} \left(\frac{T_{200}}{T_{\text{lum}}} \right)^3 \frac{1}{H(z)}. \quad (81)$$

In other words, pure preheating steepens the L_X - T_{lum} a little more than necessary but causes high-redshift clusters to be less luminous than one would expect from their temperatures because the entropy floor K_{\min} is a larger proportion of K_{200} earlier in time. This prediction appears to conflict with recent observations indicating evolution in the opposite direction (e.g., Vikhlinin *et al.*, 2002). Tying the minimum entropy scale to the cooling threshold $K_c \propto T_{\text{lum}}^{2/3} t^{2/3}$ helps to solve this problem because it leads to

$$L_X \propto T_{\text{lum}}^{2.5} \left(\frac{T_{200}}{T_{\text{lum}}} \right)^3 \frac{1}{H(z)t(z)}. \quad (82)$$

In this case, a little bit of tilt in the T_{lum}/T_{200} relation, consistent with observations (see Table I), is needed to sufficiently steepen the L_X - T_{lum} relation, and the sense of the evolution agrees with observations. In a Λ CDM universe, the redshift dependence of the luminosity normalization is $H^{-1}t^{-1} \sim (1+z)^{0.5} \sim H^{0.75}$ out to $z \sim 0.5$. Altered similarity linked to the cooling threshold is in better agreement with the slope but produces less evolution. Assuming density profiles that scale as $\rho_g(r/r_{200}) \propto (T_{\text{lum}}/K_c)^{3/2}$ yields

$$L_X \propto T_{\text{lum}}^3 \left(\frac{T_{200}}{T_{\text{lum}}} \right)^3 \frac{1}{H^3(z)t^2(z)}. \quad (83)$$

The normalization of luminosity in this relation varies as $H^{-3}t^{-2} \sim (1+z)^{0.3} \sim H^{0.5}$ to $z \sim 0.5$.

Observations of evolution in the luminosity-temperature relation are not yet precise enough to

distinguish between these latter two possibilities. The usual procedure is to compare the L_X - T_{lum} relation measured in a significantly redshifted cluster sample to the relation measured at $z \approx 0$. Vikhlinin *et al.* (2002) were the first to detect evolution, finding $L_X(T_{\text{lum}}) \propto (1+z)^{b_{LT}}$ with $b_{LT} = 1.5 \pm 0.3$, assuming a Λ CDM cosmology. These authors compared the low-redshift sample of Markevitch (1998) to a collection of 22 clusters in the redshift range $0.4 < z < 0.8$. Lumb *et al.* (2003) found a similar amount of evolution, $b_{LT} = 1.52^{+0.24}_{-0.27}$, using a smaller sample of eight clusters at $z \approx 0.4$, but not all studies find such strong evolution, which exceeds the predictions of the basic models outlined above. For example, Ettori *et al.* (2003) find $b_{LT} = 0.62 \pm 0.28$ for a sample of 28 clusters at $z > 0.4$ using the Markevitch (1998) sample as the low-redshift baseline and $b_{LT} = 0.98 \pm 0.20$ relative to the Arnaud and Evrard (1999) low-redshift baseline. Furthermore, the strength of the evolution found by Ettori *et al.* (2003) becomes *smaller* for high-redshift clusters, consistent with no evolution at all ($b_{LT} = 0.04 \pm 0.33$) when they include only their 16 clusters with $z > 0.6$ in the comparison with the Markevitch sample. Apparently, there are some systematic uncertainties in these evolution measurements that need to be accounted for.

D. Constraints on Baryon Condensation

The ultimate test for feedback models is that they must account for both the proportion of condensed baryons to hot gas in clusters and any dependence of that proportion on cluster mass. In order to apply that test, we would like to have firm numbers for the amount of condensed baryons in clusters, but such measurements can be difficult. Even if the amount of starlight were perfectly measured, converting integrated starlight to stellar mass involves uncertain assumptions about both the star-formation history of a cluster and the distribution function of stellar masses at birth, a quantity known as the initial mass function. Any variation in the star-formation history or initial mass function with cluster mass can lead to spurious systematic trends in the cluster mass function inferred from cluster richness.

Baryons contained in cold clouds are even harder to constrain because gaseous matter in this form can be nearly invisible, if it is sufficiently cold (Ferland *et al.*, 1994, 2002). However, it seems unlikely that large amounts of baryons exist in such a form, at least in rich clusters. Adding the amount of baryons inferred from starlight to the amount of hot gas observed in rich clusters accounts for nearly all the baryons expected from the global ratio of baryons to dark matter, leaving little room left in the baryon budget for cold gas clouds.

The situation is less clear in lower-mass clusters and groups of galaxies, in which the proportion of hot gas to dark matter is significantly smaller. Summing the masses

of stars and hot gas accounts for only about half the expected number of baryons in some cases, yet there is no observational evidence for large quantities of cold gas (e.g., Waugh *et al.*, 2002). Circumstantial evidence argues against there being large reservoirs of cold baryons in groups. Presumably, the rich clusters in which we now see virtually all the baryons were hierarchically assembled from objects like the baryon-poor groups of galaxies we observe today. If large amounts of baryons in their higher-redshift counterparts were locked away in some cold, condensed form, then how were they released when these groups of galaxies merged to form large clusters?

A more complete accounting of intracluster baryons, especially in low-mass systems, is sorely needed in order test the various feedback models described in Sec. IV.B. The rest of this section summarizes some of the recent work on constraining the amount of condensed intracluster baryons in the form of stars, the prospects for measuring baryon condensation through the S-Z effect, and X-ray observations of nearby clusters that may help solve the puzzles surrounding condensation and feedback.

1. Mass and Light in Clusters

Inferences of stellar mass from the observed starlight are generally based on a mass-to-light ratio expressed in solar units. That is, the mass-to-light ratio of the Sun in all wavebands equals unity. Because young stellar populations tend to emit large amounts of blue light that quickly dies out as the population ages, most recent assessments of the stellar mass in clusters have concentrated on measurements of infrared starlight in the K band at roughly 2 microns. Observing starlight in this band minimizes the uncertainties owing to a cluster's star formation history. The old stellar populations characteristic of elliptical galaxies tend to have a K -band mass-to-light ratio $\Upsilon_K \approx 0.8 h_{70}^{-1}$, and mass-to-light ratios in spiral and irregular galaxies can be up to a factor of two smaller (Bell and de Jong, 2000). For the mix of galaxies seen in clusters, Lin *et al.* (2003) estimate that the mean mass-to-light ratio ranges from $\Upsilon_K = 0.7 h_{70}^{-1}$ to $0.8 h_{70}^{-1}$ as cluster temperature climbs from 2 keV to 10 keV. From this mass-to-light ratio, they infer that the fraction of intracluster baryons in stellar form is $f_* \approx 0.1$ for rich clusters (see also Balogh *et al.*, 2001b). Notice that this value is about half that predicted by current simulations of cluster formation including strong feedback, a discrepancy that could become even larger with higher-resolution simulations (Borgani *et al.*, 2003).

Many studies, but not all of them, suggest that the fraction of condensed baryons in stars may be a function of cluster mass. The ratio of K -band light to *total* cluster mass within r_{500} found by Lin *et al.* (2003) is $\Upsilon_K = (47 \pm 3) h_{70} (M_{500}/3 \times 10^{14} h_{70}^{-1} M_\odot)^{0.31}$, which translates to a temperature dependence $\Upsilon_K \propto T_{\text{lum}}^{0.5 \pm 0.1}$. The ratio of stellar mass to total mass in this study therefore ranges from $\sim 2.2\%$ at $10^{14} h_{70}^{-1} M_\odot$ to $\sim 1.2\%$

at $10^{15} h_{70}^{-1}, M_\odot$. Similar trends with shallower slopes are seen at other wavelengths. Bahcall and Comerford (2002) find the ratio of total mass to starlight in the heart of the visible spectrum (V -band) is $\Upsilon_V \propto T_{\text{lum}}^{0.3 \pm 0.1}$. In blue light (B -band), Girardi *et al.* (2002) find $\Upsilon_B \propto M^{0.25}$. However, other studies have found no significant dependence on mass. According to Kochanek *et al.* (2003) the K -band mass-to-light ratio inside r_{200} scales as $\Upsilon_K \propto M_{200}^{-0.10 \pm 0.09}$.

2. Intergalactic Stars

Measurements of the total stellar luminosity in clusters generally focus on the light from galaxies, but what about stars that are not in galaxies? At least some of a cluster's stars float unmoored in the spaces between a cluster's galaxies (Ferguson *et al.*, 1998). These stars are thought to have originated in galaxies but were later stripped from their homes by tidal forces during a close encounter with another galaxy. Current observational limits, however, indicate that no more than 10-20% of a cluster's stars are outside of galaxies (Durrell *et al.*, 2002), implying that failing to account for intergalactic stars does not lead to large errors in measured mass-to-light ratios.

3. Global S-Z Effect

If the baryons missing in low-mass clusters are not in condensed form, then they must be in the form of hot gas beyond the regions detectable with X-ray telescopes. If that is indeed the case, then the best way of finding them may be through the Sunyaev-Zeldovich effect. Section II.C.1 showed that the integrated microwave distortion from a cluster scales with the electron temperature of the cluster and the overall mass in hot electrons. If a significant proportion of baryons have condensed, then the associated electrons are also locked away in cold clouds, where they don't contribute to the S-Z signal.

Simulations of cluster formation that include cooling indicate how the mean value of the y -distortion owing to clusters depends on cooling and feedback processes. Models by da Silva *et al.* (2001) produce $y = 3.2 \times 10^{-6}$ in the non-radiative case, dropping to $y = 2.3 \times 10^{-6}$ in the case of radiative cooling without feedback. The difference between the radiative case and non-radiative case is somewhat smaller when feedback is implemented. White *et al.* (2002) find $y = 2.5 \times 10^{-6}$ in the non-radiative case and $y = 2.1 \times 10^{-6}$ when both cooling and feedback are turned on. Testing for baryon condensation in this way may eventually be possible, but the mean value of the S-Z distortion is also very sensitive to other cosmological parameters, such as σ_8 , which will have to be very well constrained before we can use the global y parameter to test feedback models.

4. Cooling Flows in Clusters

Cores of present-day clusters are among the best places in the universe to observe the interplay between condensation and feedback. Gas at the centers of many clusters can radiate an amount of energy equal to its thermal energy in less than a billion years, yet the majority of that gas is not condensing (see Donahue and Voit, 2004, for a recent review). Early interpretations of clusters with central cooling times less than the age of the universe suggested that the core gas should gradually condense and be replaced by the surrounding material in an orderly flow of cooling gas (Cowie and Binney, 1977; Fabian and Nulsen, 1977; Mathews and Bregman, 1978). The mass condensation rates inferred from X-ray imaging ranged as high as $\sim 10^2$ to $10^3 M_{\odot} \text{ yr}^{-1}$ implying that the cores of these “cooling-flow” clusters should contain $\gtrsim 10^{12} M_{\odot}$ in the form of condensed baryons. However, exhaustive searches for this mass sink generally have not found stars forming at such a high rate (McNamara and O’Connell, 1989; O’Connell and McNamara, 1989), nor have they found sufficiently large collections of cold baryonic clouds to account for the deposited mass (Braine and Dupraz, 1994; McNamara and Jaffe, 1994; O’Dea *et al.*, 1994, 1998; Voit and Donahue, 1995).

Now X-ray spectroscopy itself is showing that condensation proceeds at a considerably slower rate, if it happens at all. The central gas in clusters with short cooling times appears to reach temperatures $\sim T_{\text{lum}}/2$, but very little X-ray line emission is seen from gas at $\lesssim T_{\text{lum}}/3$ (Peterson *et al.*, 2003, 2001). Some sort of heating mechanism seems to be inhibiting condensation below this temperature. There are plenty of candidates for resupplying the radiated heat energy—supernovae, outflows from active galactic nuclei, electron thermal conduction, and turbulent mixing have all been suggested (see Sec. IV.B)—but there is still no consensus on the relative importance of these mechanisms.

A reduced amount of condensation still appears to be occurring. For example, plenty of circumstantial evidence links short central cooling times with star formation at the centers of clusters. Objects whose central cooling time is less than the age of the universe frequently contain emission-line nebulae whose properties suggest that they are energized primarily by hot, young stars (Johnstone *et al.*, 1987; Voit and Donahue, 1997). Nebulae like these are never seen in clusters where the central cooling time is greater than the universe’s age (Hu *et al.*, 1985). Also, objects with prominent nebulae tend to have abundant cool molecular hydrogen gas, the seed material for star formation (Donahue *et al.*, 2000; Edge, 2001; Edge and Frayer, 2003). Efforts to estimate the star formation rate from the ultraviolet light emanating from the centers of clusters indicate that it may be consistent with the current upper limits on the condensation rate drawn from X-ray spectroscopy (McNamara *et al.*, 2004).

An understanding of what regulates condensation and

star formation at the centers of present-day clusters will help to solve more than just the overcooling problem of galaxy formation. It is also relevant to an aspect of bright galaxies that remains difficult to understand. The luminosity distribution function of galaxies cuts off very sharply at the high-luminosity end, far more sharply than called for in standard models of galaxy formation. Extremely powerful feedback can produce a sharp cutoff, but the amount of energy input required seems to implicate active galactic nuclei as the primary feedback source (e.g., Benson *et al.*, 2003; Scappapieco and Oh, 2004). Alternatively, thermal conduction might produce a sharp cutoff because its efficiency rises so rapidly with temperature (Benson *et al.*, 2003; Fabian *et al.*, 2002). As the halo of a massive galaxy grows and its characteristic temperature rises through a critical threshold $\sim 10^7$ K, conduction can strongly suppress further cooling and star formation, if it is not inhibited by magnetic fields. Detailed studies of cluster cores will be needed to test these possibilities. Early efforts are indicating that conduction might not be efficient enough to prevent overcooling (Dolag *et al.*, 2004; Jubelgas *et al.*, 2004).

V. CONCLUDING REMARKS

The next decade of research into cluster evolution promises to be very exciting. Large optical surveys like the Sloan Digital Sky Survey are greatly increasing the number of well-studied clusters of galaxies in the low-redshift universe. Deep surveys looking for the Sunyaev-Zeldovich effect will be finding thousands of clusters to distances well beyond a redshift of $z = 1$. The *Chandra* and *XMM-Newton* X-ray observatories are providing our most detailed look yet at the intracluster medium, its thermodynamical state, and some of the feedback processes that regulate condensation of intergalactic gas into galaxies and stars. Also, dedicated X-ray satellite missions to survey a large fraction of the sky for distant clusters are currently being planned.

Making the most of these opportunities will require cooperation between observers in those different wavebands, and theoretical modeling that closely links those cluster observables to cosmological parameters. Optical and infrared followup of S-Z surveys will be critical in order to determine the redshifts of the cluster candidates. X-ray followup of a subset of the S-Z clusters will also be necessary to establish how the thermodynamics of galaxy formation affects evolution of the mass-observable relations in the microwave band. Concentrated efforts to observe a calibration set of clusters in all of these wavebands will be very valuable in helping to establish how well the various observables trace mass and the scatter in each of these observables at a given mass.

If the Λ CDM concordance model is indeed a good description of the overall architecture of the universe and its initial perturbation spectrum, then the parameters describing the cosmological context in which galaxy for-

mation happens ought to be quite precisely established within this decade. Studies of cluster evolution will be just one part of this overall effort, which also includes distance determinations to high-redshift supernovae, increasingly sensitive observations of the cosmic microwave background, and the mapping of large-scale structure. However, consistency between the dynamics of cluster evolution and the geometry of the universe, as measured with supernova and microwave observations, will stand as a particularly critical test of the overall model. With success, most of the remaining secrets about galaxy formation—other than what dark matter and dark energy actually are—will concern baryons and their complex cooling and feedback processes.

Our understanding of what baryons do is rapidly progressing, thanks in large part to large-scale cosmological simulations on massively parallel computers. Clusters and their evolution place unique constraints on those models because clusters are the only places in the universe where the majority of the baryons emit detectable radiation, revealing their thermodynamic state and elemental abundances. Galaxy formation has clearly left its mark in the intracluster medium, but we are just beginning to decipher what it has written there in the gases between the galaxies. Perhaps in ten more years there will be as much optimism about understanding the baryonic side of galaxy formation as there is now about the understanding the darker side.

Acknowledgments

Many colleagues have helped me to understand the issues covered in this review, particularly M. Donahue, G. Bryan, A. Evrard, J. Mohr, E. Ellingson, S. Borgani, T. Ponman, M. Balogh, R. Bower, C. Lacey, S. Kay, M. Postman, H. Ebeling, A. Fabian, C. Scharf, P. Tozzi, A. Finoguenov, S. Allen, B. McNamara, R. Mushotzky, C. Sarazin, R. Wechsler, H. Hoekstra, and M. White. Thanks also to A. Vikhlinin, P. Henry, and C. Mullis for providing some of the figures. Parts of this article were written at the very hospitable Aspen Center for Physics.

References

- Abell, G. O., 1958, *Astroph. J. Suppl.* **3**, 211.
 Abell, G. O., H. G. Corwin, and R. P. Olowin, 1989, *Astroph. J. Suppl.* **70**, 1.
 Adelberger, K. L., C. C. Steidel, A. E. Shapley, and M. Pettini, 2003, *Astroph. J.* **584**, 45.
 Allen, S. W., and A. C. Fabian, 1998, *Mot. Not. Roy. Ast. Soc.* **297**, L57.
 Allen, S. W., R. W. Schmidt, H. Ebeling, A. C. Fabian, and L. van Speybroeck, 2004, eprint astro-ph/0405340.
 Allen, S. W., R. W. Schmidt, and A. C. Fabian, 2001, *Mot. Not. Roy. Ast. Soc.* **328**, L37.
 Allen, S. W., R. W. Schmidt, and A. C. Fabian, 2002, *Mot. Not. Roy. Ast. Soc.* **334**, L11.
 Arnaud, M., N. Aghanim, and D. M. Neumann, 2002, *Astron. Astroph.* **389**, 1.
 Arnaud, M., and A. E. Evrard, 1999, *Mot. Not. Roy. Ast. Soc.* **305**, 631.
 Böhringer, H., C. A. Collins, L. Guzzo, P. Schuecker, W. Voges, D. M. Neumann, S. Schindler, G. Chincarini, S. De Grandi, R. G. Cruddace, A. C. Edge, T. H. Reiprich, *et al.*, 2002, *Astroph. J.* **566**, 93.
 Babul, A., M. L. Balogh, G. F. Lewis, and G. B. Poole, 2002, *Mot. Not. Roy. Ast. Soc.* **330**, 329.
 Bacon, D. J., R. J. Massey, A. R. Refregier, and R. S. Ellis, 2003, *Mot. Not. Roy. Ast. Soc.* **344**, 673.
 Bahcall, J. N., and S. Tremaine, 1981, *Astroph. J.* **244**, 805.
 Bahcall, N. A., 1977, *Ann. Rev. Astron. Astroph.* **15**, 505.
 Bahcall, N. A., and J. M. Comerford, 2002, *Astroph. J. Lett.* **565**, L5.
 Bahcall, N. A., F. Dong, L. Hao, P. Bode, J. Annis, J. E. Gunn, and D. P. Schneider, 2003a, *Astroph. J.* **599**, 814.
 Bahcall, N. A., and X. Fan, 1998, *Astroph. J.* **504**, 1.
 Bahcall, N. A., and L. M. Lubin, 1994, *Astroph. J.* **426**, 513.
 Bahcall, N. A., T. A. McKay, J. Annis, R. S. J. Kim, F. Dong, S. Hansen, T. Goto, J. E. Gunn, C. Miller, R. C. Nichol, M. Postman, D. Schneider, *et al.*, 2003b, *Astroph. J. Suppl.* **148**, 243.
 Bahcall, N. A., J. P. Ostriker, S. Perlmutter, and P. J. Steinhardt, 1999, *Science* **284**, 1481.
 Bahcall, N. A., and R. M. Soneira, 1983, *Astroph. J.* **270**, 20.
 Balogh, M. L., A. Babul, and D. R. Patton, 1999, *Mot. Not. Roy. Ast. Soc.* **307**, 463.
 Balogh, M. L., D. Christlein, A. I. Zabludoff, and D. Zaritsky, 2001a, *Astroph. J.* **557**, 117.
 Balogh, M. L., F. R. Pearce, R. G. Bower, and S. T. Kay, 2001b, *Mot. Not. Roy. Ast. Soc.* **326**, 1228.
 Bardeen, J. M., J. R. Bond, N. Kaiser, and A. S. Szalay, 1986, *Astroph. J.* **304**, 15.
 Bartelmann, M., and P. Schneider, 2001, *Phys. Reports* **340**, 291.
 Battye, R. A., and J. Weller, 2003, *Phys. Rev. D* **68**, 083506.
 Bell, E. F., and R. S. de Jong, 2000, *Mot. Not. Roy. Ast. Soc.* **312**, 497.
 Benson, A. J., R. G. Bower, C. S. Frenk, C. G. Lacey, C. M. Baugh, and S. Cole, 2003, *Astroph. J.* **599**, 38.
 Benson, A. J., and P. Madau, 2003, *Mot. Not. Roy. Ast. Soc.* **344**, 835.
 Bertschinger, E., 1985, *Astroph. J. Suppl.* **58**, 39.
 Bertschinger, E., and A. Meiksin, 1986, *Astroph. J. Lett.* **306**, L1.
 Bialek, J. J., A. E. Evrard, and J. J. Mohr, 2001, *Astroph. J.* **555**, 597.
 Binney, J., and L. L. Cowie, 1981, *Astroph. J.* **247**, 464.
 Binney, J., and S. Tremaine, 1987, *Galactic dynamics* (Princeton, NJ, Princeton University Press, 1987, 747 p.).
 Birkinshaw, M., 1991, in *Physical Cosmology*, pp. 177–+.
 Birkinshaw, M., 1999, *Phys. Rept.* **310**, 97.
 Birkinshaw, M., J. P. Hughes, and K. A. Arnaud, 1991, *Astroph. J.* **379**, 466.
 Biviano, A., 2000, in *Constructing the Universe with Clusters of Galaxies*.
 Biviano, A., and M. Girardi, 2003, *Astroph. J.* **585**, 205.
 Blanchard, A., D. Valls-Gabaud, and G. A. Mamon, 1992, *Astron. Astroph.* **264**, 365.
 Bode, P., N. A. Bahcall, E. B. Ford, and J. P. Ostriker, 2001, *Astroph. J.* **551**, 15.
 Bond, J. R., S. Cole, G. Efstathiou, and N. Kaiser, 1991,

- Astroph. J. **379**, 440.
- Borgani, S., F. Governato, J. Wadsley, N. Menci, P. Tozzi, T. Quinn, J. Stadel, and G. Lake, 2002a, Mot. Not. Roy. Ast. Soc. **336**, 409.
- Borgani, S., M. Mezzetti, and R. Valdarnini (eds.), 2002b, *Tracing Cosmic Evolution with Galaxy Clusters*.
- Borgani, S., P. Rosati, P. Tozzi, and C. Norman, 1999, Astroph. J. **517**, 40.
- Borgani, S., P. Rosati, P. Tozzi, S. A. Stanford, P. R. Eisenhardt, C. Lidman, B. Holden, R. Della Ceca, C. Norman, and G. Squires, 2001, Astroph. J. **561**, 13.
- Borgani, S., et al., 2003, astro-ph/0310794 .
- Bower, R. G., 1991, Mot. Not. Roy. Ast. Soc. **248**, 332.
- Brüggen, M., and C. R. Kaiser, 2002, Nature (London) **418**, 301.
- Braine, J., and C. Dupraz, 1994, Astron. Astroph. **283**, 407.
- Bregman, J. N., and L. P. David, 1988, Astroph. J. **326**, 639.
- Brightenti, F., and W. G. Mathews, 2003, Astroph. J. **587**, 580.
- Bryan, G. L., 2000, Astroph. J. Lett. **544**, L1.
- Bryan, G. L., and M. L. Norman, 1998, Astroph. J. **495**, 80.
- Bullock, J. S., T. S. Kolatt, Y. Sigad, R. S. Somerville, A. V. Kravtsov, A. A. Klypin, J. R. Primack, and A. Dekel, 2001, Mot. Not. Roy. Ast. Soc. **321**, 559.
- Buote, D. A., A. D. Lewis, F. Brightenti, and W. G. Mathews, 2003, Astroph. J. **595**, 151.
- Burles, S., K. M. Nollett, and M. S. Turner, 2001, Astroph. J. Lett. **552**, L1.
- Burns, J. O., 1990, Astron. J. **99**, 14.
- Carlberg, R. G., H. K. C. Yee, and E. Ellingson, 1997a, Astroph. J. **478**, 462.
- Carlberg, R. G., H. K. C. Yee, E. Ellingson, S. L. Morris, R. Abraham, P. Gravel, C. J. Pritchett, T. Smecker-Hane, F. D. A. Hartwick, J. E. Hesser, J. B. Hutchings, and J. B. Oke, 1997b, Astroph. J. Lett. **485**, L13+.
- Carlstrom, J. E., G. P. Holder, and E. D. Reese, 2002, Ann. Rev. Astron. Astroph. **40**, 643.
- Carlstrom, J. E., M. K. Joy, L. Grego, G. P. Holder, W. L. Holzapfel, J. J. Mohr, S. Patel, and E. D. Reese, 2000, Physica Scripta Volume T **85**, 148.
- Carroll, S. M., W. H. Press, and E. L. Turner, 1992, Ann. Rev. Astron. Astroph. **30**, 499.
- Cavaliere, A., and R. Fusco-Femiano, 1976, Astron. Astroph. **49**, 137.
- Cavaliere, A., A. Lapi, and N. Menci, 2002, Astroph. J. Lett. **581**, L1.
- Cavaliere, A., N. Menci, and P. Tozzi, 1997, Astroph. J. Lett. **484**, L21+.
- Cen, R., N. Y. Gnedin, L. A. Kofman, and J. P. Ostriker, 1992, Astroph. J. Lett. **399**, L11.
- Cen, R., and J. P. Ostriker, 1999, Astroph. J. **514**, 1.
- Cole, S., 1991, Astroph. J. **367**, 45.
- Cole, S., and N. Kaiser, 1989, Mot. Not. Roy. Ast. Soc. **237**, 1127.
- Collins, C. A., L. Guzzo, H. Böhringer, P. Schuecker, G. Chin-carini, R. Cruddace, S. De Grandi, H. T. MacGillivray, D. M. Neumann, S. Schindler, P. Shaver, and W. Voges, 2000, Mot. Not. Roy. Ast. Soc. **319**, 939.
- Cowie, L. L., and J. Binney, 1977, Astroph. J. **215**, 723.
- da Silva, A. C., S. T. Kay, A. R. Liddle, P. A. Thomas, F. R. Pearce, and D. Barbosa, 2001, Astroph. J. Lett. **561**, L15.
- Dahle, H., K. Pedersen, P. B. Lilje, S. J. Maddox, and N. Kaiser, 2003, Astroph. J. **591**, 662.
- Dalton, G. B., S. J. Maddox, W. J. Sutherland, and G. Efstratiou, 1997, Mot. Not. Roy. Ast. Soc. **289**, 263.
- Davé, R., N. Katz, and D. H. Weinberg, 2002, Astroph. J. **579**, 23.
- David, L. P., W. Forman, and C. Jones, 1991, Astroph. J. **380**, 39.
- David, L. P., C. Jones, and W. Forman, 1995, Astroph. J. **445**, 578.
- David, L. P., P. E. J. Nulsen, B. R. McNamara, W. Forman, C. Jones, T. Ponman, B. Robertson, and M. Wise, 2001, Astroph. J. **557**, 546.
- David, L. P., A. Slyz, C. Jones, W. Forman, S. D. Vrtilek, and K. A. Arnaud, 1993, Astroph. J. **412**, 479.
- Davis, M., G. Efstathiou, C. S. Frenk, and S. D. M. White, 1985, Astroph. J. **292**, 371.
- Davis, M., and P. J. E. Peebles, 1983, Astroph. J. **267**, 465.
- De Grandi, S., S. Ettori, M. Longhetti, and S. Molendi, 2003, eprint astro-ph/0310828.
- De Grandi, S., and S. Molendi, 2002, Astroph. J. **567**, 163.
- Diaferio, A., and M. J. Geller, 1997, Astroph. J. **481**, 633.
- Dolag, K., M. Jubelgas, V. Springel, S. Borgani, and E. Rasia, 2004, Astroph. J. Lett. **606**, L97.
- Donahue, M., 1996, Astroph. J. **468**, 79.
- Donahue, M., J. Mack, C. Scharf, P. Lee, M. Postman, P. Rosati, M. Dickinson, G. M. Voit, and J. T. Stocke, 2001, Astroph. J. Lett. **552**, L93.
- Donahue, M., J. Mack, G. M. Voit, W. Sparks, R. Elston, and P. R. Maloney, 2000, Astroph. J. **545**, 670.
- Donahue, M., C. A. Scharf, J. Mack, Y. P. Lee, M. Postman, P. Rosati, M. Dickinson, G. M. Voit, and J. T. Stocke, 2002, Astroph. J. **569**, 689.
- Donahue, M., and G. M. Voit, 1999, Astroph. J. Lett. **523**, L137.
- Donahue, M., and G. M. Voit, 2004, in *Carnegie Astrophysics Series, Vol. 3: Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution*, edited by J. S. Mulchaey, A. Dressler, and A. Oemler, p. in press.
- Donahue, M., G. M. Voit, I. Gioia, G. Lupino, J. P. Hughes, and J. T. Stocke, 1998, Astroph. J. **502**, 550.
- Donahue, M., G. M. Voit, C. A. Scharf, I. M. Gioia, C. R. Mullis, J. P. Hughes, and J. T. Stocke, 1999, Astroph. J. **527**, 525.
- Dos Santos, S., and O. Doré, 2002, Astron. Astroph. **383**, 450.
- Durrell, P. R., R. Ciardullo, J. J. Feldmeier, G. H. Jacoby, and S. Sigurdsson, 2002, Astroph. J. **570**, 119.
- Edge, A. C., 2001, Mot. Not. Roy. Ast. Soc. **328**, 762.
- Edge, A. C., and D. T. Frayer, 2003, Astroph. J. Lett. **594**, L13.
- Edge, A. C., and G. C. Stewart, 1991, Mot. Not. Roy. Ast. Soc. **252**, 414.
- Eisenstein, D. J., and W. Hu, 1998, Astroph. J. **496**, 605.
- Eisenstein, D. J., and W. Hu, 1999, Astroph. J. **511**, 5.
- Eke, V. R., S. Cole, and C. S. Frenk, 1996, Mot. Not. Roy. Ast. Soc. **282**, 263.
- Eke, V. R., S. Cole, C. S. Frenk, and J. P. Henry, 1998, Mot. Not. Roy. Ast. Soc. **298**, 1145.
- Eke, V. R., J. F. Navarro, and M. Steinmetz, 2001, Astroph. J. **554**, 114.
- Ettori, S., S. De Grandi, and S. Molendi, 2002, Astron. Astroph. **391**, 841.
- Ettori, S., P. Tozzi, S. Borgani, and P. Rosati, 2003, eprint astro-ph/0312239.
- Ettori, S., P. Tozzi, and P. Rosati, 2003, Astron. Astroph. **398**, 879.

- Evrard, A. E., 1989, *Astroph. J. Lett.* **341**, L71.
- Evrard, A. E., 1990, *Astroph. J.* **363**, 349.
- Evrard, A. E., 1997, *Mot. Not. Roy. Ast. Soc.* **292**, 289.
- Evrard, A. E., 2004, in *Carnegie Astrophysics Series, Vol. 3: Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution*, edited by J. S. Mulchaey, A. Dressler, and A. Oemler, p. in press.
- Evrard, A. E., and J. P. Henry, 1991, *Astroph. J.* **383**, 95.
- Evrard, A. E., T. J. MacFarland, H. M. P. Couchman, J. M. Colberg, N. Yoshida, S. D. M. White, A. Jenkins, C. S. Frenk, F. R. Pearce, J. A. Peacock, and P. A. Thomas, 2002, *Astroph. J.* **573**, 7.
- Evrard, A. E., C. A. Metzler, and J. F. Navarro, 1996, *Astroph. J.* **469**, 494.
- Fabian, A. C., 1994, *Ann. Rev. Astron. Astroph.* **32**, 277.
- Fabian, A. C., C. S. Crawford, A. C. Edge, and R. F. Mushotzky, 1994, *Mot. Not. Roy. Ast. Soc.* **267**, 779.
- Fabian, A. C., E. M. Hu, L. L. Cowie, and J. Grindlay, 1981, *Astroph. J.* **248**, 47.
- Fabian, A. C., and P. E. J. Nulsen, 1977, *Mot. Not. Roy. Ast. Soc.* **180**, 479.
- Fabian, A. C., J. S. Sanders, S. W. Allen, C. S. Crawford, K. Iwasawa, R. M. Johnstone, R. W. Schmidt, and G. B. Taylor, 2003, *Mot. Not. Roy. Ast. Soc.* **344**, L43.
- Fabian, A. C., J. S. Sanders, S. Ettori, G. B. Taylor, S. W. Allen, C. S. Crawford, K. Iwasawa, R. M. Johnstone, and P. M. Ogle, 2000, *Mot. Not. Roy. Ast. Soc.* **318**, L65.
- Fabian, A. C., L. M. Voigt, and R. G. Morris, 2002, *Mot. Not. Roy. Ast. Soc.* **335**, L71.
- Felten, J. E., R. J. Gould, W. A. Stein, and N. J. Woolf, 1966, *Astroph. J.* **146**, 955.
- Ferguson, H. C., N. R. Tanvir, and T. von Hippel, 1998, *Nature (London)* **391**, 461.
- Ferland, G. J., A. C. Fabian, and R. M. Johnstone, 1994, *Mot. Not. Roy. Ast. Soc.* **266**, 399.
- Ferland, G. J., A. C. Fabian, and R. M. Johnstone, 2002, *Mot. Not. Roy. Ast. Soc.* **333**, 876.
- Ferrarese, L., and D. Merritt, 2000, *Astroph. J. Lett.* **539**, L9.
- Fillmore, J. A., and P. Goldreich, 1984, *Astroph. J.* **281**, 1.
- Finoguenov, A., M. Arnaud, and L. P. David, 2001a, *Astroph. J.* **555**, 191.
- Finoguenov, A., A. Burkert, and H. Böhringer, 2003, *Astroph. J.* **594**, 136.
- Finoguenov, A., T. H. Reiprich, and H. Böhringer, 2001b, *Astron. Astroph.* **368**, 749.
- Fixsen, D. J., E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright, 1996, *Astroph. J.* **473**, 576.
- Forman, W., E. Kellogg, H. Gursky, H. Tananbaum, and R. Giacconi, 1972, *Astroph. J.* **178**, 309.
- Freedman, W. L., B. F. Madore, B. K. Gibson, L. Ferrarese, D. D. Kelson, S. Sakai, J. R. Mould, R. C. Kennicutt, H. C. Ford, J. A. Graham, J. P. Huchra, S. M. G. Hughes, *et al.*, 2001, *Astroph. J.* **553**, 47.
- Frenk, C. S., S. D. M. White, P. Bode, J. R. Bond, G. L. Bryan, R. Cen, H. M. P. Couchman, A. E. Evrard, N. Gnedin, A. Jenkins, A. M. Khokhlov, A. Klypin, *et al.*, 1999, *Astroph. J.* **525**, 554.
- Gal, R. R., and L. M. Lubin, 2004, *Astrophys. J.* **607**, L1.
- Gebhardt, K., R. Bender, G. Bower, A. Dressler, S. M. Faber, A. V. Filippenko, R. Green, C. Grillmair, L. C. Ho, J. Kormendy, T. R. Lauer, J. Magorrian, *et al.*, 2000, *Astroph. J. Lett.* **539**, L13.
- Gibson, B. K., and F. Matteucci, 1997, *Mot. Not. Roy. Ast. Soc.* **291**, L8.
- Gilbank, D. G., R. G. Bower, F. J. Castander, and B. L. Ziegler, 2003, eprint astro-ph/0310809.
- Girardi, M., S. Borgani, G. Giuricin, F. Mardirossian, and M. Mezzetti, 1998, *Astroph. J.* **506**, 45.
- Girardi, M., P. Manzato, M. Mezzetti, G. Giuricin, and F. Limboz, 2002, *Astroph. J.* **569**, 720.
- Gladders, M. D., and H. K. C. Yee, 2000, *Astron. J.* **120**, 2148.
- Gunn, J. E., 1977, *Astroph. J.* **218**, 592.
- Gunn, J. E., and J. R. I. Gott, 1972, *Astroph. J.* **176**, 1.
- Gursky, H., E. Kellogg, S. Murray, C. Leong, H. Tananbaum, and R. Giacconi, 1971, *Astroph. J. Lett.* **167**, L81+.
- Guth, A. H., and S.-Y. Pi, 1982, *Physical Review Letters* **49**, 1110.
- Haiman, Z., J. J. Mohr, and G. P. Holder, 2001, *Astroph. J.* **553**, 545.
- Hamilton, A. J. S., 2001, *Mot. Not. Roy. Ast. Soc.* **322**, 419.
- Harrison, E. R., 1970, *Phys. Rev. D* **1**, 2726.
- Heath, D. J., 1977, *Mot. Not. Roy. Ast. Soc.* **179**, 351.
- Heckman, T. M., 2002, in *ASP Conf. Ser. 254: Extragalactic Gas at Low Redshift*, pp. 292–+.
- Heckman, T. M., L. Armus, and G. K. Miley, 1990, *Astroph. J. Suppl.* **74**, 833.
- Heisler, J., S. Tremaine, and J. N. Bahcall, 1985, *Astroph. J.* **298**, 8.
- Helsdon, S. F., and T. J. Ponman, 2000, *Mot. Not. Roy. Ast. Soc.* **315**, 356.
- Henry, J. P., 1997, *Astroph. J. Lett.* **489**, L1+.
- Henry, J. P., 2000, *Astroph. J.* **534**, 565.
- Henry, J. P., and K. A. Arnaud, 1991, *Astroph. J.* **372**, 410.
- Herschel, F. W., 1785, *Phil. Trans.* **75**, 213.
- Hoekstra, H., 2001, *Astron. Astroph.* **370**, 743.
- Hoekstra, H., M. Franx, K. Kuijken, and G. Squires, 1998, *Astroph. J.* **504**, 636.
- Holder, G., Z. Haiman, and J. J. Mohr, 2001, *Astroph. J. Lett.* **560**, L111.
- Holder, G. P., and J. E. Carlstrom, 2001, *Astroph. J.* **558**, 515.
- Horner, D. J., M. E. Donahue, and G. M. Voit, 2004, in preparation .
- Horner, D. J., R. F. Mushotzky, and C. A. Scharf, 1999, *Astroph. J.* **520**, 78.
- Hu, E. M., L. L. Cowie, and Z. Wang, 1985, *Astroph. J. Suppl.* **59**, 447.
- Hu, W., 2003, *Phys. Rev. D* **67**, 081304.
- Hu, W., and A. V. Kravtsov, 2003, *Astroph. J.* **584**, 702.
- Huterer, D., and M. White, 2002, *Astroph. J. Lett.* **578**, L95.
- Ikebe, Y., T. H. Reiprich, H. Böhringer, Y. Tanaka, and T. Kitayama, 2002, *Astron. Astroph.* **383**, 773.
- Inoue, S., and S. Sasaki, 2001, *Astroph. J.* **562**, 618.
- Irwin, J. A., and J. N. Bregman, 2000, *Astroph. J.* **538**, 543.
- Jarvis, M., G. M. Bernstein, P. Fischer, D. Smith, B. Jain, J. A. Tyson, and D. Wittman, 2003, *Astron. J.* **125**, 1014.
- Jenkins, A., C. S. Frenk, F. R. Pearce, P. A. Thomas, J. M. Colberg, S. D. M. White, H. M. P. Couchman, J. A. Peacock, G. Efstathiou, and A. H. Nelson, 1998, *Astroph. J.* **499**, 20.
- Jenkins, A., C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, and N. Yoshida, 2001, *Mot. Not. Roy. Ast. Soc.* **321**, 372.
- Jing, Y. P., 2000, *Astroph. J.* **535**, 30.
- Johnstone, R. M., A. C. Fabian, and P. E. J. Nulsen, 1987, *Mot. Not. Roy. Ast. Soc.* **224**, 75.

- Jones, C., and W. Forman, 1984, *Astroph. J.* **276**, 38.
- Jubelgas, M., V. Springel, and K. Dolag, 2004, *Mot. Not. Roy. Ast. Soc.* **351**, 423.
- Kaiser, C. R., and J. Binney, 2003, *Mot. Not. Roy. Ast. Soc.* **338**, 837.
- Kaiser, N., 1984, *Astroph. J. Lett.* **284**, L9.
- Kaiser, N., 1986, *Mot. Not. Roy. Ast. Soc.* **222**, 323.
- Kaiser, N., 1987, *Mot. Not. Roy. Ast. Soc.* **227**, 1.
- Kaiser, N., 1991, *Astroph. J.* **383**, 104.
- Kaiser, N., and G. Squires, 1993, *Astroph. J.* **404**, 441.
- Kay, S. T., 2004, *Mot. Not. Roy. Ast. Soc.* **347**, L13.
- Kay, S. T., P. A. Thomas, and T. Theuns, 2003, *Mot. Not. Roy. Ast. Soc.* **343**, 608.
- Kellogg, E., H. Gursky, H. Tananbaum, R. Giacconi, and K. Pounds, 1972, *Astroph. J. Lett.* **174**, L65+.
- Kent, S. M., and J. E. Gunn, 1982, *Astron. J.* **87**, 945.
- Kim, W.-T., and R. Narayan, 2003, *Astrophys. J.* **596**, L139.
- King, I., 1962, *Astron. J.* **67**, 471.
- Klypin, A. A., and A. I. Kopylov, 1983, *Soviet Astronomy Letters* **9**, 41.
- Knight, P. A., and T. J. Ponman, 1997, *Mot. Not. Roy. Ast. Soc.* **289**, 955.
- Kochanek, C. S., M. White, J. Huchra, L. Macri, T. H. Jarrett, S. E. Schneider, and J. Mader, 2003, *Astroph. J.* **585**, 161.
- Komatsu, E., and U. Seljak, 2001, *Mot. Not. Roy. Ast. Soc.* **327**, 1353.
- Kravtsov, A. V., and G. Yepes, 2000, *Mot. Not. Roy. Ast. Soc.* **318**, 227.
- Kriss, G. A., D. F. Cioffi, and C. R. Canizares, 1983, *Astroph. J.* **272**, 439.
- Kuo, C. L., P. A. R. Ade, J. J. Bock, C. Cantalupo, M. D. Daub, J. Goldstein, W. L. Holzapfel, A. E. Lange, M. Lueker, M. Newcomb, J. B. Peterson, J. Ruhl, *et al.*, 2004, *Astroph. J.* **600**, 32.
- Lacey, C., and S. Cole, 1993, *Mot. Not. Roy. Ast. Soc.* **262**, 627.
- Lahav, O., M. J. Rees, P. B. Lilje, and J. R. Primack, 1991, *Mot. Not. Roy. Ast. Soc.* **251**, 128.
- Landau, L. D., and E. M. Lifshitz, 1959, *Fluid mechanics* (Course of theoretical physics, Oxford: Pergamon Press, 1959).
- Larson, R. B., and H. L. Dinerstein, 1975, *Proc. Ast. Soc. Pacific* **87**, 911.
- Levine, E. S., A. E. Schulz, and M. White, 2002, *Astroph. J.* **577**, 569.
- Lewis, A. D., D. A. Buote, and J. T. Stocke, 2003, *Astroph. J.* **586**, 135.
- Lewis, A. D., E. Ellingson, S. L. Morris, and R. G. Carlberg, 1999, *Astroph. J.* **517**, 587.
- Lewis, G. F., A. Babul, N. Katz, T. Quinn, L. Hernquist, and D. H. Weinberg, 2000, *Astroph. J.* **536**, 623.
- Limber, D. N., and W. G. Mathews, 1960, *Astroph. J.* **132**, 286.
- Lin, Y., J. J. Mohr, and S. A. Stanford, 2003, *Astroph. J.* **591**, 749.
- Lloyd-Davies, E. J., T. J. Ponman, and D. B. Cannon, 2000, *Mot. Not. Roy. Ast. Soc.* **315**, 689.
- Loewenstein, M., 2001, *Astroph. J.* **557**, 573.
- Loewenstein, M., and R. F. Mushotzky, 1996, *Astroph. J.* **466**, 695.
- Longair, M. S., and M. Seldner, 1979, *Mot. Not. Roy. Ast. Soc.* **189**, 433.
- Lubin, L. M., and N. A. Bahcall, 1993, *Astroph. J. Lett.* **415**, L17.
- Lubin, L. M., J. S. Mulchaey, and M. Postman, 2004, *Astroph. J. Lett.* **601**, L9.
- Lumb, D. H., *et al.*, 2003, eprint astro-ph/0311344.
- Lumsden, S. L., R. C. Nichol, C. A. Collins, and L. Guzzo, 1992, *Mot. Not. Roy. Ast. Soc.* **258**, 1.
- Luppino, G. A., and I. M. Gioia, 1995, *Astroph. J. Lett.* **445**, L77.
- Lynden-Bell, D., 1967, *Mot. Not. Roy. Ast. Soc.* **136**, 101.
- Mac Low, M., and A. Ferrara, 1999, *Astroph. J.* **513**, 142.
- Magorrian, J., S. Tremaine, D. Richstone, R. Bender, G. Bower, A. Dressler, S. M. Faber, K. Gebhardt, R. Green, C. Grillmair, J. Kormendy, and T. Lauer, 1998, *Astron. J.* **115**, 2285.
- Majumdar, S., and J. J. Mohr, 2003, *Astroph. J.* **585**, 603.
- Majumdar, S., and J. J. Mohr, 2004, eprint astro-ph/0305341.
- Malyshkin, L., 2001, *Astroph. J.* **554**, 561.
- Malyshkin, L., and R. Kulsrud, 2001, *Astroph. J.* **549**, 402.
- Markevitch, M., 1998, *Astroph. J.* **504**, 27.
- Markevitch, M., W. R. Forman, C. L. Sarazin, and A. Vikhlinin, 1998, *Astroph. J.* **503**, 77.
- Martin, C. L., 1999, *Astroph. J.* **513**, 156.
- Mather, J. C., E. S. Cheng, R. E. Eplee, R. B. Isaacman, S. S. Meyer, R. A. Shafer, R. Weiss, E. L. Wright, C. L. Bennett, N. W. Boggess, E. Dwek, S. Gulkis, *et al.*, 1990, *Astroph. J. Lett.* **354**, L37.
- Mathews, W. G., and J. N. Bregman, 1978, *Astroph. J.* **224**, 308.
- Mathiesen, B. F., and A. E. Evrard, 2001, *Astroph. J.* **546**, 100.
- Matteucci, F., and B. K. Gibson, 1995, *Astron. Astroph.* **304**, 11.
- Mazzotta, P., E. Rasia, L. Moscardini, and G. Tormen, 2004, eprint astro-ph/0404425.
- McNamara, B. R., and W. Jaffe, 1994, *Astron. Astroph.* **281**, 673.
- McNamara, B. R., and R. W. O'Connell, 1989, *Astron. J.* **98**, 2018.
- McNamara, B. R., M. Wise, P. E. J. Nulsen, L. P. David, C. L. Sarazin, M. Bautz, M. Markevitch, A. Vikhlinin, W. R. Forman, C. Jones, and D. E. Harris, 2000, *Astroph. J. Lett.* **534**, L135.
- McNamara, B. R., M. W. Wise, and S. S. Murray, 2004, *Astroph. J.* **601**, 173.
- Mellier, Y., 1999, *Ann. Rev. Astron. Astroph.* **37**, 127.
- Messier, C., 1784, *Connaissance des Temps* (Paris).
- Metzler, C. A., M. White, and C. Loken, 2001, *Astroph. J.* **547**, 560.
- Metzler, C. A., M. White, M. Norman, and C. Loken, 1999, *Astroph. J. Lett.* **520**, L9.
- Mo, H. J., and S. D. M. White, 1996, *Mot. Not. Roy. Ast. Soc.* **282**, 347.
- Mohr, J. J., B. Mathiesen, and A. E. Evrard, 1999, *Astroph. J.* **517**, 627.
- Mohr, J. J., E. D. Reese, E. Ellingson, A. D. Lewis, and A. E. Evrard, 2000, *Astroph. J.* **544**, 109.
- Mota, D. F., and C. van de Bruck, 2004, eprint astro-ph/0401504.
- Muanwong, O., P. A. Thomas, S. T. Kay, and F. R. Pearce, 2002, *Mot. Not. Roy. Ast. Soc.* **336**, 527.
- Muanwong, O., P. A. Thomas, S. T. Kay, F. R. Pearce, and H. M. P. Couchman, 2001, *Astroph. J. Lett.* **552**, L27.
- Mulchaey, J. S., A. Dressler, and A. Oemler (eds.), 2004, *Carnegie Astrophysics Series, Vol. 3: Clusters of Galaxies*.

- ies: Probes of Cosmological Structure and Galaxy Evolution* (Cambridge: Cambridge University Press).
- Mullis, C. R., *et al.*, 2004, eprint astro-ph/0401605.
- Mushotzky, R. F., 1984, Physica Scripta Volume T **7**, 157.
- Mushotzky, R. F., 2004, in *Carnegie Astrophysics Series, Vol. 3: Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution*, edited by J. S. Mulchaey, A. Dressler, and A. Oemler, p. in press.
- Mushotzky, R. F., and C. A. Scharf, 1997, *Astroph. J. Lett.* **482**, L13+.
- Narayan, R., and M. V. Medvedev, 2001, *Astroph. J. Lett.* **562**, L129.
- Nath, B. B., and S. Roychowdhury, 2002, *Mot. Not. Roy. Ast. Soc.* **333**, 145.
- Navarro, J. F., C. S. Frenk, and S. D. M. White, 1995, *Mot. Not. Roy. Ast. Soc.* **275**, 720.
- Navarro, J. F., C. S. Frenk, and S. D. M. White, 1997, *Astroph. J.* **490**, 493.
- Navarro, J. F., and S. D. M. White, 1993, *Mot. Not. Roy. Ast. Soc.* **265**, 271.
- Navarro, J. F., *et al.*, 2003, eprint astro-ph/0311231.
- Neumann, D. M., and M. Arnaud, 2001, *Astron. Astroph.* **373**, L33.
- Nevalainen, J., M. Markevitch, and W. Forman, 2000, *Astroph. J.* **536**, 73.
- Nichol, R. C., 2004, in *Carnegie Astrophysics Series, Vol. 3: Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution*, edited by J. S. Mulchaey, A. Dressler, and A. Oemler, p. in press.
- Novicki, M. C., M. Sornig, and J. P. Henry, 2002, *Astron. J.* **124**, 2413.
- O'Connell, R. W., and B. R. McNamara, 1989, *Astron. J.* **98**, 180.
- O'Dea, C. P., S. A. Baum, P. R. Maloney, L. J. Tacconi, and W. B. Sparks, 1994, *Astroph. J.* **422**, 467.
- O'Dea, C. P., H. E. Payne, and D. Kocevski, 1998, *Astron. J.* **116**, 623.
- Oh, S. P., and A. J. Benson, 2003, *Mot. Not. Roy. Ast. Soc.* **342**, 664.
- Osmond, J. P. F., and T. J. Ponman, 2004, *Mot. Not. Roy. Ast. Soc.* **350**, 1511.
- Oukbir, J., and A. Blanchard, 1992, *Astron. Astroph.* **262**, L21.
- Peacock, J. A., 1999, *Cosmological Physics* (Cosmological physics. Publisher: Cambridge, UK: Cambridge University Press, 1999. ISBN: 0521422701).
- Peacock, J. A., and S. J. Dodds, 1994, *Mot. Not. Roy. Ast. Soc.* **267**, 1020.
- Pearce, F. R., P. A. Thomas, H. M. P. Couchman, and A. C. Edge, 2000, *Mot. Not. Roy. Ast. Soc.* **317**, 1029.
- Pearson, T. J., B. S. Mason, A. C. S. Readhead, M. C. Shepherd, J. L. Sievers, P. S. Udomprasert, J. K. Cartwright, A. J. Farmer, S. Padin, S. T. Myers, J. R. Bond, C. R. Contaldi, *et al.*, 2003, *Astroph. J.* **591**, 556.
- Peebles, P. J., and B. Ratra, 2003, *Reviews of Modern Physics* **75**, 559.
- Peebles, P. J. E., 1993, *Principles of physical cosmology* (Princeton Series in Physics, Princeton, NJ: Princeton University Press, —c1993).
- Peebles, P. J. E., R. A. Daly, and R. Juszkiewicz, 1989, *Astroph. J.* **347**, 563.
- Peebles, P. J. E., and J. T. Yu, 1970, *Astroph. J.* **162**, 815.
- Pen, U., 1997, *New Astronomy* **2**, 309.
- Penton, S. V., J. T. Stocke, and J. M. Shull, 2002, *Astroph. J.* **565**, 720.
- Perlmutter, S., G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, *et al.*, 1999, *Astroph. J.* **517**, 565.
- Peterson, J. R., S. M. Kahn, F. B. S. Paerels, J. S. Kaasstra, T. Tamura, J. A. M. Bleeker, C. Ferrigno, and J. G. Jernigan, 2003, *Astroph. J.* **590**, 207.
- Peterson, J. R., F. B. S. Paerels, J. S. Kaasstra, M. Arnaud, T. H. Reiprich, A. C. Fabian, R. F. Mushotzky, J. G. Jernigan, and I. Sakelliou, 2001, *Astron. Astroph.* **365**, L104.
- Pettini, M., A. E. Shapley, C. C. Steidel, J. Cuby, M. Dickinson, A. F. M. Moorwood, K. L. Adelberger, and M. Giavalisco, 2001, *Astroph. J.* **554**, 981.
- Pettini, M., C. C. Steidel, K. L. Adelberger, M. Dickinson, and M. Giavalisco, 2000, *Astroph. J.* **528**, 96.
- Pierpaoli, E., S. Borgani, D. Scott, and M. White, 2003, *Mot. Not. Roy. Ast. Soc.* **342**, 163.
- Pipino, A., F. Matteucci, S. Borgani, and A. Biviano, 2002, *New Astronomy* **7**, 227.
- Pizzolato, F., S. Molendi, S. Ghizzardi, and S. De Grandi, 2003, *Astroph. J.* **592**, 62.
- Ponman, T. J., D. B. Cannon, and J. F. Navarro, 1999, *Nature (London)* **397**, 135.
- Ponman, T. J., A. J. R. Sanderson, and A. Finoguenov, 2003, *Mot. Not. Roy. Ast. Soc.* **343**, 331.
- Portinari, L., A. Moretti, C. Chiosi, and J. Sommer-Larsen, 2003, eprint astro-ph/0312360.
- Postman, M., J. P. Huchra, and M. J. Geller, 1992, *Astroph. J.* **384**, 404.
- Postman, M., T. R. Lauer, W. Oegerle, and M. Donahue, 2002, *Astroph. J.* **579**, 93.
- Postman, M., L. M. Lubin, J. E. Gunn, J. B. Oke, J. G. Hoesel, D. P. Schneider, and J. A. Christensen, 1996, *Astron. J.* **111**, 615.
- Pratt, G. W., and M. Arnaud, 2002, *Astron. Astroph.* **394**, 375.
- Pratt, G. W., and M. Arnaud, 2003, *Astron. Astroph.* **408**, 1.
- Press, W. H., and P. Schechter, 1974, *Astroph. J.* **187**, 425.
- Randall, S. W., C. L. Sarazin, and P. M. Ricker, 2002, *Astroph. J.* **577**, 579.
- Rasia, E., G. Tormen, and L. Moscardini, 2003, eprint astro-ph/0309405.
- Raymond, J. C., and B. W. Smith, 1977, *Astroph. J. Suppl.* **35**, 419.
- Regos, E., and M. J. Geller, 1989, *Astron. J.* **98**, 755.
- Reiprich, T. H., and H. Böhringer, 2002, *Astroph. J.* **567**, 716.
- Renzini, A., 1997, *Astroph. J.* **488**, 35.
- Renzini, A., 2004, in *Carnegie Astrophysics Series, Vol. 3: Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution*, edited by J. S. Mulchaey, A. Dressler, and A. Oemler, p. in press.
- Ricker, P. M., and C. L. Sarazin, 2001, *Astroph. J.* **561**, 621.
- Riess, A. G., A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, *et al.*, 1998, *Astron. J.* **116**, 1009.
- Riess, A. G., L. Strolger, J. Tonry, S. Casertano, H. C. Ferguson, B. Mobasher, P. Challis, A. V. Filippenko, S. Jha, W. Li, R. Chornock, R. P. Kirshner, *et al.*, 2004, *Astroph. J.* **607**, 665.
- Rines, K., M. J. Geller, M. J. Kurtz, and A. Diaferio, 2003,

- Astron. J. **126**, 2152.
- Robinson, J., E. Gawiser, and J. Silk, 2000, Astroph. J. **532**, 1.
- Rood, H. J., T. L. Page, E. C. Kintner, and I. R. King, 1972, Astroph. J. **175**, 627.
- Rosati, P., S. Borgani, and C. Norman, 2002, Ann. Rev. Astron. Astroph. **40**, 539.
- Rosati, P., P. Tozzi, S. Ettori, V. Mainieri, R. Demarco, S. A. Stanford, C. Lidman, M. Nonino, S. Borgani, R. Della Ceca, P. Eisenhardt, B. P. Holden, *et al.*, 2004, Astron. J. **127**, 230.
- Rosner, R., and W. H. Tucker, 1989, Astroph. J. **338**, 761.
- Ruszkowski, M., and M. C. Begelman, 2002, Astroph. J. **581**, 223.
- Sanders, R. H., 2003, Mot. Not. Roy. Ast. Soc. **342**, 901.
- Sanderson, A. J. R., T. J. Ponman, A. Finoguenov, E. J. Lloyd-Davies, and M. Markevitch, 2003, Mot. Not. Roy. Ast. Soc. **340**, 989.
- Sarazin, C. L., 1988, *X-ray emission from clusters of galaxies* (Cambridge Astrophysics Series, Cambridge: Cambridge University Press, 1988).
- Sasaki, S., 1996, Pub. Ast. Soc. Japan **48**, L119.
- Scannapieco, E., and S. P. Oh, 2004, eprint astro-ph/0401087.
- Schechter, P., 1976, Astroph. J. **203**, 297.
- Schirmer, M., T. Erben, P. Schneider, G. Pietrzynski, W. Gieren, S. Carpano, A. Miclo, and F. Pierfederici, 2003, Astron. Astroph. **407**, 869.
- Schuecker, P., H. Böhringer, L. Guzzo, C. A. Collins, D. M. Neumann, S. Schindler, W. Voges, S. De Grandi, G. Chincarini, R. Crudace, V. Müller, T. H. Reiprich, *et al.*, 2001, Astron. Astroph. **368**, 86.
- Schuecker, P., R. R. Caldwell, H. Böhringer, C. A. Collins, L. Guzzo, and N. N. Weinberg, 2003, Astron. Astroph. **402**, 53.
- Seljak, U., 2002, Mot. Not. Roy. Ast. Soc. **337**, 769.
- Seljak, U., J. Burwell, and U. Pen, 2001, Phys. Rev. D **63**, 063001.
- Sheth, R. K., H. J. Mo, and G. Tormen, 2001, Mot. Not. Roy. Ast. Soc. **323**, 1.
- Sheth, R. K., and G. Tormen, 1999, Mot. Not. Roy. Ast. Soc. **308**, 119.
- Shull, J. M., J. T. Stocke, and S. Penton, 1996, Astron. J. **111**, 72.
- Shull, J. M., J. Tumlinson, and M. L. Giroux, 2003, Astroph. J. Lett. **594**, L107.
- Smith, G. P., A. C. Edge, V. R. Eke, R. C. Nichol, I. Smail, and J. Kneib, 2003, Astroph. J. Lett. **590**, L79.
- Smith, S., 1936, Astroph. J. **83**, 23.
- Soker, N., 2003, Mot. Not. Roy. Ast. Soc. **342**, 463.
- Sparks, W. B., 1992, Astroph. J. **399**, 66.
- Spergel, D. N., L. Verde, H. V. Peiris, E. Komatsu, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, *et al.*, 2003, Astroph. J. Suppl. **148**, 175.
- Spitzer, L., 1962, *Physics of Fully Ionized Gases* (Physics of Fully Ionized Gases, New York: Interscience (2nd edition), 1962).
- Stanford, S. A., B. Holden, P. Rosati, P. Tozzi, S. Borgani, P. R. Eisenhardt, and H. Spinrad, 2001, Astroph. J. **552**, 504.
- Suginohara, T., and J. P. Ostriker, 1998, Astroph. J. **507**, 16.
- Sugiyama, N., 1995, Astroph. J. Suppl. **100**, 281.
- Sunyaev, R. A., and Y. B. Zeldovich, 1970, Astroph. Sp. Sci. **7**, 3.
- Sunyaev, R. A., and Y. B. Zeldovich, 1972, Comments on Astrophysics **4**, 173.
- Sutherland, R. S., and M. A. Dopita, 1993, Astroph. J. Suppl. **88**, 253.
- Szalay, A. S., B. Jain, T. Matsubara, R. Scranton, M. S. Vogeley, A. Connolly, S. Dodelson, D. Eisenstein, J. A. Frieman, J. E. Gunn, L. Hui, D. Johnston, *et al.*, 2003, Astroph. J. **591**, 1.
- The, L. S., and S. D. M. White, 1986, Astron. J. **92**, 1248.
- Thomas, P. A., O. Muanwong, F. R. Pearce, H. M. P. Couchman, A. C. Edge, A. Jenkins, and L. Onuora, 2001, Mot. Not. Roy. Ast. Soc. **324**, 450.
- Tornatore, L., S. Borgani, V. Springel, F. Matteucci, N. Menci, and G. Murante, 2003, Mot. Not. Roy. Ast. Soc. **342**, 1025.
- Tozzi, P., and C. Norman, 2001, Astroph. J. **546**, 63.
- Tozzi, P., P. Rosati, S. Ettori, S. Borgani, V. Mainieri, and C. Norman, 2003, Astroph. J. **593**, 705.
- Tran, K. H., D. D. Kelson, P. van Dokkum, M. Franx, G. D. Illingworth, and D. Magee, 1999, Astroph. J. **522**, 39.
- Tripp, T. M., B. D. Savage, and E. B. Jenkins, 2000, Astroph. J. Lett. **534**, L1.
- Tucker, W. H., and R. Rosner, 1983, Astroph. J. **267**, 547.
- Turner, M. S., and M. White, 1997, Phys. Rev. D **56**, 4439.
- Tyson, J. A., R. A. Wenk, and F. Valdes, 1990, Astroph. J. Lett. **349**, L1.
- Valageas, P., and J. Silk, 1999, Astron. Astroph. **350**, 725.
- Valdarnini, R., 2003, Mot. Not. Roy. Ast. Soc. **339**, 1117.
- van Haarlem, M. P., C. S. Frenk, and S. D. M. White, 1997, Mot. Not. Roy. Ast. Soc. **287**, 817.
- Viana, P. T. P., and A. R. Liddle, 1996, Mot. Not. Roy. Ast. Soc. **281**, 323.
- Viana, P. T. P., and A. R. Liddle, 1999, Mot. Not. Roy. Ast. Soc. **303**, 535.
- Vikhlinin, A., W. Forman, and C. Jones, 1999, Astroph. J. **525**, 47.
- Vikhlinin, A., L. VanSpeybroeck, M. Markevitch, W. R. Forman, and L. Grego, 2002, Astroph. J. Lett. **578**, L107.
- Vikhlinin, A., A. Voevodkin, C. R. Mullis, L. VanSpeybroeck, H. Quintana, B. R. McNamara, I. Gioia, A. Hornstrup, J. P. Henry, W. R. Forman, and C. Jones, 2003, Astroph. J. **590**, 15.
- Voevodkin, A., and A. Vikhlinin, 2004, Astroph. J. **601**, 610.
- Voigt, L. M., and A. C. Fabian, 2004, Mot. Not. Roy. Ast. Soc. **347**, 1130.
- Voit, G. M., 1994, Astroph. J. Lett. **432**, L19.
- Voit, G. M., 1996, Astroph. J. **465**, 548.
- Voit, G. M., 2000, Astroph. J. **543**, 113.
- Voit, G. M., M. L. Balogh, R. G. Bower, C. G. Lacey, and G. L. Bryan, 2003, Astroph. J. **593**, 272.
- Voit, G. M., and G. L. Bryan, 2001, Nature (London) **414**, 425.
- Voit, G. M., G. L. Bryan, M. L. Balogh, and R. G. Bower, 2002, Astrophys. J. **576**, 601.
- Voit, G. M., and M. Donahue, 1995, Astroph. J. **452**, 164.
- Voit, G. M., and M. Donahue, 1997, Astroph. J. **486**, 242.
- Voit, G. M., and M. Donahue, 1998, Astroph. J. Lett. **500**, L111+.
- Voit, G. M., A. E. Evrard, and G. L. Bryan, 2001, Astroph. J. Lett. **548**, L123.
- Voit, G. M., S. T. Kay, and G. L. Bryan, 2004, in preparation .
- Voit, G. M., and T. J. Ponman, 2003, Astroph. J. Lett. **594**, L75.

- Wang, L., and P. J. Steinhardt, 1998, *Astroph. J.* **508**, 483.
- Waugh, M., M. J. Drinkwater, R. L. Webster, L. Staveley-Smith, V. A. Kilborn, D. G. Barnes, R. Bhathal, W. J. G. de Blok, P. J. Boyce, M. J. Disney, R. D. Ekers, K. C. Freeman, *et al.*, 2002, *Mot. Not. Roy. Ast. Soc.* **337**, 641.
- Weinberg, N. N., and M. Kamionkowski, 2003, *Mot. Not. Roy. Ast. Soc.* **341**, 251.
- Weller, J., R. A. Battye, and R. Kneissl, 2002, *Physical Review Letters* **88**, 231301.
- Weymann, R., 1965, *Phys. Fluids* **8**, 2112.
- Weymann, R., 1966, *Astroph. J.* **145**, 560.
- White, D. A., C. Jones, and W. Forman, 1997, *Mot. Not. Roy. Ast. Soc.* **292**, 419.
- White, M., 2001, *Astron. Astroph.* **367**, 27.
- White, M., G. Gelmini, and J. Silk, 1995, *Phys. Rev. D* **51**, 2669.
- White, M., L. Hernquist, and V. Springel, 2002, *Astroph. J.* **579**, 16.
- White, S. D. M., and C. S. Frenk, 1991, *Astroph. J.* **379**, 52.
- White, S. D. M., C. S. Frenk, M. Davis, and G. Efstathiou, 1987, *Astroph. J.* **313**, 505.
- White, S. D. M., J. F. Navarro, A. E. Evrard, and C. S. Frenk, 1993, *Nature (London)* **366**, 429.
- White, S. D. M., and M. J. Rees, 1978, *Mot. Not. Roy. Ast. Soc.* **183**, 341.
- Wittman, D., V. E. Margoniner, J. A. Tyson, J. G. Cohen, A. C. Becker, and I. P. Dell'Antonio, 2003, *Astroph. J.* **597**, 218.
- Wittman, D., J. A. Tyson, V. E. Margoniner, J. G. Cohen, and I. P. Dell'Antonio, 2001, *Astroph. J. Lett.* **557**, L89.
- Wu, K. K. S., A. C. Fabian, and P. E. J. Nulsen, 2001, *Mot. Not. Roy. Ast. Soc.* **324**, 95.
- Wu, X., and Y. Xue, 2002, *Astroph. J. Lett.* **572**, L19.
- Xue, Y., and X. Wu, 2000, *Astroph. J.* **538**, 65.
- Yee, H. K. C., and E. Ellingson, 2003, *Astroph. J.* **585**, 215.
- Yee, H. K. C., and O. López-Cruz, 1999, *Astron. J.* **117**, 1985.
- Zaroubi, S., G. Squires, G. de Gasperis, A. E. Evrard, Y. Hoffman, and J. Silk, 2001, *Astroph. J.* **561**, 600.
- Zeldovich, Y. B., 1972, *Mot. Not. Roy. Ast. Soc.* **160**, 1P.
- Zwicky, F., 1933, *Helv. physica acta* **6**, 110.
- Zwicky, F., 1937, *Astroph. J.* **86**, 217.

X-ray observations of Clusters of Galaxies

M. ARNAUD

CEA/DSM/DAPNIA Service d'Astrophysique L'Orme des Merisiers, Bat. 709, 91191 Gif sur Yvette Cedex, France

Summary. — X-ray observations of clusters provide key information on the dark matter, on the formation of structures in the Universe, and can be used to constrain the cosmological parameters. I review our current knowledge, with emphasis on recent XMM and Chandra results.

1. – Introduction

In the standard CDM (Cold Dark Matter) cosmological scenario, initial density fluctuations, generated in the early Universe, grow under the influence of gravity. The Universe gets more and more structured with time. At large scale, the matter density distribution exhibits a web-like topology with expanding voids surrounded by contracting sheets and filaments. Massive clusters of galaxies, located at the crossing of filaments, define the nodes of this cosmic web. Clusters of galaxies are the largest collapsed structures, with masses ranging from $10^{13} M_{\odot}$ for small groups to $10^{15} M_{\odot}$ for the richest clusters. Because of their size, their mass content reflects that of the Universe: $\sim 85\%$ of the mass is made of Dark Matter. The main baryonic cluster component is a hot X-ray emitting intracluster gas, as shown by the first X-ray images obtained with the Einstein satellite [1]. Only a few percent of the mass in clusters lies in the optical galaxies.

In the CDM scenario, the amplitude of the initial density fluctuations decreases with increasing scale. As a result, low-mass objects form first and then merge together to form more massive objects. Clusters of galaxies are the last manifestation of this hierarchical clustering. They started to form in the recent cosmological epoch ($z \sim 2$) and the cluster population is continuously evolving. Individual clusters are not immutable objects. A cluster continuously accretes matter and smaller groups along the filaments, and, occasionally, merges with another cluster of similar mass.

Clusters of galaxies are key objects for cosmological studies (see the review [2]). By studying the properties of clusters of galaxies we can test the scenario of structure formation and better understand the gravitational collapse of the Dark Matter and the baryon specific physics. Furthermore, because the history of structure formation depends strongly on the cosmology, studies of cluster samples can also constrain cosmological parameters. In this course, I will focus on the X-ray observations relevant for these issues and on properties which can also be studied with SZ observations. In particular, I will not discuss abundance measurements, although they are important for our understanding of the history of galaxy and star formation and of the heating and enrichment of the intra-cluster medium. This course cannot be exhaustive. More information on X-ray clusters can be found in the reference book “X-ray emission from clusters of galaxies” [3] and in more recent books [4, 5] and reviews quoted in the text.

2. – Observing clusters in X-ray

2.1. X-ray emission of clusters. – The Intra-Cluster medium (ICM) is a hot, tenuous and optically thin plasma. The gas density varies from $\sim 10^{-4} \text{ cm}^{-3}$ in the outer regions of clusters to a few $\sim 10^{-2} \text{ cm}^{-3}$ in the center. The ICM has mean temperatures in the range $kT = 0.5 - 15 \text{ keV}$, reflecting the depth of the potential well ($kT \propto GM/R$). Note that the ICM is not an isothermal plasma, although temperature variations are usually small (see below). The ICM is enriched in heavy elements, with typical abundances of 1/3 the solar value. At the temperature of the ICM, H and He are fully ionized. Most electrons come from these two elements. The electron density is nearly independent of the ionization state and is given by $n_e \sim 1.2n_H$, where n_H is the hydrogen density. The ionization stage of the other elements depends on the temperature.

The X-ray emission is that of a coronal plasma at ionization equilibrium [3, for detailed description]. For a volume element dV of electronic density n_e , temperature T , abundances $[Z/H]$, the number of photons emitted by unit time in the energy range $[E, E + dE]$ can be written as

$$(1) \quad dN(E) = n_e^2 \epsilon(E, T, [Z/H]) dE dV$$

where $\epsilon(E, T, [Z/H])$ is the photon emissivity at energy E . The intensity scales as the *square* of the density, because all emission processes (like the Bremsstrahlung emission) result from collisions between electron and ions.

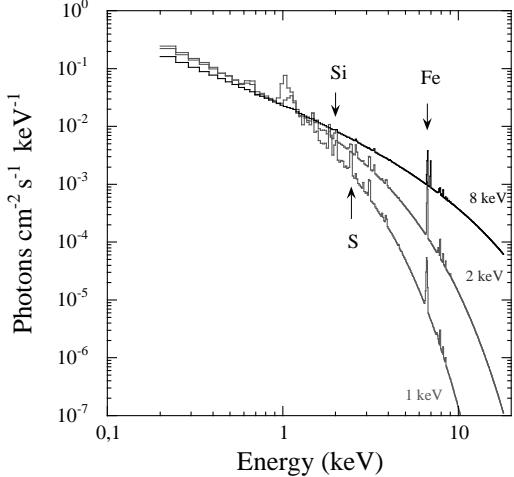


Fig. 1. – The X-ray emission from a thin plasma with 0.35 solar abundance at different temperatures, $kT = 1, 2$ and 8 keV .

Examples of X-ray spectra for typical cluster temperatures are shown in Fig. 1. Due to the high temperature, the continuum emission is dominated by thermal Bremsstrahlung, the main species by far contributing to the emission being H and He. The emissivity of this continuum is very sensitive to temperature for energies greater than kT and rather insensitive to it below. This is due to the exponential cut-off of the Bremsstrahlung emission. Indeed, it scales as $g(E, T)T^{-1/2}\exp(-E/kT)$, where $g(E, T)$, the Gaunt factor, is a gradual function of kT . The only line that clearly stands out at all temperatures is the Iron K line complex around 6.7 keV (see Fig. 1). We can also observe the K lines of other elements ($Z > 8$, H and He-like ionization states), as well as the L-shell complex of lower ionization states of Iron. However the intensity of these lines rapidly decreases with increasing temperature. Except for the cool clusters ($kT \lesssim 4 \text{ keV}$) or in the cooling core present in some clusters, one cannot expect to measure the abundance of elements other than Iron because they are completely ionized.

2.2. Extracting Physical information from X-ray observations. – From above, it is clear that X-ray observations give access to the two characteristics of the ICM, which are the density and the temperature. The shape of the spectrum determines the temperature ⁽¹⁾, whereas the normalization provides the emission measure $EM = \int n_e^2 dV$.

2.2.1. Gas temperature. Its determination requires spectroscopic data. The temperature is derived by fitting the observed spectrum (as in Fig. 5) with a thermal emission model convolved with the instrument response (i.e., taking into account how the effective area and spectral resolution vary over the energy range). The model spectrum at Earth is computed from the emitted spectrum, taking into account the cluster redshift and the

⁽¹⁾ and also the heavy element abundances from the line equivalent widths and possibly the redshift from the line positions

galactic absorption (which mostly affects the emission below 1 – 2 keV). It is important to keep in mind that:

- The temperature is constrained by the position of the exponential cut-off in the spectrum. In order to have a proper determination of the temperature, we need spectroscopic instruments sensitive up to energies greater than kT , i.e., typically 10 keV.
- The ICM is not strictly isothermal. This means that a temperature inferred from an isothermal fit to the data is actually a 'mean' value along the line of sight and in the considered cluster region. This temperature is not simply, as often thought, the 'emission weighted' temperature [6].

2.2.2. Gas density. The emissivity is not very sensitive to kT at low energy. Therefore X-ray images or surface brightness profile (as in Fig. 4) extracted in a soft energy band ($E \lesssim 2$ keV) are used to determine the gas density distribution. X-ray images reflect the ICM morphology. Note that one must not forget projection effects and that density contrast are enhanced because the X-ray emission increases as the square of the density. The emission measure along the line of sight at projected radius r , $EM(r) = \int n_e^2 dl$, can be deduced from the surface brightness, $S(\theta)$:

$$(2) \quad EM(r) = \frac{4\pi(1+z)^4 S(\theta)}{\Lambda(T, z)} \quad \text{with} \quad r = d_A(z) \theta$$

where $d_A(z)$ is the angular distance at the cluster redshift z . $\Lambda(T, z)$ is the emissivity in the considered energy band, taking into account the absorption by our galaxy, the redshift, and the instrumental spectral response. Because $\Lambda(T, z)$ depends only weakly on the temperature in the soft band, it is essentially insensitive to temperature gradients (except for cool systems, [7]) and one can generally use the average cluster temperature.

The gas density radial profile $n_g(R)$ is usually derived from Eq. 2, assuming spherical symmetry. In that case $EM(r) = \int_r^\infty n_e^2(R) R dR / \sqrt{R^2 - r^2}$. One can use de-projection techniques or parametric models fitted to the data. A popular model is the so-called isothermal β -model: $n(R) = n_0 [1 + (R/R_c)^2]^{-3\beta/2}$, which gives $S(\theta) = S_0 [1 + (\theta/\theta_c)^2]^{-3\beta+1/2}$. This model fits reasonably well cluster profiles at large radii, but it underestimates the density in central cooling core of clusters (see also Sect. 4.4).

2.2.3. Cluster properties. Depending on the quality of the data, various physical cluster properties can then be derived. The easiest to determine is the X-ray luminosity, which is computed from the observed flux in the instrument energy band ⁽²⁾ A higher statistical quality is required to measure a surface brightness profile or a spectrum. From these two quantities one derives the global temperature, the gas density radial profile and therefore the gas mass. If spatially resolved spectroscopic measurements are available,

⁽²⁾ Conversion from count rates to bolometric luminosity depends on the temperature. The usual way is to use a L_X-T relation.

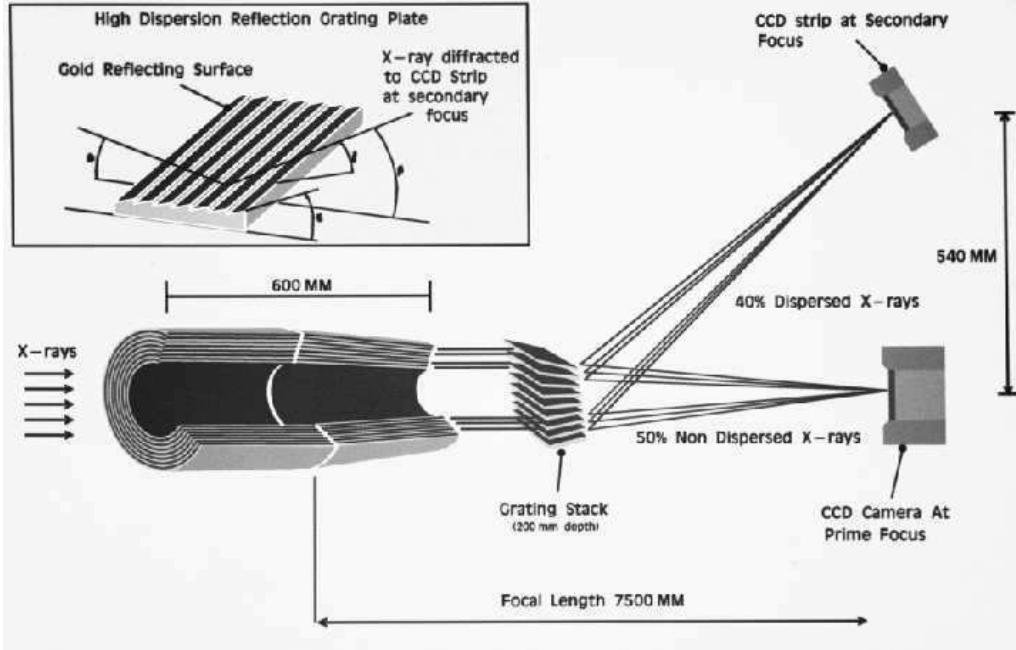


Fig. 2. – The light pass in the two XMM telescopes equipped with the EPIC and RGS experiments. After collection by the mirror module (58 Wolter shells) about 40% of the light is focused to the EPIC CCD camera, while a similar amount is dispersed by a Reflection Grating Array. The diffracted beam is imaged by a CCD camera on the Rowland circle for high resolution spectroscopy. In the third telescope, all the light is collected by the EPIC CCD experiment. The Figure was reproduced from ESA SP-1097.

one can also derive the temperature profile. This gives access to the entropy profile, key information on the history of the ICM (see Sec. 4.4). One can also deduce the total mass profile from the Hydrostatic Equilibrium Equation:

$$(3) \quad M(r) = -\frac{kT r}{G\mu m_p} \left[\frac{d \ln n}{d \ln r} + \frac{d \ln T}{d \ln r} \right]$$

where G and m_p are the gravitational constant and proton mass and $\mu \sim 0.61$. Note that an approximation of the total mass can be obtained by simply assuming isothermality.

2.3. Modern X-ray observatories. – X-rays are absorbed by the Earth's atmosphere. Therefore, X-ray observatories are put on board satellites. Two X-ray satellites are now in operation, *XMM-Newton* and *Chandra*. They are fully described in [8, 9, 10, 11] and [12]. The general concept is the same. The X-ray photons are collected and focused by grazing incidence telescopes (Fig. 2). The focal plane is equipped with CCD cameras, allowing for the measurement of the position and energy of each individual incoming photon (Fig. 2). This permits spatially resolved spectroscopy at medium resolution

$(\Delta E = 60 - 140 \text{ eV})$, in the energy range between 0.3 and 10 keV. Spectroscopy at higher resolution is performed using gratings, the 'dispersed' spectrum being read by CCDs (Fig. 2). Because this is dispersive spectroscopy, the sensitivity is degraded, the spatial information is essentially lost and the spectral resolution rapidly deteriorates as the extent of the source increases. The spectral resolution of the XMM grating instruments (XMM/RGS) is $\Delta(E)/E = 0.1 - 0.3\%$ in the $0.3 - 2 \text{ keV}$ energy band for point sources. Such spectroscopy is limited to clusters with sharply peaked emission profiles. It is restricted to the central $\theta \lesssim 2'$ region. The sensitivity of *Chandra* grating instruments is too low for cluster studies.

The *XMM-Newton* and *Chandra* observatories are complementary. *Chandra* has an extremely good spatial resolution of $\Delta\theta = 0.5''$ (compared to $8''$ for *XMM-Newton*). The strength of *XMM-Newton* is its exceptional collecting area and thus sensitivity: three high-throughput telescopes are operating in parallel. The Field of view is $30'$ in diameter, well adapted to cluster studies. *Chandra* has only one telescope, a smaller field of view of $17' \times 17'$ (for the ACIS-I instrument) and an effective area typically 3(5) times lower than *XMM-Newton* at $1.5(8) \text{ keV}$.

As compared to the previous generation of satellites, *XMM-Newton* and *Chandra* represent a giant step forward in term of sensitivity and spatial resolution. The *ROSAT* satellite [13] had good imagery capability ($\Delta\theta = 15''$) but much lower effective area and very poor spectroscopic capability. The high energy cut-off of the telescope was $E \sim 2 \text{ keV}$, so that accurate temperature measurement was limited to very cool clusters. *ASCA* was the first X-ray observatory [14] with telescopes working up to 10 keV and a CCD camera at the focal plane at one of the telescope (the other telescopes were equipped with proportional counters). As compared to spectroscopy made before with collimated spectrometers (like with the *EXOSAT* or *GINGA* satellite), the gain in sensitivity was very important. It was also the first time one could do spatially resolved spectroscopy of clusters. However, this was limited by the relatively large and energy dependent Point Spread Function. The spatial resolution of *Beppo-SAX* was better, but above all it had the capability of observing sources over more than three decades of energy, from 0.1 to 200 keV [15]. *RXTE* [16] has no imaging capability but was also used to study hard X-ray emission from clusters.

With *XMM-Newton* and *Chandra*:

- We can map the gas distribution in nearby clusters from very deep inside the core, at the scale of a few kpc with *Chandra* [17], up to very close to the virial radius with *XMM-Newton*. Temperatures profiles (and thus mass profiles) can be measured over a wide radial range, down to $0.001R_{200}$ with *Chandra* [18] up to close the virial radius with *XMM-Newton*, even in low mass systems [19]. Last but not least, we have now precise temperature maps for unrelaxed objects [20, 21] and we can resolve very sharp density features [22].
- We can measure basic cluster properties up to high z ($z \sim 1.3$) and down to the *ROSAT* detection limit (with *XMM-Newton*). This includes morphology from images, gas density radial profile, global temperature and gas mass (e.g. [23, 24]). As an example, Fig. 3, 4, 5 show the *Chandra* and *XMM-Newton* observations of RDCS 1252.9-2927 at

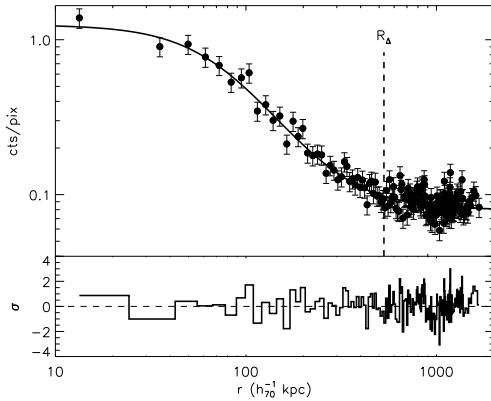
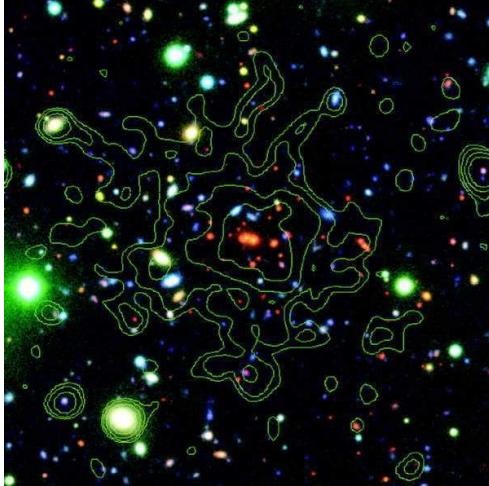


Fig. 4. – Surface brightness profile of RDCS 1252.9-2927 measured with *Chandra* (data points), with best-fit model (solid line) and residuals. Figure from [24]

Fig. 3. – Image of the cluster RDCS 1252.9-2927 at $z = 1.24$. Contours of the *Chandra* emission are overlaid on a composite VLT image. Figure from [24]

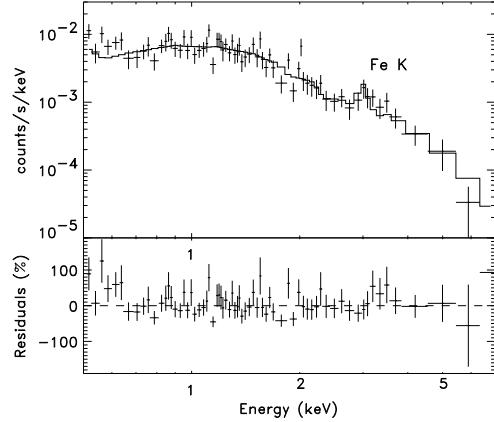


Fig. 5. – X-ray spectrum of RDCS 1252.9-2927 (data points) and best-fit thermal model (solid line) from *XMM-Newton* observation. Note the redshifted Iron line. Figure from [24]

$z = 1.24$ [24]. Total mass and entropy can be derived assuming isothermality [25]. For the brighter distant clusters, crude temperature profiles [23] or maps [26] can be obtained.

- We can find and identify new clusters [27]. Clusters at all redshifts appear as extended sources in *XMM-Newton* and *Chandra* images. However, only *Chandra* has the capability to perfectly remove point source contamination. The typical flux limit for *XMM-Newton* serendipitous surveys is 10^{-14} ergs/s/cm² in the 0.5 – 2 keV band [28], about 3 times lower than with *ROSAT*.

3. – Hierarchical Cluster Formation

3'1. *Substructures and merger events in the local Universe.* – The early Einstein images showed that clusters present a large variety of morphology, from regular clusters to very complex systems with multiple substructures [29]. Since substructures in the ICM are erased on a typical time scale of a few *Gyr*, their detection is the signature of recent dynamical evolution. This manifold morphology indicates a variety of dynamical states in the local cluster population. These observations were consistent with the idea that clusters are still forming to-day, as expected in the hierarchical formation model.

The presence of substructures by itself does not tell much about the cluster formation process. For instance, we expect substructures both if clusters form through mergers of smaller systems (CDM scenario), or by divisions of larger systems (top-down scenario). A direct support of the hierarchical CDM model was the observation of specific signatures of merger events in density and temperature maps, as predicted from numerical simulations (see the detailed reviews [30] on the physics of merger events and [31] on X-ray observations of these events with *ASCA* and *ROSAT*). The first unambiguous signature of a merger event was provided by the *ROSAT* observation of gas compression in the subcluster of A2256, as expected if this subcluster was falling towards the main cluster [32]. The *ASCA* observation of a temperature increase in the interaction region between the sub-clusters of Cygn-A, was even clearer evidence that two sub-clusters were colliding [33]. It also showed that shocks induced by mergers do contribute to the ICM heating. This is a strong prediction of numerical simulations of cluster mergers [34, 35]. When two sub-clusters collide, the interaction region is heated and compressed. As the relative motion in mergers is moderately supersonic ($v \sim 2000$ kms/s), shocks are driven in the ICM. As a result, the gas of the final cluster is heated to its virial temperature, which is higher than the virial temperatures of the initial sub-clusters.

3'2. *Cluster formation at high z .* – With *XMM-Newton* and *Chandra* we now extend the study of substructures to the distant Universe. As expected from the hierarchical formation models, we observe a variety of morphology (and thus dynamical state) up to very high z . A clear case of a double cluster is RX J1053.7+5735, observed with *XMM-Newton* at $z = 1.14$ [36]. This cluster is very probably a merger between two nearly equal mass systems. In contrast, RXJ1226.9+3332 ($z = 0.9$) observed with *XMM-Newton* is a massive cluster (the temperature is $kT = 11 \pm 1$ keV), with a very regular morphology (Fig. 6) indicative of a relaxed state [37]. Note that the existence of massive and relaxed clusters at so high z is expected in low Ω Universe but is very unlikely in a critical $\Omega = 1$ Universe. We also see unambiguous evidence of merging activity up to $z \sim 0.8$ (Fig. 7). The crude *Chandra* temperature map shows a temperature increase between the two subclusters of RX J0152.7-1357, indicating that they have started to merge [26]. Interestingly, when the two subunits will have completely merged the cluster will have the mass of Coma.

3'3. *The detailed physics of merger events and the effect of the Large Scale environment.* – With *Chandra* and *XMM-Newton*, the temperature structure in nearby clusters

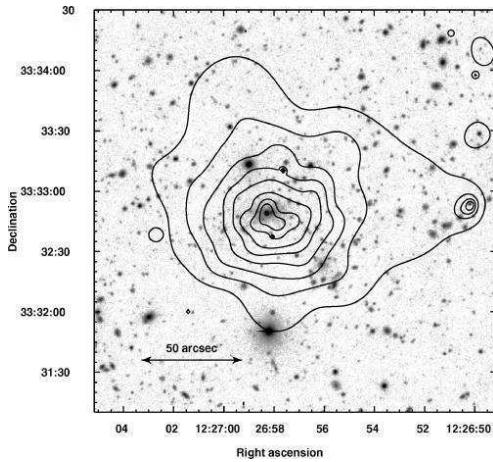


Fig. 6. – RXJ1226.9+3332, a relaxed massive cluster at $z = 0.89$. The contours of X-ray emission detected by *XMM-Newton* is overlaid on a Subaru I-band image. Figure from [37]

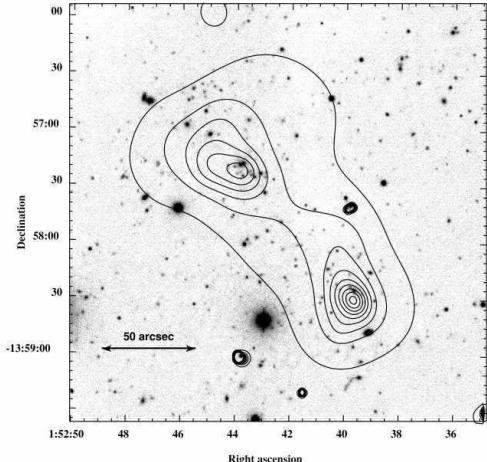


Fig. 7. – The merging cluster RX J0152.7-1357 at $z = 0.83$. The *Chandra* X-ray contours are overlaid on a Keck II I-band image. Figure from [26]. The *Chandra* temperature map shows a temperature increase between the two subclusters indicating that they have started to merge

can be mapped with unprecedented accuracy. This permits much deeper investigations of the dynamical process of cluster formation.

3.3.1. Shocks and Cold Fronts. We now understand better what happens during a merger event. The *Chandra* observation of the merging cluster 1E0657-56 provides a clear ‘text-book’ example of a shock [38] as shown in Fig. 8 and Fig. 9. The derived Mach number is about $M = 2$ [38]. Evidence of strong shocks is rare, but they might be difficult to detect due to projection effects [6]. More surprisingly, *Chandra* revealed the presence of “Cold fronts” in merging clusters, first discovered in A2142 [22] and A3667 [39]. Cold fronts are contact discontinuities between the cool core of a subcluster moving at near sonic velocity and the surrounding main cluster gas. Across the discontinuity, there is an abrupt jump of the gas density and temperature (Fig. 9). The pressure is approximately continuous (Fig. 9), which shows that the discontinuity is not a shock. 1E0657-56 exhibits both a cold front and a shock, located ahead of the cold front (Fig. 8, 9). The observations of cold fronts show that the cool core of infalling subclusters can survive the passage through the main cluster core, whereas the gas in the outer region of the subcluster is stripped by ram pressure [38]. Evidence of gas stripping during mergers is also provided by the observation of cool trails behind some merging subclusters as, for instance, in the *XMM-Newton* observation of A1644 [40]. Ultimately, the cool core is destroyed: for instance the shape of the subcluster remnant in 1E0657-56 shows that it is being actively destroyed by gas-dynamic instabilities [38].

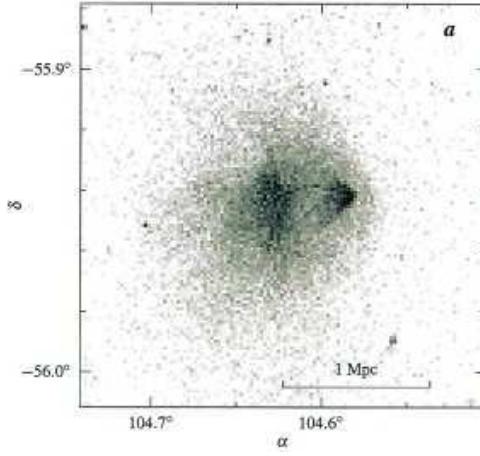


Fig. 8. – *Chandra* image of 1E0657-56. Note the "bullet" apparently just exiting the cluster core and moving westward. The bullet is preceded by an X-ray brightness edge that resembles a bow shock. Figure from [38]

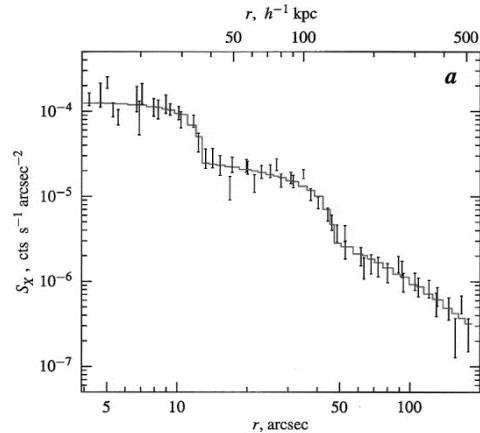


Fig. 9. – Surface brightness profile in a 120° sector centered on the bullet and directed westward (Left) and corresponding temperature and pressure profile (right). The first edge is a cold front while the second is a shock. Figure from [38]

The observation of cold fronts has further implication for the physics at play in clusters. The very steep temperature gradient and smooth surface of the cold fronts imply that thermal conduction and Kelvin-Helmotz instabilities are suppressed, probably by magnetic fields [41, 39, 42].

3.3.2. Formation history and dynamical state. It appears clearly now that the dynamical state of clusters does not always simply depend on the most recent merger event. Multiple on-going merger events are seen in several clusters, for instance in Coma [43] and A2744 [44]. The gas relaxation time can be longer than the interval between successive merger events, especially in dense environments [21]. For instance, the *XMM-Newton* observation of the double cluster A1750 [21], located in a super-cluster, revealed an increase of temperature in the region between the two sub-clusters A1750 C and A1750N

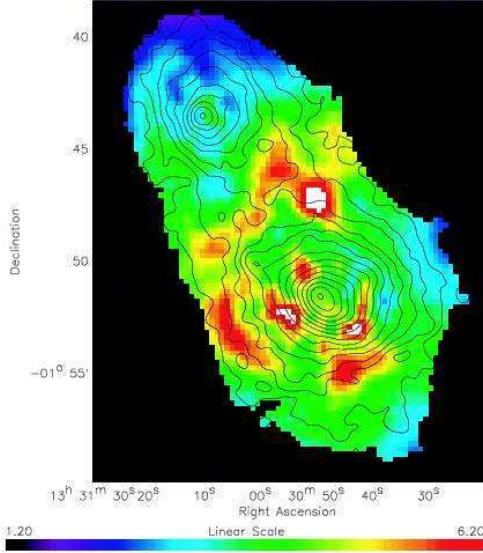


Fig. 10. – The double cluster A1750. The contours of the *XMM-Newton* emission are overlaid on the temperature map. An arc-like hot region can be seen between the two sub-clusters, A1750N in the north and A1750C in the South. The two clusters have started to collide and the gas is being shocked and compressed. The hot regions in A1750 C are regions of shocked gas from an older merger which occurred about 1-2 billion years ago. Figure from [21].

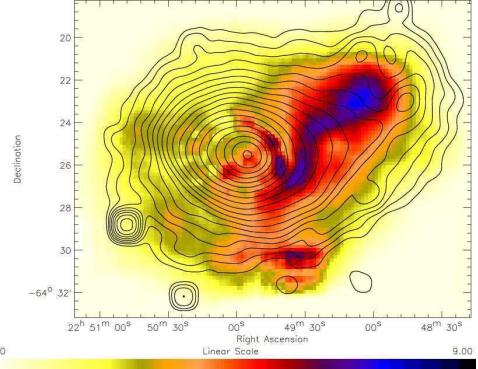


Fig. 11. – The merging cluster A3921. The contours of the *XMM-Newton* emission are overlaid on the temperature map. In contrast to A1750, the hot shocked region is not perpendicular to the line joining the centre of the two subclusters, but nearly parallel to it. This is interpreted in terms of an off-axis merger: the less massive sub-cluster infalling from the SE has already passed the core of the main cluster with a non-null impact parameter. Figure from [48]

(Fig. 10). This indicates that they have started to merge. However, the interaction is too recent to explain the temperature substructures observed in A1750C (Fig. 10). They are likely to result from a more ancient merger.

Various observations confirm the importance of the cluster large-scale environment on its formation history. We have now unambiguous evidence that sub-clusters are accreted preferentially along the direction of filament(s) connecting to the cluster, as expected in the standard scenario of cluster formation. It is provided, for instance, by the observations of A85 [45], A1367 [46] and Coma [47, 43]. The merger of sub-clusters can occur with a non-zero impact parameter, e.g. in A3921 (Fig. 11, [48]) and Coma [43], probably as a result of large-scale tidal torques.

3.4. Statistical studies of cluster morphology. – Beyond the detailed study of specific clusters, statistical analysis of cluster morphology provide important information on structure formation. Till now, such studies have been restricted to X-ray images. They will certainly be extended in the future to temperature maps, now available with

XMM-Newton and *Chandra*. These studies first require to design quantitative measurements of substructures. This is not trivial and many estimators can be used: centre-shifts [49], power-ratios [50], β and Lee statistics [51] The power-ratio method is of particular interest, because power ratios are closely related to the cluster dynamical state [50].

Substructures are common in nearby clusters. A systematic study of substructures in 470 clusters detected in the Rosat All Sky Survey show that $\sim 50\%$ of clusters have significant substructures [51]. Similar numbers are obtained from earlier Einstein observations [49, 29, 52] or *ROSAT* pointed observations [50]. A positive correlation is observed between the presence of substructures and the density of the cluster environment (the number density of surrounding clusters), as expected in the hierarchical scenario [51]. An anti-correlation is observed with the presence of a cooling core [50, 51], suggesting that cooling cores are destroyed during mergers.

The power ratio technique was used to quantify the evolution of morphology with redshift, using a sample of 40 distant clusters observed by *Chandra*. As expected from hierarchical models of structure formation, high-redshift clusters have more substructures and are dynamically more active than low-redshift clusters [53].

3.5. Mergers and non thermal emission. – The ICM is not only filled with hot thermal gas. The presence of a large scale magnetic field and relativistic electrons is revealed by radio observations of diffuse synchrotron emission: regular, centrally located, radio halos and/or irregular, peripherally located, radio relics [54, for a review]. Further evidence is provided by the detection of non thermal hard X-ray emission, interpreted as coming from the inverse Compton scattering of the cosmic microwave background by the relativistic electrons (see, e.g., the observation of A2256 with *Bepo-SAX* [55] and *RXTE* [56]). Note that the signal to noise ratio of such observations is still low and their interpretation ambiguous [56].

It has been known for some time that radio halos are associated with non relaxed clusters, as expected, e.g., if relativistic electrons are (re)-accelerated by merger shocks. For instance, there is a correlation between the dipole power ratio (a measure of the departure from the virialised state) and the power of the radio halo [57]. However, the origin and acceleration mechanism of the relativistic electrons is uncertain [58]. A recent comparison of the radio and *Chandra* temperature maps of merging clusters [59] suggests that radio halo electrons are mostly accelerated by the turbulence induced by mergers, rather than directly by shocks. However, when strongly supersonic shocks are present, they can also contribute to the acceleration [59]. First direct evidence of turbulence in the ICM was obtained recently from the *XMM-Newton* pressure spectrum of Coma [60].

4. – Structural and scaling properties of the cluster population

4.1. The cluster population. – A wide range of morphological and physical properties are observed in the galaxy cluster population. As mentioned above, substructures are common in galaxy clusters, with a rich variety in the type and scale of subclustering.

The total mass of clusters ranges from $10^{13} M_{\odot}$ for groups up to a few $10^{15} M_{\odot}$ for very rich systems, while the X-ray luminosity of the hot intra-cluster medium varies by orders of magnitude, from 10^{41} to a few 10^{45} ergs/s, and the observed temperature from typically 0.3 up to ~ 15 keV.

However, clusters do not occupy the whole parameter space of physical properties. We do see strong correlations between observed physical characteristics of nearby clusters, like luminosity, gas mass, total mass, size and temperature [61, 62, 63, 64]. Furthermore there is also some regularity in shape. If one excludes major merger events or very complex systems ($\sim 20\%$ of systems), the cluster morphology is usually dominated by a regular centrally peaked main component. In that case, the surface brightness, outside the cooling flow region, is reasonably fitted by a β -model, once minor substructures (like small secondary subclusters falling onto a main virialised component) are excised [65].

4.2. The self-similar model of cluster formation. – Regularity in the cluster population is expected on theoretical grounds. The simplest models of structure formation, purely based on gravitation, predict that galaxy clusters constitute a self-similar population. In this section, I briefly summarize the main characteristics of such models. Further enlightening discussions on cluster formation can be found in [66] and in [2] and in other reviews in this book.

As clusters are dark matter dominated objects, their formation and evolution is driven by gravity. The hierarchical collapse of initial density fluctuations of dark matter, which produces the cluster population, is a complex dynamical phenomena. It was first modeled using simple spherical collapse models, while a full treatment of the 3-D hierarchical clustering is now made with N-body simulations. From these theoretical works, key characteristics of the cluster population emerge:

- Major merger events are rare. At a given time, a cluster can be seen as a collapsed halo of dark matter, which is accreting surrounding matter and small groups. The virialized part of the cluster corresponds roughly to a fixed density contrast $\delta \sim 200$ as compared to the critical density of the Universe, $\rho_c(z)$ at the considered redshift:

$$(4) \quad \frac{M_V}{\frac{4\pi}{3} R_V^3} = \delta \rho_c(z)$$

where M_V and R_V are the 'virial' mass and radius. $\rho_c(z) = h^2(z) 3H_0^2 / (8\pi G)$, where $h(z)$ is the Hubble constant normalized to its local value: $h^2(z) = \Omega_m(1+z)^3 + \Omega_\Lambda$, where Ω_m is the cosmological density parameter and Ω_Λ the cosmological constant

- There is a strong similarity in the internal structure of virialised dark matter halos. This reflects the fact that there is no characteristic scale in the problem.

The gas properties directly follow from the dark matter properties, assuming that the gas evolution is purely driven by gravitation, i.e., by the evolution of the potential of the dark matter.

- The gas internal structure is universal, as this is the case for the dark matter (Fig. 12).

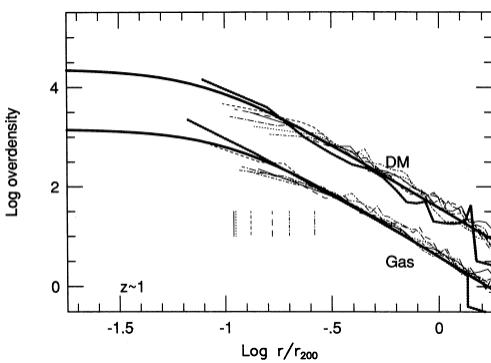


Fig. 12. – Universal cluster density profiles obtained from numerical simulations of structure formation. The radius is scaled to the virial radius, corresponding to a density contrast of 200 as compared to the mean density of the Universe in which the cluster is embedded. The Dark matter and gas densities are normalized to this density. Figure from [67].

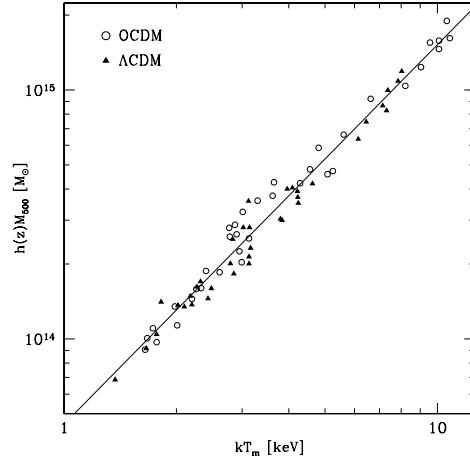


Fig. 13. – Virial mass-temperature relationship for simulated clusters. When the mass is scaled by $h(z)$, it varies as $T^{3/2}$ independent of redshift and cosmology. Figure from [73].

- The gas, within the virial radius, is roughly in hydrostatic equilibrium in the potential of the dark matter. The virial theorem then gives:

$$(5) \quad \frac{G\mu m_p M_V}{2 R_V} = \beta_T kT$$

where T is the gas mean temperature and β_T is a normalization factor, which depends on the cluster internal structure. Since this structure is universal, β_T is a constant, independent on z and cluster mass.

- The gas mass fraction f_{gas} reflects the Universe value, since the gas 'follows' the collapse of the dark matter. It is thus constant:

$$(6) \quad f_{\text{gas}} = \frac{M_{\text{gas}}}{M_V} = cst$$

Therefore, X-ray clusters are expected to exhibit self-similarity:

- Each cluster is defined by two parameters only: its mass and its redshift. Because masses are difficult to measure, it is traditional to rather use the cluster temperature. From the basic equations Eq. 4-6, one can derive a scaling law for each physical property, Q , of the form $Q \propto A(z)T^\alpha$, that relates it to the redshift and temperature. The evolution

factor, $A(z)$, in the scaling relations is due to the evolution of the mean dark matter (and thus gas) density which varies as the critical density of the Universe:

$$(7) \quad \overline{\rho_{\text{gas}}} \propto \overline{\rho_{\text{DM}}} = \delta \rho_c(z) \propto h^2(z)$$

For instance, the total and gas mass scales as $M_{\text{gas}} \propto M_V \propto h^{-1}(z)T^{3/2}$ (Fig. 13), the virial radius as $R_V \propto h^{-1}(z)T^{1/2}$. The X-ray luminosity is $L_X = \int \rho_{\text{gas}}^2 dV \Lambda(T)$, where $\Lambda(T) \propto T^{1/2}$, assuming bremsstrahlung emission. It thus scales as $L_X \propto h(z)T^2$.

- The radial profile of any physical quantity (e.g gas or dark matter density, gas temperature) can be expressed in scaled coordinates. The considered quantity is normalized according to the scaling relation estimated at the cluster temperature and redshift and the radius is expressed in units of the virial radius. Then, the scaled radial profiles are the same for all clusters, whatever their redshift or their temperature (see Fig. 12).

The self-similar model has been validated with numerical simulations of large scale structure formation that include the gas dynamics but only gravitation [67, 68, 69, 70] and by similar semi-analytical hierarchical models [71]. However, there is some ambiguity in the definition of the 'virial' mass of a cluster, or equivalently in the value of the density contrast δ used in Eq. 4. This is fully discussed in [2] (see also [72]). In the spherical top-hat model, where a cluster is considered as a spherical perturbation which has just collapsed, the boundary of a cluster is well defined. It corresponds to $\delta = 18\pi^2 \sim 200$ in a critical density Universe (SCDM model), while δ is a function $\Delta(z, \Omega_m, \Omega_\Lambda)$ of redshift and Ω_m, Ω_Λ in a Λ CDM cosmology [69]. In reality, there is no such thing as a strict boundary between a relaxed part of a cluster and an infall region. Recent numerical simulations [73, 66] indicate that tighter scaling laws are obtained if one uses a common value, $\delta \sim 200$, for all redshift and cosmologies (Fig. 13). Nevertheless, the spherical model definition is often used in the literature. In that case, note that the expected evolution factor of the scaling laws is different: $h^2(z)$ is replaced by $\Delta(z, \Omega_m, \Omega_\Lambda)h^2(z)$ (Eq. 7).

Since the first X-ray observations of clusters, the statistical properties of the observed cluster population have been compared with these theoretical predictions. This has provided valuable insight into the physics that governs the large scale structure formation and evolution of both the Baryonic and the Dark Matter components.

4.3. The Dark matter in local clusters. –

4.3.1. Theoretical predictions. Recent high resolution simulations predict that Cold Dark Matter profiles are cusped in the center [74, 75, 76, 77]. An example is the NFW profile [74] given by

$$(8) \quad \rho_{\text{DM}}(r) = \frac{\rho_c(z)\delta_c}{(cr/R_{200})(1+cr/R_{200})^2}$$

where $\rho_{\text{DM}}(r)$ is the mass density, R_{200} is the radius corresponding to a density contrast of 200 (roughly the virial radius) and c is a concentration parameter. δ_c is the characteristic

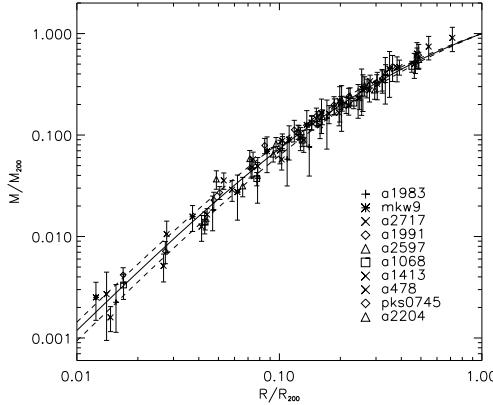


Fig. 14. – Integrated total mass profiles for a sample of clusters in the temperature range 2 – 9 keV measured with *XMM-Newton*. The mass is scaled to M_{200} , and the radius to R_{200} , both values being derived from the best fitting NFW model. The solid line corresponds to the mean scaled NFW profile and the two dashed lines are the associated standard deviation. Figure from [85].

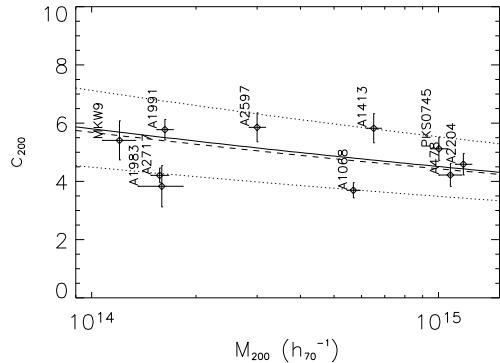


Fig. 15. – Concentration parameter c_{200} versus the cluster mass M_{200} (derived from fitting a NFW model to the data presented Fig. 14). The solid line represents the variation of c_{200} for clusters at $z = 0$ from numerical simulations [78]. The dotted lines are the standard deviation associated with this relation. The dashed line represents the same relation at a redshift of $z = 0.15$ (the maximum redshift for the sample). Figure from [85].

dimensionless density, related to the concentration parameter by $\delta_c = (200/3)c^3/[ln(1 + c) - c/(1 + c)]$. If c was a constant, the profiles of all clusters would be perfectly self-similar. Actually, it is expected to vary slightly with z and system mass [74, 78]. The corresponding integrated mass profile is of the form: $M(r) = M_{200}m(cr/R_{200})/m(c)$, where M_{200} is the mass enclosed within R_{200} (the 'virial mass'). This mass profile can be directly compared with observations.

The NFW density profile varies from $\rho_{\text{DM}} \propto r^{-1}$ at small radii to $\rho_{\text{DM}} \propto r^{-3}$ at large radii. The exact slope of the dark matter profile in the center is still debated (see [77, 76] for latest results). For instance, the simulations in [75] give a steeper slope $\rho_{\text{DM}} \propto r^{-1.5}$.

4.3.2. Observed mass profiles. With *Chandra* and *XMM-Newton* we can now measure precisely the total mass distribution in clusters (from Eq. 3) over a wide range of radius, from $\sim 0.001 R_{200}$ [18] up to $\sim 0.7 R_{200}$ [79]. The validity of the X-ray method to derive masses from the hydrostatic equilibrium equation has been validated for relaxed clusters, by comparing *Chandra* mass estimates with independent lensing mass estimate [81, and reference therein]. The observed mass profiles are well fitted using a NFW model. This is true not only for massive clusters [80, 81, 82, 79, 18, 83, 84], but also for low mass systems [7, 19]. A profile with a flat core is generally rejected with a high confidence level.

Relation	Slope	Reference	Note
L_X-T	2.64 ± 0.27	[91]	a
	2.88 ± 0.15	[61]	b
$M_{\text{gas}}-T$	1.98 ± 0.18	[62]	c
	1.71 ± 0.13	[92]	
	1.89 ± 0.20	[93]	c
	1.80 ± 0.16	[94]	c
$EM-T$	1.38	[95]	
$S-T$	0.65 ± 0.05	[108]	d
$f_{\text{gas}}-T$	0.34 ± 0.22	[62]	c
	0.66 ± 0.34	[93]	e

TABLE I. – *Logarithmic slope of local scaling laws from the literature (not exhaustive). In the standard self-similar model: $L_X \propto T^2$, $M_{\text{gas}} \propto T^{1.5}$, $EM \propto T^{0.5}$ and $S \propto T$. Notes: a: Corrected for cooling flow contribution; b: non cooling-flow clusters; c: Gas mass integrated within R_{500} ; d: entropy estimated at $0.1R_{200}$; e: f_{gas} integrated within R_{1000} .*

In a few cases [18, 83, 84], the inner slope has even been measured precisely enough⁽³⁾ to distinguish between various predictions. They favor a slope of $\alpha \sim -1$. This steep density profile in the center of clusters is incompatible with the flattened core DM profiles predicted by self-interacting Dark Matter models [82, 83]. In one case, the validity of the NFW model was tested up to the virial radius [79].

Recently, the first quantitative check of the universality of the mass profile was performed with *XMM-Newton* [19, 85]. As shown Fig. 14, the mass profiles scaled in units of R_{200} and M_{200} nearly coincide, with a dispersion of less than 15% at $0.1 R_{200}$. Furthermore, the shape is quantitatively consistent with the predictions. The derived concentration parameters are consistent with the $c-M_{200}$ relation derived from numerical simulations for a Λ CDM cosmology (Fig. 15).

This excellent *quantitative* agreement with theoretical predictions provides strong evidence in favor of the Cold Dark Matter cosmological scenario, and shows that the physics of the Dark Matter collapse is well understood, at least down to the cluster scale.

Finally, it is worth mentioning again the observation of 1E0657-56, which yields very interesting, independent constraints on the Dark Matter. A comparison of the gas, galaxies and weak-lensing maps demonstrates the presence and dominance of non-baryonic Dark Matter in this cluster and shows that the cross-section of the dark matter particles is low, excluding again most of the self-interacting dark matter models [86, 87].

4.4. Gas properties in local clusters. – It has been known for nearly 20 years that gas properties deviate from the standard self-similar model predictions: the L_X-T relation is steeper than expected. This was the first indication that the gas physics should be looked

⁽³⁾ Note that such measurements are not limited by the instrument capabilities but by the number of suitable targets. In most clusters the very center is disturbed (see below), invalidating the HE approach to compute the central mass profile.

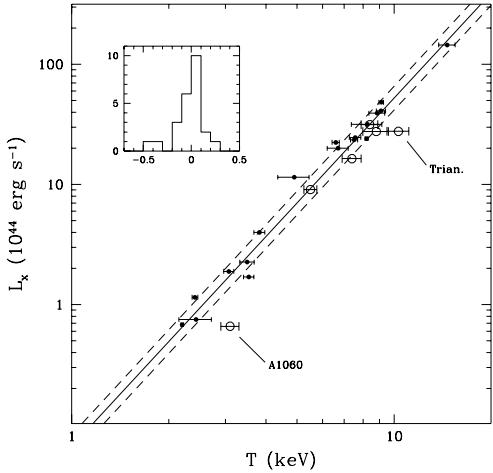


Fig. 16. – The relation between the bolometric luminosity and the temperature. Only clusters without strong cooling in the center are considered. The data are mostly from *GINGA* observations. The relation is steeper than expected in the standard self-similar model: $L_X \propto T^{2.88 \pm 0.15}$. Figure from [61].

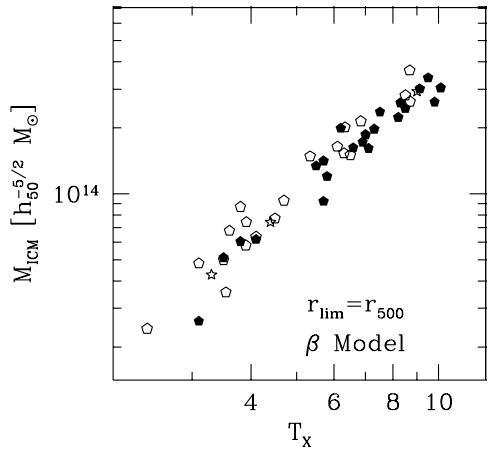


Fig. 17. – The relation between the gas mass and the temperature. The gas mass was measured with *ROSAT* and estimated within R_{500} . The temperature was measured with the *ASCA* or earlier missions. The relation is steeper than expected in the standard self-similar model: $M_{\text{gas}} \propto T^{1.98 \pm 0.18}$. Figure from [62].

at more closely and that non-gravitational processes could play a role (e.g. [88]). We have now a much more precise view of cluster properties. The new picture that emerged from recent observations is that local clusters, down to remarkably low mass ($kT \sim 2\text{keV}$), do obey self-similarity. However, the scaling laws differ from simple expectations. In this part, I summarize these observations, focusing on clusters above a typical temperature of 2 keV. Below that temperature, one typically finds groups. The latest results indicate that their properties may follow the trends observed in clusters, albeit with a significantly increased dispersion [89]. For a recent review on group properties, the reader may refer to [90].

4.4.1. Local scaling laws. Scaling laws relate various physical properties with T . This includes global gas properties like the X-ray luminosity L_X [91, 61], the gas mass M_{gas} [62, 92, 93, 94] or the gas mass fraction f_{gas} [62, 93]. The corresponding scaling relations are steeper than expected (Fig. 16, 17). However, an unambiguous picture can be obtained only by studying the internal structure of clusters. For instance, a steepening of the L_X-T relation could be due to a systematic increase of the mean gas density with T (a simple modification of the scaling laws) or to a variation of cluster shape with T (a break of self-similarity). By looking at radially averaged profiles of interesting quantities, particularly the density and the entropy, two issues can be addressed at the same time. The first is to see whether the profiles agree in *shape*; the second is to investigate the

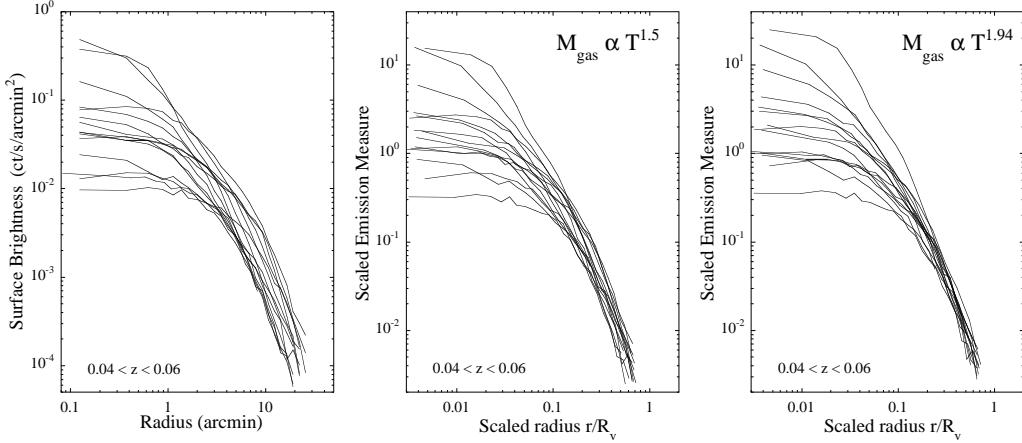


Fig. 18. – Left: Surface brightness profile of 15 nearby ($0.04 < z < 0.06$) and hot clusters ($kT > 3.5$ keV) observed with *ROSAT*. Middle: Scaled emission measure profiles. The radius is normalized to the virial radius and EM is scaled according to the classical laws ($EM \propto T^{1/2}$ or equivalently $M_{\text{gas}} \propto T^{3/2}$). Right: Same assuming $M_{\text{gas}} \propto T^{1.94}$, or $EM \propto T^{1.38}$. Figure adapted from [65, 95].

scaling of the profiles. The various scaling relations are summarized in Table I.

4.4.2. Gas density profiles. The gas content and density distribution can be studied through the emission measure along the line of sight $EM(r) = \int_r^{R_{200}} n_e^2 dl$, which is easily derived from the X-ray surface brightness profile (Sec. 2.2, Fig. 18). The scaled EM profiles of *hot* clusters measured with *ROSAT* were found to be similar in shape outside typically $0.1 - 0.2 R_{200}$ [65, 92, 95]. There is a large dispersion in the central region, linked with the presence of a cooling core (see Sec. 4.4.5). Outside that region the universal profile is well fitted by a β -model, with $\beta = 2/3$ and $R_c \sim 0.12 R_{200}$ [127]. Note however that more precise *XMM-Newton* data on hot clusters gives a steeper density profile in the center, even for weak cooling flow clusters [79]. This is a consequence of the cusped nature of the dark matter profile.

In the standard self-similar framework, the mean gas density does not depend on the temperature (Eq. 7). The emission measure is thus expected to scale as R_{200} , i.e $EM \propto T^{0.5}$. The scatter in the scaled EM profiles was found to be considerably reduced if a much steeper $EM-T$ relation, $EM \propto T^{1.38}$, was used to scale the profiles [95, see Fig. 18]. This explains the steepening of the L_X-T relation and translates into $M_{\text{gas}} \propto T^{1.94}$, which is consistent with the observed steepening of the $M_{\text{gas}}-T$ relation (Table I) as compared to the standard $M_{\text{gas}} \propto T^{1.5}$ scaling. A test case study with *XMM-Newton* of a cool cluster (A1983) suggests that clusters follow the universal scaled EM profile down to temperature as low as $kT \sim 2$ keV [7].

There is converging evidence that the gas distribution in clusters is more inflated than the dark matter distribution. The integrated gas mass fraction increases with scaled

radius [96]. The increase, measured with *Beppo-SAX* is about 25% between a density contrast of $\delta = 2500$ (about 1/3 of the virial radius) and $\delta = 500$ [97]. The results at $\delta = 500$ were largely based on extrapolations and could be biased. However, recent *XMM-Newton* and *Chandra* results seem to confirm this trend. From a sample of six massive clusters observed by *Chandra*, an average f_{gas} value of 0.113 ± 0.013 was derived at $\delta = 2500$ [98]. At this density contrast, the value derived from the *XMM-Newton* analysis of A1413 is $f_{\text{gas}} = 0.11$, perfectly consistent with Chandra results [79]. The *XMM-Newton* data extends up to $\delta = 500$ and show that f_{gas} keeps increasing with radius to reach $f_{\text{gas}} \sim 0.14$ at $\delta = 500$ (25% increase).

4.4.3. Temperature profiles. There is also a similarity in the temperature profiles of hot clusters observed with *ASCA* and *Beppo-SAX* beyond the cooling core region [99, 100, 101]. In relaxed clusters, there is usually a drop of temperature towards the center ($r \lesssim 0.1R_{200}$). This corresponds to the cooling core region (see Sec. 4.4.5). There is also a tendency for clusters with cooling cores to have flatter temperature profiles at large scale than non cooling core clusters, suggesting that the profile shape depends on the cluster dynamical state [101].

A *XMM-Newton* study of an unbiased sample of clusters shows a variety of shapes, probably linked to various dynamical states [102]. The self-similarity of shape seem to be confirmed by *Chandra* [81, 103] and *XMM-Newton* data [104, 105] for relaxed clusters. However, no consensus has been reached yet on the exact shape of the profiles. This was already the case for *ASCA* and *Beppo-SAX* studies and this is still the case with *XMM-Newton* and *Chandra*. Some studies find relatively flat profiles (within $\sim 20\%$), beyond the cooling core region, up to $0.3R_{200}$ [81] or even to $0.5R_{200}$ [104]. Other studies found steadily decreasing profiles, by 30% between $0.1R_{200}$ and $0.5R_{200}$ [105] or even by 50% [103].

4.4.4. Gas entropy. The studies of the gas density profiles suggest that the departures of the gas scaling laws from the standard self-similar model are due to a non-standard scaling of the mean gas density with temperature and not to a break of self-similarity. To understand the physical origin of these deviations, one must consider the gas entropy rather than the density. The ‘entropy’ is traditionally defined as $S = kT/n_e^{2/3}$ and is related to the true thermodynamic entropy via a logarithm and an additive constant. It is a fundamental characteristic of the ICM, because it is a probe of the thermodynamic history of the gas [106, 2]. The entropy profile of the gas and the shape of the potential well, in which it lies, fully define the X-ray properties of a relaxed cluster ⁽⁴⁾. In the standard self-similar picture, the entropy, at any scaled radius, should scale simply as $S \propto h(z)^{4/3}T$.

Since the pioneering work of [107], it is known that the entropy measured at $0.1R_{200}$ exceeds the value attainable through gravitational heating alone, an effect that is es-

⁽⁴⁾ The gas density and temperature profiles can be determined from the entropy profile and the hydrostatic equation (Eq. 3) with a boundary condition at the virial radius

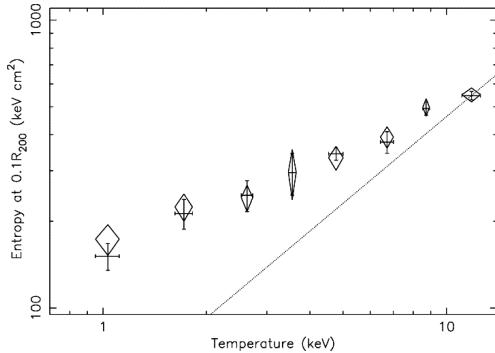


Fig. 19. – Gas entropy at $0.1 R_{200}$ versus cluster temperature. Data of individual clusters, derived from *ROSAT* and *ASCA* observations, have been grouped by temperature bins. The dotted line shows the standard self-similar prediction, $S \propto T$. The relation is shallower than expected in the standard self-similar model: $S \propto T^{0.65 \pm 0.05}$. Figure from [108].

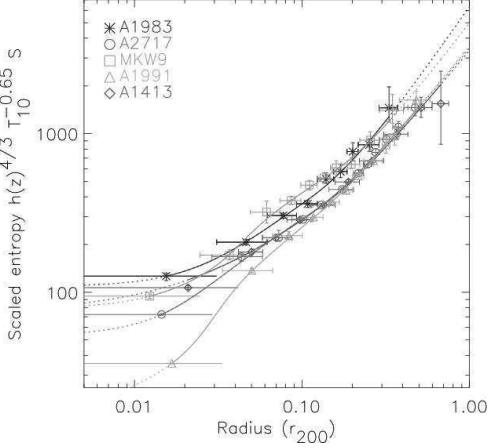


Fig. 20. – *XMM-Newton* scaled entropy profiles. The temperature ranges from 2 keV to 6.5 keV. The entropy has been scaled following the empirical relation, $S \propto h(z)^{-4/3} T^{0.65}$. The radius is scaled to the measured virial radius, R_{200} . Figure from [19].

pecially noticeable in low mass systems. Various non-gravitational processes have been proposed to explain this entropy excess, such as heating before or after collapse (from SNs or AGNs) or radiative cooling. A recent study of 66 nearby systems observed by *ASCA* and *ROSAT* [108] shows that the $S-T$ relation follows a power law but with a smaller slope than expected: the entropy measured at $0.1R_{200}$ scales as $S \propto T^{0.65}$ (see Fig. 19). High quality *XMM-Newton* observations [7, 19, see Fig. 20] show a remarkable self-similarity in the shape of the entropy profiles down to low mass ($kT \sim 2$ keV). Stacking analysis of *ROSAT* data gives the same results [108]. Except in the very centre, the *XMM-Newton* entropy profiles are self-similar in shape, with close to power law behavior in the $0.05 R_{200} < r < 0.5 R_{200}$ range. The slope is slightly shallower than predicted by shock heating models, $S(r) \propto r^{0.94 \pm 0.14}$. The normalization of the profiles is consistent with the $S \propto T^{0.65}$ scaling. Similar results were obtained more recently on a larger sample observed with *XMM-Newton* [105].

The self-similarity of shape of the entropy profile is a strong constraint for models. Simple pre-heating models, which predict large isentropic cores, must be ruled out. In addition to the gravitational effect, the gas history probably depends on the interplay between cooling and various galaxy feedback mechanisms (see [2] and other reviews in this book).

4.4.5. The complex physics in cluster core. As mentioned above, there is a very large dispersion in the core properties, within typically $0.1R_{200}$. This is linked to the complex

physics at play in the cluster center. In the center of clusters the gas density is high. The cooling time, which scales as $t_{\text{cool}} \propto T^{1/2}/n_e$ can be shorter than the 'age' of the cluster, the time since the last major merger event. We thus expect the temperature to decrease due to radiative cooling and the density to increase so that the gas stays in quasi hydrostatic equilibrium. Both *Chandra* [81] and *XMM-Newton* [109] observations clearly confirm this temperature drop in cooling clusters. These new observations have also dramatically changed our vision of cooling cores in clusters. This topic has been the subject of a recent conference [110] and is only briefly discussed here.

A major surprise is the lack of very cool gas, inconsistent with standard isobaric cooling flow models. The *XMM-Newton/RGS* high resolution spectra exhibit strong emission from cool plasma at just below the ambient temperature, T , down to $T/2$, but also exhibit a severe deficit of emission compared to the predictions at lower temperatures [111]. The standard model is also inconsistent with *XMM-Newton/EPIC* data [112, 113, 109]. In parallel, arc-second spatial imaging with *Chandra* have revealed complex interaction between AGN activity in the cluster center and the intra-cluster medium [114, for a review]. One observes X-ray cavities or 'bubbles', created by the central AGN radio lobes as they displace the X-ray gas. They are usually surrounded by cool rims and not by shocked gas. However, a shock was recently discovered at the boundary of the cavities in MS0735.6+7421 [115]. There are also 'ghost' cavities, probably bubbles that have buoyantly risen away from the cluster center. Spectacular examples of such phenomena are observed in the center of the Perseus cluster [116, 117]. In this cluster ripples are also observed in the X-ray surface brightness of the gas surrounding the central bubble. This was interpreted as resulting from the propagation of weak shocks and viscously dissipating sound waves due to repeated outbursts of the central AGN [117]. Cold fronts and sloshing gas are also observed with *Chandra* in the center of relaxed clusters [118, for a review].

Whether and how both phenomena, the absence of very cool gas and AGN/ICM interaction, are connected is still unclear [2, 119, 120, for reviews]. For instance, AGN heating may limit cooling but conduction could also play a role in heating the central region. A better understanding of cooling and AGN heating in the central part of clusters has further implications because both phenomena play a role at larger scales in clusters and during galaxy formation. Finally, the core properties have also a substantial impact on the L_X-T relation. When a cooling core is present, the luminosity is boosted (due to the peaked density profiles) and the mean temperature is decreased (due to the temperature drop in the center). This results in a large dispersion in the L_X-T relation. The dispersion is decreased when these effects are corrected for [91]. Then the L_X-T relation is the same as for clusters without strongly cooling cores [61].

4.4.6. The $M-T$ relation. The $M-T$ relation is a fundamental scaling relation. Since other scaling relations are expressed in terms of the temperature T , the $M-T$ relation provides the missing link between the gas properties and the mass. Furthermore, estimations of the cosmological parameters from cluster abundances in clusters, or from the spatial distribution of clusters, heavily rely on this relation to relate the mass to

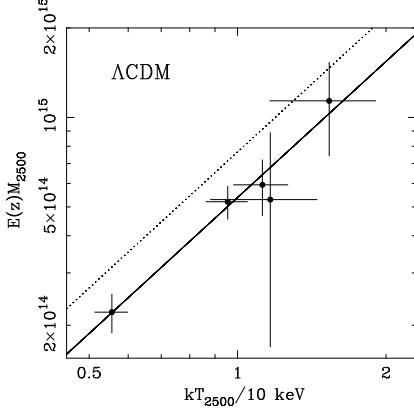


Fig. 21. – The M_{2500} - T relation observed with *Chandra* for hot clusters. Solid line: The best fit power law, $M_{2500} \propto T_{2500}^{1.51 \pm 0.27}$. Dashed line: the predicted relation from adiabatic numerical simulations [68]. Figure from [81].

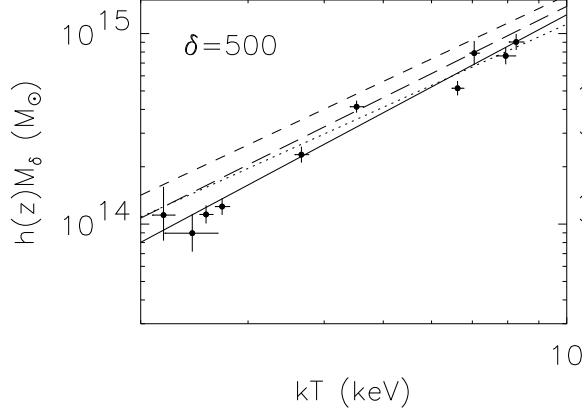


Fig. 22. – The M_{500} - T relation observed with *XMM-Newton*. Solid line: The best fit power law for the whole sample. Dotted line: same for the hot cluster subsample. Dashed line: the predicted relation from adiabatic numerical simulations [68]. Long-dashed line: the relation derived from a numerical simulation including radiative cooling, star formation and SN feedback [125]. Figure from [124].

the observables available from X-ray surveys (see Sec.5.3). A sustained observational effort to measure the local M - T relation has been undertaken using *ROSAT*, *ASCA* and *Beppo-SAX*, but no definitive picture had emerged. Does the mass scale as $T^{3/2}$ as expected [121, 93, 94]? Is this true only in the high mass range ($kT \gtrsim 4$ keV), with a steepening at lower mass [122, 123, 63]? Is the slope higher than expected over the entire mass range [96]? This was unclear. The normalizations of the M - T relation derived from *ASCA* data are generally lower than predicted by adiabatic numerical simulations [68], by typically 40% [122, 63]. On the other hand, using *Beppo-SAX* data, a normalisation consistent with the predictions was obtained, albeit with large error [93].

These studies had to rely largely on extrapolation to deduce the virial mass, and they were limited by the low resolution and the statistical quality of the temperature profiles. As shown above, significant progress on mass estimates has been made with *Chandra* and *XMM-Newton*. A first *Chandra* study [81] of five hot clusters ($kT > 5.5$ keV) derived a M - T relation slope of 1.51 ± 0.27 , consistent with the self-similar model (see Fig. 21). It confirmed the offset in normalisation. However, due to the relatively small *Chandra* field of view, the M - T relation was established at R_{2500} , i.e., about $\sim 0.3R_{200}$. More recently, the M - T relation was established down to lower density contrasts ($\delta = 2500$ to $\delta = 200$) from a sample of ten nearby relaxed galaxy clusters covering a wider temperature range, [2 – 9] keV [124]. The masses were derived from precise mass profiles measured with *XMM-Newton* at least down to $\delta = 1000$ and extrapolated beyond that radius using the best fitting NFW model. The M_{2500} - T for hot clusters is perfectly consistent with the

Chandra results. The logarithmic slope of the $M-T$ relation was well constrained. It is the same at all δ , reflecting the self-similarity of the mass profiles. At $\delta = 500$ (see Fig. 22), the slope of the relation for the sub-sample of hot clusters ($kT > 3.5$ keV) is consistent with the standard self-similar expectation: $\alpha = 1.49 \pm 0.15$. The relation steepens when the whole sample is considered, but the effect is fairly small: $\alpha = 1.71 \pm 0.09$. The normalisation of the relation differs, at all density contrasts from the prediction of gravitation based models (by $\sim 30\%$). Models that take into account radiative cooling and galaxy feedback [125, 126] are generally in better agreement with the data [124, for full discussion].

4.5. Evolution of cluster properties. – The standard self-similar model makes strong predictions for the evolution of cluster properties. Distant clusters should have the same internal structure as nearby clusters, but they should be denser, smaller and more luminous (Sec. 4.2). Physical properties derived from X-ray observations, as well as theoretical predictions, depend on the assumed cosmology. Fortunately, we have now concordant constraints (Sec. 5) on cosmological parameters, showing that we live in a flat low density Universe ($\Omega_m \sim 0.3, \Omega_\Lambda \sim 0.7$). This greatly simplifies the issue of cluster formation and evolution.

The *ROSAT* and *ASCA* observations gave the first indication that the self-similarity does hold at $z > 0$. The universal emission measure profile appears to extend to $z \sim 0.8$, with a redshift scaling consistent with the expectation for a Λ CDM cosmology [127]. A significant evolution in the normalisation of the L_X-T relation was obtained, consistent with the self-similar model. This was also the case in other studies made assuming this cosmology⁽⁵⁾ [128, 129], although more recently no significant evolution was detected in a large sample of 79 clusters [130]. However, all these observations were highly biased towards massive systems, mostly clusters discovered by the EMSS, and their statistical quality was poor.

With *XMM-Newton* and *Chandra*, we can now make high quality studies of distant clusters from X-ray samples assembled using *ROSAT* observations (Sec. 5.3.1). These new samples cover a much wider mass range than the EMSS survey. Recent *XMM-Newton* and *Chandra* evolution studies do confirm that clusters follow scaling laws up to high z [131, 132, 25]. The self-similarity of shape was also confirmed on a test case at $z = 0.6$ [23]. However, the amount of evolution remains uncertain. For instance, depending on the data considered and the analysis procedure, the normalization the L_X-T relation has been found to evolve more than expected in the standard model, as expected or less than expected. The standard self-similar model predicts that the normalization $A(z)$ varies as $h(z)$, $A(z) \equiv h(z) \sim (1+z)^{0.6-0.9}$ for the favored Λ CDM cosmology. The first studies comparing respectively *Chandra* [131] and *XMM-Newton* [132] data with the local relation measured with *ASCA* [91] gave similar evolution fac-

⁽⁵⁾ Previous studies assumed a SCDM cosmology and found no evolution of the L_X-T relation. As discussed in [127], this is an artifact due to the choice of a ‘wrong’ cosmology: the distance and therefore the luminosity are underestimated

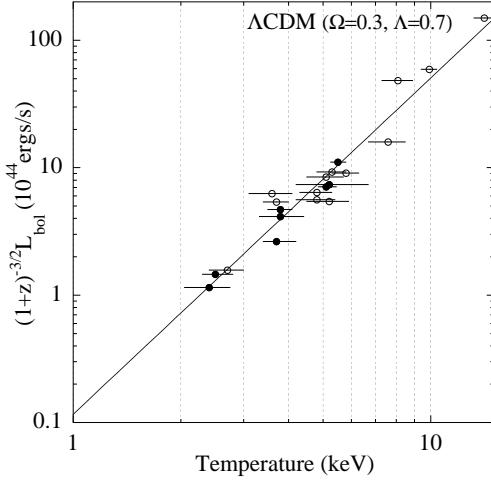


Fig. 23. – Correlation between bolometric luminosity of distant clusters and temperature. Filled circles: *XMM-Newton* data for SHARC clusters [132]; Open circles: *Chandra* data [131]. When the luminosity is scaled by $(1+z)^{3/2}$, the data points are consistent with the local L_X - T relation (solid line), indicating evolution. Figure adapted from [132, 131]

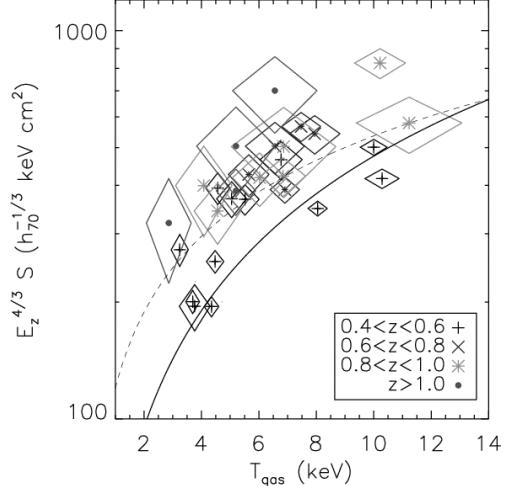


Fig. 24. – Entropy measured at $0.1R_{200}$ versus temperature for a sample of distant clusters observed with *Chandra*. Data for clusters in various z bins are indicated by different symbols. The entropy is rescaled according to the expected evolution. Dashed line: local S - T relation as measured by [108]. Figure from [25].

tors: $A(z) \equiv (1+z)^{1.52 \pm 0.26}$ and $A(z) \equiv (1+z)^{1.5 \pm 0.30}$ respectively. This is larger than expected. A more recent study [25], using a larger set of *Chandra* data and the same local reference but a different procedure, gives $A(z) \equiv (1+z)^{0.62 \pm 0.28}$, perfectly consistent with the expectation. In contrast, a smaller evolution than expected was obtained when using the local relation measured with *Beppo-SAX*: $A(z) \equiv h(z)(1+z)^{-1.04 \pm 0.32}$.

The evolution of the other scaling relations are still poorly constrained. The evolution of the M - T relation is consistent with the expectation [25]. There is some hints that the evolution of the M_{gas} - T relation is smaller than expected [131, 25], while the S - T relation would be higher than expected [25, see Fig. 24]. However, the effects are not very significant.

This illustrates the difficulty in studying cluster evolution. High precision is required because the evolution expected in the ‘reference’ standard model is small. The normalization of the key L_X - T , M - T and S - T relations evolves as $h(z)$, $h(z)^{-1}$ and $h(z)^{-4/3}$ respectively, where $h(z)$ is the Hubble constant. The evolution factor, $h(z)$, varies by 30% between $z = 0$ and $z = 0.5$ (56% at $z = 0.8$). To distinguish between various models, statistical and systematic errors have to be well below these figures. This requires the use of large unbiased cluster samples, covering a wide mass range, in both the local Universe and at high redshift. Biased results on the evolution could be obtained, for instance, by comparing the L_X - T relations at various z with non representative sub-samples of cooling

flow clusters (or by using not fully consistent treatment of the cooling center). Ideally, one also wants to compare data obtained with the same instrument, in order to minimize systematic uncertainties. For instance, a calibration error of 10% in the kT measurements is equivalent to a 30% systematic error on the luminosity ($L_{\text{bol}} \propto T^3$). Such an error could introduce a bias equivalent to the expected evolution. Ongoing *XMM-Newton* and *Chandra* large projects, aiming at studying the properties of large unbiased distant and local samples, will fulfill the above requirements. Significant progress on the evolution of cluster properties are expected from these projects.

5. – Constraining cosmological parameters with X-ray observations of clusters

Several independent methods can be used to constrain the cosmological parameters from cluster X-ray observations. This includes the baryon fraction in clusters, the cluster abundance and its evolution, and the cluster spatial distribution. All methods rely in principle or in practice on specific assumptions on the scaling and structural properties of clusters.

5.1. The baryon fraction. – In the simplest model of cluster formation, the contents of clusters is a fair sample of the Universe as a whole (Sec. 4'2). The baryon mass fraction in clusters is then $f_b = \Omega_b/\Omega_m$, where Ω_b and Ω_m are the mean baryon density and the total matter density of the Universe. f_b is the sum of the gas and galaxy mass fractions: $f_b = f_{\text{gas}} + f_{\text{gal}}$. Combined with the Ω_b value estimated from Big Bang nucleosynthesis or CMB measurements, f_b in clusters can be used to measure Ω_m [133]. The method requires an independent knowledge of the Hubble constant h , which enters in the determination of Ω_b ($\Omega_b \propto h^{-2}$), of f_{gas} ($f_{\text{gas}} \propto h^{-3/2}$) and of f_{gal} ($f_{\text{gal}} \propto h^{-1}$). Note that the baryonic mass in rich clusters is dominated by the X-ray gas. The gas mass fraction alone sets an upper limit on Ω_m .

Cluster sample studies that rely on measurements of both f_{gas} and f_{gal} are rare [134]. Ω_m is most often constrained from f_{gas} only, assuming a constant $f_{\text{gal}}/f_{\text{gas}}$ ratio, taken from other cluster studies [97, 135]. A further difficulty is that f_{gas} increases with the integration radius (Sec. 4'4.2) and numerical simulations indicate that f_{gas} within the virial radius is slightly smaller than the Universe's value [70]. Observed values must be corrected for these effects. The correction is about 20% for f_b values estimated within 1/3 of the virial radius [135]. Corrections factors are deduced from adiabatic numerical simulations [134, 135] and may not be perfectly adequate since these simulations do not include galaxy formation and fail to account for the observed ICM properties (Sec. 4'4). One also observes a significant increase of f_b with system mass, mainly resulting from the increase of f_{gas} [62, 134]. This increase of f_{gas} is likely due to non-gravitational processes, less important in high mass systems (Sec. 4'4). Therefore, most studies are restricted to massive clusters to minimize systematic errors. To firmly establish which cluster populations are fair samples of the Universe, the variation of f_{gas} with system mass must be fully understood, which remains to be done.

Present f_b data provides a tight constraint on Ω_m . All recent studies favor a low

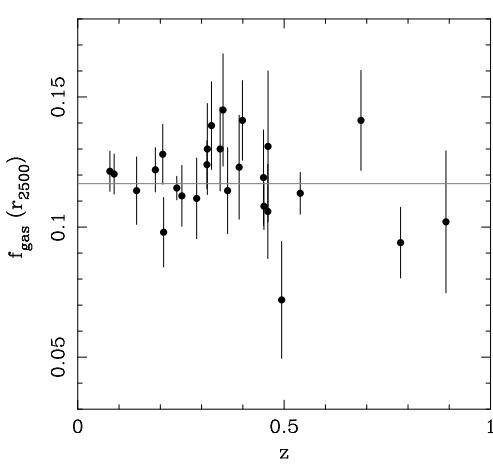


Fig. 25. – The X-ray gas mass fraction (1σ errors) versus z for a Λ CDM cosmology with $\Omega_m = 0.25$, $\Omega_\Lambda = 0.96$. In this 'best fit' cosmology f_{gas} is constant. Figure from [135].

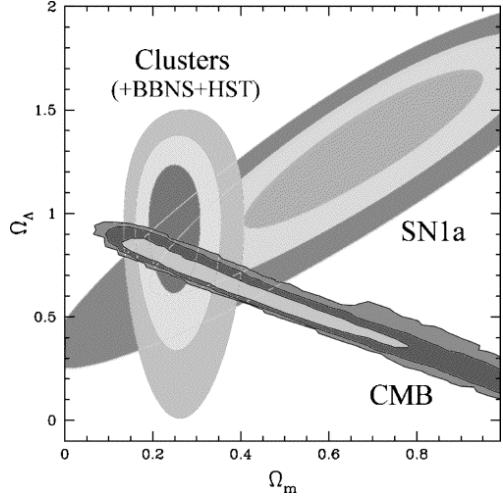


Fig. 26. – The 1, 2 and 3σ confidence constraints in the Ω_m , Ω_Λ plane obtained from the analysis of the cluster f_{gas} data. Also shown are the independent results obtained from CMB and SNI data. Figure from [135]

Ω_m Universe and are in excellent agreement, e.g. $\Omega_m = 0.37 \pm 0.08$ from *Beppo-SAX* data [97], $\Omega_m = 0.28 \pm 0.03$ from *ROSAT/ASCA* data [134] and $\Omega_m = 0.30 \pm 0.04$ from *Chandra* data [98]. The variation of f_b with mass and radius, mentioned above, are the main source of systematic uncertainties for high precision cosmology.

5.2. The gas fraction as distance indicator. – Assuming again that f_{gas} is universal and thus constant with z , $f_{\text{gas}}(z)$ can be used as distance indicator [136]. The values derived from observations depend on the angular distance as $f_{\text{gas}} \propto (D_A(z))^{3/2}$. It will be constant only for the correct underlying cosmology (Fig. 25) ⁽⁶⁾. Because the variation of D_A with redshift is controlled by the expansion rate of the Universe, $H(z)$, $f_{\text{gas}}(z)$ thus provides constraints on Ω_m and Ω_Λ for a Λ CDM cosmology, (or, more generally, on the Dark Energy present density and equation of state). Note, however, that there is a large degeneracy between the constraints on Ω_m and Ω_Λ .

Because it is a pure geometrical test, the method is straightforward. However, it requires high quality data (precise mass estimates) on a sample of clusters at various redshifts and it is not free of possible systematical errors. f_{gas} varies with radius and mass at $z = 0$. Therefore, to compare the gas fraction at various z , one estimates f_{gas} at the same fraction of the virial radius and for clusters of similar mass (preferentially

⁽⁶⁾ More generally, non evolving cluster properties can be used as distance indicators. This includes the gas mass fraction, but also the isophotal radius [137], or properties corrected for evolution, like the scaled emission measure profiles [127]

massive clusters) [97, 98]. This is sufficient to avoid biases only if cluster mass profiles are similar at all z and if f_{gas} at a given mass does not evolve with z . These assumptions must be verified from detailed observations at high z (in particular observations of the structural properties of clusters), compared with numerical simulations.

The principle of this test is similar to that of the cosmological test using SNI as 'standard' candles [138, 139, 140, and references therein]. Therefore, it probes the same domain of cosmological parameters as SNI experiments. However, there are also systematic uncertainties inherent to the use of SNI (dust extinction, possible evolution of the luminosity). Much safer constraints can thus be obtained by cross-checking the results obtained with SNI and with the gas fraction. The constraints provided by $f_{\text{gas}}(z)$ (and SNI) complement the observations of the CMB anisotropies [141, for WMAP results], which do not probe the same domain of cosmological parameters.

The recent results obtained from *Chandra* observations are very encouraging [135] (see also [98, 97]). The whole information provided by f_{gas} was used: absolute value and apparent evolution with z . For a Λ CDM cosmology, this yields: $\Omega_m = 0.245 \pm 0.04$ and $\Omega_\Lambda = 0.96 \pm 0.2$ (68% confidence level), with $\Omega_b h^2 = 0.0214 \pm 0.02$ and $h = 0.72 \pm 0.08$ as only priors (Fig. 26). Ω_Λ is positive at the 3σ level. The complementarity with CMB observations is obvious in Fig. 26. Note also the complementarity of f_{gas} and SNI measurements: the $(\Omega_m, \Omega_\Lambda)$ degeneracy intrinsic to pure distance measurement techniques (SNI or $f_{\text{gas}}(z)$ observations) has been broken because the local f_{gas} value provides independent constraints on Ω_m . On the other hand, the constraints from $f_{\text{gas}}(z)$ on Ω_Λ at a given Ω_m value are still less stringent than those provided by SNIs (Fig. 26). This can be improved using larger samples or more precise estimates of f_{gas} . Finally the consistency between f_{gas} , SNI and CMB constraints is reassuring: their respective confidence contours in the Ω_m, Ω_Λ plane intersect in a common domain.

5.3. Cosmological parameters from cluster abundance and evolution. – An independent and powerful method to constrain the cosmological parameters is to measure the growth rate of linear density perturbations, as reflected in the evolution of the galaxy cluster mass distribution function, $n(M, z)$, i.e., the comoving number density of clusters of mass M at redshift z . This cosmological test is not as direct as the previous one, since the mass function also depends on the spectrum of the initial density fluctuations.

From large N-body simulations [142] we can now accurately trace the formation of dark matter halos. Such simulations have been used to establish universal parametric formulae for the cluster mass function in various cosmological models [143, 144], improving over the original formula provided by [145]. The mass function depends mostly on Ω_m and σ_8 (i.e. the normalisation of the fluctuation power spectrum $P(k)$), but also on h , Ω_b , Ω_Λ and n (the shape of $P(k)$). The evolution of the mass function, specially of the number of high mass clusters, is extremely sensitive to Ω_m [146, 147], as illustrated Fig. 27 (see also Fig. 6 in [148]). It was first proposed in [149] to use the evolution of X-ray cluster abundance to constrain Ω_m .

In practice, this method requires to build large, well controlled, samples of local and distant clusters. Furthermore, the mass can never be directly determined and other(s)

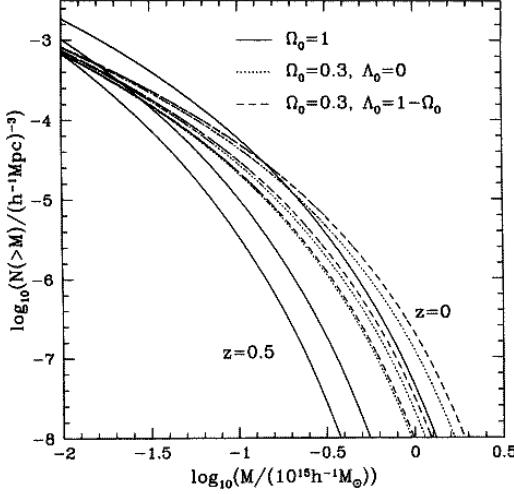


Fig. 27. – Predicted evolution of the cluster mass function. The comoving number density of clusters with mass larger than M , is plotted as a function of M , for three cosmological models: $\Omega_m = 1$. (solid lines), an open model with $\Omega_m = 0.3$ (dotted lines) and a flat model with $\Omega_m = 0.3, \Omega_\Lambda = 0.7$ (dashed lines). In each case, the predictions for $z = 0.5, 0.33, 0$ are shown. Figure from [147].

observable(s) have to be used as surrogates. I review below the available data and current constraints on the cosmological parameters. For more detailed reviews, the reader may refer to [150, 151].

5.3.1. The X-ray cluster surveys. Beyond the galactic plane, the X-ray sky is essentially populated with AGNs (about 90% of the sources at the typical ROSAT detection limit) and galaxy clusters. There is a diffuse X-ray background emission, due mostly to the local galactic emission (in the soft energy band used normally to detect clusters).

Detecting clusters in X-ray has several advantages [150]. First, the detection of X-rays is an unambiguous signature of the presence of a true potential well. Second, clusters are high contrast objects in the X-ray sky (the X-ray flux depends on the square of the density) and can be distinguished from other sources (AGNs) by their spatial extent (since ROSAT pointed observations). By contrast, the field galaxy population sets a serious problem to optical surveys, as it rapidly overwhelms the galaxy overdensity associated with clusters as z increases. X-ray surveys are also less subject to projection effects. Third, quantitative selection criteria can be defined: one searches for extended sources above a given flux limit. X-ray cluster samples are thus flux-limited samples. This allows for precise estimates of the survey volume and thus of space density of clusters.

However, there are several difficulties in constructing well controlled samples of X-ray clusters. In particular, it requires intensive optical follow-up to confirm the identification and to measure redshifts. Another critical point is to precisely estimate the 'selection function', that is, how the survey samples the true population of clusters. This includes the computation of the sky coverage $\Omega(S_X)$: the effective area covered by the survey as a function of X-ray flux. This is not trivial because the exposure time, background, PSF, efficiency etc.. vary across the telescope field of view and from pointing to pointing and so does the detection probability. The detection probability also depends on the cluster morphology. In view of these difficulties, one usually relies on Monte-Carlo simulations

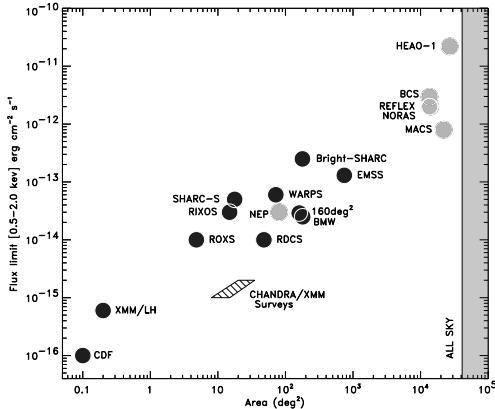


Fig. 28. – Solid angles and flux limits of various X-ray cluster surveys (references in text and in [150]). Dark filled circles: serendipitous surveys constructed from a collection of pointed observations. Light shaded circles: surveys covering contiguous areas. Figure from [150].

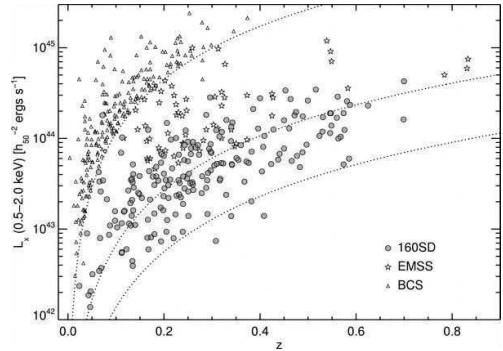


Fig. 29. – X-ray luminosity and redshift distribution of the 160SD [163], EMSS [159] and BCS [154] cluster samples. The dotted curves (left to right) are indicative flux limit of the surveys: 2.710^{-12} , 1.510^{-13} , 3.10^{-14} ergs/cm²/s in the (0.5 – 2) keV energy band. Figure from [164]

to study the survey selection function [152, 153].

Several cluster surveys have been conducted since the early HEAO-1 survey (see [150] for a complete list and descriptions of all X-ray cluster surveys). The RASS (Rosat All Sky Survey) was the first (and still unique) X-ray imaging Survey of the whole sky. Several cluster catalogs were or are still being constructed from the RASS. Published catalogs of X-ray selected clusters include the BCS [154], the NORAS [155] and the REFLEX catalogs [156]. More than 1000 clusters have been discovered, mostly at low redshift. For instance, the REFLEX cluster catalog contains ~ 450 clusters out to $z \sim 0.3$. Serendipitous search of clusters is performed using archival pointed observations: first with the EINSTEIN satellite (EMSS survey [157]), then with ROSAT, and now with XMM [158] and Chandra. The EMSS cluster catalog [159, and references therein] is still largely used for cosmological studies and several ROSAT serendipitous cluster catalogs are now available in the literature: the Bright SHARC [160], the WARPS-I [161], the Southern SHARC [162] and the 160SD [163] catalogs. The sky coverage and flux limit of various surveys are given in Fig. 28. Compared to the RASS, ROSAT serendipitous surveys go much deeper and thus cover a larger z range (up to $z \sim 1.3$). However, their sky coverage is much smaller, so that serendipitous cluster catalogs include from 12 to 100 clusters.

The mass and redshift coverages depend on the survey area and flux limit in a complex way. The lower luminosity limit (and thus mass limit) increases with z (Fig. 29), since X-ray surveys are flux-limited, at variance with SZ surveys. The abundance of clusters rapidly decreases with luminosity (Fig. 30). To detect massive (and thus rare)

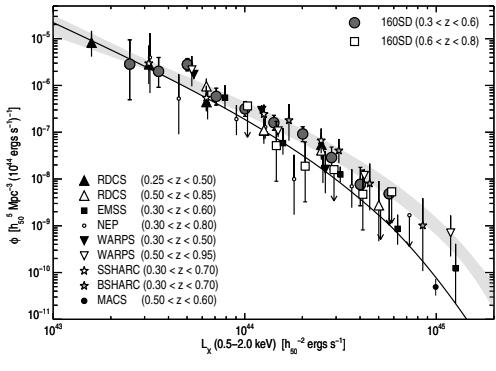


Fig. 30. – Compilation of high-redshift XLFs as measured by eight independent X-ray flux-limited surveys (Einstein de Sitter universe), compared with the local XLF (shaded area). Figure from [164].

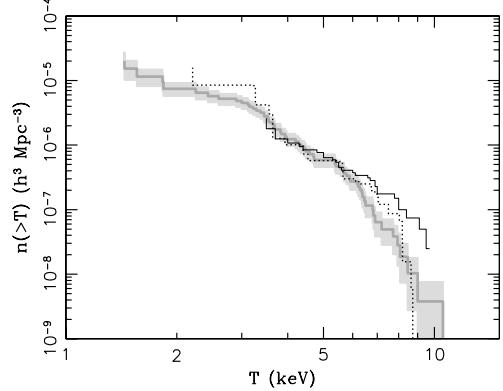


Fig. 31. – Cumulative temperature function (grey solid line with its 68% error band) established from the HIFLUGGS cluster sample [165]. Previous results from [91] and [167] are shown by the solid and dotted black lines. Figure from [165].

clusters, the sky coverage must be wide. Going deeper (decreasing the flux limit of a survey) extends the low mass coverage. However, this does not necessarily augment the z coverage. Indeed, the survey area may become too small to detect clusters above the luminosity limit at high z . These points are illustrated Fig. 29.

Cluster catalogs derived from X-ray surveys contain only basic information: X-ray luminosity and redshift. To measure the temperature (and possibly the mass) X-ray follow-up is needed, usually with the next generation satellite(s). This may change with XMM serendipitous surveys, where we expect that a fraction of the detected clusters will be bright enough to allow for the measurement of temperature [158].

5.3.2. Measures of cluster space density. The X-ray Luminosity Function (XLF) in various z bins, is a direct product of X-ray cluster surveys. Estimating the selection function in terms of luminosity is simple: one has $\Omega(L_X, z) \equiv \Omega(S_X)$ where $S_X = L_X/D_l(z)^2$. There is an excellent agreement between the local XLFs derived from various surveys [150], and between the various XLFs at high redshifts derived from *ROSAT* Surveys [164], as shown Fig. 30. A significant evolution of the XLF is found only at the bright end ($L_X \gtrsim 5 \cdot 10^{44} \text{ ergs/s}$) and above $z \sim 0.5$ [164].

The drawback of using the XLF for cosmology is that the luminosity is very sensitive to the detailed gas properties: density distribution in the core, thermodynamic history, dynamical state. The XLF is thus not easily related to the mass function. The most common method is to combine the empirical L_X-T and the $M-T$ relations. The scatter of the L_X-T relation and the uncertainty on its evolution are the major sources of systematic uncertainties.

In principle, the temperature is more directly related to the mass than the luminosity,

and the X-ray temperature function (XTF) is a better substitute for the mass function. However, because the XTF is derived from cluster samples selected according to their flux, a knowledge of the L_X-T relation (and its scatter) is still required to compute the temperature selection function from $\Omega(S_X)$. Furthermore, the XTF has generally a lower statistical quality and mass coverage than the XLF, because the temperature is only measured for sub-samples of available X-ray cluster catalogs (specially at low mass and high z).

The agreement among various local XTFs is less good than for the local XLFs. This is illustrated in Fig. 31. The XTF recently derived [165] from the HIFLUGCS cluster sample [166], which contains the 63 brightest clusters from the RASS, is compared to earlier XTFs [91] (30 clusters) and [167] (25 clusters). The XTFs agree within 20% in the [3 – 6] keV range but differ by as much as a factor 3 above $\sim 6 - 7$ keV (see also discussion in [168] and their Fig 4 and 5). The only presently available XTF at high z (Fig. 34) is derived from ASCA follow-up of the EMSS Survey [159].

The HIFLUGCS cluster sample is currently the largest local cluster sample that incorporates temperature measurements and also imagery data (from ROSAT pointed observations). Using these data, the first (and still unique) cluster mass function, XMF, for X-ray selected clusters was established [166]. It can be directly compared to the theoretical predictions, but this test is not free of possible systematic errors. The masses had to be estimated using a simple isothermal β -model for the gas distribution (and the HE equation). Moreover, the determination of the selection function is not straightforward: the L_X-M relation (and its scatter) must be known to deduce the mass selection function from $\Omega(S_X)$. This relation was estimated using an extended sample of 106 clusters.

Constraining cosmological parameters from the XTF or the XLF requires a good calibration of the $M-T$ and/or L_X-M relations, both at low and high z . This is difficult to achieve, because it requires precise total mass measurements. To avoid this problem, it was recently proposed to use the baryon mass M_b , which is easy to measure, as a proxy of the total mass [169, 170]. The Baryon mass function BMF is directly related to the theoretical mass function under the assumption that the baryon fraction in clusters is universal and equal to Ω_b/Ω_m . In practice, the small variation with mass is taken into account. The L_X-M_b and its evolution must still be known to compute the selection function in terms of M_b , but this is relatively easily measured. The local BMF, derived from a sub-sample of 52 clusters from the HIFLUGCS sample, is plotted in Fig. 36, together with the high redshift BMF derived from *Chandra* follow-up of bright 160SD clusters [169, 170].

5.3.3. Constraints from local abundances. Recent studies of the local distribution functions [168, 165, 166, 171, 172, 173, 174] indicate that $\sigma_8 \Omega_m^{0.4-0.6} = 0.4 - 0.6$, or $\sigma_8 \sim 0.8$ for $\Omega_m = 0.3$ [150]. For instance, Fig. 32 shows the constraint derived from the Reflex XLF [174], using the L_X-M relation from [166]. Note the well-known degeneracy between σ_8 and Ω_m (see [150] for explanation). Constraints from the XTF and the XLF from RASS data are consistent, but the XLF gives stricter constraints due to the wider mass coverage [173].

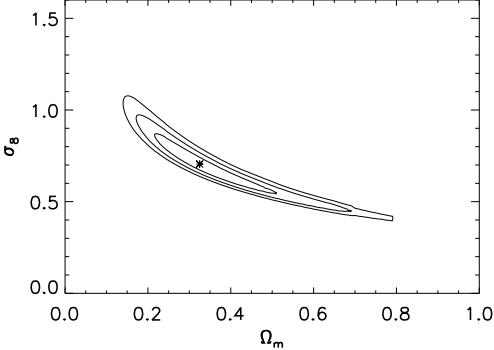


Fig. 32. – Likelihood contours ($1-3\sigma$ levels) in the $\sigma_8-\Omega_m$ plane obtained from the REFLEX cluster abundances. Note the degeneracy between σ_8 and Ω_m . Figure from [174]

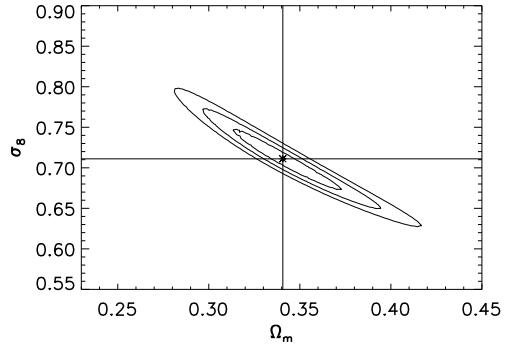


Fig. 33. – Same as Fig. 32 but obtained using both cluster abundance and large-scale distribution (scale is not the same). Note the degeneracy breaking. Figure from [174]

Errors on (σ_8, Ω_m) are currently dominated by systematic uncertainties. The major source of errors on σ_8 (at fixed Ω_m) is the uncertainty on the normalization of the $M-T$ relation [173, 175, 159]. If one assumes a higher mass for a given temperature, the amplitude of the mass function corresponding to an observed XTF is higher and a higher σ_8 value is obtained. The discrepancies among various studies on σ_8 are largely due to the different normalizations used [159]. The use of different cosmological priors and statistical methods also contributes to the observed differences [173]. Furthermore, a precise knowledge and proper treatment of the intrinsic scatter of the scaling relations ($M-T$ and L_X-T) is critical. For instance, neglecting the scatter in the L_X-M relation biases σ_8 towards high values [173].

5.3.4. Breaking the degeneracy using local cluster clustering. Galaxy clusters can also be used to trace the large-scale structure of the Universe, which depends on the cosmology and the initial density spectrum. The spatial distribution of clusters thus provide cosmological constraints, complementary to those obtained from cluster abundance. This requires to survey large contiguous regions of the sky.

The large-scale clustering and abundance of clusters was measured with unprecedented accuracy with the REFLEX survey. Recently, these data were analyzed simultaneously [174]. As shown in Fig. 33, this largely breaks the degeneracy between σ_8 and Ω_m observed when only the local XLF is used. The REFLEX sample gives $\Omega_m = 0.34 \pm 0.03$ and $\sigma_8 = 0.71 \pm 0.04$ (1σ statistical errors).

5.3.5. Constraints from evolution . As mentioned above, the evolution of the mass function is very sensitive to Ω_m . The degeneracy between σ_8 and Ω_m , obtained when using local cluster abundance, can be broken if the cluster distribution functions (XLF, XTF..) at higher redshift is known.

Till recently, cosmological tests based on the evolution of the XLF gave ambiguous

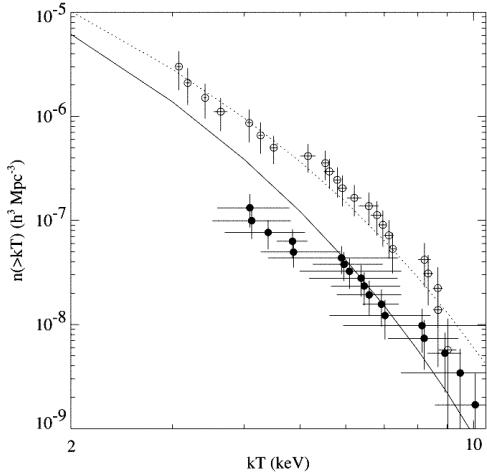


Fig. 34. – Cumulative temperature function established by [159] at low (open circle) and high (filled circles) redshift. The high redshift XTF is derived from ASCA follow-up of EMSS cluster sample. The lines are the best fit models ($\Omega_m = 0.28$, $\Omega_\Lambda = 0.94$). Figure from [159]

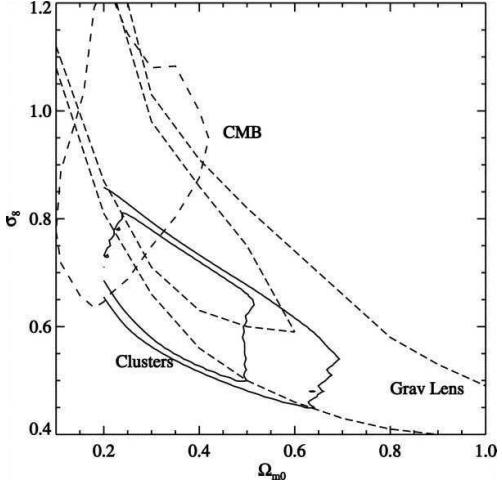


Fig. 35. – Constraints on σ_8 and Ω_m (1 and 2σ confidence contours) derived from the temperature function and its evolution, compared to constraints from CMB and weak lensing observations. Figure from [159]

results due to the large uncertainty on the evolution of the L_X-T relation [150]. This evolution is better constrained with *XMM-Newton* and *Chandra*. However, a consensus has not yet been reached. An analysis of the X-ray luminosity distribution in the RDCS sample (103 clusters out to $z \sim 0.85$) gives $\Omega_m = 0.35^{+0.13}_{-0.1}$ and $\sigma_8 = 0.66 \pm 0.06$ [176]. A SCDM Universe ($\Omega_m = 1$) is excluded with a high confidence level: $\Omega_m = 0.1 - 0.6$ at the 3σ level when systematic uncertainties are taken into account. In this study, the normalisation of the L_X-T relation was assumed to vary as $(1+z)^A$, with $A = 0 - 1$ in agreement with *Chandra* observations. On the other hand, a study of the evolution of cluster number counts [177], as measured in various surveys, using as local reference the local XTF [168], gives $\Omega_m = 0.85 - 1$. As in the previous study [176], a standard evolution of the $M-T$ relation was assumed, and a similar evolution of the L_X-T relation was used as constrained from *XMM-Newton* and *Chandra* data.

In contrast, concordant constraints are obtained from the evolution of the XTF [159, see Fig 35] or the BMF [169, see Fig 37], consistent with the value derived from the XLF RDCS study [176]. For instance, the BMF evolution [169] gives $\Omega_m = 0.24 \pm 0.12$ for a flat Universe (68% confidence level). The XTF evolution [159] gives a best fit $\Omega_m = 0.33$ value. It was also used to constrain the equation of state of the Dark Energy: $w = -0.42 \pm 0.21$ (68% confidence level).

53.6. Conclusion. Clusters of galaxies are powerful cosmological tools. Several *independent* methods can be used to constrain the cosmological parameters from cluster

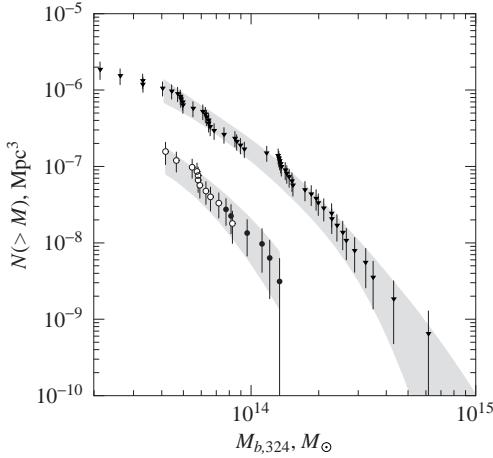


Fig. 36. – Baryon mass function for the 160SD survey sample in the redshift interval $0.4 < z < 0.8$ (filled circles) assuming $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ [169]. The local mass function from [170] is shown by triangles. Figure from [169]

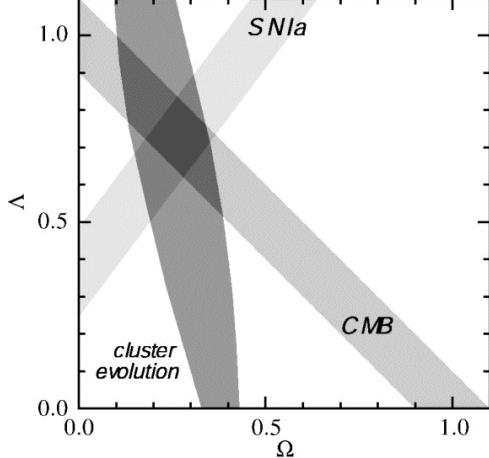


Fig. 37. – Comparison of constraints derived from the baryon mass function (Fig. 36) with those from distant supernovae and the CMB (68% confidence regions). Figure from [169]

X-ray (or S-Z) observations. They are complementary to constraints provided by CMB, SNI and weak lensing observations. Nearly all studies indicate a low Ω_m universe. σ_8 and Ω_m are typically determined with $\pm 20\%$ accuracy, whereas Ω_Λ (or the equation of state of the Dark Energy) is not yet well constrained. High precision cosmology with clusters is a priori possible. However, the key issue is to control and decrease the systematic uncertainties due to our imperfect knowledge of the physics that govern cluster formation and evolution. This includes a better knowledge of the intrinsic scatter and evolution of the scaling laws. A better precision of the abundance of massive clusters at high z should also constrain much more tightly Ω_Λ and Ω_Λ . This requires a survey with a very large sky coverage.

6. – Perspectives

The *XMM-Newton* and *Chandra* observatories are designed to operate until at least 2010. In the coming years, large efforts will be devoted on statistical analysis of known cluster samples, using archival data or large projects. Several Large Projects are in progress, to follow-up unbiased samples of local or distant clusters discovered by *ROSAT*. In parallel, the *XMM-Newton* (and to a lesser extent *Chandra*) satellites will provide new cluster samples for cosmology, extending to lower mass and possibly higher z . This includes serendipitous surveys, like the XCS [28], or contiguous surveys, like the XMM-LSS [178].

For understanding the physics of the intra-cluster medium, key information is still missing. We cannot measure the velocity structure of the gas. Moreover, the high

resolution and spatially resolved spectroscopic data required to investigate the complex cluster core is not yet available. This information can only be provided by bolometer array type instruments, as will be on board Astro-E2, planned to be launched in a few months [179].

Understanding non thermal processes in clusters (e.g electron acceleration by shocks or turbulence during merger events, effect on the cluster evolution) is another open issue. Astro-E2 will greatly improve our capability to observe high energy emission from clusters. Further progresses will require spatially resolved spectroscopy at high energies (up to $E \sim 80$ keV). When combined with radio data, this will allow us to map both the magnetic field and the non thermal particle population. Several projects are under study, like *SIMBOL-X* [180] and *NeXT* [179].

Planck (to be launched in 2007) should detect about 10000 clusters via the SZ effect. The statistical study of this sample, unique by its size, depth and sky coverage, will constrain cosmological parameters and provide information on the physics of structure formation. It will be extremely useful (and sometimes mandatory) to combine the SZ data with X-ray data, like those obtained by *XMM-Newton*. Planck and *XMM-Newton* surveys have not the same sky, mass and redshift coverage and it is unrealistic to think of a complete X-ray follow-up of the Planck sample. However, sub-samples can still be used, to calibrate for instance the $Y-M$ relation. This relation must be known to constrain cosmology on the basis of SZ cluster abundances.

The next generation of X-ray observatories, Constellation-X and XEUS, is already on the horizon for the years 2010+. They should allow us to probe the hot Universe at much higher z than presently. We will study early Black Holes, groups of galaxies at $z \sim 2$ and their evolution to the massive clusters of today, and investigate nucleosynthesis down to the present epoch. These satellites will perfectly complement observatories like ALMA, JWST that will look at the 'cool' component of the Universe.

* * *

REFERENCES

- [1] JONES C. and FORMAN B., *ApJ*, **276** (1984) 38
- [2] VOIT G.M., *Rev. Mod. Phys.*, **in press** (2004) astro-ph/0410173
- [3] SARAZIN C., *X-ray emission from clusters of galaxies*, (Cambridge University Press, Cambridge) 1988
- [4] FERETTI L., GIOIA I.M. and GIOVANNINI G., (Editors) *ASSL*, Vol. 272: *Merging Processes in Galaxy Clusters*, (Kluwer, Dordrecht) 2002
- [5] MULCHAHEY J.S., DRESSLER A. and OEMLER A., (Editors) *Carnegie Observatories Astrophysics Series*, Vol. 3: *Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution*, (Cambridge University Press, Cambridge) 2004, *in press*
- [6] MAZZOTTA P., RASIA E., MOSCARDINI L. and TORMEN G., *MNRAS*, **354** (2004) 10
- [7] PRATT G.W. and ARNAUD M., *A&A*, **408** (2003) 1
- [8] JANSEN F., LUMB D., ALTIERI B. ET AL., *A&A*, **365** (2001) L1

- [9] DEN HERDER J.W., BRINKLAN A.C., KAHN S.M., ET AL., *A&A*, **365** (2001) L7
- [10] STRÜDER L., BRIEL U., DENNER K. ET AL., *A&A*, **365** (2001) L18
- [11] TURNER M.J.L., ABBEY A., ARNAUD M. ET AL., *A&A*, **365** (2001) L27
- [12] WEISSKOPF M.C., BRINKMAN B., CANIZARES C. ET AL., *PASP*, **114** (2002) 1
- [13] TRÜMPER J., *Phys B1*, **46** (1990) 137
- [14] TANAKA Y., INOUE H. and HOLT S., *PASJ*, **46** (1994) L37
- [15] DI COCCO G., *NComC*, **20** (1997) 789
- [16] BRADT H.V., ROTHSCHILD R.E. and SWANK J.H., *A&A Suppl. Ser.*, **97** (1993) 355
- [17] FABIAN A.C., SANDERS J.S., ETTORI S. ET AL., *MNRAS*, **323** (2001) L33
- [18] LEWIS A.D., BUOTE D.A., and STOCKE J.T., *ApJ*, **586** (2003) 135
- [19] PRATT G.W. and ARNAUD M., *A&A*, **429** (2005) 791
- [20] SUN M., MURRAY S., MARKEVITCH M. and VIKHLININ A., *ApJ*, **565** (2002) 867
- [21] BELSOLE E., PRATT G.W., SAUVAGEOT J.L. and BOURDIN H., *A&A*, **515** (2004) 821
- [22] MARKEVITCH M., PONMAN T., NULSEN P.E.J ET AL., *ApJ*, **541** (2000) 542
- [23] ARNAUD M., MAJEROWICZ S., LUMB D. ET AL., *A&A*, **390** (2002) 27
- [24] ROSATI P., TOZZI P., ETTORI S. ET AL., *Astron. J.*, **127** (2004) 230
- [25] ETTORI S., TOZZI P., BORGANI S., and ROSATI P., *A&A*, **417** (2004) 13
- [26] MAUGHAN B.J., JONES L.R., EBELING H. ET AL., *ApJ*, **587** (2003) 589
- [27] VALTCHANOV I., PIERRE M., WILLIS J. ET AL., *A&A*, **423** (2004) 75
- [28] LAND K.R., NICHOL R., DAVIDSON M. ET AL., *MNRAS*, **in press** (2004) astro-ph/0405225
- [29] JONES C. and FORMAN B., in *NATO ASI Vol 366: Clusters and Superclusters of Galaxies*, edited by A.C. FABIAN 1992, p.49
- [30] SARAZIN C., in *ASSL, Vol. 272: Merging Processes in Galaxy Clusters*, edited by FERETTI L., GIOIA I.M. and GIOVANNINI G. (Kluwer:Dordrecht) 2002, p. 1
- [31] BUOTE D., in *ASSL, Vol. 272: Merging Processes in Galaxy Clusters*, edited by FERETTI L., GIOIA I.M. and GIOVANNINI G., (Kluwer:Dordrecht) 2002, p. 79
- [32] BRIEL U., HENRY J., SCHWARZ R. ET AL., *A&A*, **246** (1991) L10
- [33] MARKEVITCH M., SARAZIN C. and VIKHLININ A., *ApJ*, **521** (1999) 526
- [34] SCHINDLER S. and MÜLLER E., *A&A*, **272** (1993) 137
- [35] RICKER P.M. and SARAZIN C., *ApJ*, **561** (2001) 621
- [36] HASHIMOTO Y., BARCONS X., BÖHRINGER H. ET AL., *A&A*, **417** (2004) 819
- [37] MAUGHAN B.J., JONES L.R., EBELING H. and SCHARF C., *MNRAS*, **351** (2004) 1193
- [38] MARKEVITCH M., GONZALES A.H., DAVID L. ET AL., *ApJ*, **567** (2002) L27
- [39] VIKHLININ A., MARKEVITCH M. and MURRAY S., *ApJ*, **549** (2001) L47
- [40] REIPRICH T.H., SARAZIN C.L., KEMPNER J.C. and TITTLEY E., *ApJ*, **608** (2004) 179
- [41] ETTORI S. and FABIAN A., *MNRAS*, **317** (2000) L57
- [42] VIKHLININ A., MARKEVITCH M. and MURRAY S., *ApJ*, **551** (2001) 160
- [43] NEUMANN D.M., LUMB D., PRATT G.W. and BRIEL U.G., *A&A*, **400** (2003) 811
- [44] KEMPNER J. and DAVID J., *MNRAS*, **349** (2004) 385
- [45] DURRET F., LIMANETO G.B., FORMAN W. and CHURAZOV E., *A&A*, **403** (2003) L29
- [46] FORMAN W., CHURAZOV E., DAVID L. ET AL., 2003 astro-ph/0301476
- [47] ARNAUD M., AGHANIM A., GASTAUD R. ET AL., *A&A*, **365** (2001) L67
- [48] BELSOLE E., SAUVAGEOT J.L., PRATT G.W. and BOURDIN H., *A&A*, **430** (2005) 385
- [49] MOHR J.J., EVRARD A.E., FABRICANT D.G. and GELLER M.J., *ApJ*, **447** (1995) 8
- [50] BUOTE D.A and TSAI J.C., *ApJ*, **458** (1996) 27
- [51] SCHUECKER P., BHRINGER H., REIPRICH T.H. and FERETTI L., *A&A*, **378** (2001) 408
- [52] JONES C. and FORMAN B., *ApJ*, **511** (1999) 65
- [53] JELTEMA T.E., CANIZARES C., BAUTZ M. and BUOTE D., *ApJ*, **in press** (2005) astro-ph/0501360

- [54] GIOVANNINI G. and FERETTI L., in *ASSL, Vol. 272: Merging Processes in Galaxy Clusters*, edited by L. FERETTI I.M., GIOIA I. and GIOVANNINI G. (Kluwer:Dordrecht) 2002, p. 197
- [55] FUSCO-FEMIANO R., DAL FUIME D., DE GRANDI S. ET AL., *ApJ*, **534** (2000) L7
- [56] REPHAEIL Y. and GRUBER D., *ApJ*, **595** (2003) 137
- [57] BUOTE D., *ApJ*, **553** (2001) L15
- [58] BRUNETTI G., in *ASP Conf. Ser. 268: Matter and Energy in Clusters of Galaxies*, edited by S.BOWYER and C-Y. HWANG 2002, p. 349
- [59] GOVONI F., MARKEVITCH M., VIKHININ A. ET AL., *ApJ*, **605** (2004) 695
- [60] SCHUECKER P., FINOGUENOV A., MINIATI F. ET AL., *A&A*, **426** (2004) 387
- [61] ARNAUD M. and EVRARD A.E., *MNRAS*, **305** (1999) 631
- [62] MOHR J.J., MATHIESEN B. and EVRARD A.E., *ApJ*, **517** (1999) 627
- [63] FINOGUENOV A., REIPRICH T. H. and BÖHRINGER H., *A&A*, **368** (2001) 749
- [64] MOHR J.J. and EVRARD A.E., *ApJ*, **491** (1997) 38
- [65] NEUMANN D.M. and ARNAUD, M., *A&A*, **348** (1999) 711
- [66] EVRARD A.E. and GIOIA I., 2002, in *ASSL Vol. 272: Merging Processes in Galaxy Clusters*, edited by L. FERETTI I.M., GIOIA I. and GIOVANNINI G. (Kluwer:Dordrecht) 2002, p. 253
- [67] NAVARRO J.F., FRENK C.S. and WHITE S.D.M., *MNRAS*, **274** (1995) 720
- [68] EVRARD A.E., METZLER C.A. and NAVARRO J.F., *ApJ*, **469** (1996) 494
- [69] BRYAN G.L. and NORMAN M.L., *ApJ*, **495** (1998) 80
- [70] EKE V.R., NAVARRO J.F. and FRENK C.S., *ApJ*, **503** (1998) 569
- [71] CAVALIERE A., MENCI N. and TOZZI P., *MNRAS*, **308** (1999) 599
- [72] WHITE M., *A&A*, **367** (2001) 27
- [73] MATHIESEN B.F. and EVRARD A.E., *ApJ*, **546** (2001) 100
- [74] NAVARRO J.F., FRENK C.S., and WHITE S.D.M., *ApJ*, **490** (1997)
- [75] MOORE B., QUINN T., GOVERNATO F. ET AL., *MNRAS*, **310** (1999) 1147
- [76] NAVARRO J.F., HAYASHI E., POWER C. ET AL., *MNRAS*, **349** (2004) 1039
- [77] DIEMAND J., MOORE B. and STADEL J., *MNRAS*, **353** (2004) 624
- [78] DOLAG K., BARTELmann M., PERROTTE F. ET AL., *A&A*, **416** (2004) 853
- [79] PRATT G.W. and ARNAUD M., *A&A*, **394** (2002) 375
- [80] DAVID L.P., NULSEN P.E.J., McMAMARA B.R. ET AL., *ApJ*, **557** (2001) 546
- [81] ALLEN S.W., SCHMIDT R.W. and FABIAN A.C., *MNRAS*, **328** (2001) L37
- [82] ARABADJIS J.S., BAUTZ M.W. and GARMIRE G.P., *ApJ*, **572** (2002) 66
- [83] BUOTE D.A. and LEWIS A.D., *ApJ*, **604** (2004) 116
- [84] POINTECOUTEAU E., ARNAUD M., KAASTRA J. and DE PLAA J., *A&A*, **423** (2004) 33
- [85] POINTECOUTEAU E., ARNAUD M. and PRATT G., *A&A*, in press (2005) astro-ph/0501635
- [86] CLOWE, D., GONZALEZ, A.H. and MARKEVITCH M., *ApJ*, **604** (2004) 596
- [87] MARKEVITCH M., GONZALEZ A.H., CLOWE D. ET AL., *ApJ*, **606** (2004) 819
- [88] EVRARD A.E. and HENRY J.P., *ApJ*, **383** (1991) 95
- [89] OSMOND J.P.F. and PONMAN T., *MNRAS*, **350** (2004) 1511
- [90] MULCHAHEY J.S., in *Carnegie Observatories Astrophysics Series, Vol. 3: Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution*, edited by MULCHAHEY J.S., DRESSLER A. and OEMLER A. (Cambridge University Press, Cambridge) 2004, in press
- [91] MARKEVITCH M., *ApJ*, **504** (1998) 27
- [92] VIKHININ A., FORMAN W. and JONES C., *ApJ*, **525** (1999) 47
- [93] ETTORI S., DE GRANDI S. and MOLENDI S., *A&A*, **391** (2002) 841
- [94] CASTILLO-MORALES A. and SCHINDLER S., *A&A*, **403** (2003) 433
- [95] NEUMANN D.M. and ARNAUD M., *A&A*, **373** (2001) L33
- [96] SANDERSON A.J.R. and PONMAN T.J., *MNRAS*, **345** (2003) 124
- [97] ETTORI S., TOZZI P. and ROSATI P., *A&A*, **398** (2003) 879

- [98] ALLEN S.W., SCHMIDT R. and FABIAN A.C., *MNRAS*, **334** (2002) L11
- [99] MARKEVITCH M., FORMAN W., SARAZIN C. and VIKHLININ A., *ApJ*, **503** (1998) 77
- [100] IRWIN J. and BREGMAN J., *ApJ*, **538** (2000) 543
- [101] DE GRANDI S. and MOLENDI S., *ApJ*, **567** (2002) 163
- [102] ZHANG Y.Y., FINOGUENOV A., BÖHRINGER H. ET AL., *A&A*, **413** (2004) 49
- [103] VIKHLININ A., MARKEVITCH M., MURRAY S.S. ET AL., preprint astro-ph/0412306
- [104] ARNAUD M., PRATT G.W. and PONTECOUTEAU E., *MmSAI*, **75** (2004) 529
- [105] PIFFARETTI R., JETZER P., KAASTRA J.S. and TAMURA T., preprint astro-ph/0412233
- [106] VOIT G.M., BRYAN G.L., BALOGH M.L. and BOWER R.G., *ApJ*, **576** (2002) 601
- [107] PONMAN T.J., CANNON D.B., and NAVARRO J.F., *Nature*, **397** (1999) 135
- [108] PONMAN T.J., SANDERSON A.J.R. and FINOGUENOV A., *MNRAS*, **343** (2003) 331
- [109] KAASTRA J.S., TAMURA T., PETERSON J.R. ET AL., *A&A*, **413** (2004) 415
- [110] REIPRICH T. H., KEMPNER J. C. and SOKER N., (Editors) *The Riddle of Cooling Flows in Galaxies and Clusters of Galaxies*, 2004, electronic publication URL: <http://www.astro.virginia.edu/coolflow/>
- [111] PETERSON J.R., KAHN S.M., PAEREELS F.B.S ET AL., *A&A*, **590** (2003) 207
- [112] MOLENDI S., PIZZOLATO F., *ApJ*, **560** (2001) 194
- [113] BÖHRINGER H., MATSUSHITA K., CHURAZOV E. ET AL., *A&A*, **382** (2002) 804
- [114] BLANTON E.L., in *The Riddle of Cooling Flows in Galaxies and Clusters of Galaxies*, edited by REIPRICH T. H., KEMPNER J. C. and SOKER N. 2004, electronic publication URL: <http://www.astro.virginia.edu/coolflow/>
- [115] McNAMARA B.R., NULSEN P.E.J., WISE M.W. ET AL., *Nature*, **433** (2005) 45
- [116] FABIAN A.C., SANDERS J.S., ETTORI S. ET AL., *MNRAS*, **318** (2000) L65
- [117] FABIAN A.C., SANDERS J.S., ALLEN S.W. ET AL., *MNRAS*, **344** (2003) L43
- [118] MARKEVITCH M., VIKHLININ A. and FORMAN W., in *ASP Conf. Ser. 268: Matter and Energy in Clusters of Galaxies*, edited by BOWYER S. AND HWANG C.Y. 2002, p. 37
- [119] BREGMAN J.N., in *he Riddle of Cooling Flows in Galaxies and Clusters of Galaxies*, edited by REIPRICH T. H., KEMPNER J. C. and SOKER N. 2004, electronic publication URL: <http://www.astro.virginia.edu/coolflow/>
- [120] FABIAN A.C., *RMxAC*, **17** (2003) 303
- [121] HORNER D.J., MUSHOTZKY R.F. and SCHARF C.A., *ApJ*, **520** (1999) 78
- [122] NEVALAINEN J., MARKEVITCH M. and FORMAN W., *ApJ*, **532** (2000) 694
- [123] XU H., JIN G. and WU X.P., *ApJ*, **553** (2001) 78
- [124] ARNAUD M., PONTECOUTEAU E. and PRATT G., *A&A*, subm. (2005) astro-ph/0502210
- [125] BORGANI S., MURANTE G., SPRINGEL V. ET AL., *MNRAS*, **348** (2004) 1078
- [126] KAY S.T., DA SILVA A.C., AGHANIM N. ET AL., 2004, astro-ph/0411650.
- [127] ARNAUD M., AGHANIM N. and NEUMANN D., *A&A*, **389** (2002) 1
- [128] REICHART D.E., CASTANDER F.J. and NICHOL R.C., *ApJ*, **516** (1999) 1
- [129] NOVICKI M.C., SORNIG M. and HENRY J.P., *Astron. J.*, **124** (2002) 2413
- [130] OTA N. and MITSUDA K., *A&A*, **428** (2004) 757
- [131] VIKHLININ A., VAN SPEYBROECK L., MARKEVITCH M. ET AL., *ApJ*, **578** (2002) L107
- [132] LUMB D., BARTLETT J., ROMER A.K. ET AL., *A&A*, **420** (2004) 853
- [133] WHITE S.D.M., NAVARRO J.F., EVRARD A.E. and FRENK C.S., *Nature*, **366** (1993) 429
- [134] LIN Y.T., MOHR J.J. and STANFORD S.A., *ApJ*, **591** (2003) 749
- [135] ALLEN S.W., SCHMIDT R.W., EBELING H. ET AL., *MNRAS*, **353** (2004) 457
- [136] SASAKI S., *PASJ*, **48** (1996) L119
- [137] MOHR J.J., REESE J.J., ELLINGSON E. ET AL., *ApJ*, **544** (2000) 109
- [138] RIESS A.G., FILIPPENKO A.V., CHALLIS P. ET AL., *Astron. J.*, **116** (1998) 1009
- [139] PERLMUTTER S., ALDERING G., GOLDHABER G. ET AL., *ApJ*, **517** (1999) 565

- [140] RIESS A.G., STROLGER L.G., TONRY J. ET AL., *ApJ*, **607** (2004) 665
- [141] SPERGEL D.N., VERDE L., PEIRIS H.V. ET AL., *ApJ Suppl.*, **148** (2003) 175
- [142] EVRARD A.E., MACFARLAND T.J., COUCHMAN H.M.P. ET AL., *ApJ*, **573** (2002) 7
- [143] SHETH R.K., MO H.J. and TORMEN G., *MNRAS*, **323** (2001) 1
- [144] JENKINS A., FRENK C.S., WHITE S.D.M. ET AL., *MNRAS*, **321** (2001) 372
- [145] PRESS W.H. and SCHECHTER P., *ApJ*, **187** (1974) 425
- [146] OUKBIR J., and BLANCHARD A., *A&A*, **262** (1992) L21
- [147] EKE V.R., COLE S. and FRENK C.S., *MNRAS*, **282** (1996) 263
- [148] BORGANI S. and GUZZO L., *Nature*, **409** (2001) 39
- [149] PERRENOD S.C., *ApJ*, **236** (1980) 373
- [150] ROSATI P., BORGANI S. and NORMAN C., *Ann. Rev. Astron. Ap.*, **40** (2002) 539
- [151] HENRY J.P., in *Testing Cosmological Models with Galaxy Clusters* 2003, electronic publication, <http://www.xray.mpe.mpg.de/~ringberg03/>
- [152] VIKHLININ A., McNAMARA B.R., FORMAN W. ET AL., *ApJ*, **502** (1998) 558
- [153] ADAMI C., ULMER M.P., ROMER A.K. ET AL., *ApJ Suppl.*, **131** (2000) 391
- [154] EBELING H., EDGE A.C., ALLEN S.W. ET AL., *MNRAS*, **318** (2000) 333
- [155] BÖHRINGER H., VOGES W., HUCHRA J.P. ET AL., *ApJ Suppl.*, **129** (2000) 435
- [156] BÖHRINGER H., SCHUECKER P., GUZZO L. ET AL., *A&A*, **425** (2004) 367
- [157] GIOIA I.M., MACCACARO T., SCHILD R.E. ET AL., *ApJ Suppl.*, **72** (1990) 567
- [158] ROMER A.K., VIANA P.T.P., LIDDLE A.R. and MANN R.G., *ApJ*, **547** (2001) 594
- [159] HENRY J.P., *ApJ*, **609** (2004) 603
- [160] ROMER A.K., NICHOL R.C., HOLDEN B.P. ET AL., *ApJ Suppl.*, **126** (2000) 209
- [161] PERLMAN E.S., HORNER D.J., JONES L.R. ET AL., *ApJ Suppl.*, **140** (2002) 265
- [162] BURKE D.J., COLLINS C.A., SHARPLES R.M. ET AL., *MNRAS*, **341** (2003) 1093
- [163] MULLIS C.R., McNAMARA B.R., QUINTANA H. ET AL., *ApJ*, **594** (2003) 154
- [164] MULLIS C.R., VIKHLININ A., HENRY J.P. ET AL., *ApJ*, **607** (2004) 17
- [165] IKEBE Y., REIPRICH T.H., BÖHRINGER H. ET AL., *A&A*, **383** (2002) 773
- [166] RIEPRICH T. and BÖHRINGER H., *ApJ*, **567** (2002) 716
- [167] HENRY J.P., *ApJ*, **534** (2000) 565
- [168] BLANCHARD A., SADAT R., BARTLETT J.G. and LE DOUR M., *A&A*, **362** (2000) 809
- [169] VIKHLININ A., VOEVODKIN A., MULLIS C., ET AL., *ApJ*, **590** (2003) 15
- [170] VOEVODKIN A. and VIKHLININ A., *ApJ*, **601** (2004) 610
- [171] VIANA P.T.P., NICHOL R.C., LIDDLE A.R., *ApJ*, **569** (2002) L75
- [172] ALLEN S.W., SCHMIDT R., FABIAN A.C. and EBELING H., *MNRAS*, **342** (2003) 287
- [173] PIERPAOLI E., BORGANI S., SCOTT D., and WHITE M., *MNRAS*, **342** (2003) 163
- [174] SCHUECKER P., BÖHRINGER H., COLLINS C.A. and GUZZO L., *A&A*, **402** (2003) 53
- [175] VIANA P.T.P., KAY S.T., LIDDLE A.R. ET AL., *MNRAS*, **346** (2003) 319
- [176] BORGANI S., ROSATI P., TOZZI P. ET AL., *ApJ*, **561** (2001) 13
- [177] VAUCLAIR S.C., BLANCHARD A., SADAT R. ET AL., *A&A*, **412** (2003) L37
- [178] PIERRE M., VALTCHANOV I., ALTIERI B. ET AL., *JCAP*, **9** (2004) 11
- [179] INOUE H. and KUNIEDA H., *AdSpR*, **34** (2004) 2628
- [180] FERRANDO P., ARNAUD M., CORDIER B. ET AL., *SPIE*, **5168** (2004) 65

Weak Gravitational Lensing

Peter Schneider

Institut f. Astrophysik, Universität Bonn, D-53121 Bonn, Germany

This review was written in the fall of 2004.

To appear as:

Schneider, P. 2005, in: Kochanek, C.S., Schneider, P., Wambsganss, J.: *Gravitational Lensing: Strong, Weak & Micro*. G. Meylan, P. Jetzer & P. North (eds.), Springer-Verlag: Berlin, p.273

1 Introduction

Multiple images, microlensing (with appreciable magnifications) and arcs in clusters are phenomena of *strong lensing*. In *weak gravitational lensing*, the Jacobi matrix \mathcal{A} is very close to the unit matrix, which implies weak distortions and small magnifications. Those cannot be identified in individual sources, but only in a statistical sense. Because of that, the accuracy of any weak lensing study will depend on the number of sources which can be used for the weak lensing analysis. This number can be made large either by having a large number density of sources, or to observe a large solid angle on the sky, or both. Which of these two aspects is more relevant depends on the specific application. Nearly without exception, the sources employed in weak lensing studies up to now are distant galaxies observed in the optical or near-IR passband, since they form the densest population of distant objects in the sky (which is a statement both about the source population in the Universe and the sensitivity of detectors employed in astronomical observations). To observe large number densities of sources, one needs deep observations to probe the faint (and thus more numerous) population of galaxies. Faint galaxies, however, are small, and therefore their observed shape is strongly affected by the Point Spread Function, caused by atmospheric seeing (for ground-based observations) and telescope effects. These effects need to be well understood and corrected for, which is the largest challenge of observational weak lensing studies. On the other hand, observing large regions of the sky quickly leads to large data sets, and the problems associated with handling them. We shall discuss some of the most important aspects of weak lensing observations in Sect. 3.

The effects just mentioned have prevented the detection of weak lensing effects in early studies with photographic plates (e.g., Tyson et al. 1984); they are not linear detectors (so correcting for PSF effects is not reliable), nor are

they sensitive enough for obtaining sufficiently deep images. Weak lensing research came through a number of observational and technical advances. Soon after the first giant arcs in clusters were discovered (see Sect. 1.2 of Schneider, this volume; hereafter referred to as IN) by Soucail et al. (1987) and Lynds & Petrosian (1989), Fort et al. (1988) observed objects in the lensing cluster Abell 370 which were less extremely stretched than the giant arc, but still showed a large axis ratio and was aligned in the direction tangent to its separation vector to the cluster center; they termed these images ‘arclets’. Indeed, with the spectroscopic verification (Mellier et al. 1991) of the arclet A5 in A 370 being located at much larger distance from us than the lensing cluster, the gravitational lens origin of these arclets was proven. When the images of a few background galaxies are deformed so strongly that they can be identified as distorted by lensing, there should be many more galaxy images where the distortion is much smaller, and where it can only be detected by averaging over many such images. Tyson et al. (1990) reported this statistical distortion effect in two clusters, thereby initiating the weak lensing studies of the mass distribution of clusters of galaxies. This very fruitful field of research was put on a rigorous theoretical basis by Kaiser & Squires (1993) who showed that from the measurement of the (distorted) shapes of galaxies one can obtain a parameter-free map of the projected mass distribution in clusters.

The flourishing of weak lensing in the past ten years was mainly due to three different developments. First, the potential of weak lensing was realized, and theoretical methods were worked out for using weak lensing measurements in a large number of applications, many of which will be described in later sections. This realization, reaching out of the lensing community, also slowly changed the attitude of time allocation committees, and telescope time for such studies was granted. Second, returning to the initial remark, one requires large fields-of-views for many weak lensing application, and the development of increasingly large wide-field cameras installed at the best astronomical sites has allowed large observational progress to be made. Third, quantitative methods for the correction of observations effects, like the blurring of images by the atmosphere and telescope optics, have been developed, of which the most frequently used one came from Kaiser et al. (1995). We shall describe this technique, its extensions, tests and alternative methods in Sect. 3.5.

We shall start by describing the basics of weak lensing in Sect. 2, namely how the shear, or the projected tidal gravitational field of the intervening matter distribution can be determined from measuring the shapes of images of distant galaxies. Practical aspects of observations and the measurements of image shapes are discussed in Sect. 3. The next two sections are devoted to clusters of galaxies; in Sect. 4, some general properties of clusters are described, and their strong lensing properties are considered, whereas in Sect. 5 weak lensing by clusters is treated. As already mentioned, this allows us

to obtain a parameter-free map of the projected (2-D) mass distribution of clusters.

We then turn to lensing by the inhomogeneously distributed matter distribution in the Universe, the large-scale structure. Starting with Gunn (1967), the observation of the distortion of light bundles by the inhomogeneously distributed matter in the Universe was realized as a unique probe to study the properties of the cosmological (dark) matter distribution. The theory of this cosmic shear effect, and its applications, was worked out in the early 1990's (e.g., Blandford et al. 1991). In contrast to the lensing situations studied in the rest of this book, here the deflecting mass is manifestly three-dimensional; we therefore need to generalize the theory of geometrically-thin mass distributions and consider the propagation of light in an inhomogeneous Universe. As will be shown, to leading order this situation can again be described in terms of an 'equivalent' surface mass density. The theoretical aspects of this large-scale structure lensing, or cosmic shear, are contained in Sect. 6. Although the theory of cosmic shear was well in place for quite some time, it took until the year 2000 before it was observationally discovered, independently and simultaneously by four groups. These early results, as well as the much more extensive studies carried out in the past few years, are presented and discussed in Sect. 7. In Sect. 8, we consider the weak lensing effects of galaxies, which can be used to investigate the mass profile of galaxies. As we shall see, this galaxy-galaxy lensing, first detected by Brainerd et al. (1996), is directly related to the connection between the galaxy distribution in the Universe and the underlying (dark) matter distribution; this lensing effect is therefore ideally suited to study the biasing of galaxies; we shall also describe alternative lensing effects for investigating the relation between luminous and dark matter. In the final Sect. 9 we discuss higher-order cosmic shear statistics and how lensing by the large-scale structure affects the lens properties of localized mass concentrations. Some final remarks are given in Sect. 10.

Until very recently, weak lensing has been considered by a considerable fraction of the community as 'black magic' (or to quote one member of a PhD examination committee: "You have a mass distribution about which you don't know anything, and then you observe sources which you don't know either, and then you claim to learn something about the mass distribution?"). Most likely the reason for this is that weak lensing is indeed weak. One cannot 'see' the effect, nor can it be graphically displayed easily. Only by investigating many faint galaxy images can a signal be extracted from the data, and the human eye is not sufficient to perform this analysis. This is different even from the analysis of CMB anisotropies which, similarly, need to be analyzed by statistical means, but at least one can display a temperature map of the sky. However, in recent years weak lensing has gained a lot of credibility, not only because it has contributed substantially to our knowledge about the mass distribution in the Universe, but also because different teams, with different data set and different data analysis tools, agree on their results.

Weak lensing has been reviewed before; we shall mention only five extensive reviews. Mellier (1999) provides a detailed compilation of the weak lensing results before 1999, whereas Bartelmann & Schneider (2001; hereafter BS01) present a detailed account of the theory and technical aspects of weak lensing.¹ More recent summaries of results can also be found in Wittman (2002) and Refregier (2003a), as well as the cosmic shear review by van Waerbeke & Mellier (2003).

The coverage of topics in this review has been a subject of choice; no claim is made about completeness of subjects or references. In particular, due to the lack of time during the lectures, the topic of weak lensing of the CMB temperature fluctuations has not been covered at all, and is also not included in this written version. Apart from this increasingly important subject, I hope that most of the currently actively debated aspects of weak lensing are mentioned, and the interested reader can find her way to more details through the references provided.

2 The principles of weak gravitational lensing

2.1 Distortion of faint galaxy images

Images of distant sources are distorted in shape and size, owing to the tidal gravitational field through which light bundles from these sources travel to us. Provided the angular size of a lensed image of a source is much smaller than the characteristic angular scale on which the tidal field varies, the distortion can be described by the linearized lens mapping, i.e., the Jacobi matrix \mathcal{A} . The invariance of the surface brightness by gravitational light deflection, $I(\boldsymbol{\theta}) = I^{(s)}[\boldsymbol{\beta}(\boldsymbol{\theta})]$, together with the locally linearized lens equation,

$$\boldsymbol{\beta} - \boldsymbol{\beta}_0 = \mathcal{A}(\boldsymbol{\theta}_0) \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}_0), \quad (1)$$

where $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\theta}_0)$, then describes the distortion of small lensed images as

$$I(\boldsymbol{\theta}) = I^{(s)}[\boldsymbol{\beta}_0 + \mathcal{A}(\boldsymbol{\theta}_0) \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}_0)]. \quad (2)$$

We recall (see IN) that the Jacobi matrix can be written as

$$\mathcal{A}(\boldsymbol{\theta}) = (1 - \kappa) \begin{pmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{pmatrix}, \text{ where } g(\boldsymbol{\theta}) = \frac{\gamma(\boldsymbol{\theta})}{[1 - \kappa(\boldsymbol{\theta})]} \quad (3)$$

is the reduced shear, and the g_α , $\alpha = 1, 2$, are its Cartesian components. The reduced shear describes the shape distortion of images through gravitational light deflection. The (reduced) shear is a 2-component quantity, most

¹ We follow here the notation of BS01, except that we denote the angular diameter distance explicitly by D^{ang} , whereas D is the comoving angular diameter distance, which we also write as f_K , depending on the context; see Sect. 4.3 of IN for more details. In most cases, the distance ratio D_{ds}/D_s is used, which is the same for both distance definitions.

conveniently written as a complex number,

$$\gamma = \gamma_1 + i\gamma_2 = |\gamma| e^{2i\varphi}; \quad g = g_1 + ig_2 = |g| e^{2i\varphi}; \quad (4)$$

its amplitude describes the degree of distortion, whereas its phase φ yields the direction of distortion. The reason for the factor ‘2’ in the phase is the fact that an ellipse transforms into itself after a rotation by 180° . Consider a circular source with radius R (see Fig. 1); mapped by the local Jacobi matrix, its image is an ellipse, with semi-axes

$$\frac{R}{1 - \kappa - |\gamma|} = \frac{R}{(1 - \kappa)(1 - |g|)} \quad ; \quad \frac{R}{1 - \kappa + |\gamma|} = \frac{R}{(1 - \kappa)(1 + |g|)}$$

and the major axis encloses an angle φ with the positive θ_1 -axis. Hence, if sources with circular isophotes could be identified, the measured image ellipticities would immediately yield the value of the reduced shear, through the axis ratio

$$|g| = \frac{1 - b/a}{1 + b/a} \quad \Leftrightarrow \quad \frac{b}{a} = \frac{1 - |g|}{1 + |g|}$$

and the orientation of the major axis φ . In these relations it was assumed that $b \leq a$, and $|g| < 1$. We shall discuss the case $|g| > 1$ later.

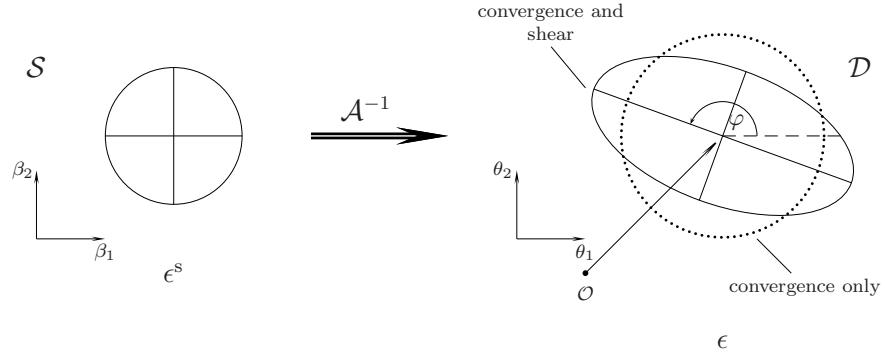


Fig. 1. A circular source, shown at the left, is mapped by the inverse Jacobian \mathcal{A}^{-1} onto an ellipse. In the absence of shear, the resulting image is a circle with modified radius, depending on κ . Shear causes an axis ratio different from unity, and the orientation of the resulting ellipse depends on the phase of the shear (source: M. Bradac)

However, faint galaxies are not intrinsically round, so that the observed image ellipticity is a combination of intrinsic ellipticity and shear. The strategy to nevertheless obtain an estimate of the (reduced) shear consists in locally averaging over many galaxy images, assuming that the intrinsic ellipticities are *randomly oriented*. In order to follow this strategy, one needs to

clarify first how to define ‘ellipticity’ for a source with arbitrary isophotes (faint galaxies are not simply elliptical); in addition, seeing by the atmospheric turbulence will blur – and thus circularize – observed images, together with other effects related to the observation procedure. We will consider these issues in turn.

2.2 Measurements of shapes and shear

Definition of image ellipticities. Let $I(\boldsymbol{\theta})$ be the brightness distribution of an image, assumed to be isolated on the sky; the center of the image can be defined as

$$\bar{\boldsymbol{\theta}} = \frac{\int d^2\theta I(\boldsymbol{\theta}) q_I[I(\boldsymbol{\theta})] \boldsymbol{\theta}}{\int d^2\theta I(\boldsymbol{\theta}) q_I[I(\boldsymbol{\theta})]} , \quad (5)$$

where $q_I(I)$ is a suitably chosen weight function; e.g., if $q_I(I) = H(I - I_{\text{th}})$, where $H(x)$ is the Heaviside step function, $\bar{\boldsymbol{\theta}}$ would be the center of light within a limiting isophote of the image. We next define the tensor of second brightness moments,

$$Q_{ij} = \frac{\int d^2\theta I(\boldsymbol{\theta}) q_I[I(\boldsymbol{\theta})] (\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j)}{\int d^2\theta I(\boldsymbol{\theta}) q_I[I(\boldsymbol{\theta})]} , \quad i, j \in \{1, 2\} . \quad (6)$$

Note that for an image with circular isophotes, $Q_{11} = Q_{22}$, and $Q_{12} = 0$. The trace of Q describes the size of the image, whereas the traceless part of Q_{ij} contains the ellipticity information. From Q_{ij} , one defines two complex ellipticities,

$$\chi \equiv \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22}} \quad \text{and} \quad \epsilon \equiv \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22} + 2(Q_{11}Q_{22} - Q_{12}^2)^{1/2}} . \quad (7)$$

Both of them have the same phase (because of the same numerator), but a different absolute value. Fig. 2 illustrates the shape of images as a function of their complex ellipticity χ . For an image with elliptical isophotes of axis ratio $r \leq 1$, one obtains

$$|\chi| = \frac{1 - r^2}{1 + r^2} \quad ; \quad |\epsilon| = \frac{1 - r}{1 + r} . \quad (8)$$

Which of these two definitions is more convenient depends on the context; one can easily transform one into the other,

$$\epsilon = \frac{\chi}{1 + (1 - |\chi|^2)^{1/2}} , \quad \chi = \frac{2\epsilon}{1 + |\epsilon|^2} . \quad (9)$$

In fact, other (but equivalent) ellipticity definitions have been used in the literature (e.g., Kochanek 1990; Miralda-Escudé 1991; Bonnet & Mellier 1995), but the two given above appear to be most convenient.

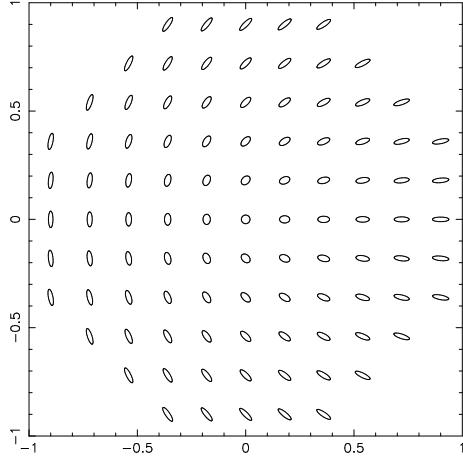


Fig. 2. The shape of image ellipses for a circular source, in dependence on their two ellipticity components χ_1 and χ_2 ; a corresponding plot in term of the ellipticity components ϵ_i would look quite similar. Note that the ellipticities are rotated by 90° when $\chi \rightarrow -\chi$ (source: D. Clowe)

From source to image ellipticities. In total analogy, one defines the second-moment brightness tensor $Q_{ij}^{(s)}$, and the complex ellipticities $\chi^{(s)}$ and $\epsilon^{(s)}$ for the unlensed source. From

$$Q_{ij}^{(s)} = \frac{\int d^2\beta I^{(s)}(\boldsymbol{\theta}) q_I[I^{(s)}(\boldsymbol{\beta})] (\beta_i - \bar{\beta}_i) (\beta_j - \bar{\beta}_j)}{\int d^2\beta I^{(s)}(\boldsymbol{\theta}) q_I[I^{(s)}(\boldsymbol{\beta})]} , \quad i, j \in \{1, 2\} , \quad (10)$$

one finds with $d^2\beta = \det \mathcal{A} d^2\theta$, $\boldsymbol{\beta} - \bar{\boldsymbol{\beta}} = \mathcal{A}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})$, that

$$Q^{(s)} = \mathcal{A} Q \mathcal{A}^T = \mathcal{A} Q \mathcal{A} , \quad (11)$$

where $\mathcal{A} \equiv \mathcal{A}(\bar{\boldsymbol{\theta}})$. Using the definitions of the complex ellipticities, one finds the transformations (e.g., Schneider & Seitz 1995; Seitz & Schneider 1997)

$$\chi^{(s)} = \frac{\chi - 2g + g^2\chi^*}{1 + |g|^2 - 2\Re(g\chi^*)} ; \quad \epsilon^{(s)} = \begin{cases} \frac{\epsilon - g}{1 - g^*\epsilon} & \text{if } |g| \leq 1 ; \\ \frac{1 - g\epsilon^*}{\epsilon^* - g^*} & \text{if } |g| > 1 . \end{cases} \quad (12)$$

The inverse transformations are obtained by interchanging source and image ellipticities, and $g \rightarrow -g$ in the foregoing equations.

Estimating the (reduced) shear. In the following we make the assumption that the intrinsic orientation of galaxies is random,

$$E(\chi^{(s)}) = 0 = E(\epsilon^{(s)}) , \quad (13)$$

which is expected to be valid since there should be no direction singled out in the Universe. This then implies that the expectation value of ϵ is [as obtained

by averaging the transformation law (12) over the intrinsic source orientation]

$$E(\epsilon) = \begin{cases} g & \text{if } |g| \leq 1 \\ 1/g^* & \text{if } |g| > 1. \end{cases} \quad (14)$$

This is a remarkable result (Schramm & Kaiser 1995; Seitz & Schneider 1997), since it shows that each image ellipticity provides an unbiased estimate of the local shear, though a very noisy one. The noise is determined by the intrinsic ellipticity dispersion

$$\sigma_\epsilon = \sqrt{\langle \epsilon^{(s)} \epsilon^{(s)*} \rangle},$$

in the sense that, when averaging over N galaxy images all subject to the same reduced shear, the $1-\sigma$ deviation of their mean ellipticity from the true shear is σ_ϵ/\sqrt{N} . A more accurate estimate of this error is

$$\sigma = \sigma_\epsilon [1 - \min(|g|^2, |g|^{-2})] / \sqrt{N} \quad (15)$$

(Schneider et al. 2000). Hence, the noise can be beaten down by averaging over many galaxy images; however, the region over which the shear can be considered roughly constant is limited, so that averaging over galaxy images is always related to a smoothing of the shear. Fortunately, we live in a Universe where the sky is ‘full of faint galaxies’, as was impressively demonstrated by the Hubble Deep Field images (Williams et al. 1996) and previously from ultra-deep ground-based observations (Tyson 1987). Therefore, the accuracy of a shear estimate depends on the local number density of galaxies for which a shape can be measured. In order to obtain a high density, one requires deep imaging observations. As a rough guide, on a 3 hour exposure with a 4-meter class telescope, about 30 galaxies per arcmin² can be used for a shape measurement.

In fact, considering (14) we conclude that the expectation value of the observed ellipticity is the same for a reduced shear g and for $g' = 1/g^*$. Schneider & Seitz (1995) have shown that one cannot distinguish between these two values of the reduced shear from a purely local measurement, and term this fact the ‘local degeneracy’; this also explains the symmetry between $|g|$ and $|g|^{-1}$ in (15). Hence, from a local weak lensing observation one cannot tell the case $|g| < 1$ (equivalent to $\det \mathcal{A} > 0$) from the one of $|g| > 1$ or $\det \mathcal{A} < 0$. This local degeneracy is, however, broken in large-field observations, as the region of negative parity of any lens is small (the Einstein radius inside of which $|g| > 1$ of massive lensing clusters is typically $\lesssim 30''$, compared to data fields of several arcminutes used for weak lensing studies of clusters), and the reduced shear must be a smooth function of position on the sky.

Whereas the transformation between source and image ellipticity appears simpler in the case of χ than ϵ – see (12), the expectation value of χ cannot be easily calculated and depends explicitly on the intrinsic ellipticity distribution

of the sources. In particular, the expectation value of χ is not simply related to the reduced shear (Schneider & Seitz 1995). However, in the weak lensing regime, $\kappa \ll 1$, $|\gamma| \ll 1$, one finds

$$\gamma \approx g \approx \langle \epsilon \rangle \approx \frac{\langle \chi \rangle}{2}. \quad (16)$$

2.3 Tangential and cross component of shear

Components of the shear. The shear components γ_1 and γ_2 are defined relative to a reference Cartesian coordinate frame. Note that the shear is *not* a vector (though it is often wrongly called that way in the literature), owing to its transformation properties under rotations: Whereas the components of a vector are multiplied by $\cos \varphi$ and $\sin \varphi$ when the coordinate frame is rotated by an angle φ , the shear components are multiplied by $\cos(2\varphi)$ and $\sin(2\varphi)$, or simply, the complex shear gets multiplied by $e^{-2i\varphi}$. The reason for this transformation behavior of the shear traces back to its original definition as the traceless part of the Jacobi matrix \mathcal{A} . This transformation behavior is the same as that of the linear polarization; the shear is therefore a *polar*. In analogy with vectors, it is often useful to consider the shear components in a rotated reference frame, that is, to measure them w.r.t. a different direction; for example, the arcs in clusters are tangentially aligned, and so their ellipticity is oriented tangent to the radius vector in the cluster.

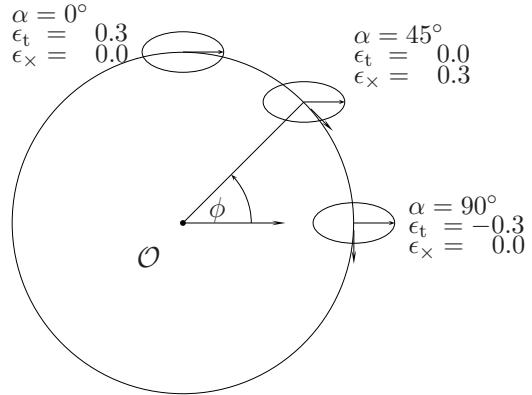


Fig. 3. Illustration of the tangential and cross-components of the shear, for an image with $\epsilon_1 = 0.3$, $\epsilon_2 = 0$, and three different directions ϕ with respect to a reference point (source: M. Bradac)

If ϕ specifies a direction, one defines the *tangential* and *cross components* of the shear *relative to this direction* as

$$\gamma_t = -\mathcal{R}\text{e} [\gamma e^{-2i\phi}] \quad , \quad \gamma_x = -\mathcal{I}\text{m} [\gamma e^{-2i\phi}] ; \quad (17)$$

For example, in case of a circularly-symmetric matter distribution, the shear at any point will be oriented tangent to the direction towards the center

of symmetry. Thus in this case choose ϕ to be the polar angle of a point; then, $\gamma_x = 0$. In full analogy to the shear, one defines the tangential and cross components of an image ellipticity, ϵ_t and ϵ_x . An illustration of these definitions is provided in Fig. 3.

The sign in (17) is easily explained (and memorized) as follows: consider a circular mass distribution and a point on the θ_1 -axis outside the Einstein radius. The image of a circular source there will be stretched in the direction of the θ_2 -axis. In this case, $\phi = 0$ in (17), the shear is real and negative, and in order to have the tangential shear positive, and thus to define tangential shear in accordance with the intuitive understanding of the word, a minus sign is introduced. Negative tangential ellipticity implies that the image is oriented in the radial direction. We warn the reader that sign conventions and notations have undergone several changes in the literature, and the current author had his share in this.

Minimum lens strength for its weak lensing detection. As a first application of this decomposition, we consider how massive a lens needs to be in order that it produces a detectable weak lensing signal. For this purpose, consider a lens modeled as an SIS with one-dimensional velocity dispersion σ_v . In the annulus $\theta_{\text{in}} \leq \theta \leq \theta_{\text{out}}$, centered on the lens, let there be N galaxy images with positions $\theta_i = \theta_i(\cos \phi_i, \sin \phi_i)$ and (complex) ellipticities ϵ_i . For each one of them, consider the tangential ellipticity

$$\epsilon_{ti} = -\mathcal{R}\text{e}(\epsilon_i e^{-2i\phi_i}) . \quad (18)$$

The weak lensing signal-to-noise for the detection of the lens obtained by considering a weighted average over the tangential ellipticity is (see BS01, Sect. 4.5)

$$\begin{aligned} \frac{S}{N} &= \frac{\theta_E}{\sigma_\epsilon} \sqrt{\pi n} \sqrt{\ln(\theta_{\text{out}}/\theta_{\text{in}})} \\ &= 8.4 \left(\frac{n}{30 \text{ arcmin}^{-2}} \right)^{1/2} \left(\frac{\sigma_\epsilon}{0.3} \right)^{-1} \left(\frac{\sigma_v}{600 \text{ km s}^{-1}} \right)^2 \\ &\times \left(\frac{\ln(\theta_{\text{out}}/\theta_{\text{in}})}{\ln 10} \right)^{1/2} \left\langle \frac{D_{\text{ds}}}{D_s} \right\rangle , \end{aligned} \quad (19)$$

where $\theta_E = 4\pi(\sigma_v/c)^2(D_{\text{ds}}/D_s)$ is the Einstein radius of an SIS, n the mean number density of galaxies, and the average of the distance ratio is taken over the source population from which the shear measurements are obtained. Hence, the S/N is proportional to the lens strength (as measured by θ_E), the square root of the number density, and inversely proportional to σ_ϵ , as expected. From this consideration we conclude that clusters of galaxies with $\sigma_v \gtrsim 600 \text{ km/s}$ can be detected with sufficiently large S/N by weak lensing, but individual galaxies ($\sigma_v \lesssim 200 \text{ km/s}$) are too weak as lenses to be detected

individually. Furthermore, the final factor in (19) implies that, for a given source population, the cluster detection will be more difficult for increasing lens redshift.

Mean tangential shear on circles. In the case of axi-symmetric mass distributions, the tangential shear is related to the surface mass density $\kappa(\theta)$ and the mean surface mass density $\bar{\kappa}(\theta)$ inside the radius θ by $\gamma_t = \bar{\kappa} - \kappa$, as can be easily shown by the relation in Sect. 3.1 of IN. It is remarkable that a very similar expression holds for general matter distributions. To see this, we start from Gauss' theorem, which states that

$$\int_0^\theta d^2\vartheta \nabla \cdot \nabla \psi = \theta \oint d\varphi \nabla \psi \cdot \mathbf{n},$$

where the integral on the left-hand side extends over the area of a circle of radius θ (with its center chosen as the origin of the coordinate system), ψ is an arbitrary scalar function, the integral on the right extends over the circle with radius θ , and \mathbf{n} is the outward directed normal on this circle. Taking ψ to be the deflection potential and noting that $\nabla^2\psi = 2\kappa$, one obtains

$$m(\theta) \equiv \frac{1}{\pi} \int_0^\theta d^2\vartheta \kappa(\vartheta) = \frac{\theta}{2\pi} \oint d\varphi \frac{\partial \psi}{\partial \theta}, \quad (20)$$

where we used that $\nabla \psi \cdot \mathbf{n} = \psi_{,\theta\theta}$. Differentiating this equation with respect to θ yields

$$\frac{dm}{d\theta} = \frac{m}{\theta} + \frac{\theta}{2\pi} \oint d\varphi \frac{\partial^2 \psi}{\partial \theta^2}. \quad (21)$$

Consider a point on the θ_1 -axis; there, $\psi_{,\theta\theta} = \psi_{11} = \kappa + \gamma_1 = \kappa - \gamma_t$. This last expression is independent on the choice of coordinates and must therefore hold for all φ . Denoting by $\langle \kappa(\theta) \rangle$ and $\langle \gamma_t(\theta) \rangle$ the mean surface mass density and mean tangential shear on the circle of radius θ , (21) becomes

$$\frac{dm}{d\theta} = \frac{m}{\theta} + \theta [\langle \kappa(\theta) \rangle - \langle \gamma_t(\theta) \rangle]. \quad (22)$$

The dimensionless mass $m(\theta)$ in the circle is related to the mean surface mass density inside the circle $\bar{\kappa}(\theta)$ by

$$m(\theta) = \theta^2 \bar{\kappa}(\theta) = 2 \int_0^\theta d\vartheta \vartheta \langle \kappa(\vartheta) \rangle. \quad (23)$$

Together with $dm/d\theta = 2\theta \langle \kappa(\theta) \rangle$, (22) becomes, after dividing through θ ,

$$\langle \gamma_t \rangle = \bar{\kappa} - \langle \kappa \rangle, \quad (24)$$

a relation which very closely matches the result mentioned above for axi-symmetric mass distributions (Bartelmann 1995). One important immediate implication of this result is that from a measurement of the tangential shear, averaged over concentric circles, one can determine the azimuthally-averaged mass profile of lenses, even if the density is not axi-symmetric.

2.4 Magnification effects

Recall from IN that a magnification μ changes source counts according to

$$n(> S, \theta, z) = \frac{1}{\mu(\theta, z)} n_0 \left(> \frac{S}{\mu(\theta, z)}, z \right), \quad (25)$$

where $n(> S, z)$ and $n_0(> S, z)$ are the lensed and unlensed cumulative number densities of sources, respectively. The first argument of n_0 accounts for the change of the flux (which implies that a magnification $\mu > 1$ allows the detection of intrinsically fainter sources), whereas the prefactor in (25) stems from the change of apparent solid angle. In the case that $n_0(S) \propto S^{-\alpha}$, this yields

$$\frac{n(> S)}{n_0(> S)} = \mu^{\alpha-1}, \quad (26)$$

and therefore, if $\alpha > 1$ (< 1), source counts are enhanced (depleted); the steeper the counts, the stronger the effect. In the case of weak lensing, where $|\mu - 1| \ll 1$, one probes the source counts only over a small range in flux, so that they can always be approximated (locally) by a power law. Provided that $\kappa \ll 1$, $|\gamma| \ll 1$, a further approximation applies,

$$\mu \approx 1 + 2\kappa; \quad \text{and} \quad \frac{n(> S)}{n_0(> S)} \approx 1 + 2(\alpha - 1)\kappa. \quad (27)$$

Thus, from a measurement of the local number density $n(> S)$ of galaxies, κ can in principle be inferred directly. It should be noted that $\alpha \sim 1$ for galaxies in the B-band, but in redder bands, $\alpha < 1$ (e.g., Ellis 1997); therefore, one expects a depletion of their counts in regions of magnification $\mu > 1$. Broadhurst et al. (1995) have discussed in detail the effects of magnification in weak lensing. Not only are the number counts affected, but since this is a redshift-dependent effect (since both κ and γ depend, for a given physical surface mass density, on the source redshift), the redshift distribution of galaxies is locally changed by magnification.

Since magnification is merely a stretching of solid angle, Bartelmann & Narayan (1995) pointed out that magnified images at fixed surface brightness have a larger solid angle than unlensed ones; in addition, the surface brightness of a galaxy is expected to be a strong function of redshift [$I \propto (1+z)^{-4}$], owing to the Tolman effect. Hence, if this effect could be harnessed, a (redshift-dependent) magnification could be measured statistically. Unfortunately, this method is hampered by observational difficulties; it seems that estimating a reliable estimate for the surface brightness from seeing-convolved images (see Sect. 3.5) is even more difficult than determining image shapes.

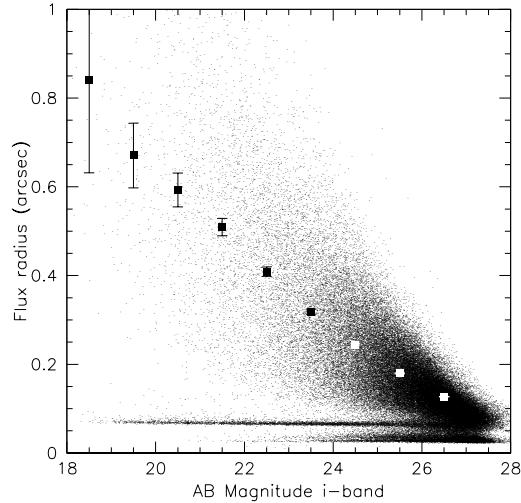


Fig. 4. The size of galaxies observed with the ACS camera onboard HST. Small dots denote the half-light radius of individual galaxies, bigger points with error bars show the mean size in a magnitude bin. The horizontal line of point at $r_h \approx 0.^{\prime\prime}08$ correspond to stellar images in the ACS fields, as they have all the same size but vary in magnitude, and points at even smaller size are noise artefacts which are not used for any lensing analysis (source: T. Schrabback)

3 Observational issues and challenges

Weak lensing, employing the shear method, relies on the shape measurements of faint galaxy images. Since the noise due to intrinsic ellipticity dispersion is $\propto \sigma_\epsilon/\sqrt{n}$, one needs a high number density n to beat this noise component down. However, the only way to increase the number density of galaxies is to observe to fainter magnitudes. As it turns out, galaxies at faint magnitudes are small, in fact typically smaller than the size of the point-spread function (PSF), or the seeing disk (see Fig. 4). Hence, for them one needs usually large correction factors between the true ellipticity and that of the seeing-convolved image. On the other hand, fainter galaxies tend to probe higher-redshift galaxies, which increases the lensing signal due to D_{ds}/D_s -dependence of the ‘lensing efficiency’.

3.1 Strategy

In the present observational situation, only the optical sky is densely populated with sources; therefore, weak lensing observations are performed with optical (or near-IR) CCD-cameras (photometric plates are not linear enough to measure these subtle effects). In order to substantiate this comment, note that the Hubble Deep Field North contains about 3000 galaxies, but only seven radio sources are detected in a very deep integration with the VLA (Richards et al. 1998).² In order to obtain a high number density of sources,

² The source density on the radio sky will become at least comparable to that currently on the optical sky with the future Square Kilometer Array (SKA).

long exposures are needed: as an illustrative example, to get a number density of useful galaxies (i.e., those for which a shape can be measured reliably) of $n \sim 20 \text{ arcmin}^{-2}$, one needs ~ 2 hours integration on a 4-m class telescope in good seeing $\sigma \lesssim 1''$.

Furthermore, large solid angles are desired, either to get large areas around clusters for their mass reconstruction, or to get good statistics of lenses on blank field surveys, such as they are needed for galaxy-galaxy lensing and cosmic shear studies. It is now possible to cover large area in reasonable amounts of observing time, since large format CCD cameras have recently become available; for example, the Wide-Field Imager (WFI) at the ESO/MPG 2.2-m telescope at La Silla has $(8\text{K})^2$ pixels and covers an area of $\sim (0.5 \text{ deg})^2$. Until recently, the CFH12K camera with $8\text{K} \times 12\text{K}$ pixels and field $\sim 30' \times 45'$ was mounted at the Canada-French-Hawaii Telescope (CFHT) on Mauna Kea and was arguably the most efficient wide-field imaging instrument hitherto. In 2003, MegaCam has been put into operation on the CFTH which has $(18\text{K})^2$ pixels and covers $\sim 1 \text{ deg}^2$. Several additional cameras of comparable size will become operational in the near future, including the 1 deg^2 instrument OmegaCAM on the newly built VLT Survey Telescope on Paranal. The largest field camera on a 10-m class telescope is SuprimeCAM, a $34' \times 27'$ multi-chip camera on the Subaru 8.2-meter telescope. Unfortunately, many optical astronomers (and decision making panels of large facilities) consider the prime use of large telescopes to be spectroscopy; for example, although the four ESO VLT unit telescopes are equipped with a total of ten instruments, the largest imagers on the VLT are the two FORS instruments, with a $\sim 6'.7$ field-of-view.³

The typical pixel size of these cameras is $\sim 0''.2$, which is needed to sample the seeing disk in times of good seeing. From Fig. 4 one concludes immediately that the seeing conditions are absolutely critical for weak lensing: an image with $0''.6$ is substantially more useful than one with taken under the more typical condition of $0''.8$ (see Fig. 5). There are two separate reasons why the seeing is such an important factor. First, seeing blurs the images and make them rounder; accordingly, to correct for the seeing effect, a larger correction factor is needed in the worse seeing conditions. In addition, since the galaxy images from which the shear is to be determined are faint, a larger seeing smears the light from these galaxies over a larger area on the sky, reducing its contrast relative to the sky noise, and therefore leads to noisier estimates of the ellipticities even before the correction.

³ Nominally, the VIMOS instrument has a four times larger f.o.v., but our analysis of early VIMOS imaging data indicates that it is totally useless for weak lensing observations, owing to its highly anisotropic PSF, which even seems to show discontinuities on chips, and its large variation of the seeing size across chips. It may be hoped that some of these image defects are improved after a complete overhaul of the instrument which occurred recently.

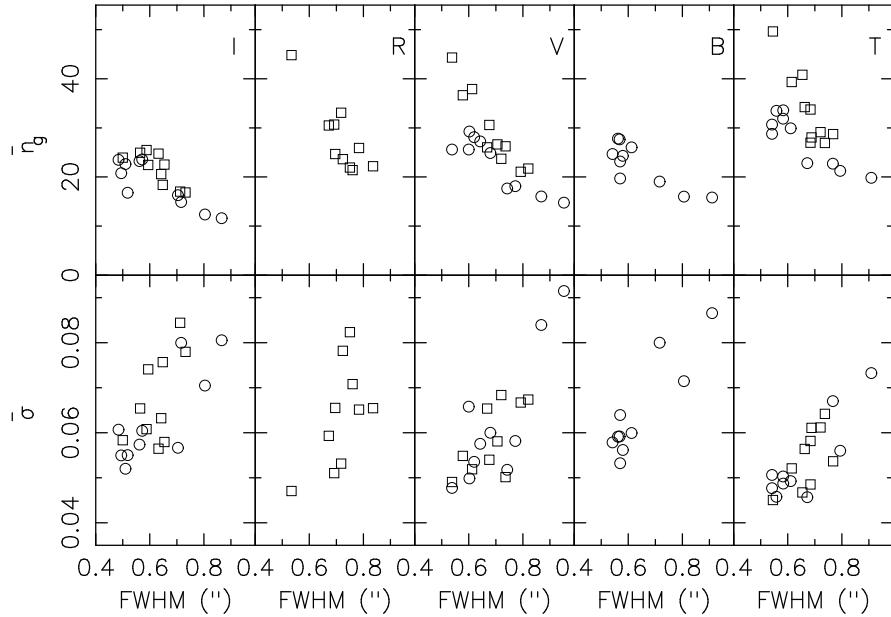


Fig. 5. Mean number density of galaxy images for which a shape can be measured (upper row) and the r.m.s. noise of a shear measurement in an area of 1 arcmin² as a function of the full width at half maximum (FWHM) of the point-spread function (PSF) – i.e., the seeing. The data were taken on 20 different fields with the FORS2 instrument at the VLT, with different filters (I, R, V and R). Squares show data taken with about 2 hours integration time, circles those with ~ 45 min exposure. The right-most panels show the coadded data of I,R,V for the long exposures, and I,V,B for the 45 min fields. The useful number of galaxy images is seen to be a strong function of the seeing, except for the I-band (which is related to the higher sky brightness and the way objects are detected). But even more dramatically, the noise due to intrinsic source ellipticity decreases strongly for better seeing conditions, which is due to (1) higher number density of galaxies for which a shape can be measured, and (2) smaller corrections for PSF blurring, reducing the associated noise of this correction. In fact, this figure shows that seeing is a more important quantity than the total exposure time (from Clowe et al. 2004b)

Deep observations of a field require multiple exposures. As a characteristic number, the exposure time for an R-band image on a 4-m class telescope is not longer than ~ 10 min to avoid the non-linear part of the CCD sensitivity curve (exposures in shorter wavelength bands can be longer, since the night sky is fainter in these filters). Therefore, these large-format cameras imply a high data rate; e.g., one night of observing with the WFI yields ~ 30 GB of science and calibration data. This number will increase by a factor ~ 6 for MegaCam. Correspondingly, handling this data requires large disk space for efficient data reduction.

3.2 Data reduction: Individual frames

We shall now consider a number of issues concerning the reduction of imaging data, starting here with the steps needed to treat individual chips on individual frames, and later consider aspects of combining them into a coadded image.

Flatfielding. The pixels of a CCD have different sensitivity, i.e., they yield different counts for a given amount of light falling onto them. In order to calibrate the pixel sensitivity, one needs flatfielding. Three standard methods for this are in use:

1. Dome-flats: a uniformly illuminated screen in the telescope dome is exposed; the counts in the pixels are then proportional to their sensitivity. The problem here is that the screen is not really of uniform brightness.
2. Twilight-flats: in the period of twilight after sunset, or before sunrise, the cloudless sky is nearly uniformly bright. Short exposures of regions of the sky without bright stars are then used to calibrate the pixel sensitivity.
3. Superflats: if many exposures with different pointings are taken with a camera during a night, then any given pixel is not covered by a source for most of the exposures (because the fraction of the sky at high galactic latitudes which is covered by objects is fairly small, as demonstrated by the deep fields taken by the HST). Hence, the (exposure-time normalized) counts of any pixel will show, in addition to a little tail due to those exposures when a source has covered it, a distribution around its sensitivity to the uniform night-sky brightness; from that distribution, the flat-field can be constructed, by taking its mode or its median.

Bad pixels. Each CCD has defects, in that some pixels are dead or show a signal unrelated to their illumination. This can occur as individual pixels, or whole pixel columns. No information of the sky image is available at these pixel positions. One therefore employs dithering: several exposures of the same field, but with slightly different pointings (dither positions) are taken. Then, any position of the field falls on bad pixels only in a small fraction of exposures, so that the full two-dimensional brightness distribution can be recovered.

Cosmic rays. Those mimic groups of bad pixels; they can be removed owing to the fact that a given point of the image will most likely be hit by a cosmic ray only once, so that by comparison between the different exposures, cosmic rays can be removed (or more precisely, masked). Another signature of a cosmic ray is that the width of its track is typically much smaller than the seeing disk, the minimum size of any real source.

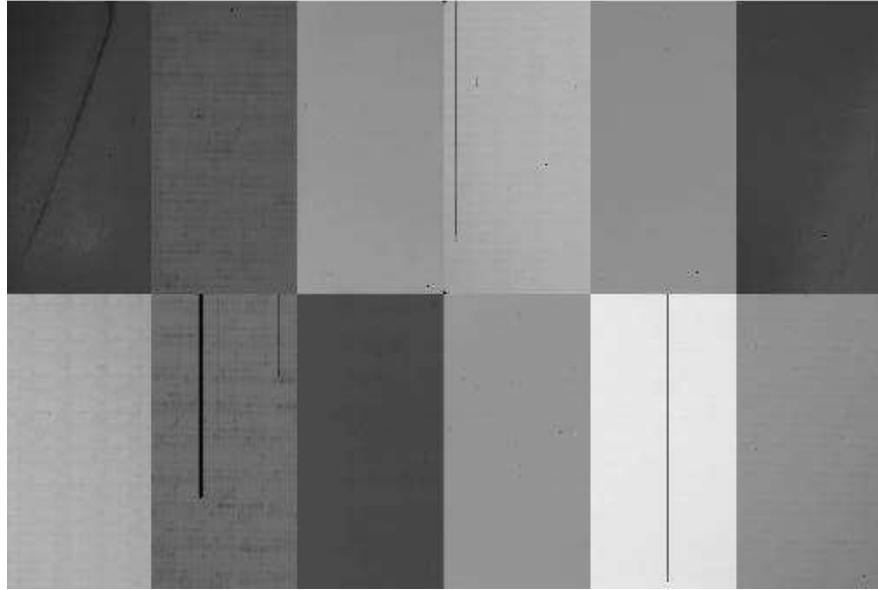


Fig. 6. A flat field for the CFH12K camera, showing the sensitivity variations between pixels and in particular between chips. Also, bad columns are clearly seen

Bright stars. Those cause large diffraction spikes, and depending on the optics and the design of the camera, reflection rings, ghost images and other unwanted features. It is therefore best to choose fields where no or very few bright stars are present. The diffraction spikes of stars need to be masked, as well as the other features just mentioned.

Fringes. Owing to light reflection within the CCD, patterns of illumination across the field can be generated (see Fig. 8); this is particularly true for thin chips when rather long wavelength filters are used. In clear nights, the fringe pattern is stable, i.e., essentially the same for all images taken during the night; in that case, it can be deduced from the images and subtracted off the individual exposures. However, if the nights are not clear, this procedure no longer works well; it is then safer to observe at shorter wavelength. For example, for the WFI, fringing is a problem for I-band images, but for the R-band filter, the amplitude of fringing is small. For the FORS instruments at the VLT, essentially no fringing occurs even in the I band (Maoli et al. 2001).

Gaps. The individual CCDs in multi-chip cameras cannot be brought together arbitrarily close; hence, there are gaps between the CCDs (see Fig. 9 for an example). In order to cover the gaps, the dither pattern can be chosen such as to cover the gaps, so that they fall on different parts of the sky in

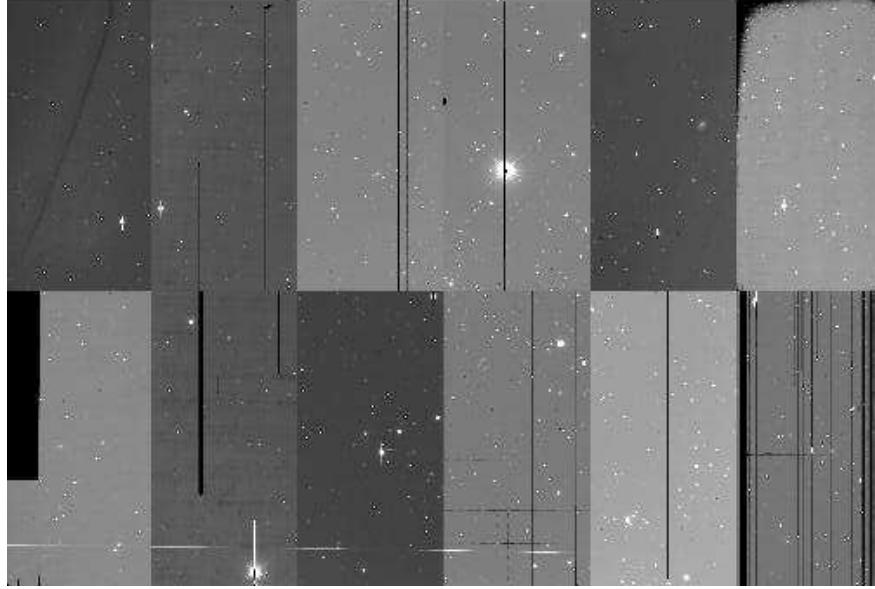


Fig. 7. A raw frame from the CFH12K camera, showing quite a number of effects mentioned in the text: bad column, saturation of bright stars, bleeding, and sensitivity variations across the field and in particular between chips

different exposures. As we shall see, such relatively large dither patterns also provide additional advantages.

Satellite trails, asteroid trails. Those have to be identified, either by visual inspection (currently the default) or by image recognition software which can detect these linear features which occur either only once, or at different positions on different exposures. These are then masked, in the same way as some of the other features mentioned above.

3.3 Data reduction: coaddition

After taking several exposures with slightly different pointing positions (for the reasons given above), frames shall be coadded to a sum-frame; some of the major steps in this coaddition procedure are:

Astrometric solution. One needs to coadd data from the same true (or sky) position, not the same pixel position. Therefore, one needs a very precise mapping from sky coordinates to pixel coordinates. Field distortions, which occur in every camera (and especially so in wide-field cameras), make this mapping non-linear (see Fig. 10). Whereas the distortion map of the telescope/camera system is to a large degree constant and therefore one of the

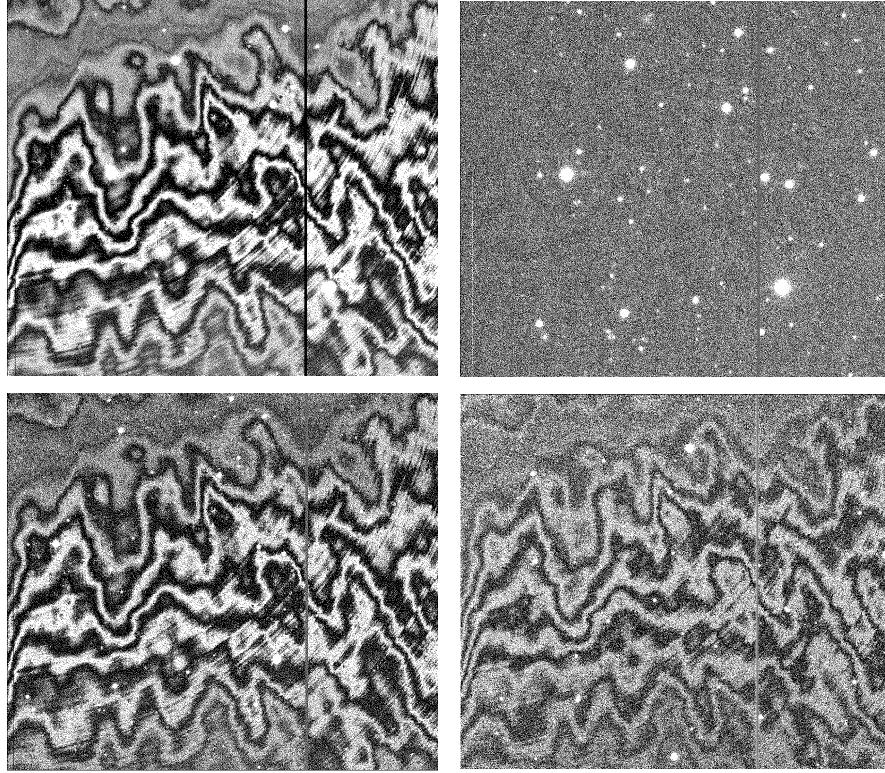


Fig. 8. The two left panels show the fringe patterns of images taken with the WFI in the I-band; the upper one was taken during photometric conditions, the lower one under non-photometric conditions. Since the fringe pattern is spatially stable, it can be corrected for (left panels), but the result is satisfactory only in the former case (source: M. Schirmer & T. Erben)

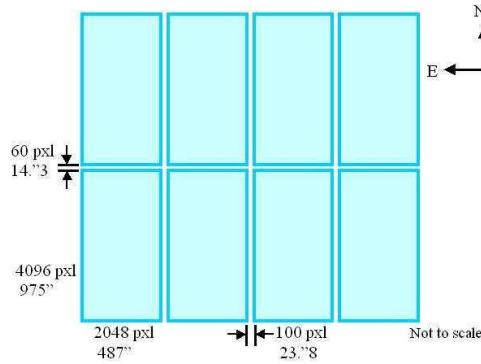


Fig. 9. Layout of the Wide Field Imager (WFI) at the ESO/MPG 2.2m telescope at La Silla. The eight chips each have $\sim 2048 \times 4096$ pixels and cover $\sim 7.5 \times 15'$

known features, it is not stable to the sub-pixel accuracy needed for weak lensing work, owing to its dependence on the zenith angle (geometrical distortions of the telescope due to gravity), temperature etc. Therefore, the pixel-to-sky mapping has to be obtained from the data itself. Two methods are used to achieve this: one of them makes use of an external reference catalog, such as the US Naval Observatory catalogue for point sources; it contains about 2 point sources per arcmin² (at high Galactic latitudes) with ~ 0.3 arcsec positional accuracy. Matching point sources on the exposures with those in the USNO catalog therefore yields the mapping with sub-arcsecond accuracy. Far higher accuracy of the relative astrometry is achieved (and needed) from internal astrometry, which is obtained by matching objects which appear at different pixel coordinates, and in particular, on different CCDs for the various dithering positions. Whereas the sky coordinates are constant, the pixel coordinates change between dithering positions. Since the distortion map can be described by a low-order polynomial, the comparison of many objects appearing at (substantially) different pixel positions yield many more constraints than the free parameters in the distortion map and thus yields the distortion map with much higher relative accuracy than external data. The corresponding astrometric solution can routinely achieve an accuracy of 0.1 pixel, or typically 0''.02 – compared with a typical field size of $\sim 30'$.

Photometric solution. Flatfielding corrects for the different sensitivities of the pixels and therefore yields accurate relative photometry across individual exposures. The different exposures are tied together by matching the brightness of joint objects, in particular across chip boundaries. To achieve an absolute photometric calibration, one needs external data (e.g., standard star observations).

The coaddition process. Coaddition has to happen with sub-pixel accuracy; hence, one cannot just shift pixels from different exposures on top of each other, although this procedure is still used by some groups. The by-now standard method is drizzling (Fruchter & Hook 2002), in which a new pixel frame is defined which usually has smaller pixel size than the original image pixels (typically by a factor of two) and which is linearly related to the sky coordinates. The astrometrically and photometrically calibrated individual frames are now remapped onto this new pixel grid, and the pixel values are summed up into the sub-pixel grid, according to the overlap area between exposure pixel and drizzle pixel (see Fig. 11). By that, drizzling automatically is flux conserving. In the coaddition process, weights are assigned, accounting for the noise properties of the individual exposures (including the masks, of course).

The result of the coaddition procedure is then a science frame, plus a weight map which contains information about the pixel noise, which is of course spatially varying, owing to the masks, CCD gaps, removed cosmic

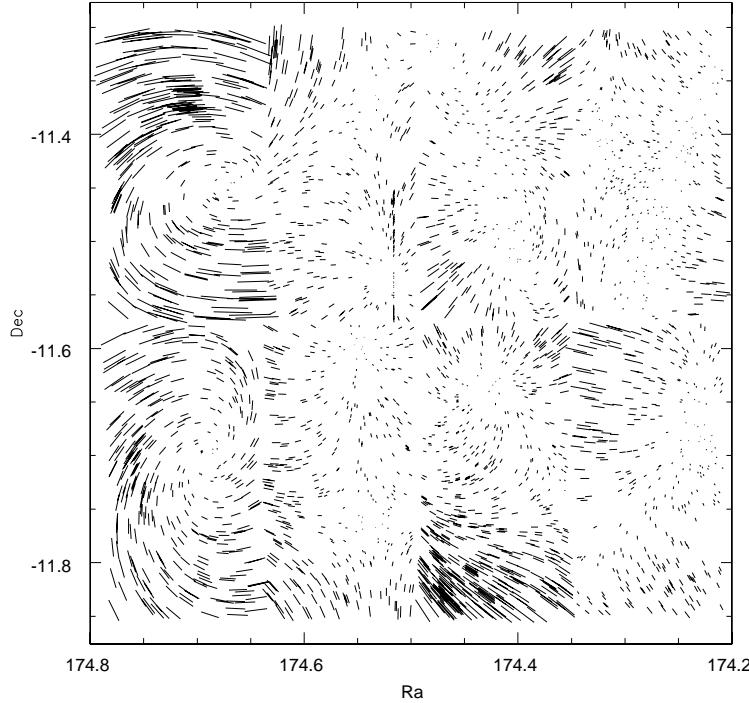


Fig. 10. This figure shows the geometric distortion of the WFI. Plotted is the difference of the positions of stars as obtained from a simple translation, and a third-order astrometric correction obtained in the process of image reduction. The patterns in the two left chips is due to their rotation relative to the other six chips. Whereas this effect looks dramatic at first sight, the maximum length of the sticks corresponds to about 6 pixels, or $\sim 1''.2$. Given that the WFI covers a field of $\sim 33''$, the geometrical distortions are remarkably small – however, they are sufficiently large that they have to be taken into account in the coaddition process (source: T. Erben & M. Schirmer)

rays and bad pixels. Fig. 12 shows a typical example of a coadded image and its corresponding weight map.

The quality of the coadded image can be checked in a number of ways. Coaddition should not erase information contained in the original exposures (except, of course, the variability of sources). This means that the PSF of the coadded image should not be larger than the weighted mean of the PSFs of the individual frames. Insufficient relative astrometry would lead to a blurring of images in the coaddition. Furthermore, the anisotropy of the PSF should be similar to the weighted mean of the PSF anisotropies of the individual frames; again, insufficient astrometry could induce an artificial anisotropy of

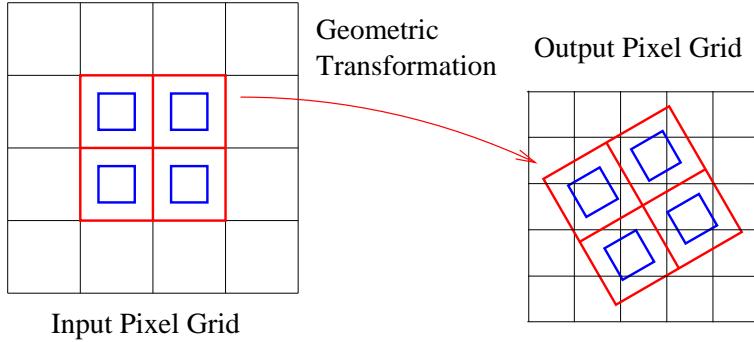


Fig. 11. The principle of drizzling in the process of coaddition is shown. The pixel grid of each individual exposure is mapped onto an output grid, where the shifts and geometric distortions obtained during the astrometric solutions are applied. The counts of the input pixel, multiplied by the relative weight of this pixel, are then dropped onto the output pixels, according to the relative overlap area, where the output pixels can be chosen smaller than the input pixels. The same procedure is applied to the weight maps of the individual exposures. If many exposures are coadded, the input pixel can also be shrunk before dropping onto the output pixel. After processing all individual exposures in this way, a coadded image and a coadded weight map is obtained (source: T. Schrabbach)

the PSF in the coaddition (which can be easily visualized, by adding two round images with a slight center offset, where a finite ellipticity would be induced).

Probably, there does not exist the ‘best’ coadded image from a given set of individual exposures. This can be seen by considering a set of exposures with fairly different individual seeing. If one is mainly interested in photometric properties of rather large galaxies, one would prefer a coaddition which puts all the individual exposures together, in order to maximize the total exposure time and therefore to minimize the photometric noise of the coadded sources. For weak lensing purposes, such a coaddition is certainly not optimal, as adding exposures with bad seeing together with those of good seeing creates a coadded image with a seeing intermediate between the good and the bad. Since seeing is a much more important quantity than depth for the shape determination of faint and small galaxy images, it would be better to coadd only the images with the good seeing. In this respect, the fact that large imaging instruments are operated predominantly in service observing more employing queue scheduling is a very valuable asset: data for weak lensing studies are then taken only if the seeing is better than a specified limit; in this way one has a good chance to get images of homogeneously good seeing conditions.

As a specific example, we show in Fig. 13 the ‘deepest wide-field image in the Southern sky’, targeted towards the Chandra Deep Field South, one

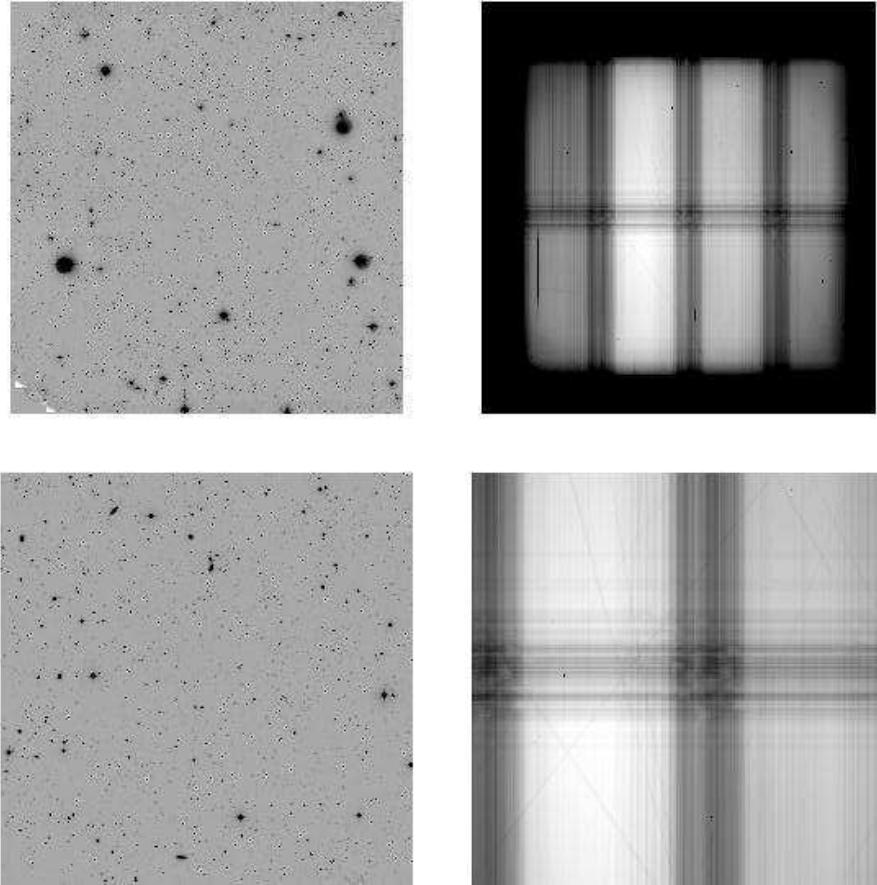


Fig. 12. A final coadded frame from a large number of individual exposures with the WFI is shown in the upper left panel, with the corresponding weight map at the upper right. The latter clearly shows the large-scale inhomogeneity of the chip sensitivity and the illumination, together with the different number of exposures contributing to various regions in the output image due to dithering and the gaps between CCDs. The two lower panels show a blow-up of the central part. Despite the highly inhomogeneous weight, the coadded image apparently shows no tracer of the gaps, which indicates that a highly accurate relative photometric solution was obtained (source: T. Erben & M. Schirmer)

of regions in the sky in which all major observatories have agreed to obtain, and make publically available, very deep images for a detailed multi-band study. For example, the Hubble Ultra Deep Field (Beckwith et al. 2003) is located in the CDFS, the deepest Chandra X-ray exposures are taken in this field, as well as two ACS@HST mosaic images, one called the GOODS field (Great Observatories Origins Deep Survey; cf. Giavalisco & Mobasher 2004), the other the GEMS survey (Rix et al. 2004).

3.4 Image analysis

The final outcome of the data reduction steps described above is an image of the sky, together with a weight map providing the noise properties of the image. The next step is the scientific exploitation of this image, which in the case of weak lensing includes the identification of sources, and to measure their magnitude, size and shape.

As a first step, individual sources on the image need to be identified, to obtain a catalog of sources for which the ellipticities, sizes and magnitudes are to be determined later. This can done with by-now standard software, like SExtractor (Bertin & Arnouts 1996), or may be part of specialized software packages developed specifically for weak lensing, such as IMCAT, developed by Nick Kaiser (see below). Although this first step seems straightforward at first glance, it is not: images of sources can be overlapping, the brightness distribution of many galaxies (in particular those with active star formation) tends to be highly structured, with a collection of bright spots, and therefore the software must be taught whether or not these are to be split into different sources, or be taken as one (composite) source. This is not only a software problem; in many cases, even visual inspection cannot decide whether a given light distribution corresponds to one or several sources. The shape and size of the images are affected by the point-spread function (PSF), which results from the telescope optics, but for ground-based images, is dominated by the blurring caused by the atmospheric turbulence; furthermore, the PSF may be affected by telescope guiding and the coaddition process described earlier.

The point-spread function. Atmospheric turbulence and the other effects mentioned above smear the image of the sky, according to

$$I^{\text{obs}}(\boldsymbol{\theta}) = \int d^2\boldsymbol{\vartheta} I(\boldsymbol{\vartheta}) P(\boldsymbol{\theta} - \boldsymbol{\vartheta}), \quad (28)$$

where $I(\boldsymbol{\vartheta})$ is the brightness profile outside the atmosphere, $I^{\text{obs}}(\boldsymbol{\theta})$ the observed brightness profile, and P is the PSF; it describes how point sources would appear on the image. To first approximation, the PSF is a bell-shaped function; its full width at half maximum (FWHM) is called the ‘seeing’ of the image. At excellent sites, and excellent telescopes, the seeing has a median of $\sim 0''.7 - 0''.8$; exceptionally, images with a seeing of $\sim 0''.5$ can be obtained.



Chandra Deep Field South (CDF-S)
(MPG/ESO 2.2-m + WFI)
ESO PR Photo 02a/03 (10 January 2003)

© European Southern Observatory



Fig. 13. A multi-color WFI image of the CDFS; the field is slightly larger than one-half degree on the side. To obtain this image, about 450 different WFI exposures were combined, resulting in a total exposure time of 15.8 hours in B, 15.6 hours in V, and 17.8 hours in R. The data were obtained in the frame of three different projects – the GOODS project, the public ESO Imaging Survey, and the COMBO-17 survey. These data were reduced and coadded by Mischa Schirmer & Thomas Erben; more than 2 TB of disk space were needed for the reduction.

Recall that typical faint galaxies are considerably smaller than this seeing size, hence their appearance is dominated by the PSF.

The main effect of seeing on image shapes is that it makes an elliptical source rounder: a small source with a large ellipticity will nevertheless appear as a fairly round image if its size is considerably smaller than the PSF. If not properly corrected for, this smearing effect would lead to a serious underestimate of ellipticities, and thus of the shear estimates. Furthermore, the PSF is not fully isotropic; small anisotropies can be introduced by guiding errors, the coaddition, the telescope optics, bad focusing etc. An anisotropic PSF makes round sources elliptical, and therefore mimics a shear. Also here, the effect of the PSF anisotropy depends on the image size and is strongest for the smallest sources. PSF anisotropies of several percent are typical; hence, if not corrected for, its effect can be larger than the shear to be measured.

The PSF can be measured at the position of stars (point sources) on the field; if it is a smooth function of position, it can be fitted by a low-order polynomial, which then yields a model for the PSF at all points, in particular at every image position, and one can correct for the effects of the PSF. A potential problem occurs if the PSF jumps between chips boundaries in multi-chip cameras, since then the coaddition produces PSF jumps on the coadded frame; this happens in cameras where the chips are not sufficiently planar, and thus not in focus simultaneously. For the WFI@ESO/MPG 2.2-m, this however is not a problem, but for some other cameras this problem exists and is severe. There is an obvious way to deal with that problem, namely to coadd data only from the same CCD chip. In this case, the gaps between chips cannot be closed in the coadded image, but for most weak lensing purposes this is not a very serious issue. In order not to lose too much area in this coaddition, the dither pattern, i.e., the pointing differences in the individual exposures, should be kept small; however, it should not be smaller than, say, 20'', since otherwise some pixels may always fall onto a few larger galaxies in the field, which then causes problems in constructing a superflat. Furthermore, small shifts between exposures means that the number of objects falling onto different chips in different exposures is small, thus reducing the accuracy of the astrometric solution. In any case, the dither strategy shall be constructed for each camera individually, taken into account its detailed properties.

3.5 Shape measurements

Specific software has been developed to deal with the issues mentioned above; the one that is most in use currently has been developed by Kaiser et al. (1995; hereafter KSB), with substantial additions by Luppino & Kaiser (1997), and later modifications by Hoekstra et al. (1998). The numerical implementation of this method is called IMCAT and is publically available. The basic features of this method shall be outlined next.

First one notes that the definition (6) of the second-order moments of the image brightness is not very practical for applying it to real data. As the effective range of integration depends on the surface brightness of the image (through the weight function q_I) the presence of noise enters the definition the Q_{ij} in a non-linear fashion. Furthermore, neighboring images can lead to very irregularly shaped integration ranges. In addition, this definition is hampered by the discreteness of pixels. For these reasons, the definition is modified by introducing a weight function $q_\theta(\boldsymbol{\theta})$ which depends explicitly on the image coordinates,

$$Q_{ij} = \frac{\int d^2\theta q_\theta(\boldsymbol{\theta}) I(\boldsymbol{\theta}) (\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j)}{\int d^2\theta q_\theta(\boldsymbol{\theta}) I(\boldsymbol{\theta})}, \quad i, j \in \{1, 2\}, \quad (29)$$

where the size of the weight function q_θ is adapted to the size of the galaxy image (for optimal S/N measurement). One typically chooses q_θ to be circular Gaussian. The image center $\bar{\boldsymbol{\theta}}$ is defined as before, but also with the new weight function $q_\theta(\boldsymbol{\theta})$, instead of $q_I(I)$. However, with this definition, the transformation between image and source brightness moments is no longer simple; in particular, the relation (11) between the second-order brightness moments of source and image no longer holds. The explicit spatial dependence of the weight, introduced for very good practical reasons, destroys the convenient relations that we derived earlier – welcome to reality.

In KSB, the anisotropy of the PSF is characterized by its (complex) ellipticity q , measured at the positions of the stars, and fitted by a low-order polynomial. Assume that the (reduced) shear g and the PSF anisotropy q are small; then they both will have a small effect on the measured ellipticity. Linearizing these two effects, one can write (employing the Einstein summation convention)

$$\hat{\chi}_\alpha^{\text{obs}} = \chi_\alpha^0 + P_{\alpha\beta}^{\text{sm}} q_\beta + P_{\alpha\beta}^g g_\beta. \quad (30)$$

The interpretation of the various terms is found as follows: First consider an image in the absence of shear and the case of an isotropic PSF; then $\hat{\chi}^{\text{obs}} = \chi^0$; thus, χ^0 is the image ellipticity one would obtain for $q = 0$ and $g = 0$; it is the source smeared by an isotropic PSF. It is important to note that $E(\chi^0) = 0$, due to the random orientation of sources. The tensor P^{sm} describes how the image ellipticity responds to the presence of a PSF anisotropy; similarly, the tensor P^g describes the response of the image ellipticity to shear in the presence of smearing by the seeing disk. Both, P^{sm} and P^g have to be calculated for each image individually; they depend on higher-order moments of the brightness distribution and the size of the PSF. A full derivation of the explicit equations can be found in Sect. 4.6.2 of BS01.

Given that $\langle \chi^0 \rangle = 0$, an estimate of the (reduced) shear is provided by

$$\epsilon = (P^g)^{-1} (\hat{\chi}^{\text{obs}} - P^{\text{sm}} q). \quad (31)$$

If the source size is much smaller than the PSF, the magnitude of P^g can be very small, i.e., the correction factor in (31) can be very large. Given that

the measured ellipticity $\hat{\chi}^{\text{obs}}$ is affected by noise, this noise then also gets multiplied by a large factor. Therefore, depending on the magnitude of P^g , the error of the shear estimates differ between images; this can be accounted for by specifically weighting these estimates when using them for statistical purposes (e.g., in the estimate of the mean shear in a given region). Different authors use different weighting schemes when applying KSB. Also, the tensors P^{sm} and P^g are expected to depend mainly on the size of the image and their signal-to-noise; therefore, it is advantageous to average these tensors over images having the same size and S/N, instead of using the individual tensor values which are of course also affected by noise. Erben et al. (2001) and Bacon et al. (2001) have tested the KSB scheme on simulated data and in particular investigated various schemes for weighting shear estimates and for determining the tensors in (30); they concluded that simulated shear values can be recovered with a systematic uncertainty of about 10%.

Maybe by now you are confused – what is ‘real ellipticity’ of an image, independent of weights etc.? Well, this question has no answer, since only images with conformal elliptical isophotes have a ‘real ellipticity’. By the way, not necessarily the one that is the outcome of the KSB procedure. The KSB process does not aim toward measuring ‘the’ ellipticity of any individual galaxy image; it tries to measure ‘a’ ellipticity which, when averaged over a random intrinsic orientation of the source, yields an unbiased estimate of the reduced shear.

Given that the shape measurements of faint galaxies and their correction for PSF effects is central for weak lensing, several different schemes for measuring shear have been developed (e.g., Valdes et al. 1983; Bonnet & Mellier 1995; Kuijken 1999; Kaiser 2000; Refregier 2003b; Bernstein & Jarvis 2002). In the shapelet method of Refregier (2003b; see also Refregier & Bacon 2003), the brightness distribution of galaxy images is expanded in a set of basis functions (‘shapelets’) whose mathematical properties are particularly convenient. With a corresponding decomposition of the PSF (the shape of stars) into these shapelets and their low-order polynomial fit across the image, a partial deconvolution of the measured images becomes possible, using linear algebraic relations between the shapelet coefficients. The effect of a shear on the shapelet coefficients can be calculated, yielding then an estimate of the reduced shear. In contrast to the KSB scheme, higher-order brightness moments, and not just the quadrupoles, of the images are used for the shear estimate.

These alternative methods for measuring image ellipticities (in the sense mentioned above, namely to provide an unbiased estimate of the local reduced shear) have not been tested yet to the same extent as is true for the KSB method. Before they become a standard in the field of weak lensing, several groups need to independently apply these techniques to real and synthetic data sets to evaluate their strengths and weaknesses. In this regard, one needs to note that weak lensing has, until recently, been regarded by many

researchers as a field where the observational results are difficult to ‘believe’ (and sure, not all colleagues have given up this view, yet). The difficulty to display the directly measured quantities graphically so that they can be directly ‘seen’ makes it difficult to convince others about the reliability of the measurements. The fact that the way from the coadded imaging data to the final result is, except for the researchers who actually do the analysis, close to a black box with hardly any opportunity to display intermediate results (which would provide others with a quality check) implies that the methods employed should be standardized and well checked.

Surprisingly enough, there are very few (published) attempts where the same data set is analyzed by several groups independently, and intermediate and final results being compared. Kleinheinrich (2003) in her dissertation has taken several subsets of the data that led to the deep image shown in Fig. 13 and compared the individual image ellipticities between the various subsets. If the subsets had comparable seeing, the measured ellipticities could be fairly well reproduced, with an rms difference of about 0.15, which is small compared to the dispersion of the image ellipticities $\sigma_\epsilon \sim 0.35$. Hence, these differences, which presumably are due to the different noise realizations on the different images, are small compared to the ‘shape noise’ coming from the finite intrinsic ellipticities of galaxies. If the subsets had fairly different seeing, the smearing correction turns out to lead to a systematic bias in the measured ellipticities. From the size of this bias, the conclusions obtained from the simulations are confirmed – measuring a shear with better than $\sim 10\%$ accuracy will be difficult with the KSB method, where the main problem lies in the smearing correction.

Shear observations from space. We conclude this section with a few comments on weak lensing observations from space. Since the PSF is the largest problem in shear measurements, one might be tempted to use observations from space which are not affected by the atmosphere. At present, the Hubble Space Telescope (HST) is the only spacecraft that can be considered for this purpose. Weak lensing observations have been carried out using two of its instruments, WFPC2 and STIS. The former has a field-of-view of about 5 arcmin², whereas STIS has a field of 51''. These small fields imply that the number of stars that can be found on any given exposure at high galactic latitude is very small, in fact typically zero for STIS. Therefore, the PSF cannot be measured from these exposures themselves. Given that an instrument in space is expected to be much more stable than one on the ground, one might expect that the PSF is stable in time; then, it can be investigated by analyzing exposures which contain many stars (e.g., from a star cluster). In fact, Hoekstra et al. (1998) and Hämerle et al. (2002) have shown that the PSFs of WFPC2 and STIS are approximately constant in time. The situation is improved with the new camera ACS onboard HST, where the field size of $\sim 3'4$ is large enough to contain about a dozen stars even for high galactic

latitude, and where some control over the PSF behavior on individual images is obtained. We shall discuss the PSF stability of the ACS in Sect. 7.3 below.

The PSF of a diffraction-limited telescope is much more complex than that of the seeing-dominated one for ground-based observations. The assumption underlying the KSB method, namely that the PSF can be described by a axi-symmetric function convolved with a small anisotropic kernel, is strongly violated for the HST PSF; it is therefore less obvious how well the shear measurements with the KSB method work in space. In addition, the HST PSF is not well sampled with the current imaging instruments, even though STIS and ACS have a pixel scale of $0''.05$. The number density of cosmic rays is much larger in space, so their removal can be more cumbersome than for ground-based observations. The intense particle bombardment also leads to aging of the CCD, which lose their sensitivity and attain charge-transfer efficiency problems. Despite these potential problems, a number of highly interesting weak lensing results obtained with the HST have been reported, in particular on clusters, and we shall discuss some of them in later sections. The new Advanced Camera for Surveys (ACS) on-board HST has a considerably larger field-of-view than previous instruments and will most likely become a highly valuable tool for weak lensing studies.

4 Clusters of galaxies: Introduction, and strong lensing

4.1 Introduction

Galaxies are not distributed randomly, but they cluster together, forming groups and clusters of galaxies. Those can be identified as overdensities of galaxies projected onto the sky, and this has of course been the original method for the detection of clusters, e.g., leading to the famous and still heavily used Abell (1958) catalog and its later Southern extension (Abell et al. 1989; ACO). Only later – with the exception of Zwicky’s early insight in 1933 that the Coma cluster must contain a lot of missing mass – it was realized that the visible galaxies are but a minor contribution to the clusters since they are dominated by dark matter. From X-ray observations we know that clusters contain a very hot intracluster gas which emits via free-free and atomic line radiation. Many galaxies are members of a cluster or a group; indeed, the Milky Way is one of them, being one of two luminous galaxies of the Local Group (the other one is M31, the Andromeda galaxy), of which ~ 35 member galaxies are known, most of them dwarfs.

In the first part of this section we shall describe general properties of galaxy clusters, in particular methods to determine their masses, before turning to their strong lensing properties, such as show up in the spectacular giant luminous arcs. Very useful reviews on clusters of galaxies are from Sarazin (1986) and in a recent proceedings volume (Mulchaey et al. 2004).

4.2 General properties of clusters

Clusters of galaxies contain tens to hundreds of bright galaxies; their galaxy population is dominated by early-type galaxies (E's and S0's), i.e. galaxies without active star formation. Often a very massive cD galaxy is located at their center; these galaxies differ from normal ellipticals in that they have a much more extended brightness profile – they are the largest galaxies. The morphology of clusters as seen in their distribution of galaxies can vary a lot, from regular, compact clusters (often dominated by a central cD galaxy) to a bimodal distribution, or highly irregular morphologies with strong sub-structure. Since clusters are at the top of the mass scale of virialized objects, the hierarchical merging scenario of structure growth predicts that many of them have formed only recently through the merging of two or more lower-mass sub-clusters, and so the irregular morphology just indicates that this happened.

X-ray observations reveal the presence of a hot (several keV) intracluster medium (ICM) which is highly enriched in heavy elements; hence, this gas has been processed through star-formation cycles in galaxies. The mass of the ICM surpasses that of the baryons in the cluster galaxies; the mass balance in clusters is approximately as follows: stars in cluster galaxies contribute $\sim 3\%$ of the total mass, the ICM another $\sim 15\%$, and the rest ($\gtrsim 80\%$) is dark matter. Hence, clusters are dominated by dark matter; as discussed below (Sect. 4.3), the mass of clusters can be determined with three vastly different methods which overall yield consistent results, leading to the aforementioned mass ratio.

We shall now quote a few characteristic values which apply to rich, massive clusters. Their virial radius, i.e., the radius inside of which the mass distribution is in approximate virial equilibrium (or the radius inside of which the mean mass density of clusters is ~ 200 times the critical density of the Universe – cf. Sect. 4.5 of IN) is $r_{\text{vir}} \sim 1.5 h^{-1} \text{ Mpc}$. A typical value for the one-dimensional velocity dispersion of the member galaxies is $\sigma_v \sim 1000 \text{ km/s}$. In equilibrium, this equals the thermal velocity of the ICM, corresponding to a temperature of $T \sim 10^{7.5} \text{ K} \sim 3 \text{ keV}$. The mass of massive clusters within the virial radius (i.e., the *virial mass*) is $\sim 10^{15} M_\odot$. The mass-to-light ratio of clusters (as measured from the B-band luminosity) is typically of order $(M/L) \sim 300 h^{-1} (M_\odot/L_\odot)$. Of course, the much more numerous typical clusters have smaller masses (and temperatures).

Cosmological interest for clusters. Clusters are the most massive bound and virialized structures in the Universe; this, together with the (related) fact that their dynamical time scale (e.g., the crossing time $\sim r_{\text{vir}}/\sigma_v$) is not much smaller than the Hubble time H_0^{-1} – so that they retain a ‘memory’ of their formation – render them of particular interest for cosmologists. The evolution of their abundance, i.e., their comoving number density as a function of mass and redshift, is an important probe for cosmological models

and traces the growth of structure; massive clusters are expected to be much rarer at high redshift than today. Their present-day abundance provides one of the measures for the normalization of the power spectrum of cosmological density fluctuations. Furthermore, they form (highly biased) signposts of the dark matter distribution in the Universe, so their spatial distribution traces the large-scale mass distribution in the Universe. Clusters act as laboratories for studying the evolution of galaxies and baryons in the Universe. Since the galaxy number density is highest in clusters, mergers of their member galaxies and, more importantly, other interactions between them occur frequently. Therefore, the evolution of galaxies with redshift is most easily studied in clusters. For example, the Butcher–Oemler effect (the fact that the fraction of blue galaxies in clusters is larger at higher redshifts than today) is a clear sign of galaxy evolution which indicates that star formation in galaxies is suppressed once they have become cluster members. More generally, there exists a density-morphology relation for galaxies, with an increasing fraction of early-types with increasing spatial number density, with clusters being on the extreme for the latter. Finally, clusters were (arguably) the first objects for which the presence of dark matter has been concluded (by Zwicky in 1933). Since they are so large, and present the gravitational collapse of a region in space with initial comoving radius of $\sim 8h^{-1}$ Mpc, one expects that their mixture of baryonic and dark matter is characteristic for the mean mass fraction in the Universe (White et al. 1993). With the baryon fraction of $\sim 15\%$ mentioned above, and the density parameter in baryons determined from big-bang nucleosynthesis in connection to the determination of the deuterium abundance in Ly α QSO absorption systems, $\Omega_b \approx 0.02h^{-2}$, one obtains a density parameter for matter of $\Omega_m \sim 0.3$, in agreement with results from other methods, most notably from the recent WMAP CMB measurements (e.g., Spergel et al. 2003).

4.3 The mass of galaxy clusters

Cosmologists can predict the abundance of clusters as a function of their mass (e.g., using numerical simulations); however, the mass of a cluster is not directly observable, but only its luminosity, or the temperature of the X-ray emitting intra-cluster medium. Therefore, in order to compare observed clusters with the cosmological predictions, one needs a way to determine their masses. Three principal methods for determining the mass of galaxy clusters are in use:

- Assuming virial equilibrium, the observed velocity distribution of galaxies in clusters can be converted into a mass estimate, employing the virial theorem; this method typically requires assumptions about the statistical distribution of the anisotropy of the galaxy orbits.
- The hot intra-cluster gas, as visible through its Bremsstrahlung in X-rays, traces the gravitational potential of the cluster. Under certain assumptions,

tions (see below), the mass profile can be constructed from the X-ray emission.

- Weak and strong gravitational lensing probes the projected mass profile of clusters, with strong lensing confined to the central regions of clusters, whereas weak lensing can yield mass measurements for larger radii.

All three methods are complementary; lensing yields the line-of-sight projected density of clusters, in contrast to the other two methods which probe the mass inside spheres. On the other hand, those rely on equilibrium (and symmetry) conditions; e.g., the virial method assumes virial equilibrium (that the cluster is dynamically relaxed) and the degree of anisotropy of the galaxy orbit distribution.

Dynamical mass estimates. Estimating the mass of clusters based on the virial theorem,

$$2E_{\text{kin}} + E_{\text{pot}} = 0 , \quad (32)$$

has been the traditional method, employed by Zwicky in 1933 to find strong hints for the presence of dark matter in the Coma cluster. The specific kinetic energy of a galaxy is $v^2/2$, whereas the potential energy is determined by the cluster mass profile, which can thus be determined using (32). One should note that only the line-of-sight component of the galaxy velocities can be measured; hence, in order to derive the specific kinetic energy of galaxies, one needs to make an assumption on the distribution of orbit anisotropies in the cluster potential. Assuming an isotropic distribution of orbits, the l.o.s. velocity distribution can then be related to the 3-D velocity dispersion, which in turn can be transformed into a mass estimate if spherical symmetry is assumed. This method requires many redshifts for an accurate mass estimate, which are available only for a few clusters. However, a revival of this method is expected and already seen by now, owing to the new high-multiplex optical spectrographs.

X-ray mass determination of clusters. The intracluster gas emits via Bremsstrahlung; the emissivity depends on the gas density and temperature, and, at lower T , also on its chemical composition, since at $T \lesssim 1 \text{ keV}$ the line radiation from highly ionized atomic species starts to dominate the total emissivity of a hot gas. Investigating the properties of the ICM with X-ray observations have revealed a wealth of information on the properties of clusters (see Sarazin 1986). Assuming that the gas is in hydrostatic equilibrium in the potential well of the cluster, the gas pressure P must balance gravity, or

$$\nabla P = -\rho_g \nabla \Phi ,$$

where ρ_g is the gas density. In the case of spherical symmetry, this becomes

$$\frac{1}{\rho_g} \frac{dP}{dr} = -\frac{d\Phi}{dr} = -\frac{GM(r)}{r^2} .$$

From the X-ray brightness profile and temperature measurement, $M(r)$, the mass inside r , both dark and luminous, can then be determined,

$$M(r) = -\frac{k_B T r^2}{G \mu m_p} \left(\frac{d \ln \rho_g}{dr} + \frac{d \ln T}{dr} \right), \quad (33)$$

where μm_p is the mean particle mass in the gas. Only for relatively few clusters are detailed X-ray brightness and temperature profile measurements available. In the absence of a temperature profile measurement, one often assumes that T does not vary with distance from the cluster center. In this case, assuming that the dark matter particles also have an isothermal distribution (with velocity traced by the galaxy velocities), one can show that

$$\rho_g(r) \propto [\rho_{\text{tot}}(r)]^\beta; \quad \text{with} \quad \beta = \frac{\mu m_p \sigma_v^2}{k_B T_g}. \quad (34)$$

Hence, β is the ratio between kinetic and thermal energy. The mass profile corresponding to the isothermality assumption follows from the Lame–Emden equation which, however, has no closed-form solution. In the King approximation, the density and X-ray brightness profile (which is obtained by a line-of-sight integral at projected distance R from the cluster center over the emissivity, which in turn is proportional to the square of the electron density, or $\propto \rho_g^2$, for an isothermal gas) become

$$\rho_g(r) = \rho_{g0} \left[1 + \left(\frac{r}{r_c} \right)^2 \right]^{-3\beta/2}; \quad I(R) \propto \left[1 + \left(\frac{R}{r_c} \right)^2 \right]^{-3\beta/2+1/2}$$

where r_c is the core radius. The observed brightness profile can now be fitted with these β -models, yielding estimates of β and r_c from which the cluster mass follows. Typical values for r_c range from 0.1 to $0.3 h^{-1}$ Mpc; and $\beta = \beta_{\text{fit}} \sim 0.65$. On the other hand, one can determine β from the temperature T and the galaxy velocity dispersion using (34), which yields $\beta_{\text{spec}} \approx 1$. The discrepancy between these two estimates of β is not well understood and probably indicates that one of assumptions underlying this ‘ β -models’ fails in many clusters, which is not too surprising (see below).

The hot ICM loses energy through its thermal radiation; the cooling time t_{cool} of the gas, i.e., the ratio between the thermal energy density and the X-ray emissivity, is larger than the Hubble time $\sim H_0^{-1}$ for all but the innermost regions. In the center of clusters, the gas density can be high enough to have $t_{\text{cool}} < H_0^{-1}$, so that there the gas can no longer be in hydrostatic equilibrium. One expects that the gas flows towards the cluster center, thereby being compressed and therefore maintain approximate pressure balance. Such ‘cooling flows’ (see, e.g., Fabian 1994) are observed indirectly, through highly peaked X-ray emission in cluster centers which indicates a strong increase of the gas density; furthermore, these cooling-flow clusters show a decrease of

T towards the center. The mass-flow rate in these clusters can be as high as $100M_{\odot} \text{ yr}^{-1}$ or even more, so that the total cooled mass can be larger than the baryonic mass of a massive galaxy. However, the fate of the cooled gas is unknown.

New results from Chandra & XMM. The two X-ray satellites Chandra and XMM, launched in 1999, have greatly increased our view of the X-ray Universe, and have led to a number of surprising results about clusters. X-ray spectroscopy verified the presence of cool gas near the center of cooling-flow clusters, but no indication for gas with temperature below $\sim 1 \text{ keV}$ has been seen, whereas the cooling is expected to rapidly proceed to very low temperatures, as the cooling function increases for lower T where atomic transitions become increasingly important. Furthermore, the new observations have revealed that at least the inner regions of clusters often show a considerably more complicated structure than implied by hydrostatic equilibrium. In some cases, the intracluster medium is obviously affected by a central AGN, which produces additional energy and entropy input, which might explain why no sub-keV gas has been detected. As the AGN activity of a galaxy may be switched on and off, depending on the fueling of the central black hole, even in clusters without a currently active AGN such heating might have occurred in the recent past, as indicated in some cases by radio relics. Cold fronts with very sharp edges (discontinuities in density and temperature, but such that $P \propto \rho T$ is approximately constant across the front), and shocks have been discovered, most likely showing ongoing or recent merger events. In many clusters, the temperature and metalicity appears to be strongly varying functions of position which invalidates the assumption of isothermality underlying the β -model. Therefore, mass estimates of central parts of clusters from X-ray observations require special care, and one needs to revise the simplified models used in the pre-Chandra era. In fact, has there ever been the believe that the β -model provides an adequate description of the gas in a cluster, the results from Chandra and XMM show that this is unjustified. The physics of the intracluster gas appears to be considerably more complicated than that.

4.4 Luminous arcs & multiple images

Strong lensing effects in cluster show up in the form of giant luminous arcs, strongly distorted arclets, and multiple images of background galaxies. Since strong lensing only occurs in the central part of clusters, it can be used only to probe their inner mass structure. However, strong lensing yields by far the most accurate central mass determinations in those cases where several strong lensing features can be identified. For a detailed account of strong lensing in clusters, the reader is referred to the review by Fort & Mellier (1994).

Furthermore, clusters thus act as a ‘natural telescope’; many of the most distant galaxies have been found by searching behind clusters, employing the

lensing magnification. For example, the recently discovered very high redshift galaxies at $z \approx 7$ (Kneib et al. 2004) and $z = 10$ (Pelló et al. 2004) were found through a search in the direction of the high-magnification region in the clusters A2218 and A1835, respectively. In the first of these two cases, the multiple imaging of the background galaxy provides not only the magnification, but also an estimate of the redshift of the source (which is not determined by any spectral line), whereas in the latter case, only the implied high magnification makes the source visible on deep HST images and allows its spectroscopy, yielding a spectral line which most likely is due to Ly α . The magnification is indeed a very important asset, as can be seen from a simple example: a value of $\mu = 5$ reduces the observing time for obtaining a spectrum by a factor 25 (in the case where the noise is sky background dominated) – which is the difference of being doable or not. Recognizing the power of natural telescopes, the deepest SCUBA surveys for faint sub-millimeter sources have been conducted (e.g., Blain et al. 1999) around clusters with well-constrained (from lensing) mass distribution to reach further down the (unlensed) flux scale.

First go: $M(\leq \theta_E)$. Giant arcs occur where the distortion (and magnification) is very large, that is near critical curves. To a first approximation, assuming a spherical mass distribution, the location of the arc from the cluster center (which usually is assumed to coincide with the brightest cluster galaxy) yields the Einstein radius of the cluster, so that the mass estimate (see IN, Eq. 43) can be applied.

$$M(\theta_{\text{arc}}) \approx \pi (D_d^{\text{ang}} \theta_{\text{arc}})^2 \Sigma_{\text{cr}} . \quad (35)$$

Therefore, this simple estimate yields the mass inside the arc radius. However, this estimate is not very accurate, perhaps good to within $\sim 30\%$ (Bartelmann & Steinmetz 1996). Its reliability depends on the level of asymmetry and substructure in the cluster mass distribution. Furthermore, it is likely to overestimate the mass in the mean, since arcs preferentially occur along the major axis of clusters. Of course, the method is very difficult to apply if the center of the cluster is not readily identified or if the cluster is obviously bimodal. For these reasons, this simple method for mass estimates is not regarded as particularly accurate.

Detailed modeling. The mass determination in cluster centers becomes much more accurate if several arcs and/or multiple images are present, since in this case, detailed modeling can be done. This typically proceeds in an interactive way: First, multiple images have to be identified (based on their colors and/or detailed morphology, as available with HST imaging). Simple (plausible) mass models are then assumed, with parameters fixed by matching the multiple images, and requiring the distortion at the arc location(s) to be

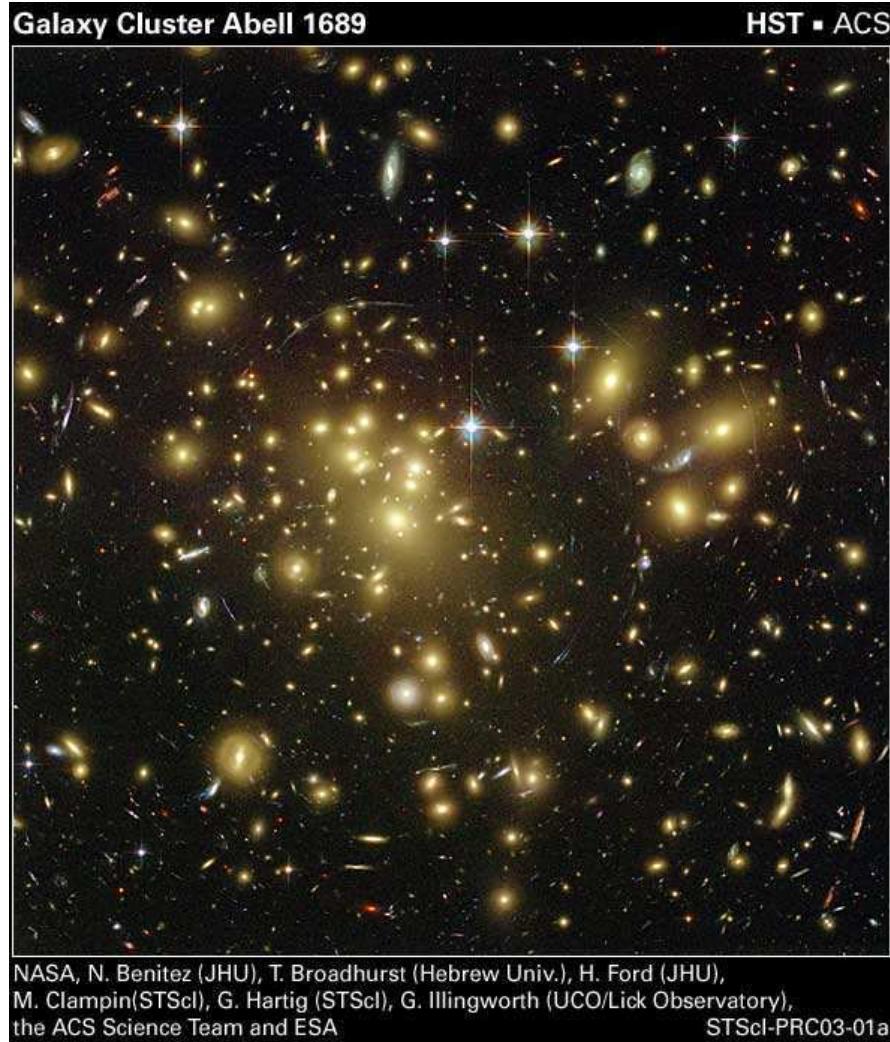


Fig. 14. The galaxy cluster Abell 1689 is the most impressive lensing cluster yet found. This image has been taken with the new Advanced Camera for Surveys (ACS) onboard HST. Numerous arcs are seen. A simple estimate for the mass of the center of the cluster, obtained by identifying the arcs radius with the Einstein radius, yields an extremely large equivalent velocity dispersion. The distribution of the arcs shown here indicates that such a simple assumption is misleading, and more detailed modeling required

strong and to have the correct orientation. This model then predicts the presence of possible further multiple images; they can be checked for through morphology, surface brightness (in particular if HST images of the cluster are available) and color. If confirmed, a new, refined model is constructed including these new additional strong lensing constraints, which yields further strong lensing predictions etc. As is the case for galaxy lensing (see SL), the components of the mass models are not arbitrary, but chosen to be physically motivated. Typically, as major component a ellipsoidal isothermal or NWF distribution is used to describe the overall mass distribution of the cluster. Refinements of the mass distribution are introduced as mass components centered on bright cluster member galaxies or on subgroups of such galaxies, describing massive subhalos which survived a previous merger. Such models have predictive power and can be trusted in quite some detail; the accuracy of mass estimates in some favorable cases can be as high as a few percent.

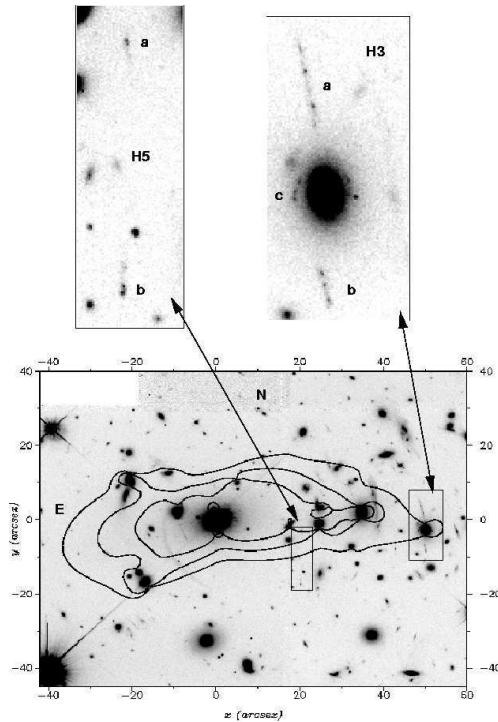


Fig. 15. The lower panel shows the critical curves of the cluster A2390 (cluster redshift $z_d = 0.231$), for three different source redshifts of $z_s = 1, 2.5$ and 4 (from inner to outer). The lens model is based on the detailed HST image shown here. Identified are two sets of multiple images, shown in the upper two panels, which obviously need to be at very high redshift. Indeed, spectroscopy shows that they have $z_s = 4.04$ and $z_s = 4.05$ (from Pelló et al. 1999)

In fact, these models can be used to predict the redshift of arcs and arclets. As an example, we mention the strong lensing analysis of the cluster Abell 2390 based on HST imaging (Pelló et al. 1999). Two pairs of multiple images were identified (see Fig. 15) which then implies that the critical curve has to pass between the individual components. The location of the critical curves

depends, however, on the source redshift. As shown in the figure, the sources have to be at a high redshift in order for the corresponding critical curves to have the correct location. In fact, spectroscopy placed the two sources at $z_s = 4.04$ and $z_s = 4.05$, as predicted by the lens model.

Since the distortion of a lens also depends on the source redshift, once a detailed mass model is available from arcs with known redshifts for at least some of them, one can estimate the value of the lens strength $\propto D_{ds}/D_s$ and thus infer the redshift of arclets. This method has been successfully applied to HST observations of clusters (Ebbels et al. 1998). Of course, having spectroscopic redshifts of the arcs available increases the calibration of the mass models; they are therefore very useful.

Lens properties from Fourier transforms. Before discussing results from these detailed models, a brief technical section shall be placed here, related to calculating lens properties of general mass distributions. A general method to obtain the lensing quantities of a mass distribution is through Fourier transformation. We assume that we have a mass distribution of finite mass; this is not a serious restriction even for models with formally infinite total mass, because we can truncate them on large scales, thus making the total mass finite, without affected any lensing properties at smaller scales. We define the Fourier transform $\hat{\kappa}(\ell)$ of the surface mass density as⁴

$$\hat{\kappa}(\ell) = \int_{\mathbb{R}^2} d^2\theta \kappa(\theta) \exp(i\ell \cdot \theta), \quad (36)$$

and its inverse by

$$\kappa(\theta) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} d^2\ell \hat{\kappa}(\ell) \exp(-i\ell \cdot \theta). \quad (37)$$

Similarly, we define the Fourier transforms of the deflection potential, $\hat{\psi}(\ell)$, of the deflection angle, $\hat{\alpha}(\ell)$, and of the complex shear, $\hat{\gamma}(\ell)$. Differentiation by θ_i in real space is replaced by multiplication by $-i\ell_i$ in Fourier space. Therefore, the Fourier transform of $\partial\psi/\partial\theta_j$ is $-i\ell_j \hat{\psi}(\ell)$. Hence, the Poisson equation as given in Sect. 2.2 of IN becomes in Fourier space

$$-|\ell|^2 \hat{\psi}(\ell) = 2\hat{\kappa}(\ell). \quad (38)$$

Thus, for $\ell \neq \mathbf{0}$, the Fourier transform of the potential which satisfies the Poisson equation can be readily determined. The $\ell = \mathbf{0}$ mode remains undetermined; however, since this mode corresponds to a constant in ψ , it is unimportant and can be set to zero. Once $\hat{\psi}$ is determined, the Fourier transform of the deflection angle and the shear follows from their definitions in

⁴ We denote the Fourier variable of three-dimensional space as \mathbf{k} , that of angular position by ℓ .

terms of the deflection potential, given in Sect. 2.2 of IN,

$$\hat{\alpha}(\ell) = -i\ell\hat{\psi}(\ell), \quad (39)$$

$$\hat{\gamma}(\ell) = -\left(\frac{\ell_1^2 - \ell_2^2}{2} + i\ell_1\ell_2\right)\hat{\psi}(\ell). \quad (40)$$

Thus, in principle, one determines the relevant quantities by Fourier transforming κ , then calculating the Fourier transforms of the potential, deflection, and shear, whose real-space counterparts are then obtained from an inverse Fourier transform, like in (37).

Up to now we have not gained anything; the Fourier transforms as defined above are two-dimensional integrals, as are the real-space relations between deflection angle and shear, and the surface-mass density. However, provided κ becomes ‘small enough’ for large values of $|\theta|$, the integral in (36) may be approximated by one over a finite region in θ -space. This finite integral is further approximated as a sum over gridpoints, with a regular grid covering the lens plane. Consider a square in the lens plane of side L , and let N be the number of gridpoints per dimension, so that $\Delta\theta = L/N$ is the size of a gridcell. The inverse grid, i.e., the ℓ -grid, has a gridcell of size $\Delta\ell = 2\pi/L$. The discrete Fourier transform then uses the values of κ on the θ -grid to calculate $\hat{\kappa}$ on the ℓ -grid. The latter, in fact, is then the Fourier transform of the periodic continuation of the mass distribution in θ -space. Because of this periodic continuation, the deflection angle as calculated from the discrete Fourier transform, which is performed by the Fast Fourier Transform (FFT) method, is the sum of the input mass distribution, plus all of its periodic continuation. Here, finally, is why we have considered the Fourier method: the FFT is a very efficient and quick procedure (see, e.g., Press et al. 1992), and arguably the best one in cases of mass distributions for which no analytical progress can be made. The lensing properties are calculated on a grid; if needed, they can be obtained for other points by interpolation.

Because of the periodic continuation, the mass distribution has to decreases sufficiently quickly for large $|\theta|$, or be truncated at large radii. In any case, L should be taken sufficiently large to minimize these periodicity effects.

Another point to mention is that a periodic mass distribution, each element of which has positive total mass, has an infinite mass, so that the deflection potential has to diverge; on the other hand, the deflection potential is enforced to be periodic. This apparent contradiction can be resolved by noting that the $\ell = 0$ mode of $\hat{\kappa}$ is not used in the calculation of $\hat{\alpha}$ and $\hat{\gamma}$. Indeed, if $\hat{\psi}$ and ψ are calculated from the above equations, then the resulting ψ does not satisfy the Poisson equation; the ψ resulting from this procedure is the one corresponding to $\kappa - \bar{\kappa}$, where $\bar{\kappa}$ is the average of κ on the θ -grid. A similar remark is true for the deflection angle. Thus, at the end, one has to add a term $\bar{\kappa}|\theta|^2/2$ to ψ , and a term $\bar{\kappa}\theta$ to α .

Since the FFT is very fast, one can choose N and L large, and then consider only the central part of the θ -grid needed for the actual lens modeling.

4.5 Results from strong lensing in clusters

The main results of the strong lensing investigations of clusters can be summarized as follows:

- The mass in cluster centers is much more concentrated than predicted by (simple) models based on X-ray observations. The latter usually predict a relatively large core of the mass distribution. These large cores would render clusters sub-critical to lensing, i.e., they would be unable to produce giant arcs or multiple images. In fact, when arcs were first discovered they came as a big surprise because of these expectations. By now we know that the intracluster medium is much more complicated than assumed in these ‘ β -model’ fits for the X-ray emission.
- The mass distribution in the inner region of clusters often shows strong substructure, or multiple mass peaks. These are also seen in the galaxy distribution of clusters, but with the arcs can be verified to also correspond to mass peaks (examples of this include the cluster Abell 2218 where arcs also curve around a secondary concentration of bright galaxies, clearly indicating the presence of a mass concentration, or the obviously bimodal cluster A 370). These are easily understood in the frame of hierarchical mergers in a CDM model; the merged clusters retain their multiple peaks for a dynamical time or even longer, and are therefore not in virial equilibrium.
- The orientation of the (dark) matter appears to follow closely the orientation of the light in the cD galaxy; this supports the idea that the growth of the cD galaxy is related to the cluster as a whole, through repeated accretion of lower-mass member galaxies. In that case, the cD galaxy ‘knows’ the orientation of the cluster.
- There is in general good agreement between lensing and X-ray mass estimates (e.g., Ettori & Lombardi 2003; Donahue et al. 2003) for those clusters where a ‘cooling flow’ indicates that they are in dynamical equilibrium, provided the X-ray analysis takes the presence of the cooling flow into account (Allen 1998).

Probably our ‘favourite’ clusters in which strong lensing effects are investigated in detail are biased in favor of having strong substructure, as this increases the lensing cross section for the occurrence of giant arcs (see below). Hence, it may be that the most detailed results obtained from strong lensing in clusters apply to a class of clusters which are especially selected because of their ability to produce spectacular arcs, and thus of their asymmetric mass distribution. Therefore, one must be careful in generalizing conclusions drawn from the ‘arc clusters’ to the cluster population as a whole.

Discrepancies. There are a few clusters where the lensing results and those obtained from analyzing the X-ray observations or cluster dynamics are in strong apparent conflict. Two of the most prominent ones shall be mentioned here. The cluster A1689 (see Fig. 14) has arcs more than $\sim 40''$ away from the cluster center, which would imply a huge mass in this cluster center. This high mass is apparently confirmed by the high velocity dispersion of its member galaxies, although their distribution in redshift makes it likely that the cluster consists of several subcomponents (see Clowe & Schneider 2001 for a summary of these results). Several weak lensing results of this cluster have been published, and they are not all in agreement: whereas Tyson & Fischer (1995) from weak shear, and Taylor et al. (1998) and Dye et al. (2001) from the magnification method (that will be discussed in the next Section) find also a very high mass for this cluster, the weak lensing analysis of Clowe & Schneider (2001; see also King et al. 2002b), based on deep wide-field imaging data of this cluster, finds a more moderate mass (or equivalent velocity dispersion) for this cluster. A new XMM-Newton X-ray observation of this cluster (Andersson & Madejski 2004) lends support for the smaller mass; in fact, their estimate of the virial mass of the cluster agrees with that obtained by Clowe & Schneider (2001). However, the discrepancy with the strong lensing mass in the cluster center remains at present; a quantitative analysis of the ACS data shown in Fig. 14 will hopefully shed light on this issue.

A second clear example for discrepant results in the cluster Cl0024+17. It has a prominent arc system, indicating an Einstein radius of $\sim 30''$, and thus a high mass. The X-ray properties of this cluster, however, indicate a much smaller mass (Soucail et al. 2000), roughly by a factor of three. This discrepancy has been reaffirmed by recent Chandra observations, which confirmed this factor-of-three problem (Ota et al. 2004). The resolution of this discrepancy has probably been found by Czoske et al. (2001, 2002), who performed an extensive spectroscopic survey of cluster galaxies. Their result is best interpreted such that Cl0024+17 presents a merger of two clusters along our line-of-sight, which implies that the measured velocity dispersion cannot be easily turned into a mass, as this system is not in virial equilibrium, and that the X-ray data cannot be converted to a mass either, due to the likely strong deviation from spherical symmetry and equilibrium. A wide field sparsely sampled HST observation of this cluster (Kneib et al. 2003) also indicates the presence of a second mass concentration about $3'$ away from the main peak. As will be mentioned below, clusters undergoing mergers have particularly high cross sections for producing arcs (Torri et al. 2004); hence, our ‘favourites’ are most likely selected for these non-equilibrium clusters.

Arc statistics. The abundance of arcs is expected to be a strong function of the cosmological parameters: they not only determine the abundance of massive clusters (through the mass function discussed in Sect. 4.5 of IN), but

also the degree of relaxation of clusters, which in turn affects their strong lensing cross section (Bartelmann et al. 1998). It is therefore interesting to consider the expected abundance of arcs as a function of cosmological parameters and compare this to the observed abundance. In a series of papers, M. Bartelmann and his colleagues have studied the expected giant arc abundance, using analytical as well as numerical techniques (e.g., Bartelmann & Weiss 1994; Bartelmann et al. 1995, 1998, 2002; Meneghetti et al. 2004; see also Dalal et al. 2003; Oguri et al. 2003; Wambsganss et al. 2004). Some of the findings of these studies can be summarized as follows:

- The formation of arcs depends very sensitively on the deviation from spherical symmetry and the detailed substructure of the mass distribution in the cluster. For this reason, analytical models which cannot describe this substructure with sufficient realism (see Bergmann & Petrosian 1993) do not provide reliable predictions for the arc statistics (in particular, axisymmetric mass models are essentially useless for estimating arc statistics), and one needs to refer to numerical simulations of structure formation. Since the substructure and triaxiality plays such an important role, these simulations have to be of high spatial and mass resolution.
- The frequency of arcs depends of course on the abundance of clusters, which in turn depends on the cosmological model and the fluctuation spectrum of the matter, in particular its normalization σ_8 . Furthermore, clusters at a given redshift have different mean ages in different cosmological models, as the history of structure growth, and thus the merging history, depends on Ω_m and Ω_Λ . Since the age of a cluster is one of the determining parameters for its level of substructure – younger clusters do not have had enough time to fully relax – this affects the lensing cross section of the clusters for arc formation. In fact, during epochs of mergers, the arc cross-section can have temporary excursions by large factors. Even the same cluster at the same epoch can have arc forming cross sections that vary by more than an order-of-magnitude between different projection directions of the cluster. For fixed cluster abundance today, low-density models form clusters earlier than high-density models.
- Since the largest contribution of the total cross section for arc formation comes from clusters at intermediate redshift ($z \sim 0.4$), also the equation-of-state of the dark energy matters; as shown in Meneghetti et al. (2004), what matters is the dark energy density at the epoch of cluster formation. In addition, the earlier clusters form, the higher their characteristic density, which then makes them more efficient lenses for arc formation.

Taking these effects together, a low-density open model produces a larger number of arcs than a flat low-density model, which in turn has more arcs than a high-density model, for a given cluster abundance today. Whereas the differences between these models obtained by Meneghetti et al. (2004) are smaller than claimed in Bartelmann et al. (1998), they in principle allow

constraining the cosmological parameters, provided they can be compared with the observed number of arcs.

Unfortunately, there are only a few systematic studies of clusters with regards to their strong lensing contents. Luppino et al. (1999) report on 8 giant arcs in their sample of the 38 most massive clusters found in the Einstein Medium Sensitivity Survey. Zaritsky & Gonzalez (2003) surveyed clusters in the redshift range $0.5 \lesssim z \lesssim 0.7$ over 69 deg^2 and found two giant arcs with $R < 21.5$ and a length $\theta_1 > 10''$. Gladders et al. (2003) found 5 arc candidates in their Red Cluster Sequence survey of 90 deg^2 , all of them being associated with high-redshift clusters. In contrast to the claim by Bartelmann et al. (1998), these observed arc frequencies can be accounted for in a standard ΛCDM Universe, as shown by Dalal et al. (2003). There are several differences between these two studies, which are based on different assumptions about the number density of clusters and the source redshift distribution, which Dalal et al. (2003) took from the Hubble Deep Field, whereas Bartelmann et al. (1998) assumed all sources having $z_s = 1$.

The strong dependence on the source redshift distribution has been pointed out by Wambsganss et al. (2004). In contrast to the other studies, they investigated the arc statistics using ray tracing through a three-dimensional mass distribution obtained from cosmological simulations, whereas the other studies mentioned considered the lensing effect of individual clusters found in these simulations. Although the former approach is more realistic, the assumption of Wambsganss et al. (2004) that the magnification of a light ray is a good measure for the length-to-width ratio of a corresponding arc is certainly not justified in detail, as shown in Dalal et al. (2003). The agreement of the lensing probability between Wambsganss et al. (2004) and Bartelmann et al. (1998) for all $z_s = 1$ is therefore most likely a coincidence.

There are further difficulties in obtaining realistic predictions for the occurrence of giant arcs that can be compared with observations. First, the question of whether an image counts as an arc depends on a combination of source size, lens magnification, and seeing. Seeing makes arcs rounder and therefore reduces their length-to-width ratio. An impressive demonstration of this effect is provided by the magnificent system of arcs in the cluster A1689 observed with the ACS onboard the HST, as shown in Fig. 14, compared to earlier ground-based images of this cluster. Second, several of the above-mentioned papers assume the source size to be $\theta = 1''$, whereas many arcs observed with HST are essentially unresolved in width, implying much smaller source sizes (and accordingly, a much higher sensitivity to seeing effects). Third, magnification bias is usually not taken into account in these theoretical studies. In fact, accounting properly for the magnification bias is quite difficult, as the surveys reporting on arc statistics are not really flux-limited. One might argue that they are surface brightness-limited, but even if this were true, the surface brightness of an arc coming from a small source depends very much on the seeing.

Therefore at present, the abundance of arcs seem to be not in conflict with a Λ CDM model, but more realistic simulations which take the aforementioned effects into account are certainly needed for a definite conclusion on this issue. On the observational side, increasing the number of clusters for which high-quality imaging is performed is of great importance, and the survey of luminous X-ray clusters imaged either with the ACS@HST or with ground-based telescopes during periods of excellent seeing would improve the observational situation dramatically. Blank-field surveys, such as they are conducted for cosmic shear research (see Sect. 7), could be used for blind searches of arcs (that is, not restricted to regions around known clusters). It may turn out, however, that the number of ‘false positives’ is unacceptably high, e.g., by misidentification of edge-on spirals, or blends of sources that yield apparent images with a high length-to-width ratio.

Constraints on collisional dark matter. Spergel & Steinhardt (2000) suggested the possibility that dark matter particles are not only weakly interacting, but may have a larger elastic scattering cross-section. If this cross-section of such self-interacting dark matter is sufficiently large, it may help to explain two of the remaining apparent discrepancies between the predictions of the Cold Dark Matter model and observations: The slowly rising rotation curves of dwarf galaxies (e.g., de Blok et al. 2001) and the substructure of galaxy-scale dark matter halos (see Sect. 8 of SL). Self-interacting may soften the strength of the central density concentration as compared to the NFW profile, and could destroy most of the subclumps. However, there are other consequence of such an interaction, in that the shapes of the inner parts of dark matter halos tend to be more spherical. Meneghetti et al. (2001) have investigated the influence of self interaction of dark matter particles on clusters of galaxies, in particular their ability to form giant arcs. From their numerical simulations of clusters with varying cross-sections of particles, they showed that even a relatively small cross-section is sufficient to reduce the ability of clusters to produce giant arcs by an order of magnitude. This is mainly due to two effects, the reduced asymmetry of the resulting mass distribution and the shallower central density profile. Furthermore, self-interactions destroy the ability of clusters to form radial arcs. Therefore, the ‘desired’ effect of self-interaction – to smooth the mass distribution of galaxies – has the same consequence for clusters, and can therefore probably be ruled out as a possible mechanism to cure the aforementioned apparent problems of the CDM model. From combining X-ray and lensing data of the cluster 0657–56, Markevitch et al. (2004) obtained upper limits on the self-interaction cross section of dark matter.

Do clusters follow the universal NFW profile? The CDM paradigm of structure formation predict a universal density profile of dark matter halos. One might therefore investigate whether the strong lensing properties of

clusters are compatible with this mass profile. Of particular value for such an investigation are clusters which contain several strong lensing features, and in particular a radial arc, as it probes the inner critical curve of the cluster. Sand et al. (2004; see also Sand et al. 2002) claim from a sample of three clusters with radial arcs, that the slope of the inner mass profile must be considerably flatter than predicted by the NFW model. However, this conclusion is derived under the assumption of an axially-symmetric lens model. As is true for strong lensing by galaxies (see SL), axisymmetric mass model are not generic, and therefore conclusions derived from them are prone to the systematic of the symmetry assumption. That was demonstrated by Bartelmann & Meneghetti (2004) who showed that, as expected, the conclusion about the inner slope changes radically once a finite ellipticity of the mass distribution is allowed for, removing the apparent discrepancy with the predictions from CDM models.

Cosmological parameters from strong lensing systems. The lens strength, at given physical surface mass density Σ , depends on the redshifts of lens and source, as well as on the geometry of the Universe which enters the distance-redshift relation. Therefore, it has been suggested that a cluster which contains a large number of strong lensing features can be used to constrain cosmological parameters, provided the sources of the arcs and multiple image systems cover a large range of redshifts (Link & Pierce 1998). Simulations of this effect, using realistic cluster models, confirmed that such purely geometrical constraints can in principle be derived (Golse et al. 2002). One of the best studied strong-lensing cluster up to now is A2218, for which four multiple-image systems with measured (spectroscopic) redshift have been identified which allows very tight constraints on the mass distribution in this cluster. Soucail et al. (2004) applied the aforementioned method to this cluster and obtained first constraints on the density parameter Ω_m , assuming a flat cosmological model. This work can be viewed as a proof of concept; the new ACS camera onboard HST will allow the identification of even richer strong lensing systems in clusters, of which the one in A1689 (see Fig. 14) is a particularly impressive example.

5 Mass reconstructions from weak lensing

Whereas strong lensing probes the mass distribution in the inner part of clusters, weak lensing can be used to study the mass distribution at much larger angular separations from the cluster center. In fact, as we shall see, weak lensing can provide a parameter-free reconstruction of the projected two-dimensional mass distribution in clusters – and hence offers the prospect of mapping the dark matter distribution of clusters directly. This discovery (Kaiser & Squires 1993) can be viewed to mark the beginning of quantitative weak lensing research. But even before this discovery, weak lensing by

clusters has been observed in a number of cases. Fort et al. (1988) found that in addition to the giant arc in A 370, there are a number of images stretched in the direction tangent to the center of the cluster, but with much less spectacular axis ratios than the giant arc in this cluster; they termed these new features ‘arclets’. Tyson et al. (1990) found a statistically significant tangential alignment of faint galaxy images relative to the center of the clusters A 1689 and Cl 1409+52, and obtained a mass profile from these lens distortion maps. Comparison with numerical simulations yielded an estimate of the cluster velocity dispersion, assuming an isothermal sphere profile.

In this section we consider the parameter-free mass reconstruction technique, first the original Kaiser & Squires method, and then a number of improvements of this method. We then turn to the magnification effects; the change of the number density of background sources, as predicted from (26), can be turned into a local estimate of the surface mass density, and this method has been employed in a number of clusters. Next we shall consider inverse methods for the reconstruction of the mass distribution, which on the one hand are more difficult to apply than the ‘direct’ methods, but on the other hand are expected to yield more satisfactory results. Whereas the two-dimensional maps yield a good visual impression on the mass distribution in clusters, it is hard to extract quantitative information from them. In order to get quantities that describe the mass and that can be compared between clusters, often parameterized mass models are more useful, which are considered next. Finally, we consider aperture mass measures, which have been introduced originally to obtain a mass quantity that is unaffected by the mass-sheet degeneracy, but as will be shown, has a number of other useful features. In particular, employing the aperture mass, one can device a method to systematically search for mass concentrations on cluster-mass scales, using their shear properties only, i.e. without referring to their luminous properties.

5.1 The Kaiser–Squires inversion

Weak lensing yields an estimate of the local (reduced) shear, as discussed in Sect. 2.2. Here we shall discuss how to derive the surface mass density from a measurement of the (reduced) shear. Recalling eq. (IN-26), the relation between shear and surface mass density is

$$\begin{aligned}\gamma(\boldsymbol{\theta}) &= \frac{1}{\pi} \int_{\mathbb{R}^2} d^2\boldsymbol{\theta}' \mathcal{D}(\boldsymbol{\theta} - \boldsymbol{\theta}') \kappa(\boldsymbol{\theta}') , \quad \text{with} \\ \mathcal{D}(\boldsymbol{\theta}) &\equiv -\frac{\theta_1^2 - \theta_2^2 + 2i\theta_1\theta_2}{|\boldsymbol{\theta}|^4} = \frac{-1}{(\theta_1 - i\theta_2)^2} .\end{aligned}\quad (41)$$

Hence, the complex shear γ is a convolution of κ with the kernel \mathcal{D} , or, in other words, \mathcal{D} describes the shear generated by a point mass. This relation can be inverted: in Fourier space this convolution becomes a multiplication,

$$\hat{\gamma}(\boldsymbol{\ell}) = \pi^{-1} \hat{\mathcal{D}}(\boldsymbol{\ell}) \hat{\kappa}(\boldsymbol{\ell}) \quad \text{for } \boldsymbol{\ell} \neq \mathbf{0} ,$$

which can be inverted to yield

$$\hat{\kappa}(\boldsymbol{\ell}) = \pi^{-1} \hat{\gamma}(\boldsymbol{\ell}) \hat{\mathcal{D}}^*(\boldsymbol{\ell}) \quad \text{for } \boldsymbol{\ell} \neq \mathbf{0}, \quad (42)$$

where the Fourier transform of \mathcal{D} is⁵

$$\hat{\mathcal{D}}(\boldsymbol{\ell}) = \pi \frac{(\ell_1^2 - \ell_2^2 + 2i\ell_1\ell_2)}{|\boldsymbol{\ell}|^2}; \quad (43)$$

note that this implies that $\hat{\mathcal{D}}(\boldsymbol{\ell})\hat{\mathcal{D}}^*(\boldsymbol{\ell}) = \pi^2$, which has been used in obtaining (42). It is obvious that $\hat{\mathcal{D}}$ is undefined for $\boldsymbol{\ell} = \mathbf{0}$, which has been indicated in the foregoing equations. Fourier back-transformation of (42) then yields

$$\begin{aligned} \kappa(\boldsymbol{\theta}) - \kappa_0 &= \frac{1}{\pi} \int_{\mathbb{R}^2} d^2\boldsymbol{\theta}' \mathcal{D}^*(\boldsymbol{\theta} - \boldsymbol{\theta}') \gamma(\boldsymbol{\theta}') \\ &= \frac{1}{\pi} \int_{\mathbb{R}^2} d^2\boldsymbol{\theta}' \operatorname{Re} [\mathcal{D}^*(\boldsymbol{\theta} - \boldsymbol{\theta}') \gamma(\boldsymbol{\theta}')]. \end{aligned} \quad (44)$$

Note that the constant κ_0 occurs since the $\boldsymbol{\ell} = \mathbf{0}$ -mode is undetermined. Physically, this is related to the fact that a uniform surface mass density yields no shear. Furthermore, it is obvious (physically, though not so easily seen mathematically) that κ must be real; for this reason, the imaginary part of the integral should be zero, and taking the real-part only [as in the second line of (44)] makes no difference. However, in practice this is different, since noisy data, when inserted into the inversion formula, will produce a non-zero imaginary part. What (44) shows is that if γ can be measured, κ can be determined.

Before looking at this in more detail, we briefly mention some difficulties with the inversion formula as given above:

- Since γ can at best be estimated at discrete points (galaxy images), smoothing is required. One might be tempted to replace the integral in (44) by a discrete sum over galaxy positions, but as shown by Kaiser & Squires (1993), the resulting mass density estimator has infinite noise (due to the θ^{-2} -behavior of the kernel \mathcal{D}).
- It is not the shear γ , but the reduced shear g that can be determined from the galaxy ellipticities; hence, one needs to obtain a mass density estimator in terms of g . In the case of ‘weak’ weak lensing, i.e., where $\kappa \ll 1$ and $|\gamma| \ll 1$, then $\gamma \approx g$.
- The integral in (44) extends over \mathbb{R}^2 , whereas data are available only on a finite field; therefore, it needs to be seen whether modifications allow the construction of an estimator for the surface mass density from finite-field shear data.

⁵ The form of $\hat{\mathcal{D}}$ can be obtained most easily by using the relations between the surface mass density and the shear components in terms of the deflection potential ψ , given in (IN-18). Fourier transforming those immediately yields $\hat{\kappa} = -|\boldsymbol{\ell}|^2 \hat{\psi}/2$, $\hat{\gamma}_1 = -(\ell_1^2 - \ell_2^2)\hat{\psi}/2$, $\hat{\gamma}_2 = -\ell_1\ell_2\hat{\psi}$. Eliminating $\hat{\psi}$ from the foregoing relations, the expression for $\hat{\mathcal{D}}$ is obtained.

- To get absolute values for the surface mass density, the additive constant κ_0 is of course a nuisance. As will be explained soon, this indeed is the largest problem in mass reconstructions, and is the *mass-sheet degeneracy* discussed in Sect. 2.5 of IN.

5.2 Improvements and generalizations

Smoothing. Smoothing of data is needed to get a shear field from discrete data points. Consider first the case that we transform (44) into a sum over galaxy images (ignoring the constant κ_0 for a moment, and also assuming the weak lensing case, $\kappa \ll 1$, so that the expectation value of ϵ is the shear γ),

$$\kappa_{\text{disc}}(\boldsymbol{\theta}) = \frac{1}{n \pi} \sum_i \mathcal{R}\text{e} [\mathcal{D}(\boldsymbol{\theta} - \boldsymbol{\theta}_i) \epsilon_i] , \quad (45)$$

where the sum extends over all galaxy images at positions $\boldsymbol{\theta}_i$ and complex ellipticity ϵ_i , and n is the number density of background galaxies. As shown by Kaiser & Squires (1993), the variance of this estimator for κ diverges. However, one can smooth this estimator, using a weight function $W(\Delta\theta)$ (assumed to be normalized to unity), to obtain

$$\kappa_{\text{smooth}}(\boldsymbol{\theta}) = \int d^2\theta' W(|\boldsymbol{\theta} - \boldsymbol{\theta}'|) \kappa_{\text{disc}}(\boldsymbol{\theta}') , \quad (46)$$

which now has a finite variance. One might expect that, since (i) smoothing can be represented by a convolution, (ii) the relation between κ and γ is a convolution, and (iii) convolution operations are transitive, it does not matter whether the shear field is smoothed first and inserted into (44), or one uses (46) directly. This statement is true if the smoothing of the shear is performed as

$$\gamma_{\text{smooth};1}(\boldsymbol{\theta}) = \frac{1}{n} \sum_i W(|\boldsymbol{\theta} - \boldsymbol{\theta}_i|) \epsilon_i . \quad (47)$$

If this expression is inserted into (44), one indeed recovers the estimate (46). However, this is not a particularly good method for smoothing, as can be seen as follows: the background galaxy positions will at least have Poisson noise; in fact, since the angular correlation function even of faint galaxies is non-zero, local number density fluctuations will be larger than predicted from a Poisson distribution. However, in the estimator (45) and in the smoothing procedure (47), these local variations of the number density are not taken into account. A much better way (Seitz & Schneider 1995) to smooth the shear is given by

$$\gamma_{\text{smooth};2}(\boldsymbol{\theta}) = \left[\sum_i W(|\boldsymbol{\theta} - \boldsymbol{\theta}_i|) \right]^{-1} \sum_i W(|\boldsymbol{\theta} - \boldsymbol{\theta}_i|) \epsilon_i , \quad (48)$$

which takes these local number density fluctuations into account. Lombardi & Schneider (2001) have shown that the expectation value of the smoothed shear estimate (48) is not exactly the shear smoothed by the kernel W , but the deviation (i.e., the bias) is very small provided the effective number of galaxy images inside the smoothing function W is substantially larger than unity, which will always be the case for realistic applications. Lombardi & Schneider (2002) then have demonstrated that the variance of (48) is indeed substantially reduced compared to that of (47), in agreement with the finding of Seitz & Schneider (1995).

When smoothed with a Gaussian kernel of angular scale θ_s , the covariance of the resulting mass map is finite, and given by (Lombardi & Bertin 1998; van Waerbeke 2000)

$$\text{Cov}(\kappa(\boldsymbol{\theta}), \kappa(\boldsymbol{\theta}')) = \frac{\sigma_\epsilon^2}{4\pi\theta_s^2 n} \exp\left(-\frac{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2}{2\theta_s^2}\right). \quad (49)$$

Thus, the larger the smoothing scale, the less noisy is the corresponding mass map; on the other hand, the more are features washed out. Choosing the appropriate smoothing scale is not easy; we shall come back to this issue in Sect. 5.3 below.

The non-linear case, $g \neq \gamma$. Noting that the reduced shear $g = \gamma/(1 - \kappa)$ can be estimated from the ellipticity of images (assuming that we avoid the potentially critical inner region of the cluster, where $|g| > 1$; indeed, this case can also be taken into account, at the price of somewhat increased complexity), one can write:

$$\kappa(\boldsymbol{\theta}) - \kappa_0 = \frac{1}{\pi} \int_{\mathbb{R}^2} d^2\boldsymbol{\theta}' [1 - \kappa(\boldsymbol{\theta}')] \text{Re}[\mathcal{D}^*(\boldsymbol{\theta} - \boldsymbol{\theta}') g(\boldsymbol{\theta}')] ; \quad (50)$$

this integral equation for κ can be solved by iteration, and it converges quickly (Seitz & Schneider 1995). Note that in this case, the undetermined constant κ_0 no longer corresponds to adding a uniform mass sheet. What the arbitrary value of κ_0 corresponds to can be seen as follows: The transformation

$$\begin{aligned} \kappa(\boldsymbol{\theta}) \rightarrow \kappa'(\boldsymbol{\theta}) &= \lambda\kappa(\boldsymbol{\theta}) + (1 - \lambda) \quad \text{or} \\ [1 - \kappa'(\boldsymbol{\theta})] &= \lambda[1 - \kappa(\boldsymbol{\theta})] \end{aligned} \quad (51)$$

changes the shear $\gamma \rightarrow \gamma' = \lambda\gamma$, and thus leaves g invariant; this is the mass-sheet degeneracy! It can be broken if magnification information can be obtained, since $\mathcal{A} \rightarrow \mathcal{A}' = \lambda\mathcal{A}$, so that

$$\mu \rightarrow \mu' = \lambda^{-2}\mu .$$

Magnification information can be obtained from the number counts of images (Broadhurst et al. 1995), owing to the magnification bias, provided the unlensed number density is sufficiently well known. In principle, the mass sheet

degeneracy can also be broken if redshift information of the source galaxies is available and if the sources are widely distributed in redshift; this can be seen as follows: let

$$Z(z_s) = \frac{D_{ds}/D_s}{\lim_{z_s \rightarrow \infty} D_{ds}/D_s} H(z_s - z_d) \quad (52)$$

(H being the Heaviside step function) be the ratio of the lens strength of a source at z_s to that of a fiducial source at infinite redshift (see Fig. 16); then, if κ_∞ and γ_∞ denote the surface mass density and shear for such a fiducial source, the reduced shear for a source at z_s is

$$g = \frac{Z\gamma_\infty}{1 - Z\kappa_\infty}, \quad (53)$$

and there is no global transformation of κ_∞ that leaves g invariant for sources at all redshifts, showing the validity of the above statement. However, even in this case the mass-sheet degeneracy is only mildly broken (see Bradac et al. 2004). In particular, only those regions in the cluster where the non-linearity (i.e., the difference between γ and g) is noticeably can contribute to the degeneracy breaking, that is, the region near the critical curves where $|g| \sim 1$.

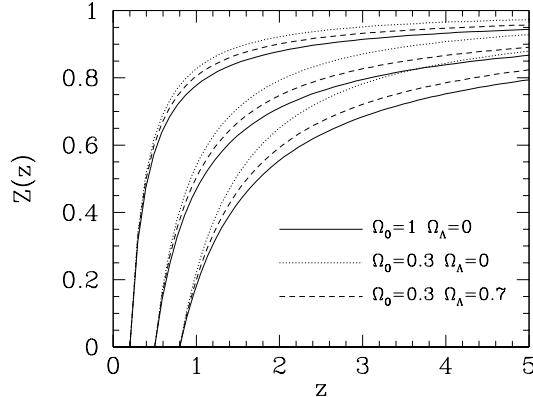


Fig. 16. The redshift weight function $Z(z_s)$, defined in (52), for three different values of the lens redshift $z_d = 0.2, 0.5$, and 0.8 , and three different geometries of the Universe, as indicated in the labels (here, Ω_m is denoted as Ω_0). Asymptotically for $z_s \rightarrow \infty$, all curves tend to $Z = 1$ (from Bartelmann & Schneider 2001)

In the non-linear case ($\gamma \neq g$) the reduced shear needs to be obtained from smoothing the galaxy ellipticities in the first place. Since the relation between g and κ is non-linear, the ‘transitivity of convolutions’ no longer applies; one thus cannot start from a discretization of an integral over image ellipticities and smooth the resulting mass map later. We also note that the accuracy with which the (reduced) shear is estimated can be improved provided redshift estimates of individual source galaxies are available (see

Fig. 17). In particular for high-redshift clusters, redshift information on individual source galaxies becomes highly valuable. This can be understood by considering a high-redshift lens, where an appreciable fraction of faint ‘source’ galaxies are located in front of the lens, and thus do not contribute to the lensing signal. However, they do contribute to the noise of the measurement. Redshift information allows the elimination of these foreground galaxies in the shear estimate and thus the reduction of noise.

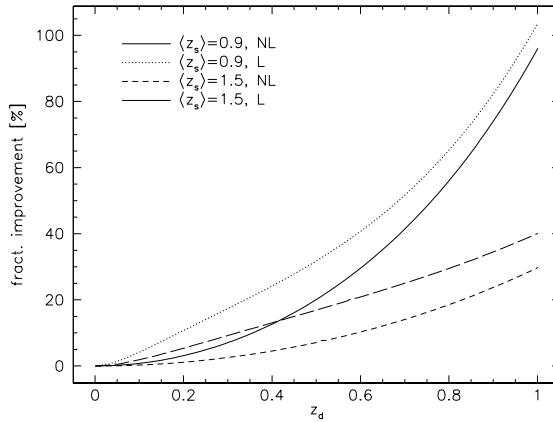


Fig. 17. The fractional gain in accuracy of the shear estimate when using redshift information of individual source galaxies, relative to the case where only the redshift distribution of the population is known, plotted as a function of the lens redshift. It is assumed that the sources have a broad redshift distribution, with a mean of $\langle z_s \rangle = 0.9$ (solid and dotted curves) or $\langle z_s \rangle = 1.5$ (short-dashed and long-dashed curves). The gain of accuracy also depends on the lens strength; the dotted and long-dashed curves assume local lens parameters of $\gamma_\infty = 0.3 = \kappa_\infty$, whereas the solid and short-dashed curves assume only very weak lensing, here approximated by $\gamma_\infty = 0 = \kappa_\infty$. One sees that the gain is dramatic once the lens redshift becomes comparable to the mean redshift of the source galaxies and is therefore of great importance for high-redshift clusters (from Bartelmann & Schneider 2001)

Finite-field mass reconstruction. In order to obtain a mass map from a finite data field, one starts from the relation (Kaiser 1995)

$$\nabla \kappa = \begin{pmatrix} \gamma_{1,1} + \gamma_{2,2} \\ \gamma_{2,1} - \gamma_{1,2} \end{pmatrix} \equiv \mathbf{u}_\gamma(\boldsymbol{\theta}), \quad (54)$$

which is a *local* relation between shear and surface mass density; it can easily be derived from the definitions of κ and γ in terms of ψ_{ij} . A similar relation

can be obtained in terms of reduced shear,

$$\nabla K(\boldsymbol{\theta}) = \frac{-1}{1 - g_1^2 - g_2^2} \begin{pmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{pmatrix} \begin{pmatrix} g_{1,1} + g_{2,2} \\ g_{2,1} - g_{1,2} \end{pmatrix} \equiv \mathbf{u}_g(\boldsymbol{\theta}), \quad (55)$$

where

$$K(\boldsymbol{\theta}) \equiv \ln[1 - \kappa(\boldsymbol{\theta})] \quad (56)$$

is a non-linear function of κ . Based on these local relations, finite-field inversion relations can be derived, and several of them appeared in the literature right after the foregoing equations have been published. For example, it is possible to obtain finite-field mass maps from line integrations (Schneider 1995; for other methods, see Squires & Kaiser 1996). Of all these finite-field methods, one can be identified as optimal, by the following reasoning: in the case of noise-free data, the imaginary part of (44) should vanish. Since one is always dealing with noisy data (at least coming from the finite intrinsic ellipticity distribution of the sources), in real life the imaginary part of (44) will not be zero. But being solely a noise component, one can choose that finite-field inversion which yields a zero imaginary component when averaged over the data field (Seitz & Schneider 1996). One way of deriving this mass map is obtained by a further differentiation of (54); this then yields a von Neumann boundary-value problem on the data field \mathcal{U} (Seitz & Schneider 2001),

$$\nabla^2 \kappa = \nabla \cdot \mathbf{u}_\gamma \quad \text{with} \quad \mathbf{n} \cdot \nabla \kappa = \mathbf{n} \cdot \mathbf{u}_\gamma \quad \text{on} \quad \partial \mathcal{U}, \quad (57)$$

where \mathbf{n} is the outward-directed normal on the boundary $\partial \mathcal{U}$ of \mathcal{U} . The analogous equation holds for K in terms of g and \mathbf{u}_g ,

$$\nabla^2 K = \nabla \cdot \mathbf{u}_g \quad \text{with} \quad \mathbf{n} \cdot \nabla K = \mathbf{n} \cdot \mathbf{u}_g \quad \text{on} \quad \partial \mathcal{U}. \quad (58)$$

Note that (57) determines the solution κ only up to an additive constant, and (58) determines K only up to an additive constant, i.e., $(1 - \kappa)$ up to a multiplicative factor. Hence, in both cases we recover the mass-sheet degeneracies for the linear and non-linear case, respectively. The numerical solution of these equations is fast, using overrelaxation (see Press et al. 1992). In fact, the foregoing formulation of the problem is equivalent (Lombardi & Bertin 1998) to the minimization of the action

$$A = \int_{\mathcal{U}} d^2\theta |\nabla \kappa(\boldsymbol{\theta}) - \mathbf{u}_\gamma(\boldsymbol{\theta})|^2, \quad (59)$$

from which the von Neumann problem can be derived as the Euler equation of the variational principle $\delta A = 0$. Furthermore, Lombardi & Bertin (1998) have shown that the solution of (57) is ‘optimal’, in that for this estimator the variance of κ is minimized.

Since (57) provides a linear relation between the shear and the surface mass density, one expects that it can also be written in the form

$$\kappa(\boldsymbol{\theta}) = \int_{\mathcal{U}} d^2\theta' \mathbf{H}(\boldsymbol{\theta}; \boldsymbol{\theta}') \cdot \mathbf{u}_\gamma(\boldsymbol{\theta}'), \quad (60)$$

where the vector field $\mathbf{H}(\boldsymbol{\theta}; \boldsymbol{\theta}')$ is the Green's function of the von Neumann problem (57). Accordingly,

$$K(\boldsymbol{\theta}) = \int_{\mathcal{U}} d^2\boldsymbol{\theta}' \mathbf{H}(\boldsymbol{\theta}; \boldsymbol{\theta}') \cdot \mathbf{u}_g(\boldsymbol{\theta}') . \quad (61)$$

Seitz & Schneider (1996) gave explicit expression for \mathbf{H} in the case of a circular and rectangular data field.

One might ask how important the changes in the resulting mass maps are compared to the Kaiser–Squires formula applied to a finite data field. For that we note that applying (44) or (50) to a finite data field is equivalent to setting the shear outside the data field to zero. Hence, the resulting mass distribution will be such as to yield a zero shear outside the data field, despite the fact that we have no indication from data that the shear indeed is zero there. This induces features in the mass map, in form of a pillow-like overall mass distribution. The amplitude of this feature depends on the strength of the lens, its location inside the data field, and in particular the size of the data field. Whereas for large data fields this amplitude is small compared to the noise amplitude of the mass map, it is nevertheless a systematic that can easily be avoided, and should be avoided, by using the finite-field inversions, which cause hardly any additional technical problems.

Various tests have been conducted in the literature as to the accuracy of the various inversions. For those, one generates artificial shear data from a known mass distribution, and compares the mass maps reconstructed with the various methods with the original (e.g., Seitz & Schneider 1996, 2001; Squires & Kaiser 1996). One of the surprising results of such comparisons is that in some cases, the Kaiser & Squires original reconstruction faired better than the explicit finite-field inversions, although it is known to yield systematics. The explanation for this apparent paradox is, however, easy: the mass models used in these test consisted of one or more localized mass peaks well inside the data field, so the shear outside the data field is very small. Noting that the KS formula applied to a finite data field is equivalent to setting $\gamma = 0$ outside the data field, this methods provides ‘information’ to the reconstruction process which is not really there, but for the mass models used in the numerical tests is in fact close to the truth. Of course, by adding this nearly correct ‘information’ to the mass reconstruction, the noise can be lowered relative to the finite-field reconstructions where no assumptions about the shear field outside the data field is made.

Constraints on the geometry of the Universe from weak lensing mass reconstructions. The strength of the lensing signal depends, for a given lens redshift, on the redshift of the sources, through the function $Z(z_s)$ (52). Suppose that the surface mass density of a cluster was well known, and that the redshifts of background sources can be determined. Then, by comparing the measured shear signal from sources at a given redshift z_s with the

one expected from the mass distribution, the value of $Z(z_s)$ can be determined. Since $Z(z)$ depends on the geometry of the Universe, parameterized through Ω_m and Ω_Λ , these cosmological parameters can in principle be determined. A similar strategy for strong lensing clusters was described at the end of Sect. 4.

Of course, the surface mass density of the cluster cannot be assumed to be known, but needs to be reconstructed from the weak lensing data itself. Consider for a moment only the amplitude of the surface mass density, assuming that its shape is obtained from the reconstruction. Changing the function $Z(z)$ by a multiplicative factor would be equivalent to changing the surface mass density Σ of the cluster by the inverse of this factor, and hence such a constant factor in Z is unobservable due to the mass-sheet degeneracy. Hence, not the amplitude of the function $Z(z)$ shown in Fig. 16 is important here, but its shape.

Lombardi & Bertin (1999) have suggested a method to perform cluster mass reconstructions and at the same time determine the cosmological parameters by minimizing the difference between the shear predicted from the reconstructed mass profile and the observed image ellipticities, where the former depends on the functional form of $Z(z)$. A nice and simple way to illustrate such a method was given in Gautret et al. (2000), called the ‘triplet method’. Consider three background galaxies which have a small separation on the sky, and assume to know the three source redshifts. Because of their closeness, one might assume that they all experience the same tidal field and surface mass density from the cluster. In that case, the shear of the three galaxies is described by five parameters, the two components of γ_∞ , κ , and Ω_m and Ω_Λ . From the six observables (two components of three galaxy ellipticities), one can minimize the difference between the predicted shear and the observed ellipticities with respect to these five parameters, and in particular obtain an estimate for the cosmological parameters. Repeating this process for a large number of triplets of background galaxies, the accuracy on the Ω 's can be improved, and results from a large number of clusters can be combined.

This procedure is probably too simple to be applied in practice; in particular, it treats κ_∞ and γ_∞ for each triplet as independent numbers, whereas the mass profile of the cluster is described by a single scalar function. However, it nicely illustrates the principle. Lombardi & Bertin (1999) have used a single density profile $\kappa_\infty(\theta)$ of the cluster, but assumed that the mass-sheet degeneracy is broken by some other means. Jain & Taylor (2003) suggested a similar technique for employing the lensing strength as a function of redshifts and cosmological parameters to infer constraints on the latter. Clearly, more work is needed in order to turn these useful ideas into a practically applicable method.

5.3 Inverse methods

In addition to these ‘direct’ methods for determining κ , inverse methods have been developed, such as a maximum-likelihood fit (Bartelmann et al. 1996; Squires & Kaiser 1996) to the data. There are a number of reasons why these are in principle preferable to the direct method discussed above. First, in the direct methods, the smoothing scale is set arbitrarily, and in general kept constant. It would be useful to obtain an objective way how this scale should be chosen, and perhaps, that the smoothing scale be a function of position: e.g., in regions with larger number densities of sources, the smoothing scale could be reduced. Second, the direct methods do not allow additional input coming from complementary observations; for example, if both shear and magnification information are available, the latter could not be incorporated into the mass reconstruction. The same is true for clusters where strong lensing constraints are known.

The shear likelihood function. In the inverse methods, one tries to fit a (very general) lens model to the observational data, such that the data agree within the estimated errors with the model. In the maximum-likelihood methods, one parameterizes the lens by the deflection potential ψ on a grid and then minimizes the regularized log-likelihood

$$-\ln \mathcal{L} = \sum_{i=1}^{N_g} \frac{|\epsilon_i - g(\boldsymbol{\theta}_i, \{\psi_n\})|^2}{\sigma_i^2(\boldsymbol{\theta}_i, \{\psi_n\})} + 2 \ln \sigma_i(\boldsymbol{\theta}_i, \{\psi_n\}) + \lambda_e S(\{\psi_n\}), \quad (62)$$

where $\sigma_i \approx \sigma_\epsilon \left(1 - |g(\boldsymbol{\theta}_i, \{\psi_n\})|^2\right)$ [see eq. (15) for the case $|g| < 1$ that was assumed here], with respect to these gridded ψ -values; this specific form of the likelihood assumes that the intrinsic ellipticity distribution follows a Gaussian with width σ_ϵ .⁶ In order to avoid overfitting, one needs a regularization term S ; entropy regularization (Seitz et al. 1998) seems very well suited (see Bridle et al. 1998; Marshall et al. 2002 for alternative regularizations). The entropy term S gets large if the mass distribution has a lot of structure; hence, in minimizing (62) one tries to match the data as closely as permitted by the entropic term (Narayan & Nityananda 1986). As a result, one obtains a model as smooth as compatible with the data, but where structure shows up where

⁶ This specific form (62) of the likelihood function assumes that the sheared ellipticity probability distribution follows a two-dimensional Gaussian with mean g and dispersion σ ; note that this assumption is not valid in general, not even when the intrinsic ellipticity distribution is Gaussian (see Geiger & Schneider 1999 for an illustration of this fact). The exact form of the lensed ellipticity distribution follows from the intrinsic distribution $p_s(\epsilon^{(s)})$ and the transformation law (12) between intrinsic and lensed ellipticity, $p(\epsilon) = p_s(\epsilon^{(s)}(\epsilon; g)) \det(\partial \epsilon^{(s)} / \partial \epsilon)$. However, in many cases the Gaussian approximation underlying (62) is sufficient and convenient for analytical considerations.

the data require it. The parameter λ_e is a Langrangean multiplier which sets the relative weight of the likelihood function and the regularization; it should be chosen such that the χ^2 per galaxy image is about unity, i.e.,

$$\sum_{i=1}^{N_g} \frac{|\epsilon_i - g(\boldsymbol{\theta}_i, \{\psi_n\})|^2}{\sigma_i^2(\boldsymbol{\theta}_i, \{\psi_n\})} \approx N_g ,$$

since then the deviation of the observed galaxy ellipticities from their expectation value g is as large as expected from the ellipticity dispersion. This choice of the regularization parameter λ_e then fixes the effective smoothing used for the reconstruction.

Strong lensing constraints can be incorporated into the inverse method by adding a term to the log-likelihood function which forces the minimum to satisfy these strong constraints nearly precisely. E.g., if a pair of multiple images at $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ is identified, one could add the term

$$\lambda_s |\boldsymbol{\beta}(\boldsymbol{\theta}_1) - \boldsymbol{\beta}(\boldsymbol{\theta}_2)|^2 = \lambda_s |[\boldsymbol{\theta}_1 - \boldsymbol{\alpha}(\boldsymbol{\theta}_1)] - [\boldsymbol{\theta}_2 - \boldsymbol{\alpha}(\boldsymbol{\theta}_2)]|^2$$

to the log-likelihood; by turning up the parameter λ_s , its minimum is guaranteed to correspond to a solution where the multiple image constraint is satisfied. Note that the form of this ‘source-plane minimization’ is simplified – see Sect. 4.6 of SL – but in the current context this approach suffices.

Magnification likelihood. Similarly, when accurate number counts of faint background galaxies are available, the magnification information can be incorporated into the log-likelihood function. If the number counts behave (locally) as a power law, $n_0(> S) \propto S^{-\alpha}$, the expected number of galaxies on the data field \mathcal{U} then is

$$\langle N \rangle = n_0 \int_{\mathcal{U}} d^2\theta |\mu(\boldsymbol{\theta})|^{\alpha-1} ; \quad (63)$$

see (26). The likelihood of observing N galaxies at the positions $\boldsymbol{\theta}_i$ can then be factorized into a term that yields the probability of observing N galaxies when the expected number is $\langle N \rangle$, and one that the N galaxies are at their observed locations. Since the probability for a galaxy to be at $\boldsymbol{\theta}_i$ is proportional to the expected number density there, $n = n_0 \mu^{\alpha-1}$, the likelihood function becomes (Seitz et al. 1998)

$$\mathcal{L}_\mu = P_N(\langle N \rangle) \prod_{i=1}^N |\mu(\boldsymbol{\theta}_i)|^{\alpha-1} , \quad (64)$$

with the first factor yielding the Poisson probability. Note that this expression assumes that the background galaxies are unclustered on the sky; in reality, where (even faint) galaxies cluster, this factorization does not strictly apply.

It should be pointed out that the deflection potential ψ , and not the surface mass density κ , should be used as variable on the grid, for two reasons:

first, shear and κ depend locally on ψ , and are thus readily calculated by finite differencing from ψ , whereas the relation between γ and κ is non-local and requires summation over all gridpoints, which is of course more time consuming. Second, and more important, the surface mass density on a finite field *does not* determine γ on this field, since mass outside the field contributes to γ as well. In fact, one can show (Schneider & Bartelmann 1997) that the shear inside a circle is fully determined by the mass distribution inside the circle and the multipole moments of the mass distribution outside the circle; in principle, the latter can thus be determined from the shear measurement.

Despite these reasons, some authors prefer to construct inverse methods in which the surface mass density on a grid serves as variables (e.g., Bridle et al. 1998; Marshall et al. 2002). The fact that the mass density on a finite field does not describe the shear in this field is accounted for in these methods by choosing a reconstruction grid that is larger than the data field and by allowing the surface mass density in this outer region to vary as well. Whereas the larger numerical grid requires a larger numerical effort, in addition to the non-local relation between κ and γ , this is of lesser importance, provided the numerical resources are available. Worse, however, is the view that the mass distribution outside the data field obtained by this method has any physical significance! It has not. This mass distribution is solely one of infinitely many that can approximately generate the shear in the data field from mass outside the data field. The fact that numerical tests show that one can indeed recover some of the mass distribution outside the data field is again a fluke, since these models are usually chosen such that all mass distribution outside the field is contained in a boundary region around the data field which is part of the numerical grid – and hence, the necessary ‘external’ shear must be generated by a mass distribution in this boundary zone which by construction is where it is. In real life, however, there is no constraint on where the ‘external’ shear contribution comes from.

5.4 Parameterized mass models

Whereas the parameter-free mass maps obtained through one of the methods discussed above provide a direct view of the mass distribution of a cluster, their quantitative interpretation is not straightforward. Peaks in the surface mass density can indicate the presence of a mass concentration, or else be a peak caused by the ellipticity noise of the galaxies. Since the estimated values for κ at different locations $\boldsymbol{\theta}$ are correlated [see eq. (49)], it is hard to imagine ‘error bars’ attached to each point. Therefore, it is often preferable to use parameterized mass models to fit the observed data; for example, fitting shear (and/or magnification) data to an NFW mass profile (see IN, Sect. 6.2) yields the virial mass M_{200} of the cluster and its concentration index c . There are basically two methods which have been used to obtain such parameterized models. The first one, assuming a spherical mass model, orders the tangential component of the observed image ellipticities into radial

bins and fits a parameterized shear profile through these bins, by minimizing a corresponding χ^2 -function. One of the disadvantages of this method is that the result of the fitting process can depend on the selected binning, but this can be largely avoided by choosing the bins fine enough. This then essentially corresponds to minimizing the first term in (62).

Alternatively, a likelihood method can be used, in which the log-likelihood function (62) – without the regularization term – is minimized, with the values of the potential on the grid $\{\psi_n\}$ replaced by a set of parameters which describe the mass profile. Schneider et al. (2000) have used this likelihood method to investigate with which accuracy the model parameters of a mass profile can be obtained, using both the shear information as well as magnification information from number counts depletion. One of the surprising findings of this study was that the slope of the fitted mass profile is highly degenerate if only shear information is used; indeed, the mass-sheet degeneracy strikes again and causes even fairly different mass profiles to have very similar reduced shear profiles, as is illustrated for a simple example in Fig. 18. In Fig. 19, the resulting degeneracy of the profile slope is seen. This degeneracy can be broken if number count information is used in addition. As seen in the middle panel of Fig. 18, the magnification profiles of the four models displayed are quite different and thus the number counts sensitive to the profile slope. Indeed, the confidence regions in the parameter fits, shown in Fig. 19, obtained from the magnification information are highly inclined relative to those from the shear measurements, implying that the combination of both methods yields much better constraints on the model parameters. Of course, as mentioned before, the mass-sheet degeneracy can also be broken if redshift information of individual background galaxies is available.

However, in order for the magnification information to yield significant constraints on the mass parameters, one needs to know the unlensed number density n_0 of sources quite accurately. In fact, even an uncertainty of less than $\sim 10\%$ in the value of n_0 renders the magnification information in relation to the shear information essentially useless (in the frame of parameterized models). Note that an accurate determination of n_0 is difficult to achieve: since n_0 corresponds to the unlensed number density of faint galaxies at the same flux limit as used for the actual data field, one requires an accurate photometric calibration. A flux calibration uncertainty of 0.1 mag corresponds to an uncertainty in n_0 of about $\sim 5\%$ for a slope of $\alpha = 0.5$, and such uncertainties are likely at the very faint flux limits needed to achieve a high number density of sources. In addition, the presence of bright cluster galaxies renders the detection and accurate brightness measurement of background galaxies difficult and requires masking of regions around them. Nevertheless, in cases where only magnification information is available, it can provide information on the mass profile by itself. Such a situation can occur for observing conditions with seeing above $\sim 1''$, when the shear method is challenged by the smallness of faint galaxies.

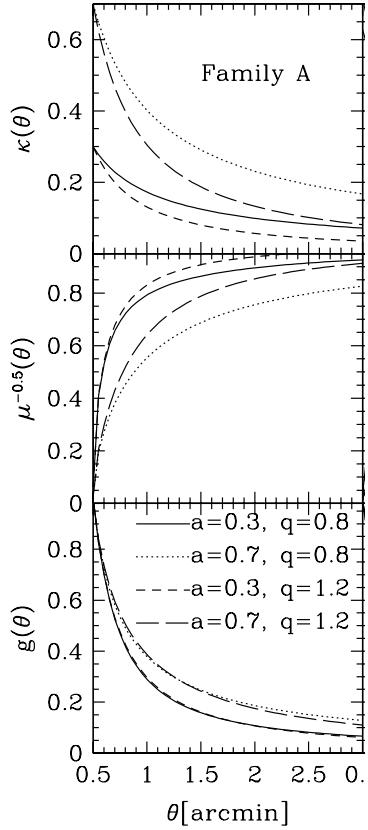


Fig. 18. The Einstein radius of a spherical mass distribution was assumed to be $\theta_E = 0.5$, and the density profile outside the Einstein radius was assumed to follow a power law, $\kappa(\theta) = a(\theta/\theta_E)^{-q}$; an SIS would have $a = 1/2$ and $q = 1$. The figure displays for four combinations of model parameters the surface mass density $\kappa(\theta)$, the function $\mu^{-1/2}$, which would be the depletion factor for source counts of slope $\beta = 1/2$, and the reduced shear $g(\theta)$. As can be seen, whereas the density profiles of the four models are quite different, the reduced shear profiles are pairwise almost fully degenerate. This is due to the mass-sheet degeneracy; it implies that it will be difficult to determine the slope q of the profiles from shear measurements alone, unless much larger fields around the cluster are used (from Schneider et al. 2000)

The result shown in Fig. 19 implies that the shape of the mass profile cannot be very well determined from the shear method, owing to the mass sheet degeneracy. This result extends to more general mass profiles than power-law models; e.g., King & Schneider (2001) considered NFW models with their two parameters c and r_{200} . A fairly strong degeneracy between these two parameters was found. Furthermore, the mass-sheet degeneracy renders it surprisingly difficult to distinguish an isothermal mass model from an NFW profile. The ability to distinguish these two families of models increases with a larger field-of-view of the observations. This expectation was indeed verified in King et al. (2002b) where the wide-field imaging data of the cluster A 1689 were analyzed with the likelihood method. Although the field size is larger than $30'$, so that the shear profile up to $\sim 15'$ from the cluster center can be measured, an NFW profile is preferred with less than 90% confidence over a power-law mass model. The determination of the mass profiles is likely to improve when strong lensing constraints are taken into account as well.

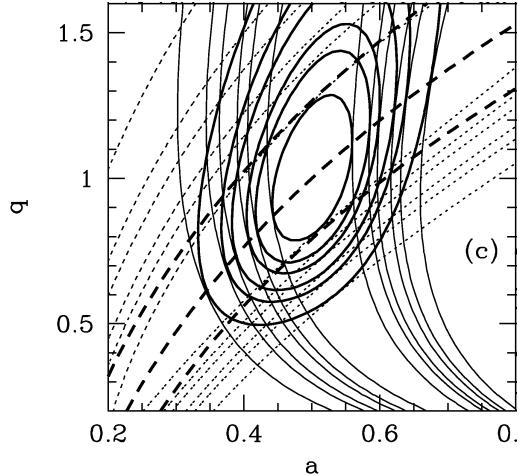


Fig. 19. For the power-law models of Fig. 18, confidence regions in the slope q and amplitude a are drawn, as derived from the shear (thin solid contours), the magnification (dotted) and their combination (thick solid). A number density of $30/\text{arcmin}^2$ for shear measurements and $120/\text{arcmin}^2$ for number counts was assumed. Thick dashed curves show models with constant total number of galaxies in the field, demonstrating that most of the constraint from magnification is due to the total counts, with little information about the detailed profile. It was assumed here that the unlensed number density of background galaxies is perfectly known; the fact that most of the magnification information comes from the total number of galaxies in the field implies that any uncertainty in the unlensed number density will quickly remove most of the magnification information (from Schneider et al. 2000)

The likelihood method for obtaining the parameters of a mass model is robust in the sense that the result is only slightly affected by substructure, as has been shown by King et al. (2001) using numerically generated cluster models. However, if a ‘wrong’ parameterization of the mass distribution is chosen, the interpretation of the resulting best-fit model must proceed carefully, and the resulting physical parameters, such as the total mass, may be biased. The principal problems with parameterized models are the same as for lens galaxies in strong lensing: unless the parameters have a well-defined physical meaning, one does not learn much, even if they are determined with good accuracy (see Sect. 4.7 of SL).

5.5 Problems of weak lensing cluster mass reconstruction and mass determination

In this section, some of the major problems of determining the mass profile of clusters from weak lensing techniques are summarized. The finite ellipticity

dispersion of galaxies generates a noise which provides a fundamental limit to the accuracy of all shear measurements. We will mention a number of additional issues here.

Number 1: The mass-sheet degeneracy. As mentioned several times, the major problem is the mass-sheet degeneracy, which implies that there is always one arbitrary constant that is undetermined from the shear data. Number count depletion can in principle lift this degeneracy, but this magnification effect has been observed in only a few clusters yet, and as mentioned above, this method has its own problems. Employing redshift information of individual source galaxies can also break this degeneracy (Bradac et al. 2004). Note that the mass-sheet degeneracy causes quite different mass profiles to have very similar reduced shear profiles.

Source redshift distribution. Since the critical surface mass density Σ_{cr} depends on the source redshift, a quantitative interpretation of the weak lensing mass reconstruction requires the knowledge of the redshift distribution of the galaxy sample used for the shear measurements. Those are typically so faint (and numerous) that it is infeasible to obtain individual spectroscopic redshifts for them. There are several ways to deal with this issue: probably the best is to obtain multi-color photometry of the fields and employ photometric redshift techniques (e.g. Connolly et al. 1995; Benítez 2000; Bolzonella et al. 2000). In order for them to be accurate, the number of bands needs to be fairly large; in addition, since much of the background galaxy population is situated at redshifts above unity, one requires near-IR images, as optical photometry alone cannot be used for photometric redshifts above $z \gtrsim 1.3$ (where the 4000 Å-break is redshifted out of the optical window). The problem with near-IR photometry is, however, that currently near-IR cameras have a substantially smaller field-of-view than optical cameras; in addition, due to the much higher sky brightness for ground-based near-IR observations, they extend to brighter flux limits (or smaller galaxy number densities) than optical images, for the same observing time. Nevertheless, upcoming wide-field near-IR cameras, such as the VISTA project on Paranal or WIRCAM at the CFHT, will bring great progress in this direction.

The alternative to individual redshift estimates of background galaxies is to use the redshift distribution obtained through spectroscopic (or detailed photometric redshift) surveys in other fields, and identify this with the faint background galaxy population at the same magnitude. In this way, the redshift distribution of the galaxies can be estimated. The issues that need to be considered here is that neither the targets for a spectroscopic survey, nor the galaxy population from which the shear is estimated, are strictly magnitude selected. Very small galaxies, for example, cannot be used for a shear estimate (or are heavily downweighted) owing to their large smearing corrections from the PSF. Similarly, for low-surface brightness galaxies it is much harder

to determine a spectroscopic redshift. Hence, in these redshift identifications, care needs to be excersized.

For cluster mass reconstructions, the physical mass scale is obtained from the average $\beta := \langle D_{ds}/D_s \rangle$ over all source galaxies. This average is fairly insensitive to the detailed redshift distribution, as long as the mean source redshift is substantially larger than the lens redshift. This is typically the case for low-redshift ($z \lesssim 0.3$) clusters. However, for higher-redshift lenses, determining β requires a good knowledge of the galaxy redshift distribution.

Contamination of the source sample. Next on the list is the contamination of the galaxy sample from which the shear is measured by cluster galaxies; a fraction of the faint galaxies will be foreground objects or faint cluster members. Whereas the foreground population is automatically taken into account in the normal lensing analysis (i.e., in determining β), the cluster members constitute an additional population of galaxies which is not included in the statistical redshift distribution. The galaxy sample used for the shear measurement is usually chosen as to be substantially fainter than the brighter cluster member galaxies; however, the abundance of dwarf galaxies in clusters (or equivalently, the shape of the cluster galaxy luminosity function) is not well known, and may vary substantially from cluster to cluster (e.g., Trentham & Tully 2002, and references therein). Including cluster members in the population from which the shear is measured weakens the lensing signal, since they are not sheared. As a consequence, a smaller shear is measured, and a lower cluster mass is derived. In addition, the dwarf contamination varies as a function of distance from the cluster center, so that the shape of the mass distribution will be affected. Color selection of faint galaxies can help in the selection of background galaxies, i.e., to obtain a cleaner set of true background galaxies. Of course, cluster dwarfs, if not properly accounted for, will also affect the magnification method. One method to deal with this problem is to use only galaxies redder than the Red Cluster Sequence of the cluster galaxies in the color-magnitude diagram, as this sequence indicates the reddest galaxies at the corresponding redshift.

Accuracy of mass determination via weak lensing. Comparing the ‘true’ mass of a cluster with that measured by weak lensing is not trivial, as one has to define what the true mass of a cluster is. Using clusters from numerical simulations, the mass is defined as the mass inside a sphere of radius r_{200} around the cluster center within which the overdensity is 200 times the critical density of the universe at the redshift considered. When comparing this mass with the projected mass inside a circle of radius $R = r_{200}$, one should not be surprised that the latter is larger (Metzler et al. 2001), since one compares apples (the mass inside a sphere) with oranges (the mass within a cylinder). Metzler et al. ascribed this to the mass in dark matter filaments

at the intersection of which massive clusters are located, but it is most likely mainly an effect of the mass definitions.

The mass-sheet degeneracy tell us there is little hope to measure the ‘total’ mass of a cluster without further assumptions. Therefore, one natural strategy is to assume a parameterized mass profile and see how accurately one can determine these parameters. The effect of ellipticity noise has already been described in Sect. 5.4. Using simulated clusters, Clowe et al. (2004a) have studied the effect of asphericity and substructure of clusters on these mass parameters, by analyzing the shear field obtained from independent projection of the clusters. They find that the non-spherical mass distribution and substructure induce uncertainties in the two parameters (r_{200} and the concentration c) of an NWF profile which are larger than those from the ellipticity noise under very good observing conditions. Among different projections of the same cluster, the value of r_{200} has a spread of 10 – 15%, corresponding to a spread in virial mass of $\sim 40\%$. Averaging over the different projections, they find that there is little bias in the mass determination, except for clusters with very large ellipticity.

Lensing by the large-scale structure. Lensing by foreground and background density inhomogeneities (i.e., the LSS), yields a fundamental limit to the accuracy of cluster mass estimates. Since lensing probes the projected density, these foreground and background inhomogeneities are present in the lensing signal. Hoekstra (2003) has investigated this effect in the determination of the parameters of an NFW mass profile; we shall return to this issue in Sect. 9.2 below when we consider lensing by the large-scale structure. In principle, the foreground and background contributions can be eliminated if the individual redshifts of the source galaxies are known, since in this case a three-dimensional mass reconstruction becomes possible (see Sect. 7.6); however, the resulting cluster mass map will be very noisy.

5.6 Results

After the first detection of a coherent alignment of galaxy images in two clusters by Tyson et al. (1990) and the development of the Kaiser & Squires (1993) mass reconstruction method, the cluster MS 1224+20 was the first for which a mass map was obtained (Fahlman et al. 1994). This investigation of the X-ray selected cluster yielded a mass map centered on the X-ray centroid of the cluster, but also a surprisingly high M/L -ratio of $\sim 800 h$ (here and in following we quote mass-to-light ratios always in Solar units). This high M/L ratio has later been confirmed in an independent analysis by Fischer (1999). This mass estimate is in strong conflict with that obtained from a virial analysis (Carlberg 1994); however, it is known that this cluster has a very complex structure, is not relaxed, and most likely a superposition of galaxy concentrations in redshift.

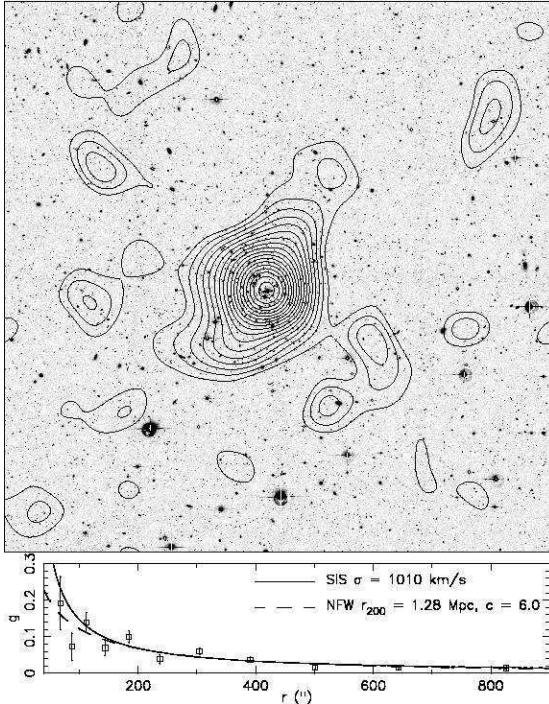


Fig. 20. Contours show the mass reconstruction of the cluster A1689, obtained from data taken with the WFI at the ESO/MPG 2.2m telescope. The image is $\sim 33'$ on a side, corresponding to $\sim 4.3 h^{-1} \text{ Mpc}$ at the cluster redshift of $z_d = 0.18$. In the lower panel, the reduced shear profile is shown, together with the best fitting SIS and NFW models. The mass reconstruction has been smoothed by a $1'.15$ Gaussian, and contour spacing is $\Delta\kappa = 0.01$. No corrections have been applied to account for contamination of the lensing signal by cluster dwarf galaxies – that would increase the mass of the best fit models by $\sim 25\%$ (taken from Clowe & Schneider 2001)

Since this pioneering work, mass reconstructions of many clusters have been performed; see Mellier (1999) and Sect. 5.4 of BS. Here, only a few recent results shall be mentioned, followed by a summary.

Wide-field mass reconstructions. The advent of large mosaic CCD cameras provides an opportunity to map large regions around clusters to be used for a mass reconstruction, and thus to measure the shear profile out to the virial radius of clusters. These large-scale observations offer the best promise to investigate the outer slope of the mass profile, and in particular distinguish between isothermal distributions and those following the NWF profile. Fig. 20 shows an example of such a mass reconstruction, that of the cluster Abell 1689 with $z_d = 0.182$. A significant shear is observed out to the virial radius. The mass peak is centered on the brightest cluster galaxy, and the overall lens signal is significant at the 13.4σ level. The shear signal is fit with two models, as shown in the lower panel of Fig. 20; the NWF profile yields a better fit than an SIS profile. Two more clusters observed with the WFI by Clowe & Schneider (2002) yield similar results, i.e., a detection of the lensing signal out to the virial radius, and a preference for an NWF mass profile, although in one of the two cases this preference is marginal. The lensing signal of such rich clusters could be contaminated by faint cluster member galaxies;

correcting for this effect would increase the estimate of the lensing strength, but requires multi-color imaging for source selection.

The cluster A1689 is (one of) the strongest lensing clusters known (see Fig. 14); in fact, it is strong enough so that a weak lensing signal can be significantly detected from near-IR images (King et al. 2002a) despite the fact that the usable number density of (background) galaxies is only $\sim 3 \text{ arcmin}^{-2}$. The estimate of its velocity dispersion from weak lensing yields an Einstein radius well below the distance of the giant arcs from the cluster center. Hence, in this cluster we see a discrepancy between the strong and weak lensing results, which cannot be easily explained by redshift differences between the arc sources and the mean redshift of the faint galaxies used for the weak lensing analysis. On the other hand, A1689 is known to be not a relaxed cluster, due to the redshift distribution of its member galaxies. This may explain the fact that the weak lensing mass estimates is also lower than that obtained from X-ray studies.

Filaments between clusters. One of the predictions of CDM models for structure formation is that clusters of galaxies are located at the intersection points of filaments formed by the dark matter distribution. In particular, this implies that a physical pair of clusters should be connected by a bridge or filament of (dark) matter, and weak lensing mass reconstructions can in principle be used to search for them. In the investigation of the $z = 0.42$ supercluster MS0302+17, Kaiser et al. (1998) found an indication of a possible filament connecting two of the three clusters, with the caveat (as pointed out by the authors) that the filament lies just along the boundary of two CCD chips; in fact, an independent analysis of this supercluster (Gavazzi et al. 2004) failed to confirm this filament. Gray et al. (2002) saw evidence for a filament connecting the two clusters A901A/901B in their mass reconstruction of the A901/902 supercluster field. Another potential filament has been found in the wide-field mass reconstruction of the field containing the pair of clusters A222/223 (Dietrich et al. 2004). Spectroscopy shows that there are also galaxies at the same redshift as the two clusters present in the ‘filament’ (Dietrich et al. 2002).

One of the problems related to the unambiguous detection of filaments is the difficulty to define what a ‘filament’ is, i.e. to devise a statistics to quantify the presence of a mass bridge. The eye easily picks up a pattern and identifies it as a ‘filament’, but quantifying such a pattern turns out to be very difficult, as shown by Dietrich et al. (2004). Because of that, it is difficult to distinguish between noise in the mass maps, the ‘elliptical’ extension of two clusters pointing towards each other, and a true filament. However, this problem is not specific to the weak investigation: even if the true projected mass distribution of a pair of clusters were known (e.g., from a cluster pair in numerical simulations), it is not straightforward to define what a filament would be.

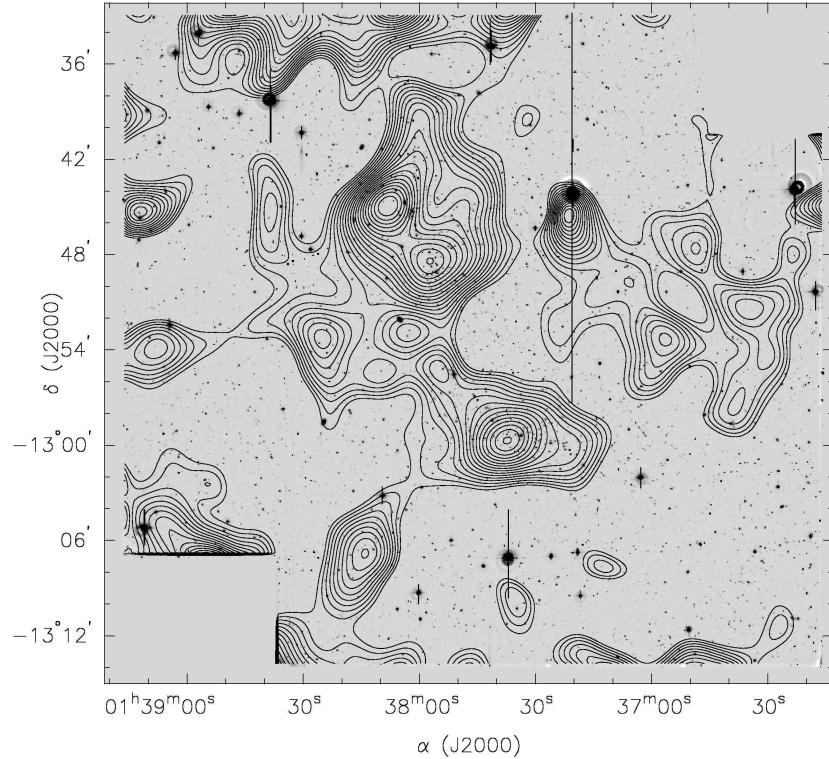


Fig. 21. A deep R-band image of the cluster pair Abell 222/223, obtained from two different pointings with the WFI@ESO/MPG 2.2m, with contours showing the reconstructed κ -map. The two clusters are in the region where the pointings overlap and thus deep imaging is available there. Both clusters are obviously detected in the mass map, with A223 (the Northern one) clearly split up into two subclusters. The mass reconstruction shows a connection between the two clusters which can be interpreted as a filament; galaxies at the clusters' redshift are present in this inter-cluster region. A further mass concentration is seen about $13'$ to the South-East of A222, which is significant at the 3.5σ level and where a clear concentration of galaxies is visible. A possible red cluster sequence indicates a substantially higher redshift for this cluster, compared to $z \approx 0.21$ of the double cluster (from Dietrich et al. 2004)

Correlation between mass and light. Mass reconstructions on wide-fields, particularly those covering supercluster regions, are ideally suited to investigate the relation between mass and galaxy light. For example, a smoothed light map of the color-selected early-type galaxies can be correlated with the reconstructed κ -map; alternatively, assuming that light traces mass, the expected shear map can be predicted from the early-type galaxies and compared to the observed shear, with the mass-to-light ratio being the essential fit parameter. Such studies have been carried out on the aforementioned supercluster fields, as well as on blank fields (Wilson et al. 2001). These studies yield very consistent results, in that the mass of clusters is very well traced by the distribution of early-type galaxies, but the mass-to-light ratio seems to vary between different fields, with $\sim 400h$ (in solar units) for the 0302 supercluster (Gavazzi et al. 2004), $\sim 200h$ for the A901/902 supercluster (Gray et al. 2002), and $\sim 300h$ for empty fields (Wilson et al. 2001) in the rest-frame B-band. When one looks in more detail at these supercluster fields, interesting additional complications appear. The three clusters in the 0302 field, as well as the three clusters in the A901/902 field (A901 is indeed a pair of clusters) have quite different properties. In terms of number density of color-selected galaxies, A901a and A902 dominate the field, whereas only A901b seems to be detected in X-rays. Considering early-type galaxies' luminosity, A901a is the most prominent of the three clusters. In contrast to this, A902 seems to be most massive as judged from the weak lensing reconstruction. Similar differences between the three clusters in the 0302 field are also seen. It therefore appears that the mass-to-light properties of clusters cover quite a range.

Cluster mass reconstructions from space. The exquisite image quality that can be achieved with the HST – imaging without the blurring effects of atmospheric seeing – suggests that such data would be ideal for weak lensing studies. This is indeed partly true: from space, the shape of smaller galaxy images can be measured than from the ground where the size of the seeing disk limits the image size of galaxies that can be used for ellipticity measurements in practice. Fig. 22 shows an HST image of the cluster A851 ($z_d = 0.41$), together with a mass reconstruction. The agreement between the mass distribution and the angular distribution of bright cluster galaxies is striking. A detailed X-ray observation of this cluster with XMM-Newton (De Filippis et al. 2003) finds two extended X-ray components coinciding with the two maxima of the bright galaxy distribution, and thus of the mass map shown in Fig. 22, in addition to several compact X-ray sources inside the HST field. Clearly, this cluster is a dynamically young system, as also seen by the inhomogeneities of the X-ray temperature and metallicity of the intracluster gas.

The drawback of cluster weak lensing studies with the HST is the small field-of-view of its WFPC2 camera, which precludes imaging of large regions

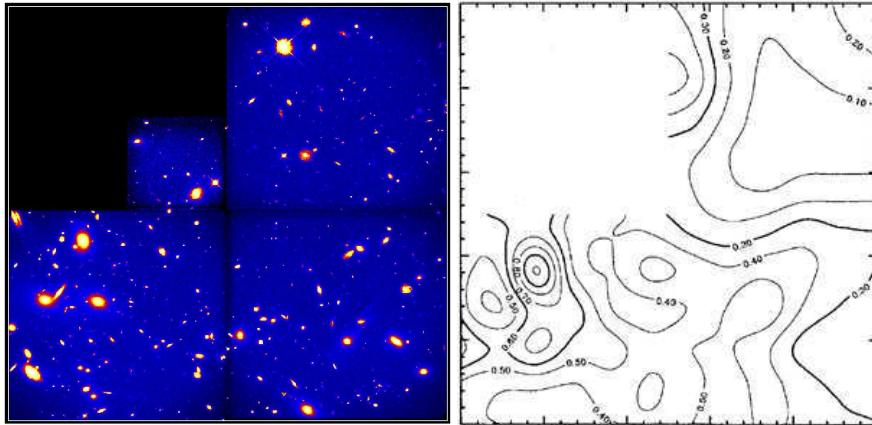


Fig. 22. The left panel shows an WFPC2@HST image of the cluster Cl0939+4713 (=Abell 851; taken from Seitz et al. 1996; the field is about 2.5 on a side), whereas the right panel shows a mass reconstruction obtained by Geiger & Schneider (1999); this was obtained using the entropy-regularized maximum likelihood method of Seitz et al. (1998). One notices the increased spatial resolution of the resulting mass map near the center of the cluster, which this method yields ‘automatically’ in those regions where the shear signal is large. Indeed, this mass map predicts that the cluster is critical in the central part, in agreement with the finding of Trager et al. (1997) that strong lensing features (multiple images plus an arc) of sources with $z \sim 4$ are seen there. The strong correlation between the distribution of mass and that of the bright cluster galaxies is obvious: Not only does the peak of the mass distribution coincide with the light center of the cluster, but also a secondary maximum in the surface mass density corresponds to a galaxy concentration (seen in the lower middle), as well as a pronounced minimum on the left where hardly any bright galaxies are visible

around the cluster center. To compensate for this, one can use multiple pointings to tile a cluster. For example, Hoekstra and collaborators have observed three X-ray selected clusters with HST mosaics; the results from this survey are summarized in Hoekstra et al. (2002d). One example is shown in Fig. 23, the high-redshift cluster MS1054–03 at $z_d = 0.83$. Also in this cluster one detects clear substructure, here consisting of three mass peaks, which is matched by the distribution of bright cluster galaxies. The shape of the mass maps indicates that this cluster is not relaxed, but perhaps in a later stage of merging, a view also supported by its hot X-ray temperature. In fact, new observations with Chandra and XMM-Newton of MS 1054 have shown that this cluster has a much lower temperature than measured earlier with ASCA (Gioia et al. 2004). Only two of the three components seen in the galaxy distribution and the mass reconstruction are seen in X-rays, with the central weak lensing component being the dominant X-ray source. The newly deter-

mined X-ray temperature is consistent with the velocity dispersion of cluster galaxies.

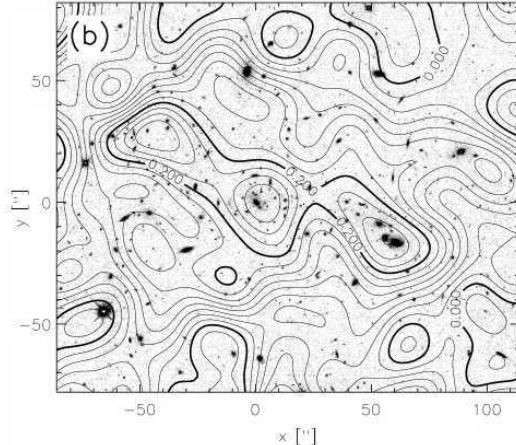


Fig. 23. Mass reconstruction (contours) of the inner part of the high-redshift ($z_d = 0.83$) cluster MS1054-03, based on a mosaic of six pointings obtained with the WFPC2@HST (from Hoekstra et al. 2000). The splitting of the cluster core into three subcomponents, also previously seen from ground-based images by Clowe et al. (2000), shows that this cluster is not yet relaxed

Magnification effects. As mentioned in Sect. 2.4, the magnification of a lens can also be used to reconstruct its surface mass density (Broadhurst et al. 1995). Provided a population of background source galaxies is identified whose number count slope α – see (26) – differs significantly from unity, local counts of these sources can be turned into an estimator of the local magnification. If the lens is weak, (27) provides a relation between the local number counts and the local surface mass density. If the lens is not weak, this relation no longer suffices, but one needs to use the full expression

$$|\mu|^{-1} = |(1 - \kappa)^2 - |\gamma|^2| , \quad (65)$$

where we have written absolute values to account for the fact that the sign of the magnification cannot be observed. There are two obvious difficulties with (65): the first comes from the sign ambiguities, namely whether μ is positive or negative, and whether $\kappa < 1$ or > 1 . Assuming that we are in the region of the cluster where $\mu > 0$ and $\kappa < 1$ (that is, outside the outer critical curve), then (65) can be rewritten as

$$\kappa = 1 - \sqrt{\mu^{-1} + |\gamma|^2} , \quad (66)$$

which shows the second difficulty: in order to estimate κ from μ , one needs to know the shear magnitude $|\gamma|$.

There are various ways to deal with this second problem. Consider first the case that the (reduced) shear is also observed, in which case one better writes

$$\kappa = 1 - [\mu (1 - |\gamma|^2)]^{-1/2} ; \quad (67)$$

but of course, if shear measurements are available, they should be combined with magnification observations in a more optimized way. A second method, using magnification only, is based on the fact that γ depends linearly on κ (ignoring finite-field problems here), and so (66) can be turned into a quadratic equation for the κ field (Dye & Taylor 1998). From numerical models of clusters, van Kampen (1998) claimed that the shear in these clusters approximately follows on average a relation of the form $|\gamma| = (1 - c)\sqrt{\kappa/c}$, with $c \sim 0.7$; however, there is (as expected) large scatter around this mean relation which by itself has little theoretical justification. Fig. 24 shows the mass reconstruction of the cluster Cl0024+17 using galaxy number counts and the two reconstruction methods just mentioned.

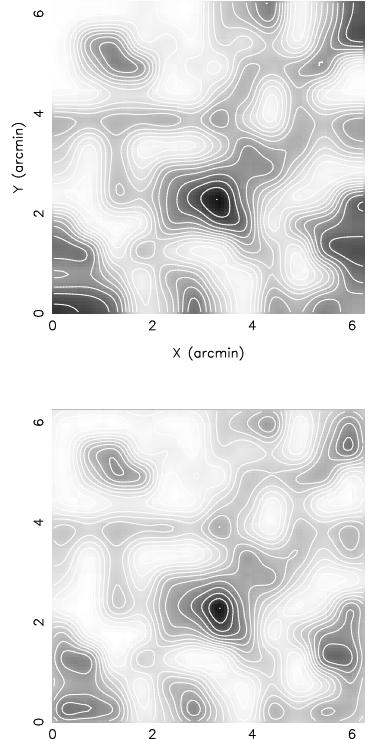


Fig. 24. Mass reconstruction of the cluster Cl0024+17 from the magnification method. The two different reconstructions are based on two different ways to turn the magnification signal – number count depletions – into a surface mass density map, as described in the text: in the upper panel, a local relation between surface mass density and shear magnitude has been used, whereas in the lower panel, the magnification was transformed into a κ map using the (non-local) quadratic dependence of the inverse magnification on the surface mass density field. Overall, these two reconstructions agree very well. To account for the presence of bright foreground galaxies, the data field had to be masked before local number densities of background galaxies were estimated – the mask is shown in Fig. 25 (from Dye et al. 2002)

Magnification effects have been observed for a few clusters, most notably Cl0024+17 (Fort et al. 1997; Rögnvaldsson et al. 2001; Dye et al. 2002) and A1689 (Taylor et al. 1998; Dye et al. 2001). We shall describe some of the results obtained for Cl0024+17 as an example (Dye et al. 2002). Since the cluster galaxies generate a local overdensity of galaxy counts, they need to

be removed first, which can be done based on a color and magnitude criterion. Comparison with extensive spectroscopy of this cluster (Czoske et al. 2001) shows that this selection is very effective for the brighter objects. For the fainter galaxies – those from which the lensing signal is actually measured – a statistical subtraction of foreground and cluster galaxies needs to be performed, which is done by subtracting galaxies according to the field luminosity function with $z < z_d$ and cluster galaxies according to the cluster luminosity function. The latter is based on the assumption that the luminosity distribution of cluster galaxies is independent from the distance to the cluster center. Next, the field of the cluster needs to be masked for bright objects, near which the photometry of fainter galaxies becomes inaccurate or impossible; Fig. 25 shows the masked data field. The number density of sources is then determined from the unmasked area. The resulting mass reconstruction is shown in Fig. 24. The results confirm the earlier finding from

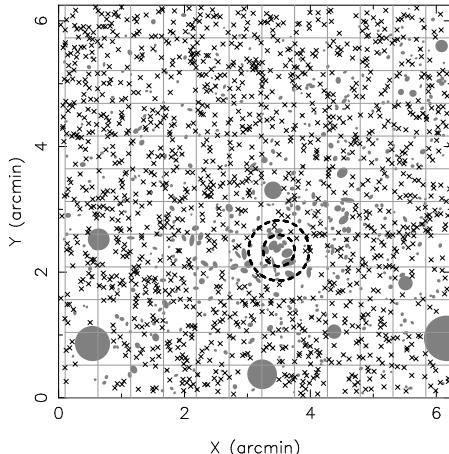


Fig. 25. The mask of the data field of the cluster Cl0024+17 (grey circles) and the location of putative background objects (crosses). The inner dashed circle shows the critical curve of the cluster as derived from the multiply imaged arc system (from Dye et al. 2002)

strong lensing (see Sect. 4.4) that the mass in the inner part of this cluster is larger by a factor ~ 3 than estimated from its X-ray emission (Soucail et al. 2000).

Magnification and shear method compared. It is interesting to consider the relative merits of shear and magnification methods for weak lensing studies. The number of clusters that have been investigated with either method are quite different, with less than a handful for which the magnification effect has been seen. The reason for this is multifold. First, the shear method does not need external calibration, as it is based on the assumption of random source ellipticity; in contrast to this, the magnification method requires the number counts of unlensed sources. Whereas this can be obtained from the same dataset, provided it covers a sufficiently large area,

this self-calibration removes one of the strongest appeals of the magnification effect, namely its potential to break the mass-sheet degeneracy. Second, the magnification method is affected by the angular correlation of galaxies, as clearly demonstrated by Athreya et al. (2002) in their study of the cluster MS 1008–1224, where the background number counts revealed the presence of a background cluster which, if not cut out of the data, would contaminate the resulting mass profile substantially. Third, the removal of foreground galaxies, and more seriously, of faint cluster members introduces an uncertainty in the results which is difficult to control. Finally, the number count method yields a lower lensing signal-to-noise than the shear method: If we consider N_γ and N_μ galaxies in a given patch of the sky, such that for the former ones the ellipticities have been measured, and for the latter ones accurate photometry is available and the galaxies are above the photometric completeness brightness, the signal-to-noise ratio from the shear – see (15) – and number count methods are

$$\left(\frac{S}{N}\right)_\gamma = \frac{|\gamma|}{\sigma_\epsilon} \sqrt{N_\gamma}; \quad \left(\frac{S}{N}\right)_\mu = 2\kappa|\alpha - 1| \sqrt{N_\mu}, \quad (68)$$

where we employed (27) in the latter case and assumed that the source galaxy positions are uncorrelated. The ratio of these two S/N values is

$$\frac{(S/N)_\gamma}{(S/N)_\mu} = \frac{|\gamma|}{\kappa} \frac{1}{2\sigma_\epsilon|1 - \alpha|} \sqrt{\frac{N_\gamma}{N_\mu}}. \quad (69)$$

For an isothermal mass profile, the first of these factors is unity. With $\sigma_\epsilon \approx 0.4$ and $\alpha \approx 0.75$ for R-band counts, the second factor is ~ 5 . The final factor depends on the quality of the data: in good seeing conditions, this ratio is of order unity. However, when the seeing is bad, the photometric completeness level can be considerably fainter than the magnitude for which the shape of galaxies can be measured reliably. Therefore, for data with relatively bad seeing, the magnification effect may provide a competitive means to extract weak lensing information. Having said all of this, the magnification method will keep its position as an alternative to shear measurements, in particular for future multi-color datasets where the separation of foreground and cluster galaxies from the background population can be made more cleanly.

Summary. The mass reconstruction of clusters using weak lensing has by now become routine; quite a few cameras at excellent sites yield data with sub-arcsecond image quality to enable this kind of work. Overall, the reconstructions have shown that the projected mass distribution is quite similar to that of the projected galaxy distribution and the shape of the X-ray emission, at least for clusters that appear relaxed. There is no strong evidence for a discrepancy between the mass obtained from weak lensing and that from X-rays, again with exceptions like for Cl0024+16 mentioned above (which most likely

is not a single cluster). The weak lensing mass profiles are considered more reliable than the ones obtained from X-ray studies, since they do not rely on symmetry or equilibrium assumptions. On the other hand, they contain contributions from foreground and background mass inhomogeneities, and are affected by the mass-sheet degeneracy. What is still lacking is a combined analysis of clusters, making use of weak lensing, X-ray, Sunyaev–Zeldovich, and galaxy dynamics measurements, although promising first attempts have been published (e.g., Zaroubi et al. 1998, 2001; Reblinsky 2000; Doré et al. 2001; Marshall et al. 2003).

5.7 Aperture mass and other aperture measures

In the weak lensing regime, $\kappa \ll 1$, the mass-sheet degeneracy corresponds to adding a uniform surface mass density κ_0 . However, one can define quantities in terms of the surface mass density which are invariant under this transformation. In addition, several of these quantities can be determined directly in terms of the locally measured shear. In this section we shall present the basic properties of the aperture measures, whereas in the following section we shall demonstrate how the aperture mass can be used to find mass concentrations based solely on their weak lensing properties.

Aperture mass. Let $U(|\boldsymbol{\theta}|)$ be a compensated weight (or filter) function, meaning $\int d\theta' \theta' U(\theta') = 0$, then the *aperture mass*

$$M_{\text{ap}}(\boldsymbol{\theta}_0) = \int d^2\theta \kappa(\boldsymbol{\theta}) U(|\boldsymbol{\theta} - \boldsymbol{\theta}_0|) \quad (70)$$

is independent of κ_0 , as can be easily seen. For example, if U has the shape of a Mexican hat, M_{ap} will have a maximum if the filter center is centered on a mass concentration. The important point to notice is that M_{ap} can be written directly in terms of the shear (Kaiser et al. 1994; Schneider 1996)

$$M_{\text{ap}}(\boldsymbol{\theta}_0) = \int d^2\theta Q(|\boldsymbol{\theta}|) \gamma_t(\boldsymbol{\theta}; \boldsymbol{\theta}_0), \quad (71)$$

where we have defined the *tangential component* γ_t of the shear relative to the point $\boldsymbol{\theta}_0$ [cf. eq. 17], and

$$Q(\theta) = \frac{2}{\theta^2} \int_0^\theta d\theta' \theta' U(\theta') - U(\theta). \quad (72)$$

These relations can be derived from (54), by rewriting the partial derivatives in polar coordinates and subsequent integration by parts (see Schneider & Bartelmann 1997); it can also be derived directly from the Kaiser & Squires

inversion formula (44), as shown in Schneider (1996). Perhaps easiest is the following derivation (Squires & Kaiser 1996): We first rewrite (70) as

$$\begin{aligned} M_{\text{ap}} &= 2\pi \int_0^{\theta_u} d\vartheta \vartheta U(\vartheta) \langle \kappa(\vartheta) \rangle \\ &= 2\pi [X(\vartheta) \langle \kappa(\vartheta) \rangle]_0^{\theta_u} - 2\pi \int_0^{\theta_u} d\vartheta X(\vartheta) \frac{d \langle \kappa \rangle}{d\vartheta}, \end{aligned} \quad (73)$$

where θ_u is the radius of the aperture, and we have defined

$$X(\theta) = \int_0^\theta d\vartheta \vartheta U(\vartheta).$$

This definition and the compensated nature of U implies that the boundary terms in (73) vanish. Making use of (24), one finds that

$$\frac{d \langle \kappa \rangle}{d\vartheta} = \frac{d\bar{\kappa}}{d\vartheta} - \frac{d \langle \gamma_t \rangle}{d\vartheta} = -\frac{2}{\vartheta} \langle \gamma_t \rangle - \frac{d \langle \gamma_t \rangle}{d\vartheta},$$

where we used (23) and (24) to obtain $d\bar{\kappa}/d\vartheta = -2 \langle \gamma_t \rangle / \vartheta$. Inserting the foregoing equation into (73), one obtains

$$\begin{aligned} M_{\text{ap}} &= 2\pi \int_0^{\theta_u} d\vartheta \vartheta \frac{2X(\vartheta)}{\vartheta^2} \langle \gamma_t(\vartheta) \rangle \\ &\quad + 2\pi [X(\vartheta) \langle \gamma_t(\vartheta) \rangle]_0^{\theta_u} - 2\pi \int_0^{\theta_u} d\vartheta \frac{dX}{d\vartheta} \langle \gamma_t(\vartheta) \rangle. \end{aligned} \quad (74)$$

The boundary term again vanishes, and one sees that the last equation has the form of (71), with the weight function $Q = 2X/\vartheta^2 - U$, reproducing (72).

We shall now consider a few properties of the aperture mass, which follow directly from (72).

- If U has finite support, then Q has finite support, which is due to the compensated nature of U . This implies that the aperture mass can be calculated on a finite data field, i.e., from the shear in the same circle where $U \neq 0$.
- If $U(\theta) = \text{const.}$ for $0 \leq \theta \leq \theta_{\text{in}}$, then $Q(\theta) = 0$ for the same interval, as is seen directly from (72). Therefore, the strong lensing regime (where γ deviates appreciably from g) can be avoided by properly choosing U (and Q).
- If $U(\theta) = (\pi\theta_{\text{in}}^2)^{-1}$ for $0 \leq \theta \leq \theta_{\text{in}}$, $U(\theta) = -[\pi(\theta_{\text{out}}^2 - \theta_{\text{in}}^2)]^{-1}$ for $\theta_{\text{in}} < \theta \leq \theta_{\text{out}}$, and $U = 0$ for $\theta > \theta_{\text{out}}$, then $Q(\theta) = \theta_{\text{out}}^2 \theta^{-2} [\pi(\theta_{\text{out}}^2 - \theta_{\text{in}}^2)]^{-1}$ for $\theta_{\text{in}} \leq \theta \leq \theta_{\text{out}}$, and $Q(\theta) = 0$ otherwise. For this special choice of U ,

$$M_{\text{ap}} = \bar{\kappa}(\theta_{\text{in}}) - \bar{\kappa}(\theta_{\text{in}}, \theta_{\text{out}}), \quad (75)$$

the mean mass density inside θ_{in} minus the mean density in the annulus $\theta_{\text{in}} \leq \theta \leq \theta_{\text{out}}$ (Kaiser 1995). Since the latter is non-negative, this yields a lower limit to $\bar{\kappa}(\theta_{\text{in}})$, and thus to $M(\theta_{\text{in}})$.

The aperture mass can be generalized to the case where the weight function U is constant on curves other than circles, e.g., on ellipses, in the sense that the corresponding expressions can be rewritten directly in terms of the shear on a finite region (see Squires & Kaiser 1996 for the case where U is constant on a set of self-similar curves, and Schneider & Bartelmann 1997 for a general set of nested curves). In general, M_{ap} is not a particularly good measure for the total mass of a cluster – since it employs a compensated filter – but it has been specifically designed that way to be immune against the mass-sheet degeneracy. However, M_{ap} is a very convenient measure for mass concentrations (see Sect. 5.8) and, as shown above, yields a robust lower limit on cluster masses.

Aperture multipoles. The aperture method can also be used to calculate multipoles of the mass distribution: define the multipoles

$$Q^{(n)} := \int d^2\theta \, |\boldsymbol{\theta}|^n \, U(|\boldsymbol{\theta}|) \, e^{ni\varphi} \, \kappa(\boldsymbol{\theta}), \quad (76)$$

then the $Q^{(n)}$ can again be expressed as an integral over the shear. Here, U is a radial weight function for which certain restrictions apply (see Schneider & Bartelmann 1997 for details), but is not required to be compensated for $n > 0$. A few cases of interest are: a weight function U which is non-zero only within an annulus $\theta_{\text{in}} \leq \theta \leq \theta_{\text{out}}$ and which continuously goes to zero as $\theta \rightarrow \theta_{\text{in,out}}$; in this case, the shear is required only within the same annulus. Likewise, if U is constant for $0 \leq \theta \leq \theta_{\text{in}}$ and then decreases smoothly to zero at θ_{out} , only the shear within the annulus is required to calculate the multipoles. Aperture multipoles can be used to calculate the multipole moments of mass concentrations like clusters directly from the shear, i.e., without obtaining first a mass map, which allows a more direct quantification of signal-to-noise properties.

The cross aperture. We have seen that the Kaiser & Squires inversion, given by the first expression in (44), must yield a real result; the imaginary part of the integral in (44) vanishes in the absence of noise. Suppose one would multiply the complex shear by $i = e^{2i\pi/4}$; this would transform the real part of the integral into the imaginary part and the imaginary part into the negative of the real part. Geometrically, multiplication by this phase factor corresponds to rotating the shear at every point by 45° . Hence, if all shears are rotated by $\pi/4$, the real part of the Kaiser & Squires inversion formula (44) yields zero. This 45-degree test has been suggested by A. Stebbins; it can be used on real data to test whether typical features in the mass map are significant, as those should have larger amplitude than spurious features obtained from the mass reconstruction in which the shear has been rotated by $\pi/4$ (the corresponding ‘mass map’ then yields a good indication of the typical noise present in the real mass map).

One can define in analogy to (71) the cross aperture by replacing the tangential component of the shear by its cross component. According to the 45-degree test, the resulting cross aperture should be exactly zero. Hence, if we define for $\boldsymbol{\theta}_0 = \mathbf{0}$

$$\begin{aligned} M := M_{\text{ap}} + iM_{\perp} &= \int d^2\theta Q(|\boldsymbol{\theta}|) [\gamma_t(\boldsymbol{\theta}) + i\gamma_x(\boldsymbol{\theta})] \\ &= - \int d^2\theta Q(|\boldsymbol{\theta}|) \gamma(\boldsymbol{\theta}) e^{-2i\phi}, \end{aligned} \quad (77)$$

where ϕ is the polar angle of $\boldsymbol{\theta}$ as in (17), then M is expected to be purely real. We shall make use of this definition and the interpretation of M in later sections.

5.8 Mass detection of clusters

Motivation. If a weak lensing mass reconstruction of a cluster has been performed and a mass peak is seen, it can also be quantified by applying the aperture mass statistics to it: placing the center of the aperture on the mass peak, and choosing the radius of the aperture to match the extent of the mass peak will give a significant positive value of M_{ap} . Now consider to observe a random field in the sky, and to determine the shear in this field. Then, one can place apertures on this field and determine M_{ap} at each point. If M_{ap} attains a significant positive value at some point, it then corresponds to a point around which the shear is tangentially oriented. Such shear patterns are generated by mass peaks according to (70) – hence, a significant peak in the M_{ap} -map corresponds to a mass concentration (which can, in principle at least, be a mass concentration just in two-dimensional projection, not necessarily in 3D). Hence, the aperture mass statistics allows us to search for mass concentrations on blank fields, using weak lensing methods (Schneider 1996). From the estimate (19), we see that the detectable mass concentrations have to have typical cluster masses.

The reason why this method is interesting is obvious: As discussed in Sect. 6 of IN, the abundance of clusters as a function of mass and redshift is an important cosmological probe. Cosmological simulations are able to predict the abundance of massive halos for a given choice of cosmological parameters. To compare these predictions with observations, cluster samples are analyzed. However, clusters are usually detected either as an overdensity in the galaxy number counts (possibly in connection with color information, to employ the red cluster sequence – see Gladders & Yee 2000), or from extended X-ray sources. In both cases, one makes use of the luminous properties of the clusters, and cosmologists find it much more difficult to predict those, as the physics of the baryonic component of the matter is much harder to handle than the dark matter. Hence, a method for cluster detection that is independent of their luminosity would provide a clean probe of cosmology. From

what was said above, the aperture mass provides such method (Schneider 1996).

To illustrate this point, we show in Fig. 26 the projected mass and the corresponding shear field as it results from studying the propagation of light rays through a numerically generated cosmological matter distribution (Jain et al. 2000; we shall return to such simulations in Sect. 6.6). From the comparison of these two panels, one sees that for each large mass concentration there is a tangential shear pattern centered on the mass peak. Thus, a systematic search for such shear patterns can reveal the presence and abundance of peaks in the mass map.

The method. The search for mass concentrations can thus be carried out by calculating the aperture mass on a grid over the data field and to identify significant peaks. A practical estimator for M_{ap} is obtained by replacing the integral in (71) by a finite sum over image ellipticities:

$$\hat{M}_{\text{ap}}(\boldsymbol{\theta}_0) = \frac{1}{n} \sum_i \epsilon_{ti}(\boldsymbol{\theta}_0) Q(|\boldsymbol{\theta}_i - \boldsymbol{\theta}_0|), \quad (78)$$

where n is the mean number density of galaxy images, and $\epsilon_{ti}(\boldsymbol{\theta}_0)$ is the ellipticity component of a galaxy at $\boldsymbol{\theta}_i$ tangent to the center $\boldsymbol{\theta}_0$ of the aperture. This estimator has easy-to-quantify signal-to-noise properties. In the absence of a lensing signal, $\langle \hat{M}_{\text{ap}} \rangle \equiv 0$, and the dispersion of $\hat{M}_{\text{ap}}(\boldsymbol{\theta}_0)$ is

$$\sigma^2(\boldsymbol{\theta}_0) = \frac{\sigma_\epsilon^2}{2n^2} \sum_i Q^2(|\boldsymbol{\theta}_i - \boldsymbol{\theta}_0|); \quad (79)$$

hence, the signal-to-noise of $\hat{M}_{\text{ap}}(\boldsymbol{\theta}_0)$ is

$$\frac{S}{N} = \frac{\sqrt{2}}{\sigma_\epsilon} \frac{\sum_i \epsilon_{ti}(\boldsymbol{\theta}_0) Q(|\boldsymbol{\theta}_i - \boldsymbol{\theta}_0|)}{\sqrt{\sum_i Q^2(|\boldsymbol{\theta}_i - \boldsymbol{\theta}_0|)}}. \quad (80)$$

The noise depends on $\boldsymbol{\theta}_0$, as the image number density can vary over the data field. The size (or radius) of the aperture shall be adapted to the mass concentrations excepted: too small aperture radii miss most of the lensing signal of real mass concentrations, but are more susceptible to noise peaks, whereas too large aperture radii include regions of very low signal which may be swamped again by noise. In addition, the shape of the filter function Q can be adapted to the expected mass profiles of mass concentrations; e.g., one can design filters which are particularly sensitive to NFW-like density profiles. In order not to prejudice the findings of a survey, it may be advantageous to use a ‘generic’ filter function, e.g., of the form

$$U(\vartheta) = \frac{9}{\pi\theta^2} \left(1 - \frac{\vartheta^2}{\theta^2}\right) \left(\frac{1}{3} - \frac{\vartheta^2}{\theta^2}\right); \quad Q(\vartheta) = \frac{6}{\pi\theta^2} \frac{\vartheta^2}{\theta^2} \left(1 - \frac{\vartheta^2}{\theta^2}\right). \quad (81)$$

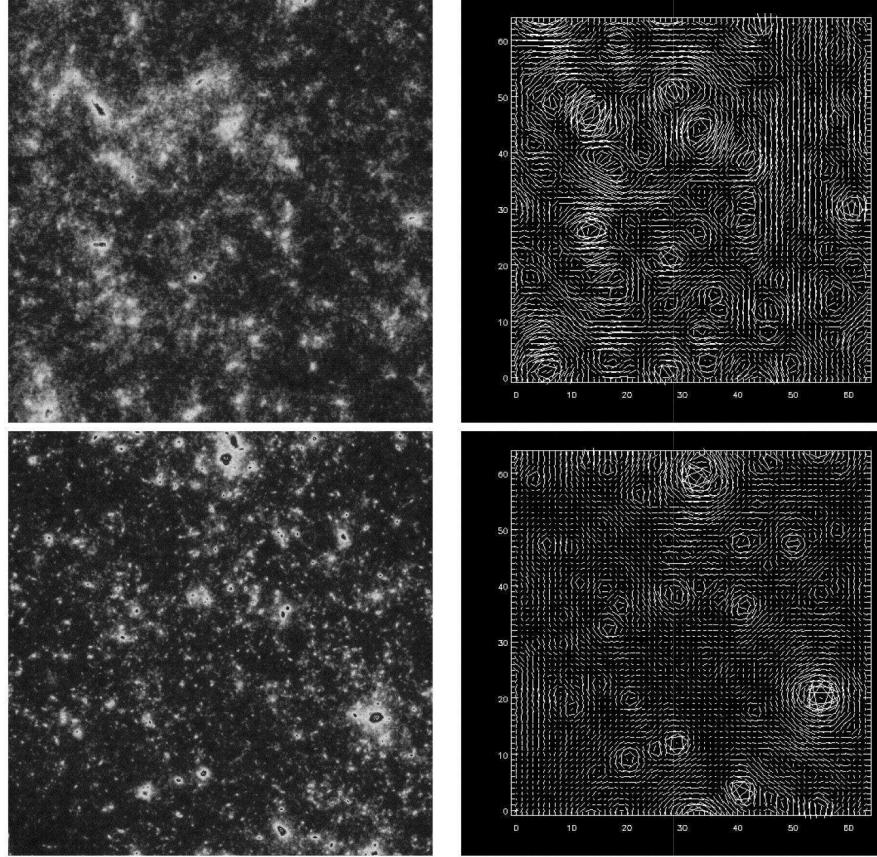


Fig. 26. Projected mass distribution of the large-scale structure (left), and the corresponding shear field (right), where the length and orientation of the sticks indicate the magnitude and direction of the local shear. The top panels correspond to an Einstein-de Sitter model of the Universe, whereas the bottom panels are for a low density open model. The size of the field is one degree on the side, and the background galaxies are assumed to all lie at the redshift $z_s = 1$. Note that each mass concentration seen in the left-hand panels generates a circular shear pattern at this position; this form the basic picture of the detection of mass concentrations from a weak lensing observation (from Jain et al. 2000)

The relation between the two expressions for M_{ap} given by (70) and (71) is only valid if the aperture lies fully inside the data field. If it does not, i.e., if the aperture crosses the boundary of the data field, these two expressions are no longer equivalent; nevertheless, the estimator (78) still measures a tangential shear alignment around the aperture center and thus signifies the presence of a mass concentration.

There are superior estimates of the significance of a detected mass peak than using the signal-to-noise ratio (80). One consists in bootstrapping; there one calculates M_{ap} at a given point (where N galaxies are in the aperture) many times by randomly drawing – with replacement – N galaxies and tests how often is signal negative. The fraction of cases with negative values corresponds to the error level of having a positive detection of M_{ap} . Alternatively, one can conduct another Monte-Carlo experiment, by randomizing all galaxy image orientations and calculating M_{ap} from these randomized samples, and ask in which fraction of realizations is the value of M_{ap} larger than the measured value? As the randomized galaxies should show no lensing signal, this fraction is again the probability of getting a value as large as that measured from random galaxy orientations. In fact, from the central limit theorem one expects that the probability distribution of M_{ap} from randomizing the image orientations will be a Gaussian of zero mean, and its dispersion can be calculated directly from (78) to be

$$\sigma^2(\boldsymbol{\theta}_0) = \frac{1}{2n^2} \sum_i |\epsilon_i|^2 Q^2(|\boldsymbol{\theta}_i - \boldsymbol{\theta}_0|), \quad (82)$$

which is similar to (79), but accounts for the moduli of the ellipticity of the individual galaxy images.

Both of the aforementioned methods take the true ellipticity distribution of galaxy images into account, and should yield very similar results for the significance. Highly significant peaks signify the presence of a mass concentration, detected solely on the basis of its mass, and therefore, it is a very promising search method for clusters.

There is nothing special about the weight function (81), except mathematical simplicity. It is therefore not clear whether these filter functions are most efficient to detect cluster-mass matter concentrations. In fact, as shown in Schneider (1996), the largest S/N is obtained if the filter function U follows the true mass profile of the lens or, equivalently, if Q follows its radial shear profile. Hennawi & Spergel (2003) and Schirmer (2004) tested a large range of filter functions, including (81), Gaussians, and those approximating an NFW profile. Based on numerical ray-tracing simulations, Hennawi & Spergel conclude that the ‘truncated’ NFW filter is most efficient for cluster detections; the same conclusion has been achieved by Schirmer (2004) based on wide-field imaging data.

Furthermore, Hennawi & Spergel have complemented their cluster search by a ‘tomographic’ component, assuming that the source galaxies have (photometric) redshift estimates available. Since the lens strength is a function of source redshift, the expected behaviour of the aperture mass signal as a function of estimated source redshift can be used as an additional search criteria. They shown that this additional information increases the sensitivity of weak lensing to find mass concentrations, in particular for higher-redshift ones; in fact, the cluster search by Wittman et al. (described below) has em-

ployed the use of redshift information. As an additional bonus, this method also provides an estimate of the lens redshift.

Results. In the past few years, a number of clusters and/or cluster candidates have been detected by the weak lensing method, and a few of them shall be discussed here. The right-hand panel of Fig. 27 shows the mass reconstruction of one of the 50 FORS1@VLT fields observed in the course of a cosmic shear survey (see Sect. 7.1). This reconstruction shows an obvious mass peak, indicated by a circle. The left panel shows the optical image, and it is obvious that the location of the mass peak coincides with a concentration of bright galaxies – this certainly is a cluster, detected by its weak lensing signal. However, no follow-up observations have been conducted yet to measure its redshift.

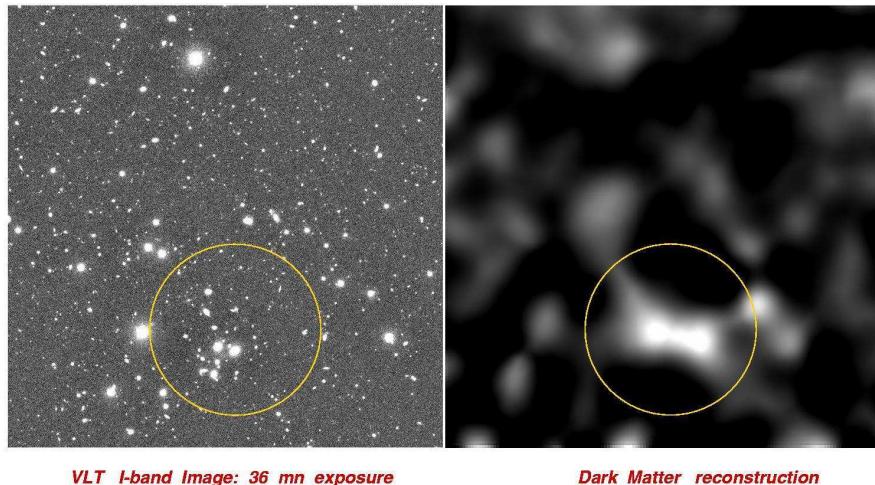


Fig. 27. A cosmic shear survey was carried out with the FORS1 instrument on the VLT (see Maoli et al. 2001 and Sect. 7.1 below). The left panel shows one of the 50 fields observed in the course of this survey, whereas the right panel shows a weak-lensing mass reconstruction of this field. Obviously, a strong mass peak is detected in this reconstruction, indicated by the circle. At the same position, one finds a strong overdensity of relatively bright galaxies on the VLT image; therefore, this mass peak corresponds to a cluster of galaxies. A reanalysis of all 50 VLT fields (Hetzterscheidt 2003) yielded no further significant cluster candidate; however, with a field size of only $\sim 6.5'$, detecting clusters in them is difficult unless these are positioned close to the field centers

Wittman et al. (2001, 2003) reported on the discovery of two clusters from their wide-field weak lensing survey; one of them is shown in Fig. 28 and discussed here. First, a peak in their mass reconstruction was identified

which has a significance of 4.5σ . The location of the mass peak is identified with a concentration of red elliptical galaxies, with the two centers separated by about $1'$ (which is about the accuracy with which the centers of mass concentrations are expected to be determined from mass reconstructions). Follow-up spectroscopy confirmed the galaxy concentration to be a cluster at redshift $z_d = 0.28$, with a velocity dispersion of $\sigma_v \sim 600$ km/s. Since multi-color photometry data are available, photometric redshift estimates of the faint galaxy population have been obtained, and the tangential shear around the mass peak has been investigated as a function of this estimated redshift. The lens signal rises as the redshift increases, as expected due to the lensing efficiency factor D_{ds}/D_s . In fact, from the source redshift dependence of the lens signal, the lens redshift can be estimated, and yields a result within ~ 0.03 of the spectroscopically measured z_d . Hence, in this case not only can the presence of a cluster be inferred from weak lensing, but at the same time a cluster redshift has been obtained from lensing observations alone. This is one example of using source redshift information to investigate the redshift structure of the lensing matter distribution; we shall return to a more general discussion of this issue in Sect. 7.6.

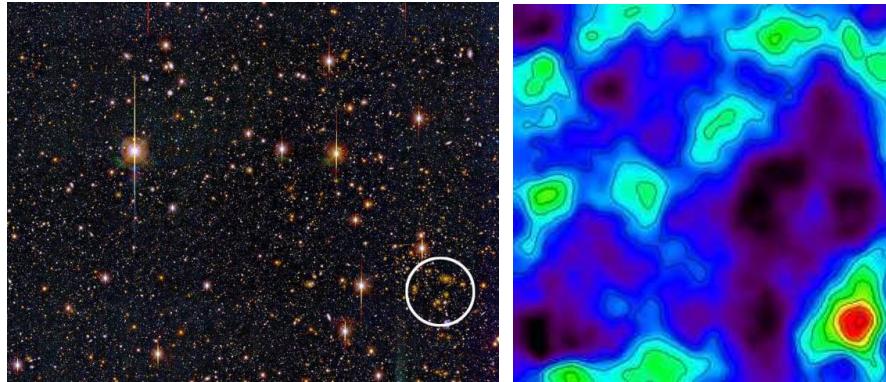


Fig. 28. Left: BTC image of a blank field, right: mass reconstruction, showing the presence of a (mass-selected) cluster near the lower right corner – spectroscopically verified to be at $z = 0.276$ (from Wittman et al. 2001)

In a wide-field imaging weak lensing survey of galaxy clusters, Dahle et al. (2003) detected three significant mass peaks away from the clusters that were targeted. One of these cases is illustrated in Fig. 29, showing the mass reconstruction in the field of the cluster A 1705. The mass peak South-West of the cluster coincides with a galaxy concentration at $z \sim 0.55$, as estimated from their color, and an arc is seen near the brightest galaxy of this cluster. A further cluster was detected in the wide-field image of the A222/223 double cluster field (Dietrich et al. 2004) which coincides with an overdensity of

galaxies. Hence, by now of order ten cluster-mass matter concentrations have been discovered by weak lensing techniques and verified as genuine clusters from optical photometry and, for some of them, spectroscopy.

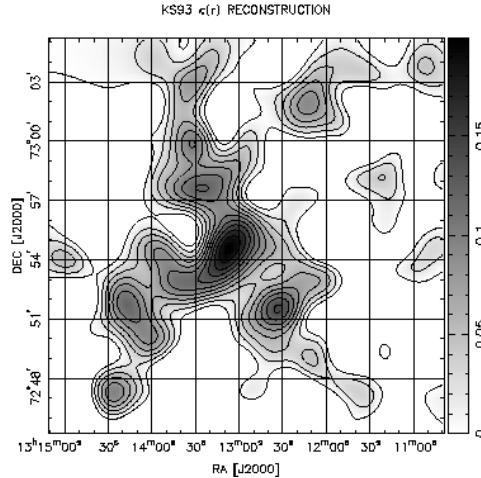


Fig. 29. Shown is the mass reconstruction of the field containing the cluster A1705, located near the center of this field. The peak $\sim 4'$ to the North-East of A1705 appears to be associated with galaxies at the same redshift as A1705. However, the peak $\sim 4'$ South-West of A1705 seems to be associated with galaxies at considerably larger redshift, at $z \sim 0.55 \pm 0.05$, as determined from the $V - I$ colors of the corresponding galaxy concentration. Indeed, an arc curving around the central galaxy of this newly detected cluster candidate is observed (from Dahle et al. 2003)

Miyazaki et al. (2002) used a 2.1 deg^2 deep image taken with the Suprime-Cam wide-field imager on Subaru to search for mass peaks. They compared their peak statistics with both, the expected peak statistics from a noise field created by intrinsic galaxy ellipticities (Jain & van Waerbeke 2000) as well as from N-body simulations, and found a broader distribution in the actual data. They interpret this as statistical evidence for the presence of mass peaks; however, their interpretation of the significant dips in the mass map as evidence for voids cannot hold, as the density contrast of voids is too small (since the fractional density contrast $\delta > -1$) to be detectable with weak lensing. They find a number density of $> 5\sigma$ peaks of about 5 deg^{-2} , well in agreement with predictions from Kruse & Schneider (1999) and Reblinsky et al. (1999). Schirmer (2004) investigated about 16 deg^2 of images taken with the WFI@ESO/MPG 2.2m, and detected $100 > 4\sigma$ -peaks, again in good agreement with theoretical expectations.

Dark clusters? In addition, however, this method has the potential to discover mass concentrations with very large mass-to-light ratio, i.e., clusters which are very faint optically and which would be missed in more conventional surveys for clusters. Two potential ‘dark clusters’ have been reported in the literature.⁷ Umetsu & Futamase (2000), using the WFPC2 onboard

⁷ A third case reported in Miralles et al. (2002) has in the meantime been considerably weakened (Erben et al. 2003).

HST detected a highly significant (4.5σ) mass concentration $1.7'$ away from the cluster Cl 1604+4304, also without an apparent overdensity of associated galaxies.

In the course of a wide-field weak lensing analysis of the cluster A 1942, Erben et al. (2000) detected a mass peak which, using the aperture mass statistics introduced previously, has been shown to be highly significant ($\sim 4.7\sigma$ on the V-band image), with the significance being obtained from the randomization and bootstrapping techniques described above. An additional I-band image confirmed the presence of a mass peak at the same location as on the V-band image, though with somewhat lower significance. No concentration of galaxies is seen near the location of the mass peak, which indicates that it either is a very dark mass concentration, or a cluster at a fairly high redshift (which, however, would imply an enormous mass for it), or, after all, a statistical fluke. It is important to note that the signal in M_{ap} comes from a range of radii (see Fig. 30); it is not dominated by a few highly flattened galaxies which happen to have a fortuitous orientation. Gray et al. (2000) have used near-IR images to search for a galaxy concentration in this direction, without finding an obvious candidate. Therefore, at present it is unclear whether the ‘dark clump’ is indeed a very unusual cluster. A low-significance X-ray source near its position, as obtained in a ROSAT observation of A 1942, certainly needs confirmation by the more sensitive X-ray observatory XMM.⁸ Of course, if there are really dark clusters, their confirmation by methods other than weak lensing would be extremely difficult; but even if we are dealing with a statistical fluke, it would be very important to find the cause for it. An HST mosaic observation of this field has been conducted; a first analysis of these data was able to confirm the findings of Erben et al., in the sense that the shear signal from galaxies seen in both, the HST images and the ground-based data, have a significant tangential alignment (von der Linden 2004). However, contrary to expectations if this was truly a lensing mass signal, there is hardly any tangential alignment from fainter galaxies, although they are expected to be located at higher redshift and thus should show a stronger shear signal. However, as a word of caution, the PSF anisotropy of WFPC2 cannot be controlled from stars on the image, owing to the small field-of-view, and no stellar cluster has been observed with the filter with which the dark clump observations were conducted, so that the PSF anisotropy cannot be accurately inferred from such calibration images. The existence of dark clusters would be highly unexpected in view of our current understanding of structure formation and galaxy evolution, and would require revisions of these models.

The search for clusters by weak lensing will certainly continue, due to the novel properties of the cluster samples obtained that way. The observational data required are the same as those used for cosmic shear studies, and

⁸ Judging from the results of several proposal submissions, people on X-ray TACs seem not to care too much about dark cluster candidates.

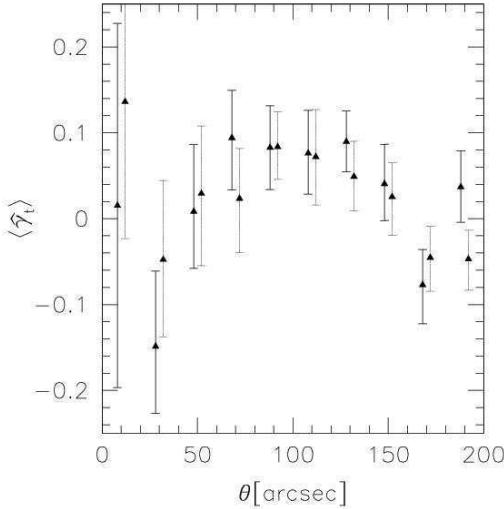


Fig. 30. Tangential shear profile from both (V- and I-band) images around the ‘dark cluster’ candidate near the cluster A1942. For each angular scales, two points (and corresponding error bars) are plotted, which are derived from two different images of the field in the V- and I-band. It can be seen that the tangential shear signal extends over quite a range in radius (from Erben et al. 2000)

several very wide-field surveys are currently conducted, as will be described in Sect. 7. Hence, we can expect to have a sizable sample of shear-selected clusters in the near future. The search for mass concentrations by weak lensing techniques is affected by foreground and background inhomogeneities, which impose fundamental limits on the reliability and completeness of such searches; we shall return to this issue in Sect. 9.2.

Expectations. Kruse & Schneider (1999) have calculated the expected number density of lensing-detected clusters, using the aperture-mass method, for different cosmological parameters; these have been verified in numerical simulations of the large-scale structure by Reblinsky et al. (1999). Depending on the cosmological model, a few clusters per deg^2 should be detected at about the 5σ level. The dependence of the expected number density of detectable mass peaks on the cosmological parameters can be used as a cosmological probe; in particular, Bartelmann et al. (2002) and Weinberg & Kamionkowski (2003) demonstrate that the observed abundance of weak lensing clusters can probe the equation-of-state of the dark energy. Bartelmann et al. (2001) argued that the abundance of weak lensing detected clusters strongly depends on their mass profile, with an order-of-magnitude difference between NFW profiles and isothermal spheres. Weinberg & Kamionkowski (2002) argued, based on the spherical collapse model of cluster formation, that a considerable fraction of such detections are expected to be due to non-virialized mass concentrations, which would then be considerably weaker X-ray emitters and may be candidates for the ‘dark clusters’.

6 Cosmic shear – lensing by the LSS

Up to now we have considered the lensing effect of localized mass concentrations, like galaxies and clusters. In addition to that, light bundles propagating through the Universe are continuously deflected and distorted by the gravitational field of the inhomogeneous mass distribution, the large-scale structure (LSS) of the cosmic matter field. This distortion of light bundles causes shape and size distortions of images of distant galaxies, and therefore, the statistics of the distortions reflect the statistical properties of the LSS (Gunn 1967; Blandford et al. 1991; Miralda-Escudé 1991; Kaiser 1992).

Cosmic shear deals with the investigation of this connection, from the measurement of the correlated image distortions to the inference of cosmological information from this distortion statistics. As we shall see, cosmic shear has become a very important tool in observational cosmology. From a technical point-of-view, it is quite challenging, first because the distortions are indeed very weak and therefore difficult to measure, and second, in contrast to ‘ordinary’ lensing, here the light deflection does not occur in a ‘lens plane’ but by a 3-D matter distribution, implying the need for a different description of the lensing optics. We start by looking at the description of light propagating through the Universe, and then consider the second-order statistical properties of the cosmic shear which reflect the second-order statistical properties of the cosmic matter field, i.e., the power spectrum. Observational results from cosmic shear surveys are presented in Sect. 7, whereas higher-order statistical properties of the shear field will be treated in Sect. 9.

6.1 Light propagation in an inhomogeneous Universe

In this brief, but rather technical section, we outline the derivation of the lensing effects of the three-dimensional mass distribution between the faint background galaxy population and us; the reader is referred to Bartelmann & Schneider (2001) for a more detailed discussion. The final result of this consideration has a very simple interpretation: in the lowest-order approximation, the 3-D cosmological mass distribution can be considered, for sources at a single redshift z_s , as an effective surface mass density κ , just like in ordinary lensing. The resulting κ is obtained as a line-of-sight integral of the density contrast $\Delta\rho$, weighted by the usual geometrical factor entering the lens equations.

The laws of light propagation follow from Einstein’s General Relativity; according to it, light propagates along the null-geodesics of the space-time metric. As shown in SEF (see also Seitz et al. 1994), one can derive from General Relativity that the governing equation for the propagation of thin light bundles through an arbitrary space-time is the equation of geodesic deviation,

$$\frac{d^2\boldsymbol{\xi}}{d\lambda^2} = \mathcal{T}\boldsymbol{\xi}, \quad (83)$$

where ξ is the separation vector of two neighboring light rays, λ the affine parameter along the central ray of the bundle, and \mathcal{T} is the *optical tidal matrix* which describes the influence of space-time curvature on the propagation of light. \mathcal{T} can be expressed directly in terms of the Riemann curvature tensor.

For the case of a weakly inhomogeneous Universe, the tidal matrix can be explicitly calculated in terms of the peculiar Newtonian potential. For that, we write the slightly perturbed metric of the Universe in the form

$$ds^2 = a^2(\tau) \left[\left(1 + \frac{2\Phi}{c^2} \right) c^2 d\tau^2 - \left(1 - \frac{2\Phi}{c^2} \right) (dw^2 + f_K^2(w)d\omega^2) \right], \quad (84)$$

where w is the comoving radial distance, $a = (1+z)^{-1}$ the scale factor, normalized to unity today, τ is the conformal time, related to the cosmic time t through $dt = a d\tau$, $f_K(w)$ is the comoving angular diameter distance, which equals w in a spatially flat model, and $\Phi(\mathbf{x}, w)$ denotes the Newtonian peculiar gravitational potential which depends on the comoving position vector \mathbf{x} and cosmic time, here expressed in terms of the comoving distance w (see Sect. 4 of IN for a more detailed description of the various cosmological terms). In this metric, the tidal matrix \mathcal{T} can be calculated in terms of the Newtonian potential Φ , and correspondingly, the equation of geodesic deviation (83) yields the evolution equation for the comoving separation vector $\mathbf{x}(\theta, w)$ between a ray separated by an angle θ at the observer from a fiducial ray

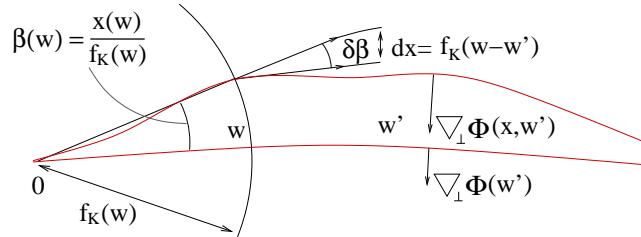


Fig. 31. Illustration of the evolution of the separation between two light rays in a curved space-time (source: T. Schrabback)

$$\frac{d^2\mathbf{x}}{dw^2} + K \mathbf{x} = -\frac{2}{c^2} [\nabla_{\perp} \Phi(\mathbf{x}(\theta, w), w) - \nabla_{\perp} \Phi^{(0)}(w)], \quad (85)$$

where $K = (H_0/c)^2 (\Omega_m + \Omega_\Lambda - 1)$ is the spatial curvature of the Universe, $\nabla_{\perp} = (\partial/\partial x_1, \partial/\partial x_2)$ is the transverse *comoving* gradient operator, and $\Phi^{(0)}(w)$ is the potential along the fiducial ray.⁹ The formal solution of

⁹ In some of the literature, this transport equation is written without the term accounting for the potential along the fiducial ray. The idea behind this is to

this transport equation is obtained by the method of Green's function, to yield

$$\mathbf{x}(\boldsymbol{\theta}, w) = f_K(w)\boldsymbol{\theta} - \frac{2}{c^2} \int_0^w dw' f_K(w-w') \left[\nabla_{\perp} \Phi(\mathbf{x}(\boldsymbol{\theta}, w'), w') - \nabla_{\perp} \Phi^{(0)}(w') \right]. \quad (86)$$

A source at comoving distance w with comoving separation \mathbf{x} from the fiducial light ray would be seen, in the absence of lensing, at the angular separation $\boldsymbol{\beta} = \mathbf{x}/f_K(w)$ from the fiducial ray (this statement is nothing but the definition of the comoving angular diameter distance). Hence, $\boldsymbol{\beta}$ is the unlensed angular position in the ‘comoving source plane’ at distance w , where the origin of this source plane is given by the intersection point with the fiducial ray. Therefore, in analogy with standard lens theory, we define the Jacobian matrix

$$\mathcal{A}(\boldsymbol{\theta}, w) = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \frac{1}{f_K(w)} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}}, \quad (87)$$

and obtain from (86)

$$\mathcal{A}_{ij}(\boldsymbol{\theta}, w) = \delta_{ij} - \frac{2}{c^2} \int_0^w dw' \frac{f_K(w-w')f_K(w')}{f_K(w)} \Phi_{,ik}(\mathbf{x}(\boldsymbol{\theta}, w'), w') \mathcal{A}_{kj}(\boldsymbol{\theta}, w'), \quad (88)$$

which describes the locally linearized mapping introduced by LSS lensing. To derive (88), we noted that $\nabla_{\perp} \Phi^{(0)}$ does not depend on $\boldsymbol{\theta}$, and used the chain rule in the derivative of Φ . This equation still is exact in the limit of validity of the weak-field metric. Next, we expand \mathcal{A} in powers of Φ , and truncate the series after the linear term:

$$\mathcal{A}_{ij}(\boldsymbol{\theta}, w) = \delta_{ij} - \frac{2}{c^2} \int_0^w dw' \frac{f_K(w-w')f_K(w')}{f_K(w)} \Phi_{,ij}(f_K(w')\boldsymbol{\theta}, w'). \quad (89)$$

Hence, to linear order, the distortion can be obtained by integrating along the unperturbed ray $\mathbf{x} = f_K(w)\boldsymbol{\theta}$; this is also called the Born approximation. Corrections to the Born approximation are necessarily of order Φ^2 . Throughout this article, we will employ the Born approximation; later, we will comment on its accuracy. If we now define the deflection potential

$$\psi(\boldsymbol{\theta}, w) := \frac{2}{c^2} \int_0^w dw' \frac{f_K(w-w')}{f_K(w) f_K(w')} \Phi(f_K(w')\boldsymbol{\theta}, w'), \quad (90)$$

compare a light ray in the inhomogeneous universe with one in the homogeneous, unperturbed universe. Apart from the conceptual difficulty, this ‘first-order expansion’ is not justified, as the light rays in an inhomogeneous universe can deviate quite significantly from straight rays in the homogeneous reference universe – much more than the length scale of typical density fluctuations. These difficulties are all avoided if one starts from the exact equation of geodesic deviation, as done here.

then $\mathcal{A}_{ij} = \delta_{ij} - \psi_{,ij}$, just as in ordinary lens theory. *In this approximation, lensing by the 3-D matter distribution can be treated as an equivalent lens plane with deflection potential ψ , mass density $\kappa = \nabla^2\psi/2$, and shear $\gamma = (\psi_{,11} - \psi_{,22})/2 + i\psi_{,12}$.*

6.2 Cosmic shear: the principle

The effective surface mass density. Next, we relate κ to fractional density contrast δ of matter fluctuations in the Universe; this is done in a number of steps:

1. To obtain $\kappa = \nabla^2\psi/2$, take the 2-D Laplacian of ψ , and add the term $\Phi_{,33}$ in the resulting integrand; this latter term vanishes in the line-of-sight integration, as can be seen by integration by parts.
2. We make use of the 3-D Poisson equation in comoving coordinates

$$\nabla^2\Phi = \frac{3H_0^2\Omega_m}{2a}\delta \quad (91)$$

to obtain

$$\kappa(\boldsymbol{\theta}, w) = \frac{3H_0^2\Omega_m}{2c^2} \int_0^w dw' \frac{f_K(w')f_K(w-w')}{f_K(w)} \frac{\delta(f_K(w')\boldsymbol{\theta}, w')}{a(w')} . \quad (92)$$

Note that κ is proportional to Ω_m , since lensing is sensitive to $\Delta\rho \propto \Omega_m \delta$, not just to the density contrast $\delta = \Delta\rho/\bar{\rho}$ itself.

3. For a redshift distribution of sources with $p_z(z) dz = p_w(w) dw$, the effective surface mass density becomes

$$\begin{aligned} \kappa(\boldsymbol{\theta}) &= \int dw p_w(w) \kappa(\boldsymbol{\theta}, w) \\ &= \frac{3H_0^2\Omega_m}{2c^2} \int_0^{w_h} dw g(w) f_K(w) \frac{\delta(f_K(w)\boldsymbol{\theta}, w)}{a(w)} , \end{aligned} \quad (93)$$

with

$$g(w) = \int_w^{w_h} dw' p_w(w') \frac{f_K(w' - w)}{f_K(w')} , \quad (94)$$

which is the source-redshift weighted lens efficiency factor D_{ds}/D_s for a density fluctuation at distance w , and w_h is the comoving horizon distance, obtained from $w(a)$ by letting $a \rightarrow 0$.

The expression (92) for the effective surface mass density can be interpreted in a very simple way. Consider a redshift interval of width dz around z , corresponding to the proper radial distance interval $dD_{prop} = |c dt| = H^{-1}(z)(1+z)^{-1} c dz$. The surface mass density in this interval is $\Delta\rho dD_{prop}$, where only the density contrast $\Delta\rho = \rho - \bar{\rho}$ acts as a lens (the ‘lensing effect’ of the mean

matter density of the Universe is accounted for by the relations between angular diameter distance and redshift; see Schneider & Weiss 1988a). Dividing this surface mass density by the corresponding critical surface mass density, and integrating along the line-of-sight to the sources, one finds

$$\kappa = \int_0^{z_s} dz \frac{4\pi G}{c^2} \frac{D_d^{\text{ang}} D_{ds}^{\text{ang}}}{D_s^{\text{ang}}} \frac{dD_{\text{prop}}}{dz} \Delta\rho. \quad (95)$$

This expression is equivalent to (92), as can be easily shown (by the way, this is a good excersize for practicing the use of cosmological quantities like redshift, distances etc.).

Limber's equation. The density field δ is assumed to be a realization of a random field. It is the properties of the random field that cosmologists can hope to predict, and not a specific realization of it. In particular, the second-order statistical properties of the density field are described in terms of the power spectrum (see IN, Sect. 6.1). We shall therefore look at the relation between the quantities relevant for lensing and the power spectrum $P_\delta(k)$ of the matter distribution in the Universe. The basis of this relation is formed by Limber's equation. If δ is a homogeneous and isotropic 3-D random field, then the projections

$$g_i(\boldsymbol{\theta}) = \int dw q_i(w) \delta(f_K(w)\boldsymbol{\theta}, w) \quad (96)$$

also are (2-D) homogeneous and isotropic random fields, where the q_i are weight functions. In particular, the correlation function

$$C_{12} = \langle g_1(\boldsymbol{\varphi}_1) g_2(\boldsymbol{\varphi}_2) \rangle \equiv C_{12}(|\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2|) \quad (97)$$

depends only on the modulus of the separation vector. The original form of the Limber (1953) equation relates C_{12} to the correlation function of δ which is a line-of-sight projection. Alternatively, one can consider the Fourier-space analogy of this relation: The power spectrum $P_{12}(\ell)$ – the Fourier transform of $C_{12}(\theta)$ – depends linearly on $P_\delta(k)$ (Kaiser 1992, 1998),

$$P_{12}(\ell) = \int dw \frac{q_1(w) q_2(w)}{f_K^2(w)} P_\delta \left(\frac{\ell}{f_K(w)}, w \right), \quad (98)$$

if the largest-scale structures in δ are much smaller than the effective range Δw of the projection. Hence, we obtain the (very reasonable) result that the 2-D power at angular scale $1/\ell$ is obtained from the 3-D power at length scale $f_K(w) (1/\ell)$, integrated over w .

Comparing (93) with (98), one sees that $\kappa(\boldsymbol{\theta})$ is such a projection of δ with the weights $q_1(w) = q_2(w) = (3/2)(H_0/c)^2 \Omega_m g(w) f_K(w)/a(w)$, so that

$$P_\kappa(\ell) = \frac{9H_0^4 \Omega_m^2}{4c^4} \int_0^{w_h} dw \frac{g^2(w)}{a^2(w)} P_\delta \left(\frac{\ell}{f_K(w)}, w \right). \quad (99)$$

The power spectrum P_κ , if observable, can therefore be used to constrain the 3-D power spectrum P_δ . For a number of cosmological models, the power spectrum $P_\kappa(\ell)$ is plotted in Fig. 32. Predictions of P_κ are plotted both for assuming linear growth of the density structure (see Sect. 6.1 of IN), as well as the prescription of the fully nonlinear power spectrum as given by the fitting formulae of Peacock & Dodds (1996). From this figure one infers that the nonlinear evolution of the density fluctuations becomes dominant for values of $\ell \gtrsim 200$, corresponding to an angular scale of about $30'$; the precise values depend on the cosmological model and the redshift distribution of the sources. Furthermore, the dimensionless power spectrum $\ell^2 P_\kappa(\ell)$, that is, the power per logarithmic bin, peaks at around $\ell \sim 10^4$, corresponding to an angular scale of $\sim 1'$, again somewhat depending on the source redshift distribution. Third, one notices that the shape and amplitude of P_κ depends on the values of the cosmological parameters; therefore, by measuring the power spectrum, or quantities directly related to it, one can constrain the values of the cosmological parameters. We consider next appropriate statistical measures of the cosmic shear which are directly and simply related to the power spectrum P_κ .

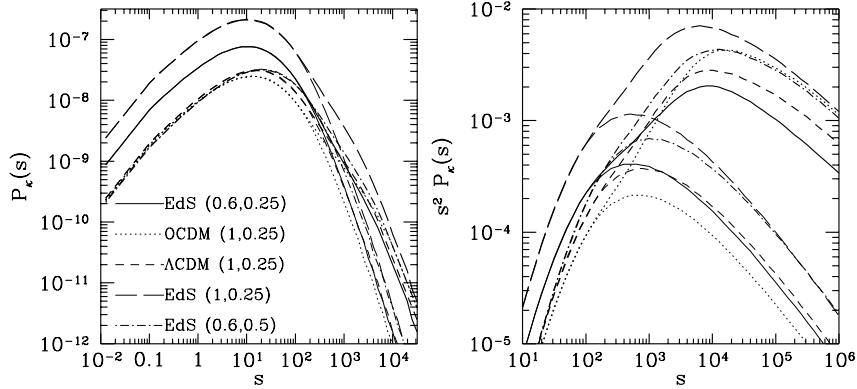


Fig. 32. The power spectrum $P_\kappa(\ell)$ (left panel) and its dimensionless form $\ell^2 P_\kappa(\ell)$ (right panel) for several cosmological models (where here, ℓ is denoted by s). Specifically, EdS denotes an $\Omega_m = 1$, $\Omega_\Lambda = 0$ Einstein-de Sitter model, OCDM an open $\Omega_m = 0.3$, $\Omega_\Lambda = 0$ Universe, and Λ CDM a flat, low-density $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ model. Numbers in parenthesis indicate $(\Gamma_{\text{spect}}, \sigma_8)$, where Γ_{spect} is the shape parameter of the power spectrum (see IN, Sect. 6.1) and σ_8 is the power-spectrum normalization. For these power spectra, the mean redshift of the galaxy distribution was assumed to be $\langle z_s \rangle = 1.5$. Thin curves show the power spectra assuming linear evolution of the density fluctuations in the Universe, and thick curves use the fully non-linear evolution, according to the prescription of Peacock & Dodds (1996). For angular scales below $\sim 30'$, corresponding to $\ell \geq 200$, the non-linear evolution of the power spectrum becomes very important (from Schneider et al. 1998a)

6.3 Second-order cosmic shear measures

We will now turn to statistical quantities of the cosmic shear field which are quadratic in the shear, i.e., to second-order shear statistics. Higher-order statistical properties, which already have been detected in cosmic shear surveys, will be considered in Sect. 9. As we shall see, all second-order statistics of the cosmic shear yield (filtered) information about, and are fully described in terms of P_κ . The most-often used second-order statistics are:

- The two-point correlation function(s) of the shear, $\xi_{\pm}(\theta)$,
- the shear dispersion in a (circular) aperture, $\langle |\hat{\gamma}|^2 \rangle(\theta)$, and
- the aperture mass dispersion, $\langle M_{\text{ap}}^2 \rangle(\theta)$.

Those will be discussed next, and their relation to $P_\kappa(\ell)$ shown. As a preparation, consider the Fourier transform of κ ,

$$\hat{\kappa}(\boldsymbol{\ell}) = \int d^2\theta e^{i\boldsymbol{\ell}\cdot\boldsymbol{\theta}} \kappa(\boldsymbol{\theta}); \quad (100)$$

then,

$$\langle \hat{\kappa}(\boldsymbol{\ell}) \hat{\kappa}^*(\boldsymbol{\ell}') \rangle = (2\pi)^2 \delta_D(\boldsymbol{\ell} - \boldsymbol{\ell}') P_\kappa(\ell), \quad (101)$$

which provides another definition of the power spectrum P_κ [compare with eq. (123) of IN]. The Fourier transform of the shear is

$$\hat{\gamma}(\boldsymbol{\ell}) = \left(\frac{\ell_1^2 - \ell_2^2 + 2i\ell_1\ell_2}{|\boldsymbol{\ell}|^2} \right) \hat{\kappa}(\boldsymbol{\ell}) = e^{2i\beta} \hat{\kappa}(\boldsymbol{\ell}), \quad (102)$$

where β is the polar angle of the vector $\boldsymbol{\ell}$; this follows directly from (42) and (43). Eq. (102) implies that

$$\langle \hat{\gamma}(\boldsymbol{\ell}) \hat{\gamma}^*(\boldsymbol{\ell}') \rangle = (2\pi)^2 \delta_D(\boldsymbol{\ell} - \boldsymbol{\ell}') P_\kappa(\ell). \quad (103)$$

Hence, the power spectrum of the shear is the same as that of the surface mass density.

Shear correlation functions. Consider a pair of points (i.e., galaxy images); their separation direction φ (i.e. the polar angle of the separation vector $\boldsymbol{\theta}$) is used to define the tangential and cross-component of the shear at these positions *for this pair*, $\gamma_t = -\mathcal{R}\text{e}(\gamma e^{-2i\varphi})$, $\gamma_x = -\mathcal{I}\text{m}(\gamma e^{-2i\varphi})$, as in (17). Using these two shear components, one can then define the correlation functions $\langle \gamma_t \gamma_t \rangle(\theta)$ and $\langle \gamma_x \gamma_x \rangle(\theta)$, as well as the mixed correlator. However, it turns out to be more convenient to define the following combinations,

$$\xi_{\pm}(\theta) = \langle \gamma_t \gamma_t \rangle(\theta) \pm \langle \gamma_x \gamma_x \rangle(\theta), \quad \xi_x(\theta) = \langle \gamma_t \gamma_x \rangle(\theta). \quad (104)$$

Due to parity symmetry, $\xi_x(\theta)$ is expected to vanish, since under such a transformation, $\gamma_t \rightarrow \gamma_t$, but $\gamma_x \rightarrow -\gamma_x$. Next we relate the shear correlation

functions to the power spectrum P_κ : Using the definition of ξ_\pm , replacing γ in terms of $\hat{\gamma}$, and making use of relation between $\hat{\gamma}$ and $\hat{\kappa}$, one finds (e.g., Kaiser 1992)

$$\xi_+(\theta) = \int_0^\infty \frac{d\ell \ell}{2\pi} J_0(\ell\theta) P_\kappa(\ell) ; \quad \xi_-(\theta) = \int_0^\infty \frac{d\ell \ell}{2\pi} J_4(\ell\theta) P_\kappa(\ell) , \quad (105)$$

where $J_n(x)$ is the n-th order Bessel function of first kind. ξ_\pm can be measured as follows: on a data field, select all pairs of faint galaxies with separation within $\Delta\theta$ of θ and then take the average $\langle \epsilon_{ti} \epsilon_{tj} \rangle$ over all these pairs; since $\epsilon_i = \epsilon_i^{(s)} + \gamma(\boldsymbol{\theta}_i)$, the expectation value of $\langle \epsilon_{ti} \epsilon_{tj} \rangle$ is $\langle \gamma_t \gamma_t \rangle(\theta)$, provided source ellipticities are uncorrelated. Similarly, the correlation for the cross-components is obtained. It is obvious how to generalize this estimator in the presence of a weight factor for the individual galaxies, as it results from the image analysis described in Sect. 3.5.

The shear dispersion. Consider a circular aperture of radius θ ; the mean shear in this aperture is $\bar{\gamma}$. Averaging over many such apertures, one defines the shear dispersion $\langle |\bar{\gamma}|^2 \rangle(\theta)$. It is related to the power spectrum through

$$\langle |\bar{\gamma}|^2 \rangle(\theta) = \frac{1}{2\pi} \int d\ell \ell P_\kappa(\ell) W_{\text{TH}}(\ell\theta) , \quad \text{where } W_{\text{TH}}(\eta) = \frac{4J_1^2(\eta)}{\eta^2} \quad (106)$$

is the top-hat filter function (see, e.g., Kaiser 1992). A practical unbiased estimator of the mean shear in the aperture is $\hat{\gamma} = N^{-1} \sum_{i=1}^N \epsilon_i$, where N is the number of galaxies in the aperture. However, the square of this expression is *not* an unbiased estimator of $\langle |\bar{\gamma}|^2 \rangle$, since the diagonal terms of the resulting double sum yield additional terms, since $E(\epsilon_i \epsilon_i^*) = |\gamma(\boldsymbol{\theta}_i)|^2 + \sigma_\epsilon^2$. An unbiased estimate for the shear dispersion is obtained by omitting the diagonal terms,

$$\widehat{\langle |\bar{\gamma}|^2 \rangle} = \frac{1}{N(N-1)} \sum_{i \neq j}^N \epsilon_i \epsilon_j^* . \quad (107)$$

This expression is then averaged over many aperture placed on the data field. Again, the generalization to allow for weighting of galaxy images is obvious. Note in particular that this estimator is not positive semi-definite.

The aperture mass. Consider a circular aperture of radius θ ; for a point inside the aperture, define the tangential and cross-components of the shear relative to the center of the aperture (as before); then define

$$M_{\text{ap}}(\theta) = \int d^2\vartheta Q(|\boldsymbol{\vartheta}|) \gamma_t(\boldsymbol{\vartheta}) , \quad (108)$$

where Q is a weight function with support $\vartheta \in [0, \theta]$. If we use the function Q given in (81), the dispersion of $M_{\text{ap}}(\theta)$ is related to power spectrum by (Schneider et al. 1998a)

$$\langle M_{\text{ap}}^2 \rangle(\theta) = \frac{1}{2\pi} \int_0^\infty d\ell \ell P_\kappa(\ell) W_{\text{ap}}(\theta\ell) , \quad \text{with } W_{\text{ap},1}(\eta) := \frac{576J_4^2(\eta)}{\eta^4} . \quad (109)$$

Crittenden et al. (2002) suggested a different pair U and Q of filter functions,

$$U(\vartheta) = \frac{1}{2\pi\theta^2} \left[1 - \left(\frac{\vartheta^2}{2\theta^2} \right) \right] \exp \left(-\frac{\vartheta^2}{2\theta^2} \right) ; \quad Q(\vartheta) = \frac{\vartheta^2}{4\pi\theta^4} \exp \left(-\frac{\vartheta^2}{2\theta^2} \right) . \quad (110)$$

These function have the disadvantage of not having finite support; however, due to the very strong fall-off for $\vartheta \gg \theta$, for many practical purposes the support can be considered effectively as finite. This little drawback is compensated by the convenient analytic properties of these filter functions, as we shall see later. For example, the relation of the corresponding aperture mass dispersion is again given by the first of eqs. (109), but the filter function simplifies to

$$W_{\text{ap},2}(\eta) = \frac{\eta^4}{4} e^{-\eta^2} . \quad (111)$$

Whereas the filter functions which relate the power spectrum to the shear correlation functions, i.e., the Bessel function appearing in (105), and to the shear dispersion, given by W_{TH} , are quite broad filters, implying that these statistics at a given angular scale depend on the power spectrum over a wide range of ℓ , the two filter function $W_{\text{ap},1,2}$ are very localized and thus the aperture mass dispersion yields highly localized information about the power spectrum (see Bartelmann & Schneider 1999, who showed that replacing the filter function W by a delta-‘function’ causes an error of only $\sim 10\%$). Hence, the shape of $\langle M_{\text{ap}}^2 \rangle(\theta)$ directly reflects the shape of the power spectrum as can also be seen in Fig. 35 below.

Interrelations. These various 2-point statistics all depend linearly on the power spectrum P_κ ; therefore, one should not be too surprised that they are all related to each other (Crittenden et al. 2002). The surprise perhaps is that these interrelations are quite simple. First, the relations between ξ_\pm and P_κ can be inverted, making use of the orthonormality relation of Bessel functions:

$$P_\kappa(\ell) = 2\pi \int_0^\infty d\theta \theta \xi_+(\theta) J_0(\ell\theta) = 2\pi \int_0^\infty d\theta \theta \xi_-(\theta) J_4(\ell\theta) . \quad (112)$$

Next, we take one of these and plug them into the relation (105) between the other correlation function and P_κ , to find:

$$\xi_+(\theta) = \xi_-(\theta) + \int_\theta^\infty \frac{d\vartheta}{\vartheta} \xi_-(\vartheta) \left(4 - 12 \frac{\theta^2}{\vartheta^2} \right) ; \quad (113)$$

$$\xi_-(\theta) = \xi_+(\theta) + \int_0^\theta \frac{d\vartheta}{\theta^2} \xi_+(\vartheta) \left(4 - 12 \frac{\vartheta^2}{\theta^2} \right). \quad (114)$$

These equations show that the two shear correlation functions are not independent of each other, the reason for that being that the shear (which itself is a two-component quantity) is derived from a single scalar field, namely the deflection potential ψ . We shall return to this issue further below. Using (112) in the equation for the shear dispersion, one finds

$$\langle |\bar{\gamma}|^2 \rangle(\theta) = \int_0^{2\theta} \frac{d\vartheta}{\theta^2} \xi_+(\vartheta) S_+ \left(\frac{\vartheta}{\theta} \right) = \int_0^\infty \frac{d\vartheta}{\theta^2} \xi_-(\vartheta) S_- \left(\frac{\vartheta}{\theta} \right),$$

where the S_\pm are simple functions, given explicitly in Schneider et al. (2002a) and plotted in Fig. 33. Finally, the same procedure for the aperture mass

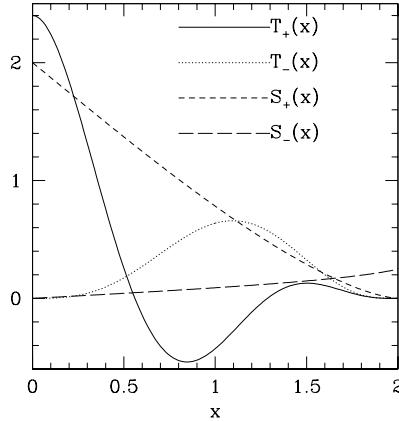


Fig. 33. The function $S_\pm(x)$ and $T_\pm(x)$ which relate the shear and aperture mass dispersion to the correlation functions. Note that S_- does not vanish for $x > 2$, as is the case for the other three functions (from Schneider et al. 2002a)

dispersion lets us write

$$\langle M_{ap}^2 \rangle(\theta) = \int_0^{2\theta} \frac{d\vartheta}{\theta^2} \xi_+(\vartheta) T_+ \left(\frac{\vartheta}{\theta} \right) = \int_0^{2\theta} \frac{d\vartheta}{\theta^2} \xi_-(\vartheta) T_- \left(\frac{\vartheta}{\theta} \right), \quad (115)$$

again with analytically known functions T_\pm , given for the filter function (81) in Schneider et al. (2002a), and for the filter function (110) in Jarvis et al. (2003b). Hence, all these 2-point statistics can be evaluated from the correlation functions $\xi_\pm(\theta)$, which is of particular interest, since they can be measured best: Real data fields contain holes and gaps (like CCD defects; bright stars; nearby galaxies, etc.) which makes the placing of apertures difficult; however, the evaluation of the correlation functions is not affected by gaps, as one uses all pairs of galaxy images with a given angular separation. Furthermore, it should be noted that the aperture mass dispersion at angular scale θ can be calculated from $\xi_\pm(\vartheta)$ in the finite range $\vartheta \in [0, 2\theta]$, and the

shear dispersion can be calculated from ξ_+ on $\vartheta \in [0, 2\theta]$, but not from ξ_- on a finite interval; this is due to the fact that ξ_- on small scales does not contain the information of the power spectrum on large scales, because of the filter function J_4 in (105).

We also note that from a cosmic shear survey, the power spectrum P_κ can be determined directly, as has been investigated by Kaiser (1998), Seljak (1998) and Hu & White (2001). This is *not* done by applying (112), as these relations would require the determination of the correlation function for all separation, but by more sophisticated methods. A simple example (though not optimal) is to consider the measured shear field on the square; Fourier transforming it and binning modes in $|\ell|$ then yields an estimate of the power spectrum, once the power from the intrinsic ellipticity dispersion is subtracted. Better methods aim at minimizing the variance of the reconstructed power spectrum (Seljak 1998; Hu & White 2001). As mentioned before, the aperture mass dispersion is a filtered version of the power spectrum with such a narrow filter, that it contains essentially the same information as P_κ over the corresponding angular scale and at $\ell \sim 5/\theta$, provided P_κ has no sharp features.

6.4 Cosmic shear and cosmology

Why cosmology from cosmic shear? Before continuing, it is worth to pause for a second and ask the question why one tries to investigate cosmological questions by using cosmic shear – since it is widely assumed that the CMB will measure ‘all’ cosmological quantities with high accuracy. Partial answers to this question are:

- Cosmic shear measures the mass distribution at much lower redshifts ($z \lesssim 1$) and at smaller physical scales [$R \sim 0.3 h^{-1} (\theta/1') \text{ Mpc}$] than the CMB; indeed, it is the only way to map out the dark matter distribution directly without any assumptions about the relation between dark and baryonic matter.
- Cosmic shear measures the non-linearly evolved mass distribution and its associated power spectrum $P_\delta(k)$; hence, in combination with the CMB it allows us to study the evolution of the power spectrum and in particular, provide a very powerful test of the gravitational instability paradigm for structure growth.
- As was demonstrated by the recent results from the WMAP satellite (Bennett et al. 2003), the strongest constraints are derived when combining CMB measurements (constraining the power spectrum on large spatial scales) with measurements on substantially smaller scales, to break parameter degeneracies remaining from the CMB results alone (see Spergel et al. 2003). Hu & Tegmark (1999) have explicitly demonstrated how much the accuracy of estimates of cosmological parameters is improved when the CMB results from missions like WMAP and later Planck is

complemented by cosmic shear measurements (see Fig. 34). In fact, as we shall see later, combinations of CMB anisotropy measurements have already been combined with cosmic shear measurements (see Fig. 47) and lead to substantially improved constraints on the cosmological parameters.

- It provides a fully independent way to probe the cosmological model; given the revolutionary claims coming from the CMB, SN Ia, and the LSS of the galaxy distribution, namely that more than 95% of the contents in the Universe is in a form that we have not the slightest idea about what it is (the names ‘dark matter’ and ‘dark energy’ reflect our ignorance about their physical nature), an additional independent verification of these claims is certainly welcome.
- For a foreseeable future, astronomical observations will provide the only possibility to probe the dark energy empirically. The equation of state of the dark energy can be probed best at relatively low redshifts, that is with SN Ia and cosmic shear observations, whereas CMB anisotropy measurements are relatively insensitive to the properties of the dark energy, as the latter was subdominant at the epoch of recombination.
- As we have seen in Sect. 5.8, cosmic shear studies provide a new and highly valuable search method for cluster-scale matter concentrations.

Expectations. The cosmic shear signal depends on the cosmological model, parameterized by Ω_m , Ω_Λ , and the shape parameter Γ_{spect} of the power spectrum, the normalization of the power spectrum, usually expressed in terms of σ_8 , and the redshift distribution of the sources. By measuring ξ_\pm over a significant range of angular scales one can derive constraints on these parameters. To first order, the amplitude of the cosmic shear signal depends on the combination $\sim \sigma_8 \Omega_m^{0.5}$, very similar to the cluster abundance. Furthermore, the cosmic shear signal shows a strong dependence on the source redshift distribution. These dependencies are easily understood qualitatively: A higher normalization σ_8 increases P_δ on all scales, thus increasing P_κ . The increase with Ω_m is mainly due to the prefactor in (99), i.e. due to the fact that the light deflection depends on $\Delta\rho$, not just merely on $\delta = \Delta\rho/\bar{\rho}$, as most other cosmological probes. Finally, increasing the redshift of sources has two effects: first, the lens efficiency $D_{\text{ds}}/D_s = f_K(w_s - w)/f_K(w_s)$ at given distance w increases as the sources are moved further away, and second, a larger source redshift implies a longer ray path through the inhomogeneous matter distribution.

In Fig. 35 the predictions of the shear dispersion and the aperture mass dispersion are shown as a function of angular scale, for several cosmological models. The dependencies of the power spectrum P_κ on cosmological parameters and ℓ is reflected in these cosmic shear measures. In particular, the narrow filter function which relates the aperture mass dispersion to the power spectrum implies that the peak in $\ell^2 P_\kappa(\ell)$ at around $\ell \sim 10^4$ (see Fig. 32)

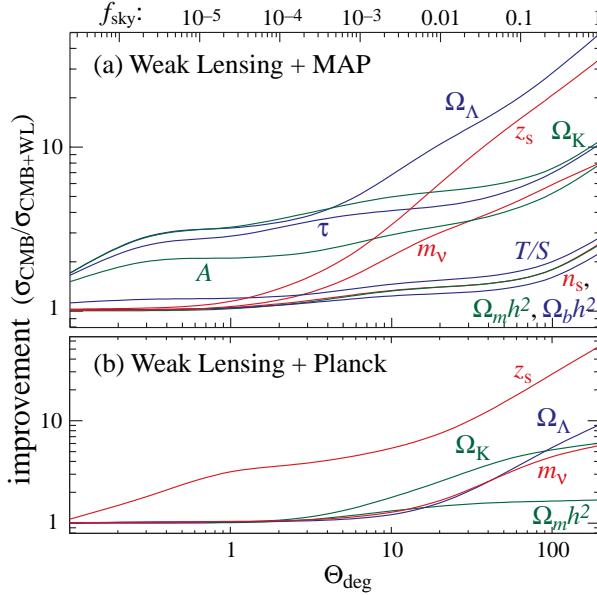


Fig. 34. The improvement of the accuracy of cosmological parameters when supplementing CMB data from WMAP (upper panel) and the Planck satellite (lower panel) by a cosmic shear survey of solid angle $\theta^2\pi$. The accuracies are significantly improved, certainly when combined with WMAP, but even in combination with Planck, the accuracies of the density parameters can be increased, when using next-generation cosmic shear surveys with hundreds of square degrees (from Hu & Tegmark 1999)

translates into a peak of $\langle M_{ap}^2 \rangle$ at around $\theta \sim 1'$. The non-linear evolution of the power spectrum is dominating the cosmic shear result for scales below $\sim 30'$; the fact that the non-linear prediction approach the linear ones at somewhat smaller scales for the shear dispersion $\langle |\bar{\gamma}|^2 \rangle$ is due to the fact that this statistics corresponds to a broad-band filter W_{TH} (106) of P_κ which includes the whole range of small ℓ values, which are less affected by non-linear evolution.

Deriving constraints. From the measured correlation functions $\xi_{\pm}(\theta)$ (or any other measure of the cosmic shear, but we will concentrate on the statistics which is most easily obtained from real data), obtaining constraints on cosmological parameters can proceed through maximizing the likelihood $\mathcal{L}(p|\xi^{\text{obs}})$, which yields the probability for the set of cosmological parameters being p , given the observed correlation function ξ^{obs} . This likelihood is given by the probability $P(\xi^{\text{obs}}|p)$ that the observed correlation function is ξ^{obs} , given the parameters p . For a given set of parameters p , the correlation func-

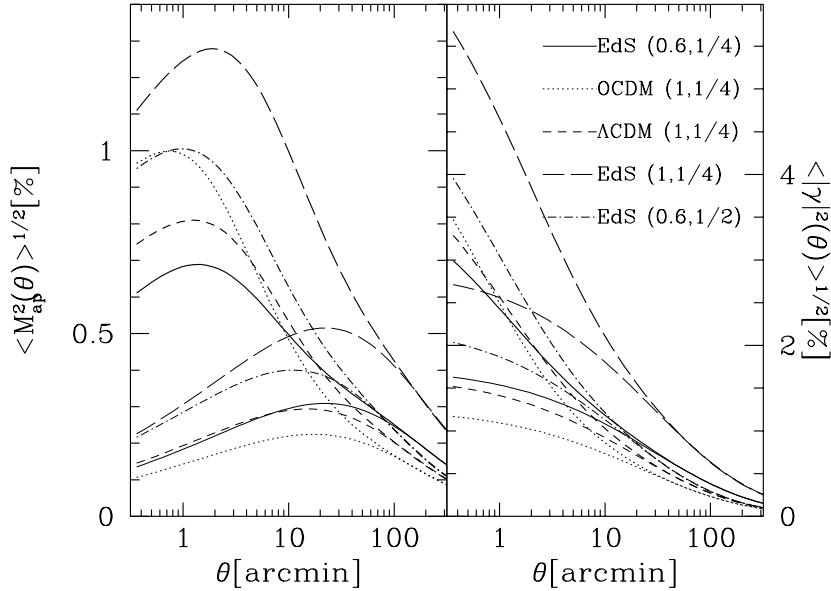


Fig. 35. The square root of the aperture mass dispersion (left) and of the shear dispersion (right), for the same cosmological models as were used for Fig. 32, again with results from assuming linear growth of structure in the Universe shown as thin curves, whereas the fully non-linear evolution was taken into account for the thick curves. One sees that the aperture mass signal is considerably smaller than that of the shear dispersion; this is due to the fact that the filter function W_{ap} is much narrower than W_{TH} ; hence, at a given angular scale, $\langle M_{\text{ap}}^2 \rangle$ samples less power than $\langle |\bar{\gamma}|^2 \rangle$. However, this also implies that the aperture mass dispersion provides much more localized information about the power spectrum than the shear dispersion and is therefore a more useful statistics to consider. Other advantages of $\langle M_{\text{ap}}^2 \rangle$ will be described further below. For scales below $\sim 30'$, the non-linear evolution of the power spectrum becomes very important (from Schneider et al. 1998a)

tion $\xi(p)$ is predicted. If one assumes that the observed correlations ξ^{obs} are drawn from a (multi-variate) Gaussian probability distribution, then

$$P(\xi^{\text{obs}}|p) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \text{Cov}}} \exp\left(\frac{-\chi^2(p, \xi^{\text{obs}})}{2}\right),$$

with

$$\chi^2(p, \xi^{\text{obs}}) = \sum_{ij} (\xi_i(p) - \xi_i^{\text{obs}}) \text{Cov}_{ij}^{-1} (\xi_j(p) - \xi_j^{\text{obs}}). \quad (116)$$

Here, the $\xi_i = \xi(\theta_i)$ are the values of the correlation function(s) (i.e., either ξ_{\pm} , or using both) in angular bins, n is the number of angular bins in case either one of the ξ_{\pm} is used, or if both are combined, twice the number of

angular bins, and Cov_{ij}^{-1} is the inverse of the covariance matrix, which is defined as

$$\text{Cov}_{ij} = \langle [\xi_i(p) - \xi_i^{\text{obs}}] [\xi_j(p) - \xi_j^{\text{obs}}] \rangle , \quad (117)$$

where the average is over multiple realizations of the cosmic shear survey under consideration. Cov_{ij} can be determined either from the ξ_{\pm} itself, from simulations, or estimated from the data in terms of the ξ_{\pm}^{obs} (see Schneider et al. 2002b; Kilbinger & Schneider 2004, Simon et al. 2004). Nevertheless, the calculation of the covariance is fairly cumbersome, and most authors have used approximate methods to derive it, such as the field-to-field variations of the measured correlation. In fact, this latter approach may be more accurate than using the analytic expressions of the covariance in terms of the correlation function, which are obtained by assuming that the shear field is Gaussian, so that the four-point correlation function can be factorized as products of two-point correlators. As it turns out, $\xi_+(\theta)$ is strongly correlated across angular bins, much less so for $\xi_-(\theta)$; this is due to the fact that the filter function that describes ξ in terms of the power spectrum P_κ is much broader for ξ_+ (namely J_0) than J_4 which applies for ξ_- .

The accuracy with which ξ_{\pm} can be measured, and thus the covariance matrix, depends on the number density of galaxies (that is, depth and quality of the images), the total solid angle covered by the survey, and its geometrical arrangement (compact survey vs. widely separated pointings). The accuracy is determined by a combination of the intrinsic ellipticity dispersion and the cosmic (or sampling) variance. The likelihood function then becomes

$$\mathcal{L}(p|\xi^{\text{obs}}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \text{Cov}}} \exp\left(\frac{-\chi^2(p, \xi^{\text{obs}})}{2}\right) P_{\text{prior}}(p) , \quad (118)$$

where $P_{\text{prior}}(p)$ contains prior information (or prejudice) about the parameters to be determined. For example, the redshift distribution of the sources (at given apparent magnitude) is fairly well known from spectroscopic redshift surveys, and so the prior probability for z_s would be chosen to be a fairly narrow function which describes this prior knowledge on the redshifts. One often assumes that all but a few parameters are known precisely, and thus considers a restricted space of parameters; this is equivalent to replacing the prior for those parameters which are fixed by a delta-'function'. If m parameters are assumed to be undetermined, but one is mainly interested in the confidence contours of $m' < m$ parameters, then the likelihood function is integrated over the remaining $m - m'$ parameters; this is called marginalization and yields the likelihood function for these m' parameters.

There are two principal contributions to the 'noise' of cosmic shear measurements. One is the contribution coming from the finite intrinsic ellipticity dispersion of the source galaxies, the other due to the finite data fields of any survey. This latter effect implies that only a *typical* part of the sky is mapped, whose properties will in general deviate from the *average* properties of such a region in the sky for a given cosmology. This effect is called cosmic variance,

or sample variance. Whereas the noise from intrinsic ellipticity dispersions dominates at small angular scales, at scales beyond a few arcminutes the cosmic variance is always the dominating effect (e.g., Kaiser 1998; White & Hu 2000).

Of course, all of what was said above can be carried over to the other second-order shear statistics, with their respective covariance matrices. The first cosmic shear measurements were made in terms of the shear dispersion and compared to theoretical prediction from a range of cosmological models. As is true for the correlation functions, the shear dispersion is strongly correlated between different angular scales. This is much less the case for the aperture mass dispersion, where the correlation quickly falls off once the angular scales differ by more than a factor ~ 1.5 (see Schneider et al. 2002b). Even less correlated is the power spectrum itself. These properties are of large interest if the results from a cosmic shear survey are displayed as a curve with error bars; for the aperture mass dispersion and the power spectrum estimates, these errors are largely uncorrelated. However, for deriving cosmological constraints, the correlation function ξ_{\pm} are most useful since they contain all second-order information in the data, in addition of being the primary observable.

6.5 E-modes, B-modes

In the derivation of the lensing properties of the LSS, we ended up with an equivalent surface mass density. In particular, this implied that \mathcal{A} is a symmetric matrix, and that the shear can be obtained in terms of κ or ψ . Now, the shear is a 2-component quantity, whereas both κ and ψ are scalar fields. This then obviously implies that the two shear components are not independent of each other!

Recall that (54) yields a relation between the gradient of κ and the first derivatives of the shear components; in particular, (54) implies that $\nabla \times \mathbf{u}_\gamma \equiv 0$, yielding a local constraint relation between the second derivatives of the shear components. The validity of this constraint equation guarantees that the imaginary part of (44) vanishes. This constraint is also present at the level of 2-point statistics, since one expects from (112) that

$$\int_0^\infty d\theta \theta \xi_+(\theta) J_0(\theta\ell) = \int_0^\infty d\theta \theta \xi_-(\theta) J_4(\theta\ell) .$$

Hence, the two correlation functions ξ_{\pm} are not independent. The observed shear field is not guaranteed to satisfy these relations, due to noise, remaining systematics, or other effects. Therefore, searching for deviations from this relation allows a check for these effects. However, there might also be a ‘shear’ component present that is not due to lensing (by a single equivalent thin matter sheet κ). Shear components which satisfy the foregoing relations are called E-modes; those which don’t are B-modes – these names are exported from

the polarization of the CMB, which has the same mathematical properties as the shear field, namely that of a polar.

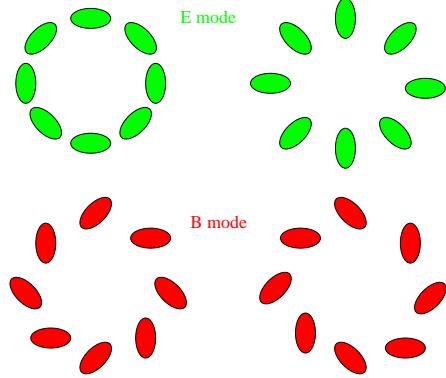


Fig. 36. Sketch of the distinction between E- and B-modes of the shear. The upper row shows a typical E-mode shear pattern coming from a mass overdensity (left) or underdensity (right), yielding tangential and radial alignment of the shear, respectively. The lower row shows a B-mode pattern, which is obtained from the E-mode pattern by rotating all shears by 45° . Those cannot be produced from gravitational lensing (from van Waerbeke & Mellier 2003)

The best way to separate these modes locally is provided by the aperture measures: $\langle M_{\text{ap}}^2(\theta) \rangle$ is sensitive *only* to E-modes. If one defines in analogy – recall (77)

$$M_\perp(\theta) = \int d^2\vartheta Q(|\vartheta|) \gamma_\times(\vartheta), \quad (119)$$

then $\langle M_\perp^2(\theta) \rangle$ is sensitive *only* to B-modes. In fact, one can show that for a pure E-mode shear field, $M_\perp \equiv 0$, and for a pure B-mode field, $M_{\text{ap}} \equiv 0$. Furthermore, in general (that is, even if a B-mode is present), $\langle M_{\text{ap}} \rangle = 0$, since $\langle \kappa \rangle = 0$, and $\langle M_\perp \rangle = 0$, owing to parity invariance: a non-zero mean value of M_\perp would introduce a net orientation into the shear field. Using the same argument, one finds that $\langle M_{\text{ap}}^m M_\perp^n \rangle = 0$ for n odd (Schneider 2003).

E/B-mode decomposition of a shear field. There are a number of (equivalent) ways to decompose a shear field into its two modes. One is provided by the Kaiser & Squires mass reconstruction (44), which yields, for a general shear field, a complex surface mass density $\kappa = \kappa^E + i\kappa^B$. Another separation is obtained by considering the vector field $\mathbf{u}_\gamma(\theta)$ (54) obtained from the first derivatives of the shear components. This vector will in general not be a gradient field; its gradient component corresponds to the E-mode field, the remaining one to the B-mode. Hence one defines

$$\nabla^2 \kappa^E = \nabla \cdot \mathbf{u}_\gamma \quad ; \quad \nabla^2 \kappa^B = \nabla \times \mathbf{u}_\gamma. \quad (120)$$

In full analogy with the ‘lensing-only’ case (i.e., a pure E-mode), one defines the (complex) potential $\psi(\theta) = \psi^E(\theta) + i\psi^B(\theta)$ by the Poisson equation $\nabla^2 \psi = 2\kappa$, and the shear is obtained in terms of the complex ψ in the usual

way,

$$\begin{aligned}\gamma &= \gamma_1 + i\gamma_2 = (\psi_{,11} - \psi_{,22})/2 + i\psi_{,12} \\ &= \left[\frac{1}{2} (\psi_{,11}^E - \psi_{,22}^E) - \psi_{,12}^B \right] + i \left[\psi_{,12}^E + \frac{1}{2} (\psi_{,11}^B - \psi_{,22}^B) \right].\end{aligned}\quad (121)$$

On the level of second-order statistics, one considers the Fourier transforms of the E- and B-mode convergence, and defines the two power spectra P_E , P_B , and the cross-power spectrum P_{EB} by

$$\begin{aligned}\langle \hat{\kappa}^E(\ell) \hat{\kappa}^{E*}(\ell') \rangle &= (2\pi)^2 \delta_D(\ell - \ell') P_E(\ell), \\ \langle \hat{\kappa}^B(\ell) \hat{\kappa}^{B*}(\ell') \rangle &= (2\pi)^2 \delta_D(\ell - \ell') P_B(\ell), \\ \langle \hat{\kappa}^E(\ell) \hat{\kappa}^{B*}(\ell') \rangle &= (2\pi)^2 \delta_D(\ell - \ell') P_{EB}(\ell).\end{aligned}\quad (122)$$

From what was said above, the cross power P_{EB} vanishes for parity-symmetric shear fields, and we shall henceforth ignore it. The shear correlation functions now depend on the power spectra of both modes, and are given as (Crittenden et al. 2002; Schneider et al. 2002a)

$$\begin{aligned}\xi_+(\theta) &= \int_0^\infty \frac{d\ell \ell}{2\pi} J_0(\ell\theta) [P_E(\ell) + P_B(\ell)], \\ \xi_-(\theta) &= \int_0^\infty \frac{d\ell \ell}{2\pi} J_4(\ell\theta) [P_E(\ell) - P_B(\ell)].\end{aligned}$$

Hence, in the presence of B-modes, the ξ_- correlation function cannot be obtained from ξ_+ , as was the case for a pure E-mode shear field. The inverse relation (112) now gets modified to

$$\begin{aligned}P_E(\ell) &= \pi \int_0^\infty d\theta \theta [\xi_+(\theta) J_0(\ell\theta) + \xi_-(\theta) J_4(\ell\theta)], \\ P_B(\ell) &= \pi \int_0^\infty d\theta \theta [\xi_+(\theta) J_0(\ell\theta) - \xi_-(\theta) J_4(\ell\theta)].\end{aligned}\quad (123)$$

Hence, the two power spectra can be obtained from the shear correlation functions. However, due to the infinite range of integration, one would need to measure the correlation functions over all angular scales to apply the previous equations for calculating the power spectra. Much more convenient for the E/B-mode decomposition is the use of the aperture measures, since one can show that

$$\begin{aligned}\langle M_{ap}^2 \rangle(\theta) &= \frac{1}{2\pi} \int_0^\infty d\ell \ell P_E(\ell) W_{ap}(\theta\ell), \\ \langle M_\perp^2 \rangle(\theta) &= \frac{1}{2\pi} \int_0^\infty d\ell \ell P_B(\ell) W_{ap}(\theta\ell),\end{aligned}\quad (124)$$

so that these two-point statistics clearly separate E- and B-modes. In addition, as mentioned before, they provide a highly localized measure of the

corresponding power spectra, since the filter function $W_{\text{ap}}(\eta)$ involved is very narrow. As was true for the E-mode only case, the aperture measures can be expressed as finite integrals over the correlation functions,

$$\begin{aligned}\langle M_{\text{ap}}^2 \rangle(\theta) &= \frac{1}{2} \int \frac{d\vartheta}{\theta^2} \left[\xi_+(\vartheta) T_+ \left(\frac{\vartheta}{\theta} \right) + \xi_-(\vartheta) T_- \left(\frac{\vartheta}{\theta} \right) \right], \\ \langle M_\perp^2 \rangle(\theta) &= \frac{1}{2} \int \frac{d\vartheta}{\theta^2} \left[\xi_+(\vartheta) T_+ \left(\frac{\vartheta}{\theta} \right) - \xi_-(\vartheta) T_- \left(\frac{\vartheta}{\theta} \right) \right],\end{aligned}\quad (125)$$

where the two functions T_{\pm} are the same as in (115) and have been given explicitly in Schneider et al. (2002a) for the weight function Q given in (81), and in Jarvis et al. (2003) for the weight function (110). Hence, the relations (125) remove the necessity to calculate the aperture measures by placing apertures on the data field which, owing to gaps and holes, would make this an inaccurate and biased determination. Instead, obtaining the correlation functions from the data is all that is needed.

The relations given above have been applied to recent cosmic shear surveys, and significant B-modes have been discovered (see Sect. 7); the question now is what are they due to? As mentioned before, the noise, which contributes to both E- and B-modes in similar strengths, could be underestimated, the cosmic variance which also determines the error bars on the aperture measures and which depends on fourth-order statistical properties of the shear field could also be underestimated, there could be remaining systematic effects, or B-modes could indeed be present. There are two possibilities known to generate a B-mode through lensing: The first-order in Φ (or ‘Born’) approximation may not be strictly valid, but as shown by ray-tracing simulations through cosmic matter fields (e.g., Jain et al. 2000), the resulting B-modes are expected to be very small. Clustering of sources also yields a finite B-mode (Schneider et al. 2002a), but again, this effect is much smaller than the observed amplitude of the B-modes (see Fig. 37).

Intrinsic alignment of source galaxies. Currently the best guess for the generation of a finite B-mode are intrinsic correlations of galaxy ellipticities. Such intrinsic alignments of galaxy ellipticities can be caused by tidal gravitational fields during galaxy formation, owing to tidal interactions between galaxies, or between galaxies and clusters. Predictions of the alignment of the projected ellipticity of the galaxy mass can be made analytically (e.g. in the frame of tidal torque theory) or from numerical simulations; however, the predictions from various groups differ by large factors (e.g., Croft & Metzler 2000; Crittenden et al. 2001; Heavens et al. 2000; Jing 2002) which means that the process is not well understood at present. For example, the results of these studies depend on whether one assumes that the light of a galaxy is aligned with the dark matter distribution, or aligned with the angular momentum vector of the dark halo. This is related to the question of whether

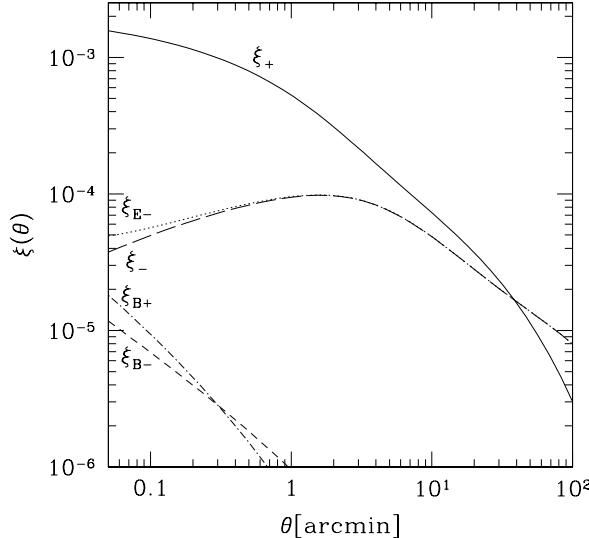


Fig. 37. The correlation functions $\xi_{\pm}(\theta)$ for a Λ CDM model with $\Gamma_{\text{spect}} = 0.21$ and $\sigma_8 = 1$, and a source population with mean redshift of $\langle z_s \rangle = 1.5$. Also plotted are the corresponding correlation functions that arise separately from the E- and B-modes, with the ξ_{E+} mode curve coinciding within the line thickness with ξ_+ . In this calculation, the clustering of the faint galaxy population was taken into account, and they give rise to a very small B-mode contribution, as can be seen from the $\xi_{B\pm}$ curves. The smallness of the B-mode due to intrinsic source clustering renders this effect not viable to explain the B-modes observed in some of the cosmic shear surveys (from Schneider et al. 2002a)

the orientation of the galaxy light (which is the issue of relevance here) is the same as that of the mass.

If intrinsic alignments play a role, then

$$\xi_+ = \langle \epsilon_i \epsilon_j^* \rangle = \left\langle \epsilon_i^{(s)} \epsilon_j^{(s)*} \right\rangle + \xi_+^{\text{lens}}, \quad (126)$$

and measured correlations ξ_{\pm} contain both components, the intrinsic correlation and the shear. Of course, there is no reason why intrinsic correlations should have only a B-mode. If a B-mode contribution is generated through this process, then the measured E-mode is most likely also contaminated by intrinsic alignments. Given that intrinsic alignments yield ellipticity correlations only for spatially close sources (i.e., close in 3-D, not merely in projection), it is clear that the deeper a cosmic shear survey is, and thus the broader the redshift distribution of source galaxies, the smaller is the relative amplitude of an intrinsic signal. Most of the theoretical investigations on the

strength of intrinsic alignments predict that the deep cosmic shear surveys (say, with mean source redshifts of $\langle z_s \rangle \sim 1$) are affected at a $\sim 10\%$ level, but that shallower cosmic shear surveys are more strongly affected; for them, the intrinsic alignment can be of same order or even larger than the lensing signal.

However, the intrinsic signal can be separated from the lensing signal if redshift information of the sources is available, owing to the fact that $\langle \epsilon_i^{(s)} \epsilon_j^{(s)*} \rangle$ will be non-zero only if the two galaxies are at essentially the same redshift. Hence, if z -information is available (e.g., photometric redshifts), then galaxy pairs which are likely to have similar redshifts are to be avoided in estimating the cosmic shear signal (King & Schneider 2002; Heymans & Heavens 2002, Takada & White 2004). This will change the expectation value of the shear correlation function, but in a controllable way, as the redshifts are assumed to be known. Indeed, using (photometric) redshifts, one can simultaneously determine the intrinsic and the lensing signal, essentially providing a cosmic shear tomography (King & Schneider 2003). This again is accomplished by employing the fact that the intrinsic correlation can only come from galaxies very close in redshift. Hence, in the presence of intrinsic alignments, the redshift dependent correlation functions $\xi_{\pm}(z_1, z_2; \theta)$ between galaxies with estimated redshifts z_i are expected to show a strong peak over the range $|z_1 - z_2| \lesssim \Delta z$, where Δz is the typical uncertainty in photometric redshifts. It is this peak that allows one to identify and subtract the intrinsic signal from the correlation functions. An efficient method to calculate the covariance of the redshift-dependent correlation functions has been developed by Simon et al. (2004), where the improvement in the constraints on cosmological parameters from redshift information has been studied, confirming the earlier results by Hu (1999) which were based on considerations of the power spectrum.

Brown et al. (2002) obtained a measurement of the intrinsic ellipticity correlation from the Super-COSMOS photographic plate data, where the galaxies are at too low a redshift for cosmic shear playing any role. Heymans et al. (2003) used the COMBO-17 data set (that will be described in Sect. 7.3 below) for which accurate photometric redshifts are available to measure the intrinsic alignment. The results from both studies is that the models predicting a large intrinsic amplitude can safely be ruled out. Nevertheless, intrinsic alignment affects cosmic shear measurements, at about the 2% level for a survey with the depth of the VIRIMOS-DESCART survey, and somewhat more for the slightly shallower COMBO-17 survey. Hence, to obtain precision measurements of cosmic shear, very important for constraining the equation of state of dark energy, these physically close pairs of galaxies need to be identified in the survey, making accurate photometric redshifts mandatory.

Correlation between intrinsic ellipticity and shear. The relation (126) above implicitly assumes that the shear is uncorrelated with the intrinsic

shape of a neighboring galaxy. However, as pointed out by Hirata & Seljak (2004), this is not necessarily the case. Hence consider galaxies at two significantly different redshifts $z_i < z_j$. For them, the first term in (126) vanishes. However, making use of $\epsilon = \epsilon^{(s)} + \gamma$, one finds

$$\langle \epsilon_i \epsilon_j^* \rangle = \left\langle \epsilon_i^{(s)} \gamma_j^* \right\rangle + \xi_+^{\text{lens}}, \quad (127)$$

where the first term on the right-hand side describes the correlation between the intrinsic ellipticity of the lower-redshift galaxy with the shear along the l.o.s. to the higher-redshift one. The correlation can in principle be non-zero: if the intrinsic alignment of the light of a galaxy is determined by the large-scale tidal gravitational field, then this tidal field at the redshift z_i causes both, an alignment of the nearer galaxy and a contribution to the shear of the more distant one (see Fig. 38). This alignment effect can therefore not be removed by considering only pairs at different redshifts.

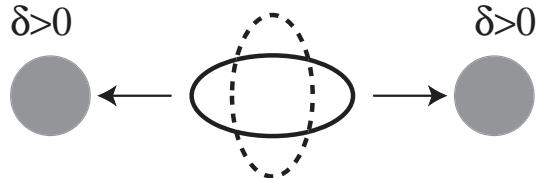


Fig. 38. A tidal gravitational field, for example caused by two matter concentrations, can produce an alignment of a galaxy situated at the same redshift (indicated by the solid ellipse), as well as contributing to the shear towards a more distant galaxy (as indicated by the dashed ellipse) (from Hirata & Seljak 2004)

The importance of this effect depends on the nature of the alignment of galaxies relative to an external tidal field. If the alignment is linear in the tidal field strength, then this effect can be a serious contaminant of the cosmic shear signal, in particular for relatively shallow surveys (where the mean source redshift is small); in particular, this effect can yield much larger contaminations than the intrinsic alignment given by the first term in (126). As can be seen from Fig. 38, the resulting contribution is negative, hence decreases the lensing signal. If, however, the intrinsic alignment depends quadratically on the tidal field, as is suggested by tidal torque theory, than this effect is negligible. Whether or not this effect is relevant needs to be checked from observations. Assuming that the matter density field is represented approximately by the galaxy distribution, the latter can be used to estimate the tidal gravitational field, in particular its direction. Alternatively, since the correlation between the intrinsic alignment and the shear towards more distant galaxies has a different redshift dependence than the lensing shear signal, these two contributions can be disentangled from the z -dependence of the signal.

It should be noted that the use of photometric redshifts also permits to study the cosmic shear measures as a function of source redshift; hence, one can probe various redshift projections $P_\kappa(\ell)$ of the underlying power spectrum $P_\delta(k; z)$ separately. This is due to the fact that the cosmic shear signal from different populations of galaxies (i.e., with different redshift distributions) lead to different weight functions $g(w)$ [see (94)], and thus to different weighting in the projection (99) of the power spectrum. Not surprisingly, uncertainties of cosmological parameters are thereby reduced (Hu 1999; Simon et al. 2004). Also, as shown by Taylor (2001), Hu & Keeton (2002) and Bacon & Taylor (2003), in principle the three-dimensional mass distribution $\delta(\mathbf{x})$ can be reconstructed if the redshifts of the source galaxies are known (see Sect. 7.6).

6.6 Predictions; ray-tracing simulations

The power spectrum of the convergence P_κ can be calculated from the power spectrum of the cosmological matter distribution P_δ , using (99); the latter in turn is determined by the cosmological model. However, since the non-linear evolution of the power spectrum is essential for making accurate quantitative predictions for the shear properties, there is no analytic method known how to calculate the necessary non-linear P_δ . As was mentioned in Sect. 6.1 of IN, fairly accurate fitting formulae exist which yield a closed-form expression for P_δ and which can be used to obtain P_κ (see, e.g., Jain & Seljak 1997). Nevertheless, there are a number of reasons why this purely analytic approach should at least be supplemented by numerical simulations.

- First, the fitting formulae for P_δ (Peacock & Dodds 1996; Smith et al. 2003) have of course only a finite accuracy, and are likely to be insufficient for comparison with results from the ongoing cosmic shear surveys which are expected to yield very accurate measurements, owing to their large solid angle.
- A second reason why simulations are needed is to test whether the various approximations that enter the foregoing analytical treatment are in fact accurate enough. To recall them, we employed the Born approximation, i.e., neglected terms of higher order than linear in the Newtonian potential when deriving the convergence, and we assumed that the shear everywhere is small, so that the difference between shear and reduced shear can be neglected, at least on average. This, however, is not guaranteed: regions in the sky with large shear are most likely also those regions where the convergence is particularly large, and therefore, there one expects a correlation between γ and κ , which can affect the dispersion of $g = \gamma/(1 - \kappa)$.
- Third, whereas fairly accurate fitting formulae exist for the power spectrum, this is not the case for higher-order statistical properties of the

matter distribution; hence, when considering higher-order shear statistics (Sect. 9), numerical simulations will most likely be the only way to obtain accurate predictions.

- The covariance of the shear correlations (and all other second-order shear measures) depends on fourth-order statistics of the shear field, for which hardly any useful analytical approximations are available. The analytical covariance estimates are all based on the Gaussian assumption for the fourth-order correlators. Therefore, simulations are invaluable for the calculation of these covariances, which can be derived for arbitrary survey geometries.

Ray-tracing simulations: The principle. The simulations proceed by following light rays through the inhomogeneous matter distribution in the Universe. The latter is generated by cosmological simulations of structure evolution. Those start at an early epoch by generating a realization of a Gaussian random field with a power spectrum according to the cosmological model considered, and follow the evolution of the density and velocity field of the matter using Newtonian gravity in an expanding Universe. The mass distribution is represented by discrete particles whose evolution in time is followed. A finite volume of the Universe is simulated this way, typically a box of comoving side-length L , for which periodic boundary conditions are applied. This allows one to use Fast Fourier Transforms (FFT) to evaluate the gravitational potential and forces from the density distribution. The box size L should be chosen such that the box contains a representative part of the real Universe, and must therefore be larger than the largest scales on which structure is expected, according to the power spectrum; a reasonable choice is $L \gtrsim 100h^{-1}$ Mpc. The number of grid points and the number of particles that can be distributed in this volume is limited by computer memory; modern simulations work typically with 256^3 points and the same number of particles, though larger simulations have also been carried out; this immediately yields the size of grid cells, of order $0.5h^{-1}$ Mpc. This comoving length, if located at a redshift of $z \sim 0.3$ (which is about the most relevant for cosmic shear), subtends an angle of roughly $2'$ on the sky. The finite number of particles yields the mass resolution of the simulations, which is typically $\sim 10^{10}h^{-1}M_\odot$, depending on cosmological parameters.

In order to obtain higher spatial resolution, force calculations are split up into near-field and far-field forces. The gravitational force due to the distant matter distribution is obtained by grid-based FFT methods, whereas the force from nearby masses is calculated from summing up the forces of individual particles; such simulations yield considerably higher resolution of the resulting mass distribution. Since the matter in these simulations is represented by massive particles, these can undergo strong interactions, leading to (unphysical) large orbital deflections. In order to avoid these unphysical strong collisions, the force between pairs of particles is modified at short dis-

tances, typically comparable to the mean separation of two particles in the simulation. This softening length defines the minimum length scale on which the results from numerical simulations can be considered reliable. Cosmological simulations consider either the dark matter only or, more recently, the hydrodynamics effects of baryons have been incorporated as well.

The outcome of such simulations, as far as they are relevant here, are the 3-dimensional positions of the matter particles at different (output) times or redshifts. In order to study the light propagation through this simulated mass distribution, one employs multiple lens-plane theory. First, the volume between us and sources at some redshift z_s is filled with boxes from the cosmological simulations. That is, the comoving distance $w_s = w(z_s)$ is split up into n intervals of length L , and the mass distribution at an output time close to $t_i = t(w = (i - 1/2)L)$ is considered to be placed at this distance. In this way, one has a light cone covered by cubes containing representative matter distributions. Since the mass distributions at the different times t_i are not independent of each other, but one is an evolved version of the earlier one, the resulting mass distribution is highly correlated over distances much larger than L . This can be avoided by making use of the statistical homogeneity and isotropy of the mass distribution: each box can be translated by an arbitrary two-dimensional vector, employing the periodicity of the mass distribution, and rotated by an arbitrary angle; furthermore, the three different projections of the box can be used for its orientation. In this way – a kind of recycling of numerical results – the worst correlations are removed.

Alternatively, one can combine the outputs from several simulations with different realizations of the initial conditions. In this case, one can use simulation boxes of different spatial extent, to match the comoving size of a big light cone as a function of redshift. That is, for a given light-cone size, only relatively small boxes are needed at low redshifts, and bigger ones at higher redshift (see White & Hu 2000).

Second, the mass in each of these boxes is projected along the line-of-sight, yielding a surface mass density at the appropriate comoving distance $w_i = (i - 1/2)L$. Each of these surface mass densities can now be considered a lens plane, and the propagation of light can be followed from one lens plane to the next; the corresponding theory was worked out in detail by Blandford & Narayan (1986; see also Chap. 9 of SEF), but applied as early as 1970 by Refsdal (1970) for a cosmological model consisting of point masses only (see also Schneider & Weiss 1988a,b). Important to note is that the surface mass density Σ in each lens plane is the projection of $\Delta\rho = \rho - \bar{\rho}$ of a box, so that for each lens plane, $\langle \Sigma \rangle = 0$. As has been shown in Seitz et al. (1994), this multiple lens-plane approach presents a well-defined discretization of the full 3-dimensional propagation equations. Light bundles are deflected and distorted in each lens plane and thus represented as piecewise straight rays. The resulting Jacobi matrix \mathcal{A} is then obtained as a sum of products of the tidal matrices in the individual lens planes, yielding a discretized version of

the form (88) for \mathcal{A} . The result of such simulations is then the matrix $\mathcal{A}(\boldsymbol{\theta})$ on a predefined angular grid, as well as the positions $\boldsymbol{\beta}(\boldsymbol{\theta})$ in the source plane. The latter will not be needed here, but have been used in studies of multiple images caused by the LSS (see Wambsganss et al. 1998).

One needs special care in applying the foregoing prescription; in particular, in the smoothing process to obtain a mass distribution from the discrete particles; Jain et al. (2000) contains a detailed discussion on these issues.¹⁰ The finite spatial resolution in the simulations translates into a redshift-dependent angular resolution, which degrades for the low redshift lens planes; on the other hand, those have a small impact on the light propagation due to the large value of Σ_{cr} for them [see eq. (10) of IN]. The discreteness of particles gives rise to a shot-noise term in the mass distribution, yielding increased power on small angular scales.

Results from ray-tracing simulations. We shall summarize here some of the results from ray-tracing simulation:

- Whereas the Jacobi matrix in this multi-deflection situation is no longer symmetric, the contribution from the asymmetry is very small. The power spectrum of the asymmetric part of \mathcal{A} is at least three orders of magnitude smaller than the power spectrum P_κ , for sources at $z_s = 1$ (Jain et al. 2000). This result is in accord with analytical expectations (e.g., Bernardeau et al. 1997; Schneider et al. 1998a), i.e., that terms quadratic in the Newtonian potential are considerably smaller than first-order terms, and supports the validity of the Born approximation. Furthermore, this result suggests that a simpler method for predicting cosmic shear distributions from numerical simulations may be legitimate, namely to project the mass distribution of all lens planes along the grid of angular positions, with the respective weighting factors, according to (92), i.e., employing the Born approximation. Of course this simplified method is computationally much faster than the full ray-tracing.
- The power spectra obtained reproduce the ones derived using (99), over the range of wavevectors which are only mildly affected by resolution and discreteness effects. This provides an additional check on the accuracy of the fitting formulae for the non-linear power spectrum.
- The simulation results give the full two-dimensional shear map, and thus can be used to study properties other than the second-order ones, e.g., higher-order statistics, or the occurrence of circular shear patterns indicating the presence of strong mass concentrations. An example of such maps is shown in Fig. 26. These shear maps can be used to simulate real surveys, e.g., including the holes in the data field resulting from masking or

¹⁰ For other recent ray-tracing simulations related to cosmic shear, see e.g. Barber et al. (2000); Hamana & Mellier (2001); Premadi et al. (2001); Taruya et al. (2002); Fluke et al. (2002); Barber (2002); Vale & White (2003).

complicated survey geometries, and thus to determine the accuracy with which the power spectra can be determined from such surveys. Note that in order to quantify the error (or covariance matrix) of any second-order statistics, one needs to know the fourth-order statistics, which in general cannot be obtained analytically when outside the linear (Gaussian) regime. Simulations are also used to obtain good survey strategies.

Higher-order correction terms. Up to now we have considered the lowest-order approximation of the Jacobi matrix (88) and have argued that this provides a sufficiently accurate description. Higher-order terms in Φ were neglected since we argued that, because the Newtonian potential is very small, these should play no important role. However, this argument is not fully correct since, whereas the potential certainly is small, its derivatives are not necessarily so. Of course, proper ray-tracing simulation take these higher-order terms automatically into account.

We can consider the terms quadratic in Φ when expanding (88) to higher order. There are two such terms, one containing the product of second-order derivatives of Φ , the other a product of first derivatives of Φ and its third derivatives. The former is due to lens-lens coupling: The shear and surface mass densities from different redshifts (or lens planes, in the discretized approximation) do not simply add, but multi lens plane theory shows that the tidal matrices from different lens planes get multiplied. The latter term comes from dropping the Born approximation and couples the deflection of a light ray (first derivative of Φ) with the change of the tidal matrix with regards to the position (third derivatives of Φ). These terms are explicitly given in the appendix of Schneider et al. (1998a), in Bernardeau et al. (1997) and in Cooray & Hu (2002) and found to be indeed small, providing corrections of at most a few percent. Furthermore, Hamana (2001) has shown that the magnification bias caused by the foreground matter inhomogeneities on the selection of background galaxies has no practical effect on second-order cosmic shear statistics.

Another effect that affects the power spectrum P_κ is the difference between shear and reduced shear, the latter being the observable. Since the correlation function of the reduced shear is the correlation function of the shear plus a term containing a product of two shears and one surface mass density, this correction depends linearly on the third-order statistical properties of the projected mass κ . Also this correction turns out to be very small; moreover, it does not give rise to any B-mode contribution (Schneider et al. 2002a).

7 Large-scale structure lensing: results

After the theory of cosmic shear was considered in some detail in the previous section, we shall summarize here the observational results that have been

obtained so far. In fact, as we will see, progress has been incredibly fast over the past \sim four years, with the first detections reported in 2000, and much larger surveys being available by now, with even larger ones ongoing or planned. Already by now, cosmic shear is one of the pillars on which our cosmological model rests.

The predictions discussed in the previous section have shown that the rms value of cosmic shear is of the order of $\sim 2\%$ on angular scales of $\sim 1'$, and considerably smaller on larger scales. These small values make the measurements of cosmic shear particularly challenging, as the observational and instrumental effects described in Sect. 3 are expected to be larger than the cosmic shear signal, and thus have to be understood and removed with great precision. For example, the PSF anisotropy of nearly all wide-field cameras is considerably larger than a few percent and thus needs to be corrected for. But, as also discussed in Sect. 3, methods have been developed and thoroughly tested which are able to do so.

7.1 Early detections of cosmic shear

Whereas the theory of cosmic shear was worked out in the early 1990's (Blandford et al. 1991; Miralda-Escudé 1991; Kaiser 1992), it took until the year 2000 before this effect was first discovered.¹¹ The reason for this evolution must be seen by a combination of instrumental developments, i.e. the wide-field CCD mosaic cameras, and the image analysis software, like IMCAT (the software package encoding the KSB method discussed in Sect. 3.5), with which shapes of galaxies can be accurately corrected for PSF effects. Then in March 2000, four groups independently announced their first discoveries of cosmic shear (Bacon et al. 2000; Kaiser et al. 2000; van Waerbeke et al. 2000, Wittman et al. 2000). In these surveys, of the order of 10^5 galaxy images have been analyzed, covering about 1 deg^2 . Later that year, Maoli et al. (2001) reported a significant cosmic shear measurement from 50 widely separated FORS1@VLT images, each of size $\sim 6'.5 \times 6'.5$, which also agreed with the earlier results. The fact that the results from four independent teams agreed within the respective error bars immediately lend credit to this new window of observational cosmology. This is also due to the fact that 4 different telescopes, 5 different cameras (the UH8K and CFH12K at CFHT, the $8' \times 16'$ -imager on WHT, the BTC at the 4m-CTIO telescope and FORS1 at the VLT), independent data reduction tools and at least two different image analysis methods have been used. These early results are displayed in Fig. 39, where the (equivalent) shear dispersion is plotted as a function of effective circular aperture radius, together with the predictions for several cosmological models. It is immediately clear that a high-normalization Einstein-de

¹¹ An early heroic attempt by Mould et al. (1994) to detect cosmic shear on a single $\sim 9' \times 9'$ field only yielded an upper limit, and the putative detection of a shear signal by Schneider et al. (1998b; see also Fort et al. 1996) in three $2' \times 2'$ fields is, due to the very small sky area, of no cosmological relevance.

Sitter model can already be excluded from these early results, but the other three models displayed are equally valid approximations to the data.

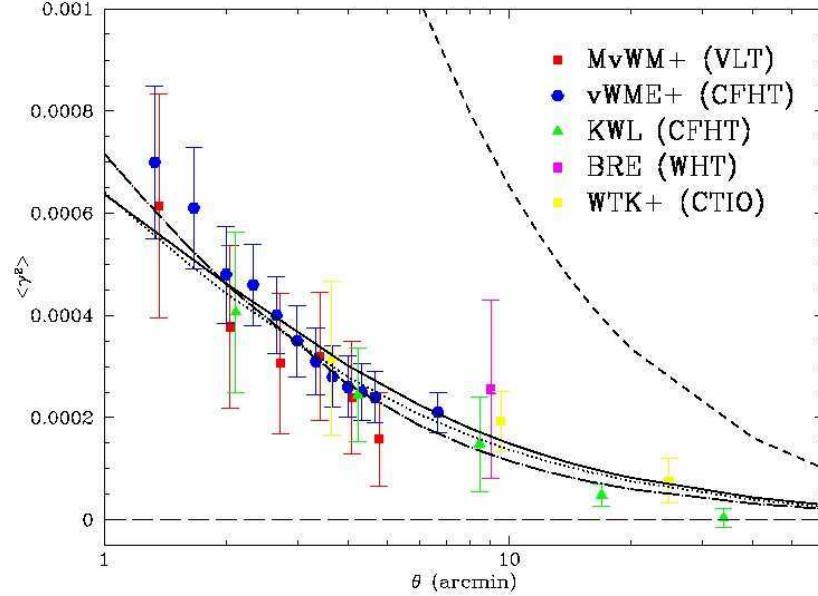


Fig. 39. Shear dispersion as a function of equivalent circular aperture radius as obtained from the first five measurements of cosmic shear (MvWM+: Maoli et al. 2001; vWME+: van Waerbeke et al. 2000; KWL: Kaiser, Wilson & Luppino 2000; BRE: Bacon, Refregier & Ellis 2000; WTK: Wittman et al. 2000). The data points within each team are not statistically independent, due to the fairly strong covariance of the shear dispersion on different angular scales, but points from different teams are independent (see text). The error bars contain the noise from the intrinsic ellipticity dispersion and, for some of the groups, also an estimate of cosmic variance. The four curves are predictions from four cosmological models; the uppermost one corresponds to an Einstein-de Sitter Universe with normalization $\sigma_8 = 1$, and can clearly be excluded by the data. The other three models are cluster normalized – see Sect. 4.4 of IN – and all provide equally good fits to these early data (courtesy: Y. Mellier)

Maoli et al. (2001) considered the constraints one obtains by combining the results from these five surveys, in terms of the normalization parameter σ_8 of the power spectrum. The confidence contours in the $\Omega_m - \sigma_8$ -plane are shown in Fig. 40. There is clearly a degeneracy between these two parameters from the data sets considered, roughly tracing $\sigma_8 \sim 0.59\Omega_m^{-0.47}$; although the best fitting model is defined by $\Omega_m = 0.26$, $\sigma_8 = 1.1$, it cannot be significantly

distinguished from, e.g., a $\Omega_m = 1$, $\sigma_8 = 0.62$ model since the error bars displayed in Fig. 39 are too large and the range of angular scales over which the shear was measured is too small. In Fig. 40, the solid curve displays the normalization as obtained from the abundance of massive clusters, which is seen to follow pretty much the valley of degeneracy from the cosmic shear analysis. This fact should not come as a surprise, since the cluster abundance probes the power spectrum on a comoving scale of about $8h^{-1}$ Mpc, which is comparable to the median scale probed by the cosmic shear measurements. However, the predictions of the cluster abundance rely on the assumption that the initial density field was Gaussian, whereas the cosmic shear prediction is independent of this assumption, which therefore can be tested by comparing the results from both methods.

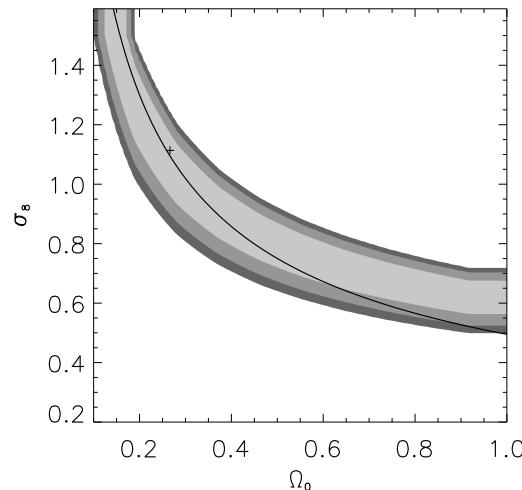


Fig. 40. Constraints on Ω_m and σ_8 from the five surveys shown in Fig. 39; shown are 1, 2 and 3- σ confidence regions. The cross denotes the best-fitting model, but as can be seen, these two parameters are highly degenerate with the data used. The solid curve displays the constraint from cluster normalization (from Maoli et al. 2001)

7.2 Integrity of the results

As mentioned before, the cosmic shear effects are smaller than many observational effects (like an anisotropic PSF) that could mimic a shear; it is therefore necessary to exclude as much as possible such systematics from the data. The early results described above were therefore accompanied by quite a large number of tests; they should be applied to all cosmic shear surveys as a sanity check. A few of those shall be mentioned here.

Stellar ellipticity fits. The ellipticity of stellar objects should be well fitted by a low-order function, so one is able to predict the PSF anisotropy at galaxy locations. After subtracting this low-order fit from the measured

stellar ellipticities, there should be no coherent spatial structure remaining, and the ellipticity dispersion of the corrected ellipticities should be considerably smaller than the original ones, essentially compatible with measurement noise.

Correlation of PSF anisotropy with corrected galaxy ellipticities.

After correcting for the anisotropy of the PSF, there should remain no correlation between the corrected galaxy ellipticities and the ellipticity of the PSF. This correlation can be measured by considering $\langle \epsilon \epsilon^* \rangle$, where ϵ is the corrected galaxy ellipticity (31), and ϵ^* the uncorrected stellar ellipticity (i.e., the PSF anisotropy). Bacon et al. (2000) found that for fairly low signal-to-noise galaxy images, this correlation was significantly different from zero, but for galaxies with high S/N (only those entered their cosmic shear analysis), no significant correlation remained. The same was found in van Waerbeke et al. (2000), except that the average $\langle \epsilon_1 \rangle$ was slightly negative, but independent of ϵ_1^* . The level of $\langle \epsilon_1 \rangle$ was much smaller than the estimated cosmic shear, and does not affect the latter by more than 10%.

Spatial dependence of mean galaxy ellipticity. When a cosmic shear survey consist of many uncorrelated fields, the mean galaxy ellipticity at a given position on the CCD chips should be zero, due to the assumed statistical isotropy of the shear field. If, on the other hand, the shear averaged over many fields shows a dependence on the chip position, most likely optical distortions and/or PSF effects have not been properly accounted for.

Parity invariance. The two-point correlation function $\xi_{\times}(\theta) = \langle \gamma_t \gamma_{\times} \rangle(\theta)$ is expected to vanish for a density distribution that is parity symmetric. More generally, every astrophysical cause for a ‘shear’ signal (such as intrinsic galaxy alignments, or higher-order lensing effects) is expected to be invariant under parity transformation. A significant cross-correlation ξ_{\times} would therefore indicate systematic effects in the observations and/or data analysis.

7.3 Recent cosmic shear surveys

Relatively soon after the announcement of the first cosmic shear detections, additional results were published. These newer surveys can roughly be classified as follows: deep surveys, shallower, but much wider surveys, and special surveys, such as obtained with the Hubble Space Telescope. We shall mention examples of each of these classes here, without providing a complete list.

Deep surveys. Currently the largest of the deep surveys from which cosmic shear results have been published is the VIRMOS-DESCART survey, carried out with the CFH12K camera at the CFHT; this camera covers about $45' \times 30'$

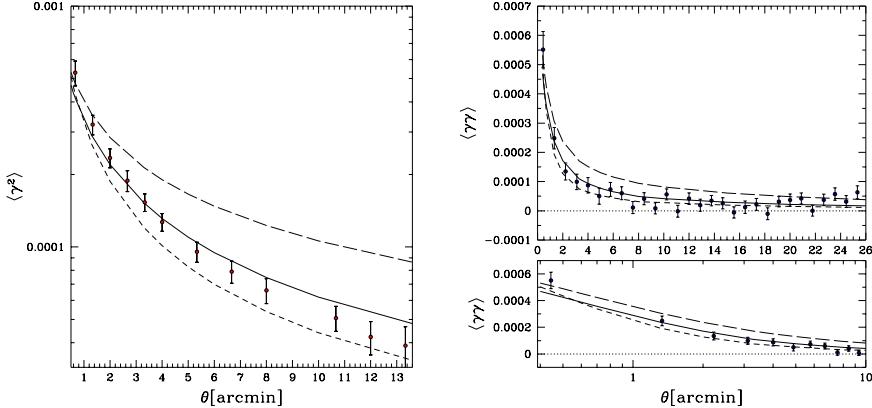


Fig. 41. The shear dispersion as a function of aperture radius (left) and the shear correlation function $\xi_+(\theta)$ (right) as measured from the VIRIMOS-DESCART survey (van Waerbeke et al. 2001). The lower panel on the right shows an enlargement with logarithmic axis of the larger figure. The error bars were calculated from simulations in which the galaxy images have been randomized in orientation. The curves show predictions from three different cosmological models, corresponding to $(\Omega_m, \Omega_\Lambda, \sigma_8) = (0.3, 0, 0.9)$ (open model, short-dashed curves), $(0.3, 0.7, 0.9)$ (low-density flat model, solid curves), and $(1, 0, 0.6)$ (Einstein-de Sitter Universe, long-dashed curves). In all cases, the shape parameter of the power spectrum was set to $\Gamma_{\text{spect}} = 0.21$. The redshift distribution of the sources was assumed to follow the law (128), with $\alpha = 2$, $\beta = 1.5$ and $z_0 = 0.8$, corresponding to a mean redshift of $\bar{z} \approx 1.2$

in one exposure. The exposure time of the images, taken in the I-band, is one hour. The survey covers four fields of $2^\circ \times 2^\circ$ each, of which roughly 8.5 deg^2 have been used for a weak lensing analysis up to now (van Waerbeke et al. 2001, 2002). About 20% of the area is masked out, to account for diffraction spikes, image defects, bright and large foreground objects etc. The number density of galaxy images used for the cosmic shear analysis is about 17 arcmin^{-2} . A small part of this survey was used for the early cosmic shear detection (van Waerbeke et al. 2000). Compared to the earlier results, the error bars on the shear measurements are greatly reduced, owing to the much better statistics. We show in Fig. 41 the shear dispersion and the correlation function as measured from this survey. Furthermore, this survey yielded the first detection of a significant $\langle M_{\text{ap}}^2 \rangle$ -signal; we shall come back to this later. In order to compare the measured shear signal with cosmological predictions, one needs to assume a redshift distribution for the galaxies; a frequently used parameterization for this is

$$p(z) = N \left(\frac{z}{z_0} \right)^\alpha \exp \left[- \left(\frac{z}{z_0} \right)^\beta \right], \quad (128)$$

where α and β determine the shape of the redshift distribution, z_0 the characteristic redshift, and N is a normalization factor, chosen such as $\int dz p(z) = 1$.

Another example of a deep survey is the Suprime-Cam survey (Hamana et al. 2003), a 2.1 deg^2 survey taken with the wide-field camera Suprime-Cam (with a $34' \times 27'$ field-of-view) at the 8.2-m Subaru telescope. With an exposure time of 30 min, the data is considerably deeper than the VIRMOS-DESCART survey, due to the much larger aperture of the telescope. After cuts in the object catalog, the resulting number density of objects used for the weak lensing analysis is $\approx 30 \text{ arcmin}^{-2}$. Fig. 42 shows how small the PSF anisotropy is, and that the correction with a fifth-order polynomial over the whole field-of-view in fact reduces the remaining stellar ellipticities considerably. This survey has detected a significant cosmic shear signal, as measured by the shear correlation functions and the aperture mass dispersion, over angular scales $2' \lesssim \theta \lesssim 40'$. The shear signal increases as fainter galaxies are used in the analysis, as expected, since fainter galaxies are expected to be at larger mean redshift and thus show a stronger shear signal.

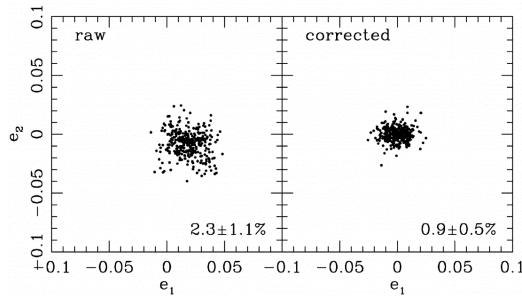


Fig. 42. Stellar ellipticities before and after correction for PSF anisotropies in the Suprime-Cam survey. Numbers give mean and dispersion of stellar ellipticities $|\chi|$ (from Hamana et al. 2003)

Bacon et al. (2003) combine images taken at the Keck II telescope and the WHT. For the former, 173 fields were used, each having a f.o.v. of $2' \times 8'$; and the data from WHT were obtained from 20 different fields, covering about 1 deg^2 in total. The large number of fields minimizes the sample variance of this particular survey, and the two instruments used allowed a cross-check of instrumental systematics.

Very wide surveys. Within a given observing time, instead of mapping a sky region to fairly deep magnitudes, one can also map larger regions with smaller exposure time; since most of the surveys have been carried out with goals in addition to cosmic shear, the survey strategy will depend on these other considerations. We shall mention two very wide surveys here.

Hoekstra et al. (2002a; also Hoekstra et al. 2002b) used the Red Cluster Sequence (RCS) survey, a survey designed to obtain a large sample of galaxy clusters using color selection techniques (Gladders & Yee 2000). The cosmic shear analysis is based on 53 deg^2 of R_C -band data, spread over 13 patches on

the sky and observed with two different instruments, the CFH12K@CFHT for Northern fields, and the Mosaic II camera at the CTIO 4m telescope in the South. The integration times are 900 s and 1200 s, respectively. The shear dispersion as measured with the two instruments are in satisfactory agreement and thus can be safely combined. Owing to the shallower magnitude, the detected shear is smaller than in the deeper surveys mentioned above: on a scale of 2.5 arcmin, the shear dispersion is $\langle |\bar{\gamma}|^2 \rangle \sim 4 \times 10^{-5}$ in the RCS survey, compared to $\sim 2 \times 10^{-4}$ in the deeper VIRMOS-DESCART survey (see Fig. 41), in accordance with expectations.

Jarvis et al. (2003) presented a cosmic shear survey of 75 deg^2 , taken with the BTC camera and the Mosaic II camera on the CTIO 4m telescope, with about half the data taken with each instrument. The survey covers 12 fields, each with sidelength of $\sim 2.5^\circ$. For each pointing, three exposures of 5 min were taken, making the depth of this survey comparable to the RCS. A total of $\sim 2 \times 10^6$ galaxies with $R \leq 23$ were used for the shear analysis. Since this survey has some peculiar properties which are very educational, it will be discussed in somewhat more detail. The first point to notice is the large pixel size of the BTC, of $0''.43$ per pixel – for comparison, the CFH12K has $\sim 0''.20$ per pixel. With a median seeing of $1''.05$, the PSF is slightly undersampled with the BTC. Second, the PSF anisotropy on the BTC is very large, as shown in Fig. 43 – a large fraction of the exposures has stellar images with ellipticities higher than 10%. Obviously, this renders the image analysis and the correction for PSF effects challenging. As shown on the right-hand part of Fig. 43, this challenge is indeed met. This fact is very nicely illustrated in Fig. 44, where the corrected stellar ellipticities are shown as a function of the PSF anisotropy; in essence, the correction reduces the PSF anisotropy by nearly a factor of 300!

The third point to notice is that the image analysis for this survey has not been carried out with IMCAT (as for most of the other surveys), but by a different image analysis method described in Bernstein & Jarvis (2002). In this respect, this survey is independent of all the others described in this section; it is important to have more than one image analysis tool to check potential systematics of either one.

One of the amazing results from the CTIO cosmic shear survey is that the shear dispersion can be measured with about a 3σ significance on each of the 12 fields. Hence, this provides a shear dispersion measurement on scales larger than 1 degree (the radius of a circle with area of the mean area of the 12 fields of $\sim 6.2 \text{ deg}^2$); the shear dispersion on these scales is $\langle |\bar{\gamma}|^2 \rangle = 0.0012 \pm 0.0003$.

Special surveys. There are a number of cosmic shear surveys which cover a much smaller total area than the ones mentioned above, and are thus not competitive in terms of statistical accuracy, but which have some special properties which give them an important complementary role. One example are surveys carried out with the Hubble Space Telescope. Since for them the

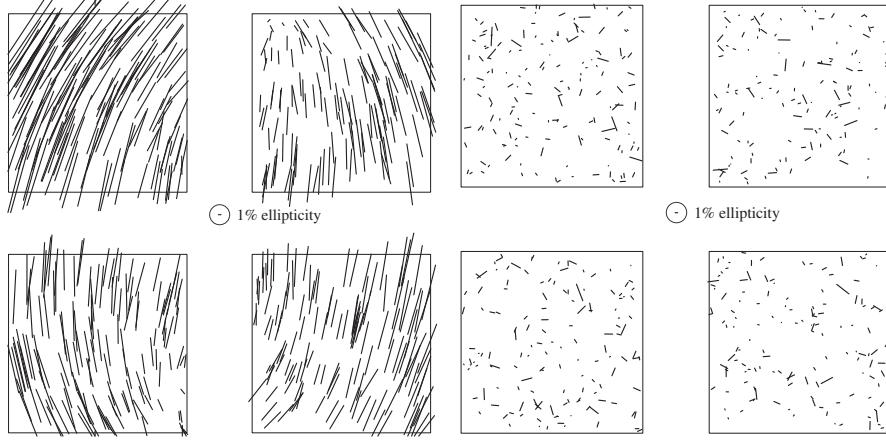


Fig. 43. On the left-hand side, the raw ellipticities of stars are shown for the four CCDs of the BTC instrument; for reference, a 1% ellipticity is indicated. After correcting for the PSF anisotropy, the remaining stellar ellipticities (shown on the right) are of order 1–2%, and essentially uncorrelated with position on the chip, i.e., they are compatible with measurement noise (from Jarvis et al. 2003)

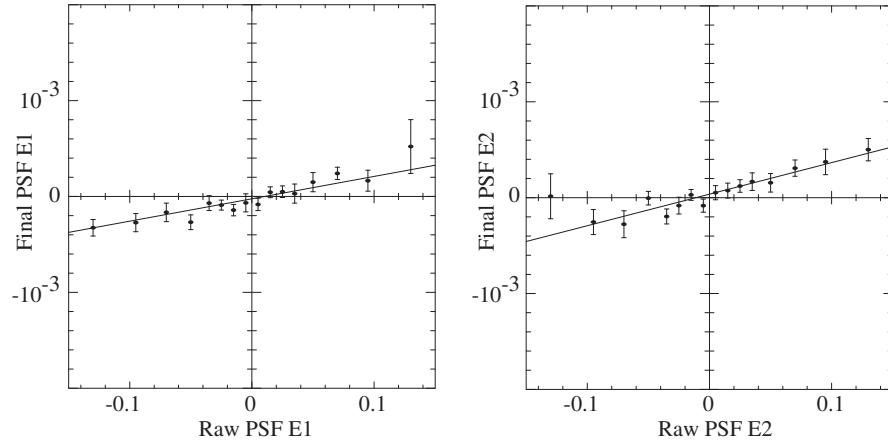


Fig. 44. These two plots show the two components of the stellar ellipticities as measured on the data (x -axis) and after correction, from the Jarvis et al. (2003) survey. The slope of the straight line is about $1/300$, meaning that the strong PSF anisotropy can be corrected for up to this very small residual. The final PSF anisotropy is well below 5×10^{-4} . This figure, together with Fig. 43, demonstrates how well the procedures for PSF corrections work (from Jarvis et al. 2003)

PSF is much smaller than for ground-based observations, PSF corrections in measuring galaxy ellipticities are expected to be correspondingly smaller. The drawback of HST observations is that its cameras, at least before the installment of the ACS, have a small field-of-view, less than 1 arcmin² for the STIS CCD, and about 5 arcmin² for WFPC2. This implies that the total area covered by HST surveys are smaller than those achievable from the ground, and that the number of stars per field are very small, so that PSF measurements are typically not possible on those frames which are used for a cosmic shear analysis. Hence, the PSF needs to be measured on different frames, e.g., taken on star clusters, and one needs to assume (this assumption can be tested, of course) that the PSF is fairly stable in time. In fact, this is not really true, as the telescopes moves in and out the Earth's shadow every orbit, thereby changing its temperature and thus changing its length (an effect called breathing). A further potential problem of HST observations is that the WFPC2 has a pixel scale of 0''.1 and thus substantially undersamples the PSF; this is likely to be a serious problem for very faint objects whose size is not much larger than the PSF size.

Cosmic shear surveys from two instruments onboard HST have been reported in the literature so far. One of the surveys uses archival data from the Medium Deep Survey, a mostly parallel survey carried out with the WFPC2. Refregier et al. (2002) used 271 WFPC2 pointings observed in the I-band, selected such that each of them is separated from the others by at least 10' to have statistically independent fields. They detected a shear dispersion on the scale of the WFC-chips (which is equivalent to a scale $\theta \sim 0''.72$) of $\langle |\bar{\gamma}|^2 \rangle \sim 3.5 \times 10^{-4}$, which is a 3.8σ detection. The measurement accuracy is lower than that, owing to cosmic variance and uncertainties in the redshift distribution of the sources. Hämerle et al. (2002) used archival parallel data taken with STIS; from the 121 fields which are deep enough, have multiple exposures, and are at sufficiently high galactic latitude, they obtained a shear dispersion of $\langle |\bar{\gamma}|^2 \rangle \sim 15 \times 10^{-4}$ on an effective scale of $\sim 30''$, a mere 1.5σ detection. This low significance is due to the small total area covered by this survey. On the other hand, since the pixel scale of STIS is half of that of WFPC2, the undersampling problem is much less in this case. A larger set of STIS parallel observations were analyzed with respect to cosmic shear by Rhodes et al. (2004) and Miralles et al. (2003). Whereas Rhodes et al. obtained a significant ($\sim 5\sigma$) detection on an angular scale of $\sim 30''$, Miralles et al. concluded that the degradation of the STIS CCD in orbit regarding the charge transfer efficiency prevents a solid measurement of weak lensing. The discrepancies between these two works, which are based to a large degree on the same data set, is unclear at present. Personally I consider this discrepancy as a warning sign that weak lensing measurement based on small fields-of-view, and correspondingly too few stars to control the PSF on the science exposures, need to be regarded with extreme caution.

The new ACS onboard HST offers better prospects for cosmic shear measurements, since it has a substantially larger field-of-view. A first result was derived by Schrabback (2004), again based on parallel data. He found that the PSF is not stable in time, but that the anisotropy pattern changes among only a few characteristic patterns. He used those as templates, and the (typically a dozen) stars in the science frames to select a linear combination of these templates for the PSF correction of individual frames, thereby obtaining a solid detection of cosmic shear from the early ACS data.

A further survey that should be mentioned here is the one conducted on COMBO17 fields (Brown et al. 2003). COMBO17 is a one square degree survey, split over four fields, taken with the WFI at the ESO/MPG 2.2m telescope on La Silla, in 5 broad-band and 12 medium-band filters. In essence, therefore, this multi-band survey produces low-resolution spectra of the objects and thus permits to determine very accurate photometric redshifts of the galaxies taken for the shear analysis. Therefore, for the analysis of Brown et al., the redshift distribution of the galaxies is assumed to be very well known and not a source of uncertainty in translating the cosmic shear measurement into a constraint on cosmological parameters. We shall return to this aspect in Sect. 7.6. The data set was reanalyzed by Heymans et al. (2004) where special care has been taken to identify and remove the signal coming from intrinsic alignment of galaxy shapes.

7.4 Detection of B-modes

The recent cosmic shear surveys have measured the aperture mass dispersion $\langle M_{\text{ap}}^2(\theta) \rangle$, as well as its counterpart $\langle M_{\perp}^2(\theta) \rangle$ for the B-modes (see Sect. 6.5). These aperture measures are obtained in terms of the directly measured shear correlation functions, using the relations (125). As an example, we show in Fig. 45 the aperture measures as obtained from the Red Cluster Sequence survey (Hoekstra et al. 2002a). A significant measurement of $\langle M_{\text{ap}}^2(\theta) \rangle$ is obtained over quite a range of angular scales, with a peak around a few arcminutes, as predicted from CDM power spectra (see Fig. 35). In addition to that, however, a significant detection of $\langle M_{\perp}^2(\theta) \rangle$ signifies the presence of B-modes. As discussed in Sect. 6.5, those cannot be due to cosmic shear. The only plausible explanation for them, apart from systematics in the observations and data analysis, is an intrinsic alignment of galaxies. If this is the cause of the B-modes, then one would expect that the relative contribution of the B-mode signal decreases as higher-redshift galaxies are used for the shear measurement. In fact, this expectation is satisfied, as shown in Fig. 45, where the galaxy sample is split into a bright and faint part, and the relative amplitude of the B-mode signal is smaller for the fainter (and thus presumably more distant) sample.

Similar detections of a B-mode signal have been obtained by the other surveys. For example, van Waerbeke et al. (2001) reported a significant B-mode signal on angular scales of a few arcminutes. In the reanalysis of the

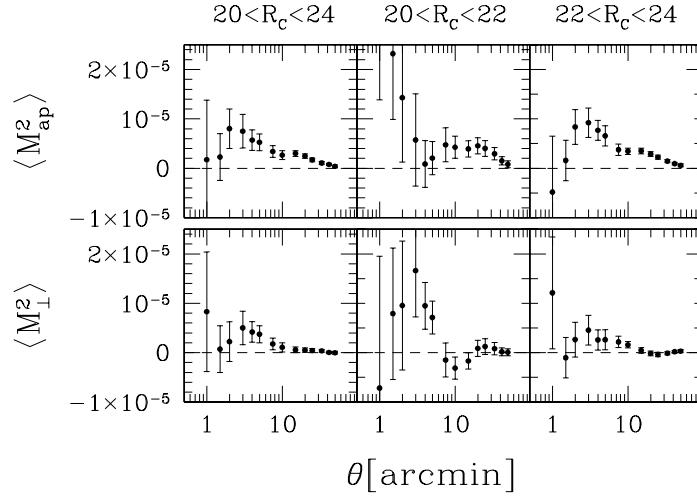


Fig. 45. The aperture mass dispersion $\langle M_{\text{ap}}^2(\theta) \rangle$ (top panels) and the cross aperture dispersion $\langle M_{\perp}^2(\theta) \rangle$ (bottom panels) from the RCS survey (Hoekstra et al. 2002a). In the left panels, all galaxies with apparent magnitude $20 \leq R_C \leq 24$ are used, the middle and right panels show the same statistics for the brighter and fainter subsamples of background galaxies, respectively. Error bars in the former are larger, owing to the smaller number of bright galaxies

VIRMOS-DESCART data, van Waerbeke et al. (2002) reported that the B-mode on these scales was caused by the polynomial PSF anisotropy fit: the third-order function (fitted for each chip individually) has its largest amplitude near the boundary of the chips and is least well constrained there, unless one finds stars close to these edges. If a second-order polynomial fit is used, the B-modes on a few arcminute scales disappear. Van Waerbeke et al. (2002) calculate the aperture statistics from the uncorrected stellar ellipticities in their survey and found that the ‘E- and B-modes’ of the PSF anisotropy have very similar amplitude and shape (as a function of θ). This similarity is unlikely to change in the course of the PSF correction procedure. Thus, they argue, that if the B-mode is due to systematics in the data analysis, a systematic error of very similar amplitude will also affect the E-mode. Jarvis et al. (2003) found a significant B-mode signal on angular scales below $\sim 30'$; hence, despite their detection of an E-mode signal over a large range of angular scales $1' \lesssim \theta \lesssim 100'$, one suspects that part of this signal might be due to non-lensing effects.

Given our lack of understanding about the origin of the B-mode signal, and the associated likelihood that any effect causing a B-mode signal also contributes a non-lensing part to the E-mode signal, one needs a prescription on how to use the detected E-mode signal for a cosmological analysis.

Depending on what one believes the B-modes are due to, this prescription varies. For example, if the B-mode is due to a residual systematic, one would add its signal in quadrature to the error bars of the E-mode signal, as done in van Waerbeke et al. (2002). On the other hand, if the B-mode signal is due to intrinsic alignments of galaxies, as is at least suggested for the RCS survey from Fig. 45 owing to its dependence on galaxy magnitudes, then it could be more reasonable to subtract the B-mode signal from the E-mode signal, if one assumes that intrinsic alignments produce similar amplitudes of both modes [which is far from clear, however; Mackey et al. (2002) find that the E-mode signal from intrinsic alignments is expected to be ~ 3.5 times higher than the corresponding B-mode signal].

Owing to the small size of the fields observed with the early HST instruments, no E/B-mode decomposition can be carried out from these surveys – the largest size of these fields is smaller than the angular scale at which the aperture mass dispersion is expected to peak (see Fig. 35). However, future cosmic shear studies carried out with ACS images will most likely be able to detect, or set upper bounds on the presence of B-modes.

In fact, it is most likely that (most of) the B-mode signal seen in the cosmic shear surveys is due to remaining systematics. Hoekstra (2004) investigated the PSF anisotropy of the CFH12k camera using fields with a high number density of stars. Randomly selecting about 100 stars per CCD, which is the typical number observed in high galactic latitude fields, he fitted a second-order polynomial to these stars representing the PSF anisotropy. Correcting with this model all the stars in the field, the remaining stellar ellipticities carry substantial E- and B-mode signals, essentially on all angular scales, but peaking at about the size of a CCD. A substantially smaller residual is obtained if the ellipticities of stars in one of the fields is corrected by a more detailed model of the PSF anisotropy as measured from a different field; this improvement indicates that the PSF anisotropy pattern in the data set used by Hoekstra is fairly stable between different exposures. This, however, is not necessarily the case in other datasets. Nevertheless, if one assumes that the PSF anisotropy is a superposition of two effects, one from the properties of the telescope and instrument itself, the other from the specific observation procedure (e.g., tracking, wind shake, etc.), and further assuming that the latter one affects mainly the large-scale properties of the anisotropy pattern, then a superposition of a PSF model (obtained from a dense stellar field and describing the small-scale properties of the anisotropy pattern) plus a low-order polynomial can be a better representation of the PSF anisotropy. This indeed was verified in the tests made by Hoekstra (2004). In their reanalysis of the VIRMOS-DESCART survey, van Waerbeke et al. (2004) have fitted the PSF anisotropy with a rational function, instead of a polynomial. This functional form was suggested by the study of Hoekstra (2004). When correcting the galaxy ellipticities with this new PSF model, essentially no more B-modes in the VIRMOS-DESCART survey are detected. Further studies on

PSF anisotropy corrections need to be conducted; possibly the optimal way of dealing with them will be instrument-specific.

7.5 Cosmological constraints

The measured cosmic shear signal can be translated into constraints on cosmological parameters, by comparing the measurements with theoretical predictions. In Sect. 6.4 we have outlined how such a comparison can be made; there, we have concentrated on the shear correlation functions as the primary observables. However, the detection of significant B-modes in the shear field makes the aperture measures the ‘better’ statistics to compare with predictions. They can be calculated from the shear correlation functions, as shown in (125). Calculating a likelihood function from the aperture mass dispersion proceeds in the same way as outlined in Sect. 6.4 for the correlation functions.

We have argued in Sect. 6.3 that $\langle M_{\text{ap}}^2(\theta) \rangle$ provides very localized information about the power spectrum $P_\kappa(\ell)$ and is thus a very useful statistic. One therefore might expect that the aperture mass dispersion as calculated from the shear correlation functions contains essentially all the second-order statistical information of the survey. This is not true, however; one needs to recall that the shear correlation function ξ_+ is a low-pass filter of the power spectrum, and thus contains information of P_κ on angular scales larger than the survey size. This information is no longer contained in the aperture mass dispersion, owing to its localized associated filter. Therefore, in order to keep this long-range information in the comparison with theoretical predictions, it is useful to complement the estimates of $\langle M_{\text{ap}}^2(\theta) \rangle$ with either the shear dispersion, or the correlation function ξ_+ , at a scale which is not much smaller than the largest scale at which $\langle M_{\text{ap}}^2(\theta) \rangle$ is measured. Note, however, that this step implicitly assumes that on these large angular scales, the shear signal is essentially free of B-mode contributions. If this assumption is not true, and cannot be justified from the survey data, then this additional constraint should probably be dropped.

The various constraints on parameters that have been derived from the cosmic shear surveys differ in the amount of prior information that has been used. As an example, we consider the analysis of van Waerbeke et al. (2002). These authors have considered a model with four free parameters: Ω_m , the normalization σ_8 , the shape parameter Γ_{spect} and the characteristic redshift z_s (or, equivalently, mean redshift \bar{z}_s) of their galaxy sample, assuming a flat Universe, i.e., $\Omega_A = 1 - \Omega_m$. They have used a flat prior for Γ_{spect} and \bar{z}_s in a fairly wide interval over which they marginalized the likelihood function (see Fig. 46). Depending on the width of these intervals, the confidence regions are more or less wide. It should be noted that the confidence contours close if Γ_{spect} and \bar{z}_s are assumed to be known (see van Waerbeke et al. 2001), but when these two parameters are kept free, Ω_m and σ_8 are degenerate.

The right panel of Fig. 46 shows the corresponding constraints as obtained from the RCS survey. Since this survey is shallower and only extends

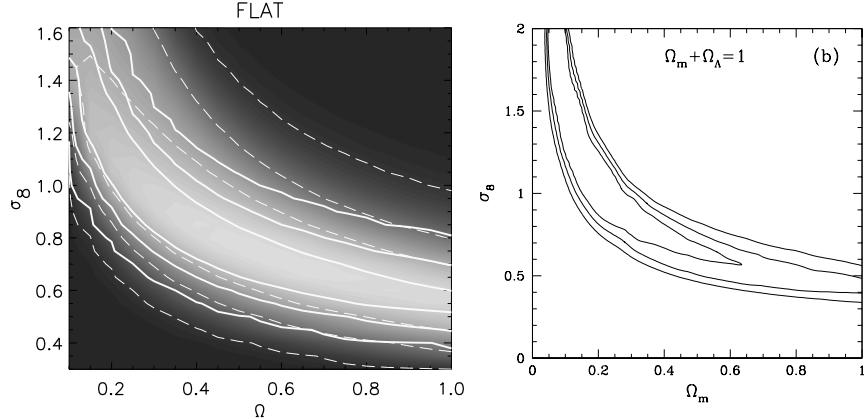


Fig. 46. Constraints on Ω_m and σ_8 from two cosmic shear surveys. Left: The VIRMOS-DESCART survey (van Waerbeke et al. 2002). The grey-scale and dashed contours show the 68%, 95% and 99.9% confidence regions with a marginalization over the range $\Gamma_{\text{spect}} \in [0.05, 0.7]$, and mean galaxy redshift in the range $\bar{z}_s \in [0.50, 1.34]$, whereas the solid contours show the same confidence regions with the stronger priors $\Gamma_{\text{spect}} \in [0.1, 0.4]$ and $\bar{z}_s \in [0.8, 1.1]$. Right: The RCS survey (Hoekstra et al. 2002), showing the 1, 2, and 3σ confidence regions for a prior $\Gamma_{\text{spect}} \in [0.05, 0.5]$ and mean redshift $\bar{z}_s \in [0.54, 0.66]$. In both cases, a flat Universe has been assumed

to magnitudes where spectroscopic surveys provide information on their redshift distribution, the range of \bar{z}_s over which the likelihood is marginalized is smaller than for the VIRMOS-DESCART survey. Correspondingly, the confidence region is slightly smaller in the case. Even smaller confidence regions are obtained if external information is used: Hoekstra et al. (2002a) considered Gaussian priors with $\Omega_m + \Omega_\Lambda = 1.02 \pm 0.06$, as follows from pre-WMAP CMB results, $\Gamma_{\text{spect}} = 0.21 \pm 0.03$, as follows from the 2dF galaxy redshift survey, and $\bar{z}_s = 0.59 \pm 0.02$, for which the width of the valley of maximum likelihood narrows considerably. Jarvis et al. (2003) used for their estimate of cosmological parameters the aperture mass dispersion at three angular scales plus the shear dispersion at $\theta = 100'$, and they considered alternatively the E-mode signal, and the E-mode signal \pm the B-mode signal, to arrive at constraints on the $\Omega_m - \sigma_8$ parameter plane. Since the CTIO survey samples a larger angular scale than the other surveys (data at small angular scales are discarded owing to the large B-mode signal there), the results are much less sensitive to Γ_{spect} ; furthermore, for the same reason the Jarvis et al. results are much less sensitive to the fit of the non-linear power spectrum according to Peacock & Dodds (1996) which van Waerbeke et al. (2002) found to be not accurate enough for some cosmological models. In fact, if instead of the Peacock & Dodds fitting formula, the fit by Smith et al. (2003) is used to

describe the non-linear power spectrum, the resulting best estimate of σ_8 is decreased by 8% for the RCS survey (as quoted in Jarvis et al. 2003).

For the RCS and the CTIO surveys, the covariance matrix was obtained from field-to-field variations, i.e., $\text{Cov}_{ij} = \langle (d_i - \mu_i)(d_j - \mu_j) \rangle$, where μ_i is the mean of the observable d_i (e.g., the aperture mass dispersion at a specific angular scale) over the independent patches of the survey, and angular brackets denote the average over all independent patches. The estimate of the covariance matrix for the VIRMOS-DESCART survey is slightly different, as it has only four independent patches.

To summarize the results from these surveys, each of them found that a combination of parameters of the form $\sigma_8 \Omega_m^\alpha$ is determined best from the data, with $\alpha \sim 0.55$, where the exact value of α depends on the survey depth. If we consider the specific case of $\Omega_m = 0.3$ which is close to the concordance value that was recently confirmed by WMAP, then the VIRMOS-DESCART survey yields $\sigma_8 = 0.94 \pm 0.12$, the RCS survey has $\sigma_8 = 0.91^{+0.05}_{-0.12}$, which improves to $\sigma_8 = 0.86^{+0.04}_{-0.05}$ if the stronger (Gaussian) priors mentioned above are used, and the CTIO survey yields $\sigma_8 = 0.71^{+0.12}_{-0.16}$, here as 2σ limits. Whereas these results are marginally in mutual agreement, the CTIO value for σ_8 is lower than the other two. The higher values are also supported by results from the WFPC2 survey by Refregier et al. (2002), who find $\sigma_8 = 0.94 \pm 0.17$, Bacon et al. (2003) with $\sigma_8 = 0.97 \pm 0.13$, and the earlier surveys discussed in Sect. 7.1. The only survey supporting the low value of the CTIO survey is COMBO17 (Brown et al. 2002; see also the reanalysis of this dataset by Heymans et al. 2004). Most likely, these remaining discrepancies will be clarified in the near future; see discussion below. It should also be noted that at least for some of the surveys, a large part of the uncertainty comes from the unknown redshift distribution of the galaxies; this situation will most likely improve, as efficient spectrographs with large multiplex capability become available at 10m-class telescopes, which will in the near future deliver large galaxy redshift surveys at very faint magnitudes. Those can be used to much better constrain the redshift distribution of the source galaxies in cosmic shear surveys.

7.6 3-D lensing

As mentioned several times before, using individual source redshift information, as will become available in future multi-color wide-field surveys, can improve the cosmological constraints obtained from weak lensing. In this section we shall therefore summarize some of the work that has been published on this so-called 3-D lensing.

Three-dimensional matter distribution. Provided the redshifts of individual source galaxies are known (or estimated from their multiple colors),

one can derive the 3-D matter distribution, not only its projection. The principle of this method can be most easily illustrated in the case of a flat Universe, for which the surface mass density $\kappa(\boldsymbol{\theta}, w)$ for sources at comoving distance w becomes – see (93)

$$\kappa(\boldsymbol{\theta}, w) = \frac{3H_0^2\Omega_m}{2c^2} \int_0^w dw' \frac{w'(w-w')}{w} \frac{\delta(w'\boldsymbol{\theta}, w')}{a(w')} . \quad (129)$$

Multiplying this expression by w and differentiating twice yields

$$\frac{d^2}{dw^2} (w \kappa(\boldsymbol{\theta}, w)) = \frac{3H_0^2\Omega_m}{2c^2} \frac{w}{a(w)} \delta(w\boldsymbol{\theta}, w) ,$$

which therefore allows one to obtain the three-dimensional density contrast δ in terms of the surface mass densities κ at different source redshifts. As we have seen in Sect. 5, there are several methods how to obtain the surface mass density from the observed shear. To illustrate the 3-D method, we use the finite-field reconstruction in the form of (60), for which one finds

$$\delta(w\boldsymbol{\theta}, w) = \frac{2c^2}{3H_0^2\Omega_m} \frac{a(w)}{w} \int d^2\boldsymbol{\theta}' \mathbf{H}(\boldsymbol{\theta}; \boldsymbol{\theta}') \cdot \frac{d^2}{dw^2} [w \mathbf{u}_\gamma(\boldsymbol{\theta}', w)] . \quad (130)$$

Taylor (2001) derived the foregoing result, but concentrated on the 3-D gravitational potential instead of the mass distribution, and Bacon & Taylor (2003) and Hu & Keeton (2003) discussed practical implementations of this relation. First to note is the notorious mass-sheet degeneracy, which in the present context implies that one can add an arbitrary function of w to the reconstructed density contrast δ . This cannot be avoided, but if the data field is sufficiently large, so that averaged over it, the density contrast is expected to vanish, this becomes a lesser practical problem. For such large data fields, the above mass reconstruction can be substituted in favour of the simpler original Kaiser & Squires (1993) method. Still more freedom is present in the reconstruction of the gravitational potential. The second problem is one of smoothing: owing to the noisiness of the observed shear field, the w -differentiation (as well as the $\boldsymbol{\theta}$ -differentiation present in the construction of the vector field \mathbf{u}_γ) needs to be carried out on the smoothed shear field. A discretization of the observed shear field, as also suggested by the finite accuracy of photometric redshifts, can be optimized with respect to this smoothing (Hu & Keeton 2003).

A first application of this methods was presented in Taylor et al. (2004) on one of the COMBO17 fields which contains the supercluster A901/902. The clusters present clearly show up also in the 3-D mass map, as well as a massive structure behind the cluster A902 at higher redshift. Already earlier, Wittman et al. (2001, 2003) estimated the redshifts of clusters found in their deep blank-field data by studying the dependence of the weak lensing signal on the estimated source redshifts, and subsequent spectroscopy showed that these estimates were fairly accurate.

Power spectrum estimates. A redshift-dependent shear field can also be used to improve on the cosmological constraints obtained from cosmic shear. Hu (1999) has pointed out that even crude information on the source redshifts can strongly reduce the uncertainties of cosmological parameters. In fact, the 3-D power spectrum can be constructed from redshift-dependent shear data (see, e.g., Heavens 2003, Hu 2002, and references therein). For illustration purposes, one can use the κ power spectrum for sources at fixed comoving distance w , which reads in a flat Universe – see (99)

$$P_\kappa(\ell, w) = \frac{9H_0^4\Omega_m^2}{4c^4} \int_0^w dw' \frac{(w-w')^2}{w^2 a^2(w')} P_\delta\left(\frac{\ell}{w'}, w'\right). \quad (131)$$

Differentiating $w^2 P_\kappa$ three times w.r.t. w then yields (Bacon et al. 2004)

$$P_\delta(k, w) = \frac{2c^4}{9H_0^4\Omega_m^2} a^2(w) \frac{d^3}{dw^3} [w^2 P_\kappa(wk, w)]. \quad (132)$$

In this way, one could obtain the three-dimensional power spectrum of the matter. However, this method is essentially useless, since it is both very noisy (due to the third-order derivatives) and throws away most of the information contained in the shear field, as it makes use only of shear correlations of galaxies having the same redshift, and not of all the pairs at different distances. A much better approach to construct the three-dimensional power spectrum is given, e.g., by Pen et al. (2003).

In my view, the best use of three-dimensional data is to construct the shear correlators $\xi_{\pm}(\theta; z_1, z_2)$, as they contain all second-order statistical information in the data and at the same time allow the identification and removal of a signal from intrinsic shape correlations of galaxies (King & Schneider 2003). From these correlation functions, one can calculate a χ^2 function as in (116) and minimize it w.r.t. the wanted parameters. One problem of this approach is the large size of the covariance matrix, which now has six arguments (two angular separations and four redshifts). However, as shown in Simon et al. (2004), it can be calculated fairly efficiently, provided one assumes that the fourth-order correlations factorize into products of two-point correlators, i.e., Gaussian fields (if this assumption is dropped, the covariance must be calculated from cosmological N-body simulations).

Bacon et al. (2004) used the COMBO17 data to derive the shape of the power spectrum, using the redshift dependent shear correlations. They parameterize the power spectrum in the form $P(k, z) \propto Ak^\alpha e^{-sz}$, so that it is described by an amplitude A , a local slope α and a growth parameter s which describes how the amplitude of the power spectrum declines towards higher redshifts. In fact, the slope $\alpha = -1.2$ was fixed to the approximate value in Λ CDM models over the relevant range of spatial scales and redshifts probed by the COMBO17 data (since the data used cover only $1/2$ deg 2 , reducing the number of free parameters by fixing α is useful). The evolution of the power spectrum is found with high significance in the data. Furthermore, the

authors show that the use of redshift information improves the accuracy in the determination of σ_8 by a factor of two compared to the 2-D cosmic shear analysis of the same data (Brown et al. 2003).

The main application of future multi-waveband cosmic shear surveys will be to derive constraints on the equation of state of dark energy, as besides lensing there are only a few methods available to probe it, most noticeably the magnitude-redshift relation of SN Ia. Since dark energy starts to dominate the expansion of the Universe only at relatively low redshifts, little information about its properties is obtainable from the CMB anisotropies alone. For that reason, quite a number of workers have considered the constraints on the dark energy equation of state that can be derived from future cosmic shear surveys (e.g., Huterer 2002; Hu 2002; Munshi & Wang 2003; Hu & Jain 2003; Abazajian & Dodelson 2003; Benabed & van Waerbeke 2003; Song & Knox 2003). The results of these are very encouraging; the sensitivity on the dark energy properties is due to its influence on structure growth. With (photometric) redshift information on the source galaxies, the evolution of the dark matter distribution can be studied by weak lensing, as shown above. Van Waerbeke & Mellier (2003) have compared the expected accuracy of the cosmic shear result from the ongoing CFHT Legacy Survey with the variation of various dark energy models and shown that the CFHTLS will be able to discriminate between some of these models, with even much better prospects from future space-based wide-field imaging surveys (e.g., Hu & Jain 2003).

7.7 Discussion

The previous sections have shown that cosmic shear research has matured; several groups have successfully presented their results, which is important in view of the fact that the effects one wants to observe are small, influenced by various effects, and therefore, independent results from different instruments, groups, and data analysis techniques are essential in this research. We have also seen that the results from the various groups tend to agree with each other, with a few very interesting discrepancies remaining whose resolution will most likely teach us even more about the accuracies of data analysis procedures.

Lessons for cosmology. A natural question to ask is, what has cosmic shear taught us so far about cosmology? The most important constraint coming from the available cosmic shear results is that on the normalization σ_8 , for which only few other accurate methods are available. We have seen that cosmic shear prefers a value of $\sigma_8 \approx 0.8 - 0.9$, which is slightly larger than current estimates from the abundance of clusters, but very much in agreement with the measurement of WMAP. The estimate from the cluster abundance is, however, not without difficulties, since it involves several scaling relations which need to be accurately calibrated; hence, different authors

arrive at different values for σ_8 (see, e.g., Pierpaoli, Scott & White 2001; Seljak 2002; Schuecker et al. 2003). The accuracy with which σ_8 is determined from CMB data alone is comparable to that of cosmic shear estimates; as shown in Spergel et al. (2003), more accurate values of σ_8 are obtained only if the CMB measurements are combined with measurements on smaller spatial scales, such as from galaxy redshift surveys and the Lyman alpha forest statistics. Thus, the σ_8 -determination from cosmic shear is certainly competitive with other measurements. Arguably, cosmic shear sticks out in this set of smaller-scale constraints due to the fewer physical assumptions needed for its interpretation.

But more importantly, it provides a fully independent method to measure cosmological parameters. Hence, at present the largest role of the cosmic shear results is that it provides an independent approach to determining these parameters; agreement with those obtained from the CMB, galaxy redshift surveys and other methods are thus foremost of interest in that they provide additional evidence for the self-consistency of our cosmological model which, taken at face value, is a pretty implausible one: we should always keep in mind that we are claiming that our Universe consists of 4.5% normal (baryonic) matter, with the rest being shared with stuff that we have given names to ('dark matter', 'dark energy'), but are pretty ignorant about what that actually is. Insofar, cosmic shear plays an essential role in shaping our cosmological view, and has become one of the pillars on which our standard model rests.

Agreement, or discrepancies? How to clarify the remaining discrepancies that were mentioned before – what are they due to? One needs to step back for a second and be amazed that these results are in fact so well in agreement as they are, given all the technical problems a cosmic shear survey has to face (see Sect. 3). Nevertheless, more investigations concerning the accuracy of the results need to be carried out, e.g., to study the influence of the different schemes for PSF corrections on the final results. For this reason, it would be very valuable if the same data set is analyzed by two independent groups and to compare the results in detail. Such comparative studies may be a prerequisite for the future when much larger surveys will turn cosmic shear into a tool for precision cosmology.

Joint constraints from CMB anisotropies and cosmic shear. As mentioned before, the full power of the CMB anisotropy measurements is achieved when these results are combined with constraints on smaller spatial scales. The tightest constraints from WMAP are obtained when it is combined with results from galaxy redshift surveys and the statistics of the Ly α forest absorption lines (Spergel et al. 2003). Instead of the latter, one can instead use results from cosmic shear, as it provides a cleaner probe of the statistical properties of the matter distribution in the Universe. As was pointed

out before (e.g., Hu & Tegmark 1999; see Fig. 34), the combination of CMB measurements with cosmic shear results is particularly powerful to break degeneracies that are left from using the former alone. Contaldi et al. (2003) used the CMB anisotropy results from WMAP (Bennett et al. 2003), supplemented by anisotropy measurements on smaller angular scales from ground-based experiments, and combined them with the cosmic shear aperture mass dispersion from the RCS survey (Hoekstra et al. 2002a). As is shown in Fig. 47, the constraints in the Ω_m - σ_8 -parameter plane are nearly mutually orthogonal for the CMB and cosmic shear, so that the combined confidence region is substantially smaller than each of the individual regions.

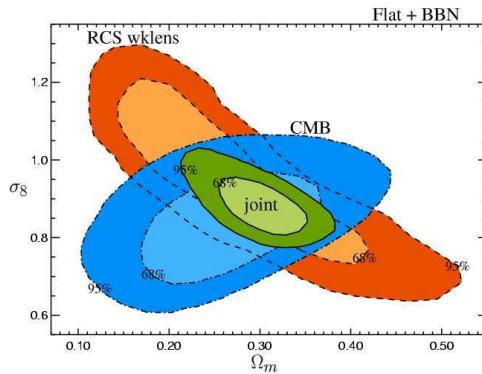


Fig. 47. The confidence region in the Ω_m - σ_8 -plane obtained from the two-dimensional marginalized likelihood. Shown are the 68% and 95% confidence regions derived individually from the CMB and the RCS cosmic shear survey, as well as those obtained by combining both constraints (Contaldi et al. 2003)

Wide vs. deep surveys. In designing future cosmic shear surveys, the survey strategy needs to decide the effective exposure time. For a given total observing time (the most important practical constraint), one needs to find a compromise between depth and area. Several issues need to be considered in this respect:

- The lensing signal increases with redshift, and therefore with increasing depth of a survey; it should therefore be easier to detect a lensing signal in deep surveys. Furthermore, by splitting the galaxy sample into subsamples according to the magnitude (and/or colors), one can study the dependence of the lensing signal on the mean source redshift, which is an important probe of the evolution of the matter power spectrum, and thus of cosmology. If one wants to probe the (dark) matter distribution at appreciable redshifts ($z \sim 0.5$), one needs to carry out deep surveys.
- A wider survey is more likely to probe the linear part of the power spectrum which is more securely predicted from cosmological models than the non-linear part; on the other hand, measurement of the latter, when compared with precise models (e.g., from numerical simulations), can probe the non-linear gravitational clustering regime.

- Depending on the intrinsic galaxy alignment, one would prefer deeper surveys, since the relative importance of the intrinsic signal decreases with increasing survey depth. Very shallow surveys may in fact be strongly affected by the intrinsic signal (e.g., Heymans & Heavens 2003). On the other hand, for precision measurements, as will become available in the near future, one needs to account for the intrinsic signal in any case, using redshift information (at least in a statistical sense), and so shallow surveys lose this potential disadvantage. In fact, the redshift estimates of shallower surveys are easier to obtain than for deeper ones.
- In this context, one needs to compromise between area and the number of filters in which exposures should be taken. Smaller area means worse statistics, e.g., larger effects of cosmic variance, but this has to be balanced against the additional redshift information. Also, if a fixed observing time is used, one needs to account for the weather, seeing and sky brightness distribution. One should then device a strategy that the best seeing periods are used to obtain images in the filter which is used for shape measurements, and bright time shall be spent on the longest wavelength bands.
- Fainter galaxies are smaller, and thus more strongly affected by the point-spread function. One therefore expects that PSF corrections are on average smaller for a shallow survey than for a deeper one. In addition, the separation between stars and galaxies is easier for brighter (hence, larger) objects.

The relative weight of these arguments is still to be decided. Whereas some of the issues could be clarified with theoretical investigations (i.e., in order to obtain the tightest constraints on cosmological parameters, what is the optimal choice of area and exposure time, with their product being fixed), others (like the importance of intrinsic alignments) still remain unclear. Since big imaging surveys will be conducted with a broad range of scientific applications in mind, this choice will also depend on those additional science goals.

Future surveys. We are currently witnessing the installment of square-degree cameras at some of the best sites, among them Megacam at the CFHT, and OmegaCAM at the newly built VLT Survey Telescope (the 2.6m VST) on Paranal (I present here European-biased prospects, as I am most familiar with these projects). Weak lensing, and in particular cosmic shear has been one of the science drivers for these instruments, and large surveys will be carried out with them. Already ongoing is the CFHT Legacy Survey, which will consist of three parts; the most interesting one in the current context is a $\sim 160 \text{ deg}^2$ survey with an exposure time of $\sim 1 \text{ h}$ in each of five optical filters. This survey will therefore yield a more than ten-fold increase over the current VIRIMOS-DESCART survey, with corresponding reductions of the statistical and cosmic variance errors on measurements. The multi-color

nature of this survey implies that one can obtain photometric redshift estimates at least for a part of the galaxies which will enable the suppression of the potential contribution to the shear signal from intrinsic alignments of galaxies. A forecast of the expected accuracy of cosmological parameter estimates from the CFHTLS combined with the WMAP CMB measurements has been obtained by Tereno et al. (2004). It is expected that a substantial fraction of the VST observing time will be spent on multi-band wide-field surveys which, if properly designed, will be extremely useful for cosmic shear research. In order to complement results from the CFHTLS, accounting for the fact that the VST has smaller aperture than the CFHT (2.6m vs. 3.6m), a somewhat shallower but wider-field survey would be most reasonable. For both of these surveys, complementary near-IR data will become available after about 2007, with the WirCam instrument on CFHT, and the newly build VISTA 4m-telescope equipped with a wide-field near-IR camera on Paranal, which will yield much better photometric redshift estimates than the optical data alone. Furthermore, with the PanStarrs project, a novel method for wide-field imaging and a great leap forward in the data access rate will be achieved.

Towards the end of the decade, a new generation of cosmic shear surveys may be started; there are two projects currently under debate which would provide a giant leap forward in terms of survey area and/or depth. One is a satellite project, SNAP/JDEM, originally designed for finding and follow-up of high-redshift supernovae to study the expansion history of the Universe and in particular to learn about the equation of state of the dark energy. With its large CCD array and multi-band imaging, SNAP will also be a wonderful instrument for cosmic shear research, yielding photometric redshift estimates for the faint background galaxies, and it is expected that the observing time of this satellite mission will be split between these two probes of dark energy. The other project under discussion is the LSST, a 8m telescope equipped with a $\sim 9 \text{ deg}^2$ camera; such an instrument, with an efficiency larger than a factor 40 over Megacam@CFHT, would allow huge cosmic shear surveys, easily obtaining a multi-band survey over all extragalactic sky (modulo the constraints from the hemisphere). Since studying the equation of state of dark energy will be done most effectively with good photometric redshifts of source galaxies, the space experiment may appear more promising, given the fact that near-IR photometry is needed for a reliable redshift estimate, and sufficiently deep near-IR observations over a significant area of sky is not possible from the ground.

8 The mass of, and associated with galaxies

8.1 Introduction

Whereas galaxies are not massive enough to show a weak lensing signal individually (see eq. 19), the signal of many galaxies can be superposed statis-

tically. Therefore, if one considers sets of foreground (lens) and background galaxies, then in the mean, in a foreground-background galaxy pair, the image ellipticity of the background galaxy will be preferentially oriented in the direction tangent to the line connecting foreground and background galaxy. The amplitude of this tangential alignment then yields a mean lensing strength that depends on the redshift distributions of foreground and background galaxies, and on the mass distribution of the former population. This effect is called galaxy-galaxy lensing and will be described in Sect. 8.2 below; it measures the mass properties of galaxies, provided the lensing signal is dominated by the galaxies themselves. This will not be the case for larger angular separations between foreground and background galaxies, since then the mass distribution in which the foreground galaxies are embedded (e.g., their host groups or clusters) starts to contribute significantly to the shear signal. The interpretation of this signal then becomes more difficult. On even larger scales, the foreground galaxies contribute negligibly to the lens signal; a spatial correlation between the lens strength and the foreground galaxy population then reveals the correlation between light (galaxies) and mass in the Universe. This correlated distribution of galaxies with respect to the underlying (dark) matter in the Universe – often called the bias of galaxies – can be studied with weak lensing, as we shall describe in Sect. 8.3 by using the shear signal, and in Sect. 8.4 employing the magnification effect. It should be pointed out here that our lack of knowledge about the relation between the spatial distribution of galaxies and that of the underlying (dark) matter is one of the major problems that hampers the quantitative interpretation of galaxy redshift surveys; hence, these lensing studies can provide highly valuable input into the conclusions drawn from these redshift surveys regarding the statistical properties of the mass distribution in the Universe.

8.2 Galaxy-galaxy lensing

The average mass profile of galaxies. Probing the mass distribution of galaxies usually proceeds with dynamical studies of luminous tracers. The best-known method is the determination of the rotation curves of spiral galaxies, measuring the rotational velocity of stars and gas as a function of distance from the galaxy's center (see Sofue & Rubin 2001 for a recent review). This then yields the mass profile of the galaxy, i.e. $M(\leq r) \propto v_{\text{rot}}^2(r) r$. For elliptical galaxies, the dynamics of stars (like velocity dispersions and higher-order moments of their velocity distribution, as a function of r) is analyzed to obtain their mass profiles; as the kinematics of stars in ellipticals is more complicated than in spirals, their mass profiles are more difficult to measure (e.g., Gerhard et al. 2001). In both cases, these dynamical methods provided unambiguous evidence for the presence of a dark matter halo in which the luminous galaxy is embedded; e.g., the rotation curves of spirals are flat out to the most distant point where they can be measured. The lack of stars or

gas prevents the measurement of the mass profile to radii beyond the luminous extent of galaxies, that is beyond $\sim 10h^{-1}$ kpc. Other luminous tracers that have been employed to study galaxy masses at larger radii include globular clusters that are found at large galacto-centric radii (Coté et al. 2003), planetary nebulae, and satellite galaxies. Determining the relative radial velocity distribution of the latter with respect to their suspected host galaxy leads to estimates of the dark matter halo out to distances of $\sim 100h^{-1}$ kpc. These studies (e.g., Zaritsky et al. 1997) have shown that the dark matter halo extends out to at least these distances.

One of the open questions regarding the dark matter profile of galaxies is the spatial extent of the halos. The dynamical studies mentioned above are all compatible with the mass profile following approximately an isothermal law ($\rho \propto r^{-2}$), which has to be truncated at a finite radius to yield a finite total mass. Over the limited range in radii, the isothermal profile cannot easily be distinguished from an NFW mass profile (see IN, Sect. 6.2), for which measurements at larger distances are needed (the mass distribution in the central parts of galaxies is affected by the baryons and thus not expected to follow the NFW profile; see Sect. 7 of SL).

Weak gravitational lensing provides a possibility to study the mass profiles of galaxies at still larger radii. Light bundles from distant background galaxies provide the ‘dynamical tracers’ that cannot be found physically associated with the galaxies. Light bundles get distorted in such a way that on average, images of background sources are oriented tangent to the transverse direction connecting foreground (lens) and background (source) galaxy. The first attempt to detect such a galaxy-galaxy lensing signal was reported in Tyson et al. (1984), but the use of photographic plates and the relatively poor seeing prevented a detection. Brainerd et al. (1996) presented the first detection and analysis of galaxy-galaxy lensing. Since then, quite a number of surveys have measured this effect, some of them using millions of galaxies.

Strategy. Consider pairs of fore- and background galaxies, with separation in a given angular separation bin. The expected lensing signal is seen as a statistical tangential alignment of background galaxy images with respect to foreground galaxies. For example, if ϕ is the angle between the major axis of the background galaxy and the connecting line, values $\pi/4 \leq \phi \leq \pi/2$ should be slightly more frequent than $0 \leq \phi \leq \pi/4$ (see Fig. 48). Using the fact that the intrinsic orientations of background galaxies are distributed isotropically, one can show (Brainerd et al. 1996) that

$$p(\phi) = \frac{2}{\pi} \left[1 - \gamma_t \left\langle \frac{1}{|\epsilon^s|} \right\rangle \cos(2\phi) \right], \quad (133)$$

where $\phi \in [0, \pi/2]$ and γ_t is the mean tangential shear in the angular bin chosen. Thus, the amplitude of the cos-wave yields the (average) strength of the shear.

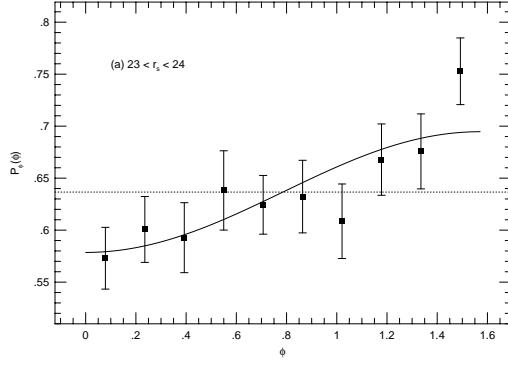


Fig. 48. The probability distribution $p(\phi)$ of the angle ϕ between the major axis of the background galaxy image and the connecting line to the foreground galaxy is plotted for the sample of Brainerd et al. (1996), together with the best fit according to (133). The galaxy pairs have separation $5'' \leq \Delta\theta \leq 34''$, and are foreground-background selected by their apparent magnitudes.

The mean tangential ellipticity $\langle \epsilon_t(\theta) \rangle$ of background galaxies relative to the direction towards foreground galaxies measures the mean tangential shear at separation θ . Since the signal is averaged over many foreground–background pairs, it measures the average mass profiles of the foreground galaxies. For sufficiently large samples of galaxies, the lens sample can be split into several subsamples, e.g., according to their color and/or morphology (early-type vs. late-type galaxies), or, if redshift estimates are available, they can be binned according to their luminosity. Then, the mass properties can be derived for each of the subsamples.

The distinction between foreground and background galaxies is ideally performed using redshift information. This is indeed the case for the galaxy-galaxy lensing studies based on the Sloan Digital Sky Survey, for which early results have been reported by McKay et al. (2001); all lens galaxies used there have spectroscopic redshifts, whereas the source galaxies are substantially fainter than the lens galaxies so that they can be considered as a background population. For other surveys, the lack of redshift information requires the separation of galaxies to be based solely on their apparent magnitudes: fainter galaxies are on average at larger distances than brighter ones. However, the resulting samples of ‘foreground’ and ‘background’ galaxies will have (often substantial) overlap in redshift, which needs to be accounted for statistically in the quantitative analysis of these surveys.

Quantitative analysis. The measurement of the galaxy-galaxy lensing signal provides the tangential shear as a function of pair separation, $\gamma_t(\theta)$. Without information about the redshifts of individual galaxies, the separation of galaxies into a ‘foreground’ and ‘background’ population has to be based on apparent magnitudes only. In the ideal case of a huge number of foreground galaxies, one could investigate the mass properties of ‘equal’ galaxies, by finely binning them according to redshift, luminosity, color, morphology etc. However, in the real world such a fine binning has not yet been possible, and therefore, to convert the lensing signal into physical parameters of the lens, a

parameterization of the lens population is needed. We shall outline here how such an analysis is performed.

The first ingredient is the redshift probability distribution $p(z|m)$ of galaxies with apparent magnitude m which is assumed to be known from redshift surveys (and/or their extrapolation to fainter magnitudes). This probability density depends on the apparent magnitude m , with a broader distribution and larger mean redshift expected for fainter m . Since the distribution of ‘foreground’ and ‘background’ galaxies in redshift is known for a given survey, the probabilities $p(z|m)$ can be employed to calculate the value of $D_{\text{ds}}/D_{\text{s}}$, averaged over all foreground–background pairs (with this ratio being set to zero if $z_{\text{s}} \leq z_{\text{d}}$). For given physical parameters of the lenses, the shear signal is proportional to this mean distance ratio.

The mass profiles of galaxies are parameterized according to their luminosity. For example, a popular parameterization is that of a truncated isothermal sphere, where the parameters are the line-of-sight velocity dispersion σ (or the equivalent circular velocity $V_c = \sqrt{2}\sigma$) and a truncation radius s at which the $\rho \propto r^{-2}$ isothermal density profile turns into a steeper $\rho \propto r^{-4}$ law. The velocity dispersion is certainly dependent on the luminosity, as follows from the Tully-Fisher and Faber-Jackson relations for late- and early-type galaxies, respectively. One therefore assumes the scaling $\sigma = \sigma_*(L/L_*)^{\beta/2}$, where L_* is a fiducial luminosity (and which conveniently can be chosen close to the characteristic luminosity of the Schechter luminosity function). Furthermore, the truncation scale s is assumed to follow the scaling $s = s_*(L/L_*)^{\eta}$. The total mass of a galaxy then is $M \propto \sigma^2 s$, or $M = M_*(L/L_*)^{\beta+\eta}$.

Suppose m and z were given; then, the luminosity of galaxy would be known, and for given values of the parameters σ_* , s_* , β and η , the mass properties of the lens galaxy would be determined. However, since z is not known, but only its probability distribution, only the probability distribution of the lens luminosities, and therefore the mass properties, are known. One could in principle determine the expected shear signal $\gamma_t(\theta)$ for a given survey by calculating the shear signal for a given set of redshifts z_i for all lens and source galaxies, and then averaging this signal over the z_i using the redshift probability distribution $p(z_i|m_i)$. However, this very-high dimensional integration cannot be performed; instead, one uses a Monte-Carlo integration method (Schneider & Rix 1997): Given the positions $\boldsymbol{\theta}_i$ and magnitudes m_i of the galaxies, one can draw for each of them a redshift according to $p(z_i|m_i)$, and then calculate the shear at all positions $\boldsymbol{\theta}_i$ corresponding to a source galaxy, for each set of parameters σ_* , s_* , β and η . This procedure can be repeated several times, yielding the expected shear $\langle \gamma_i \rangle$ and its dispersion $\sigma_{\gamma,i}$ for each source galaxy’s position. One can then calculate the likelihood function

$$\mathcal{L} = \prod_{i=1}^{N_s} \frac{1}{\pi(\sigma_\epsilon^2 + \sigma_{\gamma,i}^2)} \exp\left(-\frac{|\epsilon_i - \langle \gamma_i \rangle|^2}{\sigma_\epsilon^2 + \sigma_{\gamma,i}^2}\right), \quad (134)$$

where σ_ϵ is the intrinsic ellipticity dispersion of the galaxies. \mathcal{L} depends on the parameters of the model, and can be maximized with respect to them, thereby yielding estimates of σ_* , s_* , β and η .

First detection The galaxy-galaxy lensing effect was first found by Brainerd et al. (1996), on a single 9.6×9.6 field. They considered ‘foreground’ galaxies in the magnitude range $m \in [20, 23]$, and ‘background’ galaxies with $m \in [23, 24]$; this yielded 439 foreground and 506 background galaxies, and 3202 pairs with $\Delta\theta \in [5'', 34'']$.¹² For these pairs, the distribution of the alignment angle ϕ is plotted in Fig. 48. This distribution clearly is incompatible with the absence of a lens signal (at the 99.9% confidence level), and thus provides a solid detection.

They analyzed the lens signal $\gamma_t(\theta)$ in a way similar to the method outlined above, except that their Monte-Carlo simulations also randomized the positions of galaxies. The resulting likelihood yields $\sigma_* \approx 160_{-60}^{+50}$ km/s (90% confidence interval), whereas for s_* only a lower limit of $25h^{-1}$ kpc (1σ) is obtained; the small field size, in combination with the relative insensitivity of the lensing signal to s_* once this value is larger than the mean transverse separation of lensing galaxies, prohibited the detection of an upper bound on the halo size.

Galaxy-galaxy lensing from the Red-Sequence Cluster Survey (RCS). Several groups have published results of their galaxy-galaxy lensing surveys since its first detection. Here we shall describe the results of a recent wide-field imaging survey, the RCS; this survey was already described in the context of cosmic shear in Sect. 7.3. 45.5 square degrees of single-band imaging data were used (Hoekstra et al. 2004). Choosing lens galaxies with $19.5 \leq R_C \leq 21$, and source galaxies having $21.5 \leq R_C \leq 24$ yielded $\sim 1.2 \times 10^5$ lenses with median redshift of 0.35 and $\sim 1.5 \times 10^6$ sources with median redshift of ~ 0.53 , yielding $\langle D_{ds}/D_s \rangle = 0.29 \pm 0.01$ for the full sample of lenses and sources. Fig. 49 shows the shear signal for this survey.

The lens signal is affected by galaxies counted as lenses, but which in fact are in the foreground. As long as they are not physically associated with lens galaxies, this effect is accounted for in the analysis, i.e., in the value of $\langle D_{ds}/D_s \rangle$. However, if fainter galaxies cluster around lens galaxies, this produces an additional effect. Provided the orientation of the associated faint galaxies are random with respect to the separation vector to their bright neighbor, these physical pairs just yield a dilution of the shear signal. The amplitude of this effect can be determined from the angular correlation function of bright and faint galaxies, and easily corrected for. Once this has been

¹² The lower angular scale has been chosen to avoid overlapping isophotes of foreground and background galaxies, whereas the upper limit was selected since it gave the largest signal-to-noise for the deviation of the angular distribution shown in Fig. 48 from a uniform one.

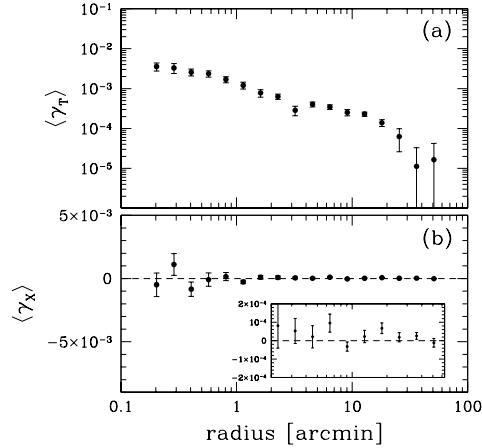


Fig. 49. (a) Tangential shear as a function of angular separation, obtained from the RCS survey; the shear signal is detected out to nearly one degree scale. (b) Cross shear signal, which is expected to vanish identically in the absence of systematic effects on the ellipticity measurements. As can be seen, the cross signal is indeed compatible with zero. The inset expands the scale, to better show the error bars (from Hoekstra et al. 2003)

done, the corrected shear signal within $10'' \leq \theta \leq 2'$ has been fitted with an SIS model, yielding a mean velocity dispersion of the lens galaxies of $\sqrt{\langle \sigma^2 \rangle} = 128 \pm 4$ km/s. If the scaling relations between galaxy luminosity and velocity dispersion as described above is employed, with $\beta = 0.6$, the result is $\sigma_* = 140 \pm 4$ km/s for $L_* = 10^{10} h^{-2} L_\odot$ in the blue passband.

To interpret the shear results on larger angular scales, the SIS model no longer suffices, and different mass models need to be employed. Using a truncated isothermal model, the best-fitting values of the scaling parameters $\beta = 0.60 \pm 0.11$ and $\eta = 0.24^{+0.26}_{-0.22}$ are obtained, when marginalizing over all other parameters. Furthermore, $\sigma_* = 137 \pm 5$ km/s, in very close agreement with the results from small θ and the SIS model; this is expected, since most of the signal comes from these smaller separations. Most interesting, the analysis also yields an estimate of the truncation scale of $s_* = (185 \pm 30)h^{-1}$ kpc, providing one of only a few estimates of the scale of the dark matter halo. Hoekstra et al. also performed the analysis in the frame of an NFW mass model.

These results can then be used to calculate the mass-to-light ratio of an L_* galaxy and, using the scaling, of the galaxy population as a whole. Considering only galaxies with $M \geq 10^{10} h^{-1} M_\odot$, the mean mass-to-light ratio inside the virial radius of galaxy halos is about 100 in solar units.

The shape of dark matter halos. In the mass models considered before, the mass distribution of galaxies was assumed to be axi-symmetric. In fact, this assumption is not crucial, since the relation between shear and surface mass density, $\gamma_t(\vartheta) = \bar{\kappa}(\vartheta) - \kappa(\vartheta)$ is true for a general mass distribution, provided γ_t and $\kappa(\vartheta)$ are interpreted as the mean tangential shear and mean surface mass density on a circle of radius ϑ , and $\bar{\kappa}(\vartheta)$ as the mean surface

mass density inside this circle (see eq. 24). However, deviations from axial symmetry are imprinted on the shear signal and can in principle be measured. If the mass distribution is ‘elliptical’, the shear along the major axis (at given distance ϑ) is larger than that along the minor axis, and therefore, an investigation of the strength of the shear signal relative to the orientation of the galaxy can reveal a finite ellipticity of the mass distribution. For that, it is necessary that the orientation of the mass distribution is (at least approximately) known. Provided the orientation of the mass distribution follows approximately the orientation of the luminous part of galaxies, one can analyze the direction dependence of the shear relative to the major axis of the light distribution (Natarajan & Refregier 2000). Hoekstra et al. (2002b) have used the RCS to search for such a direction dependence; they parameterized the lenses with a truncated isothermal profile with ellipticity $\epsilon_{\text{mass}} = f\epsilon_{\text{light}}$, where f is a free parameter. The result $f = 0.77 \pm 0.2$ indicates first that the mass distribution of galaxies is not round (which would be the case for $f = 0$, which is incompatible with the data), and second, that the mass distribution is rounder than that of the light distribution, since $f < 1$. However, it must be kept in mind that the assumption of equal orientation between light and mass is crucial for the interpretation of f ; misalignment causes a decrease of f . Note that numerical simulations of galaxy evolution predict such a misalignment between total mass and baryons, with an rms deviation of around 20° (van den Bosch et al. 2002). Given the above result on f , it is therefore not excluded that the flattening of halos is very similar to that of the light. Also note that this result yields a value averaged over all galaxies; since the lens efficiency of elliptical galaxies (at given luminosity) is larger than that of spirals, the value of f is dominated by the contributions from early-type galaxies.

Results from the Sloan Survey. The Sloan Digital Sky Survey (e.g., York et al. 2000) will map a quarter of the sky in five photometric bands, and obtain spectra of about one million galaxies. A large fraction of the data has already been taken by SDSS, and parts of this data have already been released (Abazajian et al. 2004). The huge amount of photometric data in principle is ideal for weak lensing studies, as it beats down statistical uncertainties to an unprecedented low level. However, the site of the telescope, the relatively large pixel size of $0''.4$, the relatively shallow exposures of about one minute and the drift-scan mode in which data are taken (yielding excellent flat-fielding, and thus photometric properties, somewhat at the expense of the shape of the PSF) render the data less useful for, e.g., cosmic shear studies: the small mean redshift of the galaxies yields a very small expectation value of the cosmic shear, which can easily be mimicked by residuals from PSF corrections. However, galaxy-galaxy lensing is much less sensitive to larger-scale PSF problems, since the component of the shear used in the analysis is not attached to pixel directions, but to neighboring galaxies, and thus

varies rapidly with sky position. Another way of expressing this fact is that the galaxy-galaxy lensing signal would remain unchanged if a uniform shear would be added to the data; therefore, SDSS provides an great opportunity for studying the mass profile of galaxies.

Fischer et al. (2000) reported the first results from the SDSS, and a larger fraction of the SDSS data was subsequently used in a galaxy-galaxy lensing study by McKay et al. (2001), where also the spectroscopic redshifts of the lens galaxies were used. Their sample consists of ~ 31000 lens galaxies with measured redshifts, and $\sim 3.6 \times 10^6$ source galaxies selected in the brightness range $18 \leq r \leq 22$. For this magnitude range, the redshift distribution of galaxies is fairly well known, leaving little calibration uncertainty in the interpretation of the shear signal. In particular, there is very little overlap in the redshift distribution of source and lens galaxies. The data set has been subjected to a large number of tests, to reveal systematics; e.g., null results are obtained when the source galaxies are rotated by 45° (or, equivalently, if γ_x is used instead of γ_t), or if the lens galaxies are replaced by an equal number of randomly distributed points relative to which the tangential shear component is measured. Since the redshifts of the lens galaxies are known, the shear can be measured directly in physical units, so one can determine

$$\Delta\Sigma_+ = \bar{\Sigma}(\leq R) - \Sigma(R) \quad (135)$$

in M_\odot/pc^2 as a function of R in kpc.

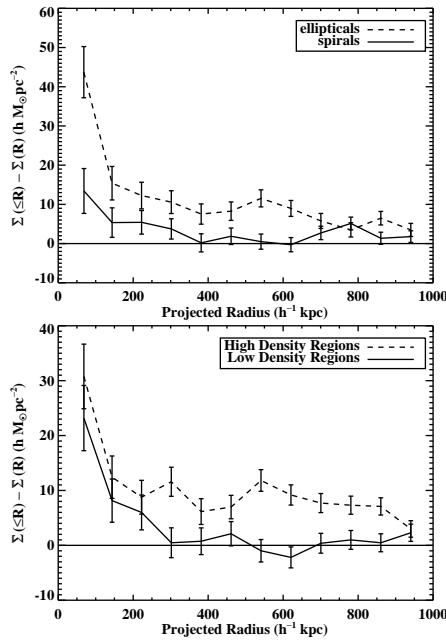


Fig. 50. The galaxy-galaxy lensing signal from the SDSS plotted against physical radius R . The lens sample has been subdivided into early- and late-type galaxies (upper panel), and in galaxies situated in dense environments vs. those with a smaller neighboring galaxy density (lower panel). The figure clearly shows that the lensing signal is dominated by elliptical galaxies, and by those located in dense environment. Owing to the morphology-density relation of galaxies, these two results are not mutually independent. Note that the lensing signal can be measured out to $1h^{-1}$ Mpc, considerably larger than the expected size of galaxy halos; therefore, the shear at these large separations is most likely caused by the larger-scale mass distribution in which the galaxies are embedded (from McKay et al. 2001)

Fig. 50 shows the lensing result from McKay et al. (2001), where the lens sample has been split according to the type of galaxy (early vs. late type) and according to the local spatial number density of galaxies, which is known owing to the spectroscopic redshifts. The fact that most of the signal on small scales is due to ellipticals is expected, as they are more massive at given luminosity than spirals. The large spatial extent of the shear signal for ellipticals relative to that of spirals can be interpreted either by ellipticals having a larger halo than spirals, or that ellipticals are preferentially found in high-density environments, which contribute to the lens signal on large scales. This latter interpretation is supported by the lower panel in Fig. 50 which shows that the signal on large scales is entirely due to lens galaxies in dense environments. This then implies that the galaxy-galaxy lensing signal on large scales no longer measures the density profile of individual galaxies, but gets more and more dominated by group and cluster halos in which these (predominantly early-type) galaxies are embedded.

A separation of these contributions from the data themselves is not possible at present, but can be achieved in the frame of a theoretical model. Guzik & Seljak (2001) employed the halo model for the distribution of matter in the universe (see Cooray & Sheth 2002) to perform this separation. There, the galaxy-galaxy lensing signal either comes from matter in the same halo in which the galaxy is embedded, or due to other halos which are physically associated (i.e., clustered) with the former. This latter contribution is negligible on the scales below $\sim 1h^{-1}$ Mpc on which the SDSS obtained a measurement. The former contribution can be split further into two terms: the first is from the dark matter around the galaxies themselves, whereas the second is due to the matter in groups and clusters to which the galaxies might belong. The relative amplitude of these two terms depends on the fraction of galaxies which are located in groups and clusters; the larger this fraction, the more important are larger-scale halos for the shear signal. Guzik & Seljak estimate from the radial dependence of the SDSS signal that about 20% of galaxies reside in groups and clusters; on scales larger than about $200h^{-1}$ kpc their contribution dominates. The virial mass of an early-type L_* galaxy is estimated to be $M_{200}(L_*) = (9.3 \pm 2.2) \times 10^{11} h^{-1} M_\odot$, and about a factor of three smaller for late-type galaxies (with luminosity measured in a red passband; the differences are substantially larger for bluer passbands, owing to the sensitivity of the luminosity to star formation activity in late types). From the mass-to-light ratio in red passbands, Guzik & Seljak estimate that an L_* galaxy converts about 10–15% of its virial mass into stars. Since this fraction is close to the baryon fraction in the universe, they conclude that most of the baryons of an L_* galaxy are transformed into stars. For more massive halos, the mass-to-light ratio increases ($M/L \propto L^{0.4 \pm 0.2}$), and therefore their conversion of baryons into stars is smaller – in agreement with what we argued about clusters, where most of the baryons are present in the form of a hot intracluster gas.

Yang et al. (2003) studied the cross-correlation between mass and galaxies using numerical simulations of structure formation and semi-analytic models of galaxy evolution. The observed dependence of the galaxy-galaxy lensing signal on galaxy luminosity, morphological type and galaxy environment, as obtained by McKay et al. (2001), is well reproduced in these simulations. The galaxy-mass correlation is affected by satellite galaxies, i.e. galaxies not situated at the center of their respective halo. Central galaxies can be selected by restricting the foreground galaxy sample to relatively isolated galaxies. The galaxy-galaxy lensing signal for such central galaxies can well be described by an NFW mass profile, whereas this no longer is true if all galaxies are considered. Combining the measurement with the simulation, they find that an L_* -galaxy typically resides in a halo with a virial mass of $\sim 2 \times 10^{12} h^{-1} M_\odot$.

With the SDSS progressing, larger datasets become available, allowing a more refined analysis of galaxy-galaxy lensing (Sheldon et al. 2004; Seljak et al. 2004). In the analysis of Seljak et al. (2004), more than 2.7×10^5 galaxies with spectroscopic redshifts have been used as foreground galaxies, and as background population those fainter galaxies for which photometric redshifts have been estimated. The resulting signal is shown in Fig. 51, for six different bins in (foreground) galaxy luminosity.

In a further test to constrain systematic effects in the data, Hirata et al. (2004) have used spectroscopic and photometric redshifts to study the question whether an alignment of satellite galaxies around the lens galaxies can affect the galaxy-galaxy lensing signal from the SDSS; they obtain an upper limit of a 15% contamination.

The SDSS already has yielded important information about the mass properties of galaxies; taken into account that only a part of the data of the complete survey have been used in the studies mentioned above, an analysis of the final survey will yield rich harvest when applied to a galaxy-galaxy lensing analysis.

Lensing by galaxies in clusters. As an extension of the method presented hitherto, one might use galaxy-galaxy lensing also to specifically target the mass profile of galaxies in the inner part of clusters. One might expect that owing to tidal stripping, their dark matter halo has a considerably smaller spatial extent than that of the galaxy population as a whole. The study of this effect with lensing is more complicated than galaxy-galaxy lensing in the field, both observationally and from theory. Observationally, the data sets that can be used need to be taken in the inner part of massive clusters; since these are rare, a single wide-field image usually contains at most one such cluster. Furthermore, the number of massive galaxies projected near the center of a cluster is fairly small. Therefore, in order to obtain good statistics, the data of different clusters should be combined. Since the cores of clusters are optically bright, measuring the shape of faint background galaxies is more difficult than in a blank field. From the theoretical side, the lensing strength

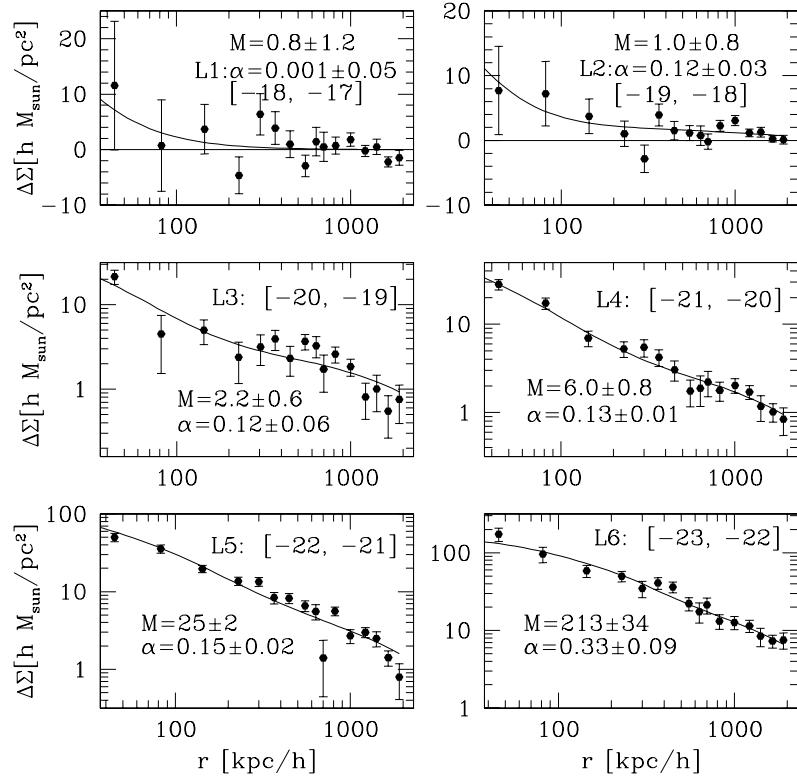


Fig. 51. The galaxy-galaxy lensing signal for six luminosity bins of foreground galaxies, as indicated by the absolute magnitude interval in each panel. The curves show a two-parameter model fitted to the data, based on the halo model, and the fit parameters are indicated: M is the virial mass of the halo (in units of $10^{11} h^{-1} M_{\odot}$) in which the galaxies reside, and α is the fraction of the galaxies which are not central inside the halo, but satellite galaxies (from Seljak et al. 2004)

of the cluster is much stronger than that of the individual cluster galaxies, and so this large-scale shear contribution needs to be accounted for in the galaxy-galaxy lensing analysis.

Methods for performing this separation between cluster and galaxy shear were developed by Natarajan & Kneib (1997) and Geiger & Schneider (1998). Perhaps the simplest approach is provided by the aperture mass methods, applied to the individual cluster galaxies; there one measures the tangential shear inside an annulus around each cluster galaxy. This measure is insensitive to the shear contribution which is linear in the angular variable θ , which is a first local approximation to the larger-scale shear caused by the

cluster. Alternatively, a mass model of the (smoothed) cluster can be obtained, either from strong or weak lensing constraints, or preferentially both, and subtracted from the shear signal around galaxies to see their signal. However, once the mass fraction in the galaxies becomes considerable, this method starts to become biased. Geiger & Schneider (1999) have suggested to simultaneously perform a weak lensing mass reconstruction of the cluster and a determination of the parameters of a conveniently parameterized mass model of cluster galaxies (e.g., the truncated isothermal sphere); since the maximum likelihood method for the mass reconstruction (see Sect. 5.3) was used, the solution results from maximizing the likelihood with respect to the mass profile parameters (the deflection potential on a grid) and the galaxy mass parameters.

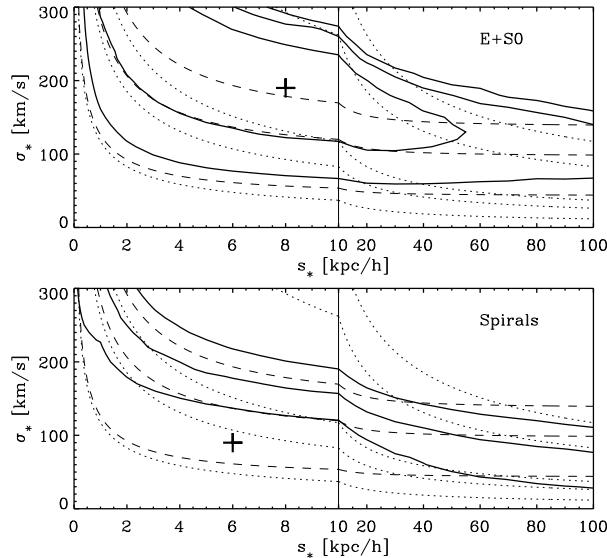


Fig. 52. Significance contours (solid) for galaxy properties obtained from galaxy-galaxy lensing of galaxies in the cluster Cl0939+4713. The parameters are the velocity dispersion σ_* and the halo truncation radius s_* of an L_* -galaxy. Based on HST data (see Fig. 22), a simultaneous reconstruction of the cluster mass profile and the determination of the galaxy mass parameters was performed. No significant lensing signal is seen from the 55 late-type galaxies (lower panel), but a clear detection and upper bound to the halo size is detected for the 56 early-types. Dashed and dotted curves connect models with the same mass inside $8h^{-1}$ kpc and total mass of an L_* -galaxy, respectively (from Geiger & Schneier 1999)

Natarajan et al. (1998), by analyzing HST data of the cluster AC114, concluded that the truncation radius of a fiducial L_* galaxy in this cluster is \sim

$15h^{-1}$ kpc; similarly, Geiger & Schneider (1999) showed that the best-fitting truncation radius for early-type galaxies in the cluster A851 is $\sim 10h^{-1}$ kpc (see Fig. 52). Although the uncertainties are fairly large, these results indicate that indeed galaxies near cluster centers have a halo size considerably smaller than the average galaxy. The sample of clusters which can be investigated using this method will dramatically increase once the cluster sample observed with the new ACS camera onboard HST becomes available and gets properly analyzed.

8.3 Galaxy biasing: shear method

On small scales, galaxy-galaxy lensing measures the mass profile of galaxies, whereas on intermediate scales the environment of galaxies starts to dominate the shear signal. On even larger scale (say, beyond $\sim 1h^{-1}$ Mpc), the host halo contribution becomes negligible. Beyond that distance, any signal must come from the correlation of galaxy positions with the mass distribution in the Universe. This correlation, and the related issue of galaxy biasing (see Sect. 6.1 of IN), can ideally be studied with weak lensing. In this section we shall outline how these quantities can be determined from shear measurements, and describe some recent results. As we shall see, this issue is intimately related to galaxy-galaxy lensing. The next section deals with the magnification of distant sources caused by mass overdensities correlated with galaxies and thereby causing an apparent correlation between high-redshift sources and low-redshift galaxies; the amplitude of this signal is again proportional to the correlation between galaxies and the underlying dark matter.

An interesting illustration of the correlation between galaxies and mass has been derived by Wilson et al. (2001). They studied 6 fields with $30' \times 30'$ each, selected bright early-type galaxies from their $V - I$ colors and I magnitudes and measured the shear from faint galaxies. Assuming that mass is strongly correlated with early-type galaxies, these can be used to predict the shear field, with an overall normalization given by the mean mass-to-light ratio of the early-type galaxies. This correlation has indeed been found, at the 5.2σ significance level, and a value of $M/L \approx 300h$ in solar units has been obtained, assuming a flat low-density Universe.

The galaxy-mass correlation and the bias parameter. First, the concept of the correlation between galaxies and mass shall be described more quantitatively. The mass density inhomogeneities are described, as before, by the dimensionless density contrast $\delta(\mathbf{x}, w)$. In analogy to this quantity, one defines the number density contrast $\delta_g(\mathbf{x}, w)$ of galaxies as

$$\delta_g(\mathbf{x}, w) := \frac{n(\mathbf{x}, w) - \bar{n}(w)}{\bar{n}(w)}, \quad (136)$$

where $n(\mathbf{x}, w)$ is the number density of galaxies at comoving position \mathbf{x} and comoving distance w (the latter providing a parameterization of cosmic time

or redshift), and $\bar{n}(w)$ is the mean number density of galaxies at that epoch. Since the galaxy distribution is discrete, the true number density is simply a sum of delta-functions. What is meant by n is that the probability of finding a galaxy in the volume dV situated at position \mathbf{x} is $n(\mathbf{x}) dV$.

The relation between δ and δ_g describes the relative distribution of galaxies and matter in the Universe. The simplest case is that of an *unbiased* distribution, for which $\delta_g = \delta$; then, the probability of finding a galaxy at any location would be just proportional to the matter density. However, one might expect that the relation between luminous and dark matter is more complicated. For example, galaxies are expected to form preferentially in the high-density peaks in the early Universe, which would imply that there are proportionally more galaxies within mass overdensities. This led to the introduction of the concept of biasing (e.g., Bardeen et al. 1986; Kaiser 1984). The simplest form of biasing, called linear deterministic biasing, is provided by setting $\delta_g = b\delta$, with b being the bias parameter. One might suspect that the relative bias is approximately constant on large scales, where the density field is still in its linear evolution (i.e., on scales $\gtrsim 10h^{-1}$ Mpc today). On smaller scales, however, b most likely is no longer simply a constant. For example, the spatial distribution of galaxies in clusters seems to deviate from the radial mass profile, and the distributions of different galaxy types are different. Furthermore, by comparing the clustering properties of galaxies of different types, one can determine their relative bias, from which it is concluded that more luminous galaxies are more strongly biased than less luminous ones, and early-type galaxies are more strongly clustered than late-types (see Norberg et al. 2001 and Zehavi et al. 2002 for recent results from the 2dFGRS and the SDSS). This is also expected from theoretical models and numerical simulations which show that more massive halos cluster more strongly (e.g., Sheth et al. 2001; Jing 1998). In order to account for a possible scale dependence of the bias, one considers the Fourier transforms of δ and δ_g and relates them according to

$$\hat{\delta}_g(\mathbf{k}, w) = b(|\mathbf{k}|, w) \hat{\delta}(\mathbf{k}, w), \quad (137)$$

thus accounting for a possible scale and redshift dependence of the bias.

Even this more general bias description is most likely too simple, as it is still deterministic. Owing to the complexity of galaxy formation and evolution, it is to be expected that the galaxy distribution is subject to stochasticity in excess to Poisson sampling (Tegmark & Peebles 1998; Dekel & Lahav 1999). To account for that, another parameter is introduced, the correlation parameter $r(|\mathbf{k}|, w)$, which in general will also depend on scale and cosmic epoch. To define it, we first consider the correlator

$$\langle \hat{\delta}(\mathbf{k}, w) \hat{\delta}_g^*(\mathbf{k}', w) \rangle = (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}') P_{\delta g}(|\mathbf{k}|, w), \quad (138)$$

where the occurrence of the delta function is due to the statistical homogeneity of the density fields, and $P_{\delta g}$ denotes the cross-power between galaxies

and matter. The correlation parameter r is then defined as

$$r(|\mathbf{k}|, w) = \frac{P_{\delta g}(|\mathbf{k}|, w)}{\sqrt{P_\delta(|\mathbf{k}|, w) P_g(|\mathbf{k}|, w)}}. \quad (139)$$

In the case of *stochastic biasing*, the definition of the bias parameter is modified to

$$P_g(|\mathbf{k}|, w) = b^2(|\mathbf{k}|, w) P_\delta(|\mathbf{k}|, w), \quad (140)$$

which agrees with the definition (137) in the case of $r \equiv 1$, but is more general since (140) no longer relates the phase of (the Fourier transform of) δ_g to that of δ . Combining the last two equations yields

$$P_{\delta g}(|\mathbf{k}|, w) = b(|\mathbf{k}|, w) r(|\mathbf{k}|, w) P_\delta(|\mathbf{k}|, w). \quad (141)$$

We point out again that galaxy redshift surveys are used to determine the two-point statistics of the galaxy distribution, and therefore P_g ; in order to relate these measurements to P_δ , assumptions on the properties of the bias have to be made. As we shall discuss next, weak lensing can determine both the bias parameter and the correlation parameter.

The principle. In order to determine b and r , the three power spectra defined above (or functions thereof) need to be measured. Second-order cosmic shear measures, as discussed in Sect. 6, are proportional to the power spectrum P_δ . The correlation function of galaxies is linearly related to P_g . In particular, the three-dimensional correlation function is just the Fourier transform of P_g , whereas the angular correlation function contains a projection of P_g along the line-of-sight and thus follows from Limber's equation as discussed in Sect. 6.2. Finally, the cross-power $P_{\delta g}$ describes the correlation between mass and light, and thus determines the relation between the lensing properties of the mass distribution in the Universe to the location of the galaxies. Galaxy-galaxy lensing on large angular scales (where the mass profile of individual galaxies no longer yields a significant contribution) provides one of the measures for such a correlation. Hence, measurements of these three statistical distributions allow a determination of r and b .

As we shall consider projected densities, we relate the density field of galaxies on the sky to the spatial distribution. Hence, consider a population of ('foreground') galaxies with spatial number density $n(\mathbf{x}, w)$. The number density of these galaxies on the sky at $\boldsymbol{\theta}$ is then $N(\boldsymbol{\theta}) = \int dw \nu(w) n(f_k(w)\boldsymbol{\theta}, w)$, where $\nu(w)$ is the redshift-dependent selection function, describing which fraction of the galaxies at comoving distance w are included in the sample. Foremost, this accounts for the fact that for large distances, only the more luminous galaxies will be in the observed galaxy sample, but ν can account also for more subtle effects, such as spectral features entering or leaving the photometric bands due to redshifting. The mean number density of galaxies on the sky is $\bar{N} = \int dw \nu(w) \bar{n}(w)$; the redshift distribution, or more

precisely, the distribution in comoving distance, of these galaxies therefore is $p_f(w) = \nu(w) \bar{n}(w)/\bar{N}$, thus relating the selection function $\nu(w)$ to the redshift distribution. Using the definition (136), one then finds that

$$N(\boldsymbol{\theta}) = \bar{N} \left[1 + \int dw p_f(w) \delta_g(f_K(w)\boldsymbol{\theta}, w) \right]. \quad (142)$$

We shall denote the fractional number density by $\kappa_g(\boldsymbol{\theta}) := [N(\boldsymbol{\theta}) - \bar{N}] / \bar{N} = \int dw p_f(w) \delta_g(f_K(w)\boldsymbol{\theta}, w)$.

Aperture measures. We have seen in Sect. 6.3 that the aperture mass dispersion provides a very convenient measure of second-order cosmic shear statistics. Therefore, it is tempting to use aperture measures also for the determination of the bias and the mass-galaxy correlation. Define in analogy to the definition of the aperture mass M_{ap} in terms of the projected mass density the aperture counts (Schneider 1998),

$$\mathcal{N}(\theta) = \int d^2\vartheta U(|\boldsymbol{\vartheta}|) \kappa_g(\boldsymbol{\vartheta}), \quad (143)$$

where the integral extends over the aperture of angular radius θ , and $\boldsymbol{\vartheta}$ measures the position relative to the center of the aperture. An unbiased estimate of the aperture counts is $\bar{N}^{-1} \sum_i U(|\boldsymbol{\theta}_i|)$, where the $\boldsymbol{\theta}_i$ are the positions of the galaxies. We now consider the dispersion of the aperture counts,

$$\langle \mathcal{N}^2(\theta) \rangle = \int d^2\vartheta U(|\boldsymbol{\vartheta}|) \int d^2\vartheta' U(|\boldsymbol{\vartheta}'|) \langle \kappa_g(\boldsymbol{\vartheta}) \kappa_g(\boldsymbol{\vartheta}') \rangle. \quad (144)$$

The correlator in the last expression is the angular two-point correlation function $\omega(\Delta\vartheta)$ of the galaxies; its Fourier transform is the angular power spectrum $P_\omega(\ell)$ of galaxies. Using the definition of κ_g together with the result (98) allows us to express P_ω in terms of the three-dimensional power spectrum of the galaxy distribution,

$$\begin{aligned} P_\omega(\ell) &= \int dw \frac{p_f^2(w)}{f_K^2(w)} b^2 \left(\frac{\ell}{f_K(w)}, w \right) P_\delta \left(\frac{\ell}{f_K(w)}, w \right) \\ &= \bar{b}^2 \int dw \frac{p_f^2(w)}{f_K^2(w)} P_\delta \left(\frac{\ell}{f_K(w)}, w \right), \end{aligned} \quad (145)$$

where we made use of (140), and in the final step we defined the mean bias parameter \bar{b} which is a weighted average of the bias parameter over the redshift distribution of the galaxies and which depends on the angular wave number ℓ . To simplify notation, we shall drop the bar on b and consider the bias factor as being conveniently averaged over redshift (and later, also over spatial scale). The aperture count dispersion then becomes

$$\langle \mathcal{N}^2(\theta) \rangle = \frac{1}{2\pi} \int d\ell \ell P_\omega(\ell) W_{ap}(\theta\ell) = 2\pi b^2 H_{gg}(\theta), \quad (146)$$

where W_{ap} is given in (109), and we have defined

$$H_{\text{gg}}(\theta) = \int dw \frac{p_f^2(w)}{f_K^2(w)} \mathcal{P}(w, \theta) , \quad (147)$$

with

$$\mathcal{P}(w, \theta) = \frac{1}{(2\pi)^2} \int d\ell \ell P_\delta \left(\frac{\ell}{f_K(w)}, w \right) W_{\text{ap}}(\theta\ell) . \quad (148)$$

Using the same notation (following Hoekstra et al. 2002c), we can write the aperture mass dispersion as

$$\langle M_{\text{ap}}^2(\theta) \rangle = \frac{9\pi}{2} \left(\frac{H_0}{c} \right)^4 \Omega_m^2 H_\kappa(\theta) , \quad (149)$$

with

$$H_\kappa(\theta) = \int dw \frac{g^2(w)}{a^2(w)} \mathcal{P}(w, \theta) , \quad (150)$$

where $g(w)$ (see eq. 94) describes the source-redshift weighted efficiency factor of a lens at distance w . One therefore obtains an expression for the bias factor,

$$b^2 = \frac{9}{4} \left(\frac{H_0}{c} \right)^4 \frac{H_\kappa(\theta)}{H_{\text{gg}}(\theta)} \Omega_m^2 \frac{\langle \mathcal{N}^2(\theta) \rangle}{\langle M_{\text{ap}}^2(\theta) \rangle} = f_b(\theta) \Omega_m^2 \frac{\langle \mathcal{N}^2(\theta) \rangle}{\langle M_{\text{ap}}^2(\theta) \rangle} . \quad (151)$$

Note that $f_b(\theta)$ depends, besides the aperture radius θ , on the cosmological parameters Ω_m and Ω_A , but for a given cosmological model, it depends only weakly on the filter scale θ and on the adopted power spectrum P_δ (van Waerbeke 1998; Hoekstra et al. 2002c). This is due to the fact that both, $\langle \mathcal{N}^2(\theta) \rangle$ and $\langle M_{\text{ap}}^2(\theta) \rangle$ are linear in the power spectrum, through the functions H , and in both cases they probe only a very narrow range of k -values, owing to the narrow width of the filter function W_{ap} . Hence, the ratio $\langle \mathcal{N}^2(\theta) \rangle / \langle M_{\text{ap}}^2(\theta) \rangle$ is expected to be very close to a constant if the bias factor b is scale independent.

Next we consider the correlation coefficient r between the dark matter distribution and the galaxy field. Correlating $M_{\text{ap}}(\theta)$ with $\mathcal{N}(\theta)$ yields

$$\begin{aligned} \langle M_{\text{ap}}(\theta) \mathcal{N}(\theta) \rangle &= \int d^2\vartheta U(|\vartheta|) \int d^2\vartheta' U(|\vartheta'|) \langle \kappa(\vartheta) \kappa_g(\vartheta') \rangle \\ &= 3\pi \left(\frac{H_0}{c} \right)^2 \Omega_m b r H_{\kappa g}(\theta) , \end{aligned} \quad (152)$$

with

$$H_{\kappa g}(\theta) = \int dw \frac{p_f(w) g(w)}{a(w) f_K(w)} \mathcal{P}(w, \theta) . \quad (153)$$

It should be noted that $\langle M_{\text{ap}}(\theta) \mathcal{N}(\theta) \rangle$ is a first-order statistics in the cosmic shear. It correlates the shear signal with the location of galaxies, which

are assumed to trace the total matter distribution. As shown in Schneider (1998), the signal-to-noise of this correlator is higher than that of $\langle M_{\text{ap}}^2 \rangle$, and therefore was introduced as a convenient statistics for the detection of cosmic shear. In fact, in their original analysis of the RCS, based on 16 deg^2 , Hoekstra et al. (2001) obtained a significant signal for $\langle M_{\text{ap}}(\theta)\mathcal{N}(\theta) \rangle$, but not for $\langle M_{\text{ap}}^2(\theta) \rangle$. Combining (146) and (149) with (152), the correlation coefficient r can be expressed as

$$r = \frac{\sqrt{H_\kappa(\theta) H_{\text{gg}}(\theta)}}{H_{\kappa g}(\theta)} \frac{\langle M_{\text{ap}}(\theta)\mathcal{N}(\theta) \rangle}{\sqrt{\langle M_{\text{ap}}^2(\theta) \rangle \langle \mathcal{N}^2(\theta) \rangle}} = f_r(\theta) \frac{\langle M_{\text{ap}}(\theta)\mathcal{N}(\theta) \rangle}{\sqrt{\langle M_{\text{ap}}^2(\theta) \rangle \langle \mathcal{N}^2(\theta) \rangle}}. \quad (154)$$

As was the case for f_b , the function f_r depends only very weakly on the filter scale and on the adopted form of the power spectrum, so that a variation of the (observable) final ratio with angular scale would indicate the scale dependence of the correlation coefficient.

Whereas the two aperture measures M_{ap} and \mathcal{N} can in principle be obtained from the data field by putting down circular apertures, and the corresponding second-order statistics can likewise be determined through unbiased estimators defined on these apertures, this is not the method of choice in practice, due to gaps and holes in the data field. Note that in our discussion of cosmic shear in Sect. 6.3, we have expressed $\langle M_{\text{ap}}^2(\theta) \rangle$ in terms of the shear two-point correlation functions $\xi_\pm(\theta)$ – see (115) – just for this reason. In close analogy, $\mathcal{N}^2(\theta)$ can be expressed in terms of the angular correlation function $\omega(\theta)$ of the projected galaxy positions, as seen by (144), or more explicitly, when replacing the power spectrum $P_\omega(\ell)$ in (146) by its Fourier transform, which is the angular correlation function, one finds

$$\langle \mathcal{N}^2(\theta) \rangle = \int_0^{2\theta} \frac{d\vartheta}{\theta^2} \omega(\vartheta) T_+ \left(\frac{\vartheta}{\theta} \right), \quad (155)$$

where the function T_+ is the same as that occurring in (115). Correspondingly, we introduce the power spectrum $P_{\kappa g}(\ell)$, which is defined as

$$\langle \hat{\kappa}(\ell) \hat{\kappa}_g^*(\ell') \rangle = (2\pi)^2 \delta_D(\ell - \ell') P_{\kappa g}(|\ell|). \quad (156)$$

Applying (98), as well as the definitions of the bias and correlation functions, this projected cross-power spectrum is related to the 3-D density contrast by

$$P_{\kappa g}(\ell) = \frac{3}{2} \left(\frac{H_0}{c} \right)^2 \Omega_m b r \int dw \frac{g(w)p_f(w)}{a(w)f_K(w)} P_\delta \left(\frac{\ell}{f_K(w)}, w \right). \quad (157)$$

The angular correlation function $\langle \kappa(\vartheta) \kappa(\vartheta') \rangle$ occurring in (152) can then be replaced by its Fourier transform $P_{\kappa g}$. On the other hand, since the Fourier transform of the surface mass density κ is simply related to that of the shear,

one can consider the correlation between the galaxy positions with the tangential shear component,

$$\begin{aligned} \langle \gamma_t(\theta) \rangle &:= \langle \kappa_g(\mathbf{0}) \gamma_t(\boldsymbol{\theta}) \rangle \\ &= - \int \frac{d^2\ell}{(2\pi)^2} \int \frac{d^2\ell'}{(2\pi)^2} e^{2i(\beta' - \varphi)} \exp(-i\boldsymbol{\theta} \cdot \boldsymbol{\ell}') \langle \hat{\kappa}_g(\boldsymbol{\ell}) \hat{\kappa}(\boldsymbol{\ell}') \rangle \\ &= \frac{1}{2\pi} \int d\ell \ell J_2(\theta\ell) P_{\kappa g}(\ell) \\ &\Rightarrow P_{\kappa g}(\ell) = 2\pi \int d\theta \theta \langle \gamma_t(\theta) \rangle J_2(\theta\ell). \end{aligned} \quad (158)$$

Note that $\langle \gamma_t(\theta) \rangle$ is just the galaxy-galaxy lensing signal discussed in Sect. 8.2; this shows very clearly that galaxy-galaxy lensing measures the correlation of mass and light in the Universe. In terms of this mean tangential shear, the aperture mass and galaxy number counts can be written as

$$\langle M_{ap}(\theta) \mathcal{N}(\theta) \rangle = \int_0^{2\theta} \frac{d\vartheta \vartheta}{\theta^2} \langle \gamma_t(\vartheta) \rangle T_2\left(\frac{\vartheta}{\theta}\right), \quad (159)$$

where the function T_2 is defined in a way similar to T_\pm and given explicitly as

$$T_2(x) = 576 \int_0^\infty \frac{dt}{t^3} J_2(xt) [J_4(t)]^2; \quad (160)$$

this function vanishes for $x > 2$, so that the integral in (159) extends over a finite interval only. Hence, all three aperture correlators can be calculated from two-point correlation functions which can be determined from the data directly, independent of possible gaps in the field geometry.

Results from the RCS. Hoekstra et al. (2002c) have applied the foregoing equations to a combination of their RCS survey and the VIRMOS-DESCART survey. The former was used to determine $\langle \mathcal{N}^2 \rangle$ and $\langle M_{ap} \mathcal{N} \rangle$, the latter for deriving $\langle M_{ap}^2 \rangle$. As pointed out by these authors, this combination of surveys is very useful, in that the power spectrum at a redshift around $z \sim 0.35$ can be probed; indeed, they demonstrate that the effective redshift distribution over which the power spectrum, and thus b and r are probed, are well matched for all three statistics for their choice of surveys. ‘Foreground’ galaxies for the measurement of $\omega(\theta)$ and $\langle \gamma_t(\theta) \rangle$ are chosen to have $19.5 \leq R_C \leq 21$, ‘background’ galaxies are those with $21.5 \leq R_C \leq 24$. In Fig. 53 the three aperture statistics are shown as a function of angular scale, as determined from their combined survey, whereas in the right panels, the ratios of these statistics as they appear in (151) and (154) are displayed. Also shown are predictions of these quantities from two cosmological models, assuming $b = 1$ and $r = 1$. The fact that these model predictions are fairly constant in the right-hand panels shows that the factors f_b and f_r are nearly independent of the radius θ of the aperture, as mentioned before.

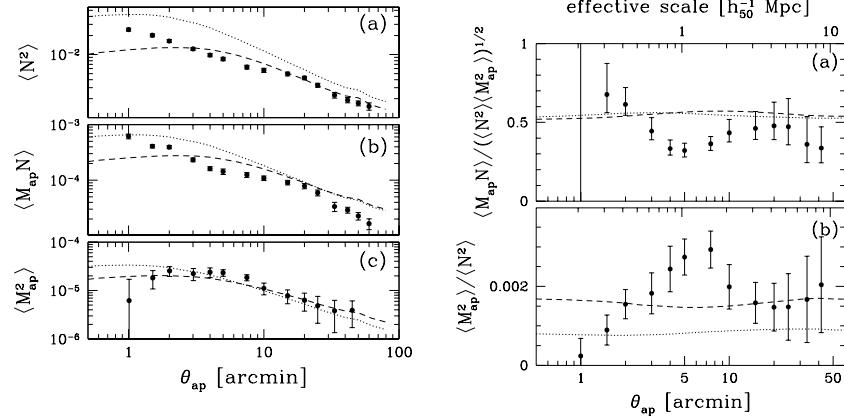


Fig. 53. The left figure displays the three aperture statistics as measured by combining the RCS and the VIRMOS-DESCART survey. Points show measured values, as determined from the correlation functions. The right panels display the ratios of the aperture statistics as they appear in (151) and (154). The dotted and dashed curves in all panels show the predictions for an OCDM and a Λ CDM model, respectively, both with $\Omega_m = 0.3$, $\sigma_8 = 0.9$, and $\Gamma_{\text{spect}} = 0.21$, for the fiducial values of $b = 1 = r$. The fact that the curves in the right panels are nearly constant show the near-independence of f_b and f_r on the filter scale. The upper axis in the right panels show the effective physical scale on which the values of b and r are measured (from Hoekstra et al. 2002c)

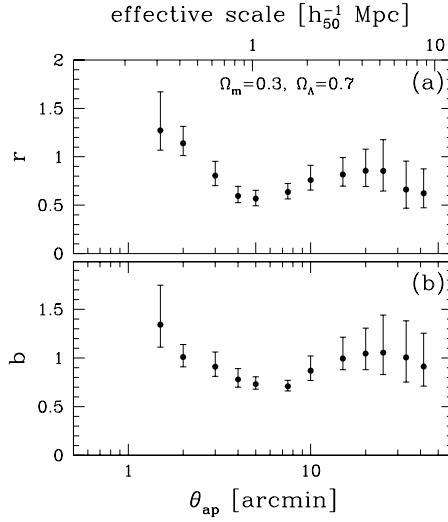


Fig. 54. The values of the bias and correlation coefficient, as determined from (151) and (154) and the results shown in Fig. 53; here, a Λ CDM model has been assumed for the cosmology dependence of the functions f_b and f_r . The upper axis indicates the effective scale on which b and r are measured (from Hoekstra et al. 2002c)

The results for the bias and correlation factor are shown in Fig. 54, as a function of angular scale and effective physical scale, corresponding to a median redshift of $z \sim 0.35$. The results indicate that the bias factor and the galaxy-mass correlation coefficient are compatible with a constant value on large scales, $\gtrsim 5h^{-1}$ Mpc, but on smaller scales both seem to change with scale. The transition between these two regimes occurs at about the scale where the density field at redshift $z \sim 0.35$ turns from linear to non-linear evolution. In fact, in the non-linear regime one does not expect a constant value of both coefficients, whereas in the linear regime, constant values for them appear natural. It is evident from the figure that the error bars are still too large to draw definite conclusions about the behavior of b and r as a function of scale, but the approach to investigate the relation between galaxies and mass is extremely promising and will certainly yield very useful insight when applied to the next generation of cosmic shear surveys. In particular, with larger surveys than currently available, different cuts in the definition of foreground and background galaxies can be used, and thus the redshift dependence of b and r can be investigated. This is of course optimized if (photometric) redshift estimates for the galaxy sample become available.

Results from the SDSS. The large sample of galaxies with spectroscopic redshifts already available now from the SDSS permits an accurate study of the biasing properties of these galaxies (see the end of Sect. 8.2). Two different approaches should be mentioned here: the first follows along the line discussed above and has been published in Sheldon et al. (2004). In short, the galaxy-galaxy signal can be translated into the galaxy-mass cross-correlation function ξ_{gm} , due to the knowledge of galaxy redshifts. The ratio of ξ_{gm} and the galaxy two-point correlation function ξ_{gg} then depends on the ratio r/b . In Fig. 55 we show the galaxy-mass correlation as a function of linear scale, as well as the ratio b/r . Note that from the SDSS no cosmic shear measurement has been obtained yet, owing to the complex PSF properties, and therefore b and r cannot be measured separately from this data set.

The galaxy-mass correlation function follows a power law over more than two orders-of-magnitude in physical scale, and its slope is very similar to the slope of the galaxy two-point correlation function. Hence, the ratio between these two is nearly scale-independent. When splitting the sample into blue and red, and early- and late-type galaxies, the correlation length is larger for the red and the early-type ones. Furthermore, as expected, the lensing signal increases with the velocity dispersion in early-type galaxies.

An alternative approach was taken by Seljak et al. (2004). Their starting point is the fact that the biasing properties of dark matter halos is very well determined from cosmological simulations. This is of course not true for the biasing of galaxies. The bias parameter of galaxies with luminosity L is given as

$$b(L) = \int dM p(M|L) b_h(M), \quad (161)$$

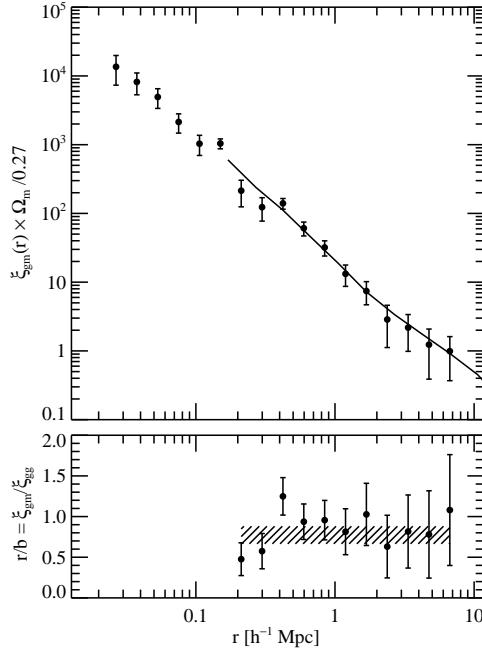


Fig. 55. The galaxy-mass cross-correlation function $\xi_{gm}(r)$, as a function of linear scale (dots with error bars), scaled to a matter density parameter of $\Omega_m = 0.27$, as well as the two-point galaxy correlation function obtained from the same set of (foreground) galaxies (solid curve). The ratio between these two is given in the lower panel, which plots b/r as a function of scale. Over the full range of scales, ξ_{gm} can be well approximated by a power law, $\xi_{gm} = (r/r_0)^{-\gamma}$, with slope $\gamma = 1.79 \pm 0.06$ and correlation length $r_0 = (5.4 \pm 0.7)(\Omega_m/0.27)^{-1/\gamma} h^{-1} \text{ Mpc}$. The ratio $r/b \approx (1.3 \pm 0.2)(\Omega_m/0.27)$ is consistent with being scale-independent

where b_h is the bias of halos of mass M relative to the large-scale matter distribution, and $p(M|L)$ is the probability that a galaxy with luminosity L resides in a halo of mass M . This latter probability distribution is then parameterized for any luminosity bin, by assuming that a fraction $1 - \alpha$ of all galaxies in the luminosity bin considered are at the center of their parent halos, whereas the remaining fraction α are satellite galaxies. For the central galaxies, a unique mass $M(L)$ is assigned, whereas for the non-central ones, a mass distribution is assumed. The values of α and M for six luminosity bins are shown in the various panels of Fig. 51; they are obtained by fitting the galaxy-galaxy lensing signal with the model just described. The main reason why the mass spectrum can be probed is that the numerous low-mass galaxy halos contribute to the lensing signal only at relatively small scales, whereas at larger scales the higher-mass halos dominate the signal; hence, different halo masses appear at different separations in the galaxy-galaxy lensing signal. In this way, $b(L)$ can be determined, which depends on the non-linear mass scale M_* (see Sect. 6.2 of IN). The bias parameter is a relatively slowly varying function of galaxy luminosity for $L \lesssim L_*$, approaching a value ~ 0.7 for very low-luminosity galaxies, but quickly rises for $L > L_*$.

Seljak et al. combined these measurements of the bias parameter with the clustering properties of the SDSS galaxies and the WMAP results on the CMB anisotropy, and derived new constraints on $\sigma_8 = 0.88 \pm 0.06$ and the bias parameter of an L_* -galaxy, $b_* = 0.99 \pm 0.07$; furthermore, the combination of these datasets is used to obtain new constraints on the standard cosmological

parameters. This work has opened up a new way on how to employ the results from galaxy-galaxy lensing as a cosmological tool.

8.4 Galaxy biasing: magnification method

High-redshift QSOs are observed to be correlated on the sky with lower-redshift galaxies and clusters. This topic has indeed an interesting history: The detection of very close associations of high- z QSOs with low- z galaxies (see Arp 1987, and references therein) has been claimed as evidence against the cosmological interpretation of the QSO redshifts, as the probabilities of observing such close pairs of objects which are physically unrelated were claimed to be vanishingly small. However, these probabilities were obtained a posteriori, and of course, any specific configuration has a vanishingly small probability. Since the cosmological interpretation of QSO redshifts is supported by overwhelming evidence, the vast majority of researchers consider these associations as a statistical fluke.

A physical possibility to generate the association of background sources with foreground objects is provided by the magnification bias caused by lensing: the number counts of background sources is changed in regions where a foreground lens yields magnifications different from unity – see Sect. 5 of IN. Thus, close to a galaxy where $\mu > 1$, the number counts of bright background QSOs can be enhanced since the slope of their counts is steeper than unity. There have been various attempts in the literature to ‘explain’ the observed QSO-galaxy associations by invoking the magnification bias, either with a smooth galaxy mass distribution or by including the effects of microlensing; see SEF for a detailed discussion of this effect. The bottom line, however, is that the magnification effect is by far not large enough to account for the small (a posteriori) probabilities of the observed individual close associations.

The topic has been revived, though in a different direction, by the finding that high-redshift AGNs are statistically associated with low-redshift galaxies. Fugmann (1990) provided evidence that radio-selected high- z AGNs from the 1-Jansky-catalog are correlated with relatively bright (and therefore low- z) galaxies taken from the Lick catalog, an analysis that later on was repeated by Bartelmann & Schneider (1993), using a slightly different statistics. Different samples of foreground and background populations have been employed in further studies, including the correlation between 1-Jansky AGN with bright IRAS galaxies (Bartelmann & Schneider 1994; Bartsch et al. 1997), high- z QSOs with clusters from the Zwicky catalog of clusters (Rodrigues-Williams & Hogan 1994; Seitz & Schneider 1995b), 1-Jansky AGNs with red galaxies from the APM catalog (Benítez & Martínez-González 1995; see also Norman & Impey 2001), to mention just a few. Radio-selected AGN are considered to be a more reliable probe since their radio flux is unaffected by extinction, an effect which could cause a bias (if the sky shows patchy extinction, both galaxies and QSOs would have correlated inhomogeneous distributions on

the sky) or anti-bias (if extinction is related to the lensing matter) for flux-limited optical surveys of AGNs, and which therefore needs to be taken into account in the correlation analysis of optically-selected AGNs. However, most radio source catalogs are not fully optically identified and lack redshifts, and using incomplete radio surveys therefore can induce a selection bias (Benítez et al. 2001). These latter authors investigated the correlation between two completely identified radio catalogs with the COSMOS galaxy catalog, and found a very significant correlation signal.

The upshot of all these analyses is that there seems to be a positive correlation between the high- z sources and the low- z objects, on angular scales between $\sim 1'$ and about 1° . The significances of these correlations are often not very large, they typically are at the $2\text{--}3\sigma$ level, essentially limited by the finite number of high-redshift radio sources with a large flux (the latter being needed for two reasons: first, only radio surveys with a high flux threshold, such as the 1-Jansky catalog, have been completely optically identified and redshifts determined, which is necessary to exclude low-redshift sources which could be physically associated with the ‘foreground’ galaxy population, and second, because the counts are steep only for high fluxes, needed to obtain a high magnification bias.) If this effect is real, it cannot be explained by lensing caused by individual galaxies; the angular region on which galaxies produce an appreciable magnification is just a few arcseconds. However, if galaxies trace the underlying (dark) matter distribution, the latter can yield magnifications (in the same way as it yields a shear) on larger scales. Thus, an obvious qualitative interpretation of the observed correlation is therefore that it is due to magnification of the large-scale matter distribution in the Universe of which the galaxies are tracers. This view is supported by the finding (Ménard & Péroux 2003) that there is a significant correlation of bright QSOs with metal absorption systems in the sense that there are relatively more bright QSOs with an absorber than without; this effect shows the expected trend from magnification bias caused by matter distributions associated with the absorbing material.

We therefore consider a flux-limited sample of AGNs, with distance probability distribution $p_Q(w)$, and a sample of galaxies with distance distribution $p_f(w)$. It will be assumed that the AGN sample has been selected such that it includes only objects with redshift larger than some threshold z_{\min} , corresponding to a minimum comoving distance w_{\min} , which is larger than the distances of all galaxies in the sample. We define the AGN-galaxy correlation function as

$$w_{Qg}(\theta) = \frac{\langle [N_g(\phi) - \bar{N}_g] [N_Q(\phi + \theta) - \bar{N}_Q] \rangle}{\bar{N}_g \bar{N}_Q}; \quad (162)$$

where $N_g(\phi)$ and $N_Q(\phi)$ are the observed number densities of galaxies and AGNs, respectively. The former is given by (142). The observed number density of AGN is affected by the magnification bias. Provided the unlensed

counts can be described (locally) as a power-law in flux, $N_{Q,0}(> S) \propto S^{-\beta}$, then from (108) of IN we find that $N_Q(\phi) = N_{Q,0} \mu^{\beta-1}(\phi)$, where $\mu(\phi)$ is the magnification in the direction ϕ . Then, if the magnifications that are relevant are small, we can approximate

$$\mu(\phi) \approx 1 + 2\kappa(\phi) = 1 + \delta\mu(\phi), \quad (163)$$

and the projected surface mass density κ is given by (93) with p_w in (94) replaced by p_Q . Assuming that the magnifications do not affect the mean source counts \bar{N}_Q , the cross-correlation becomes

$$w_{Qg}(\theta) = 2(\beta - 1)\bar{b}(\theta) \bar{r}(\theta) w_{\kappa g}(\theta), \quad (164)$$

where \bar{b} and \bar{r} are the effective bias factor of the galaxies and the mean galaxy-mass correlation function just as in Sect. 8.3, and $w_{\kappa g}$ is the correlation between the projected density field κ and the projected number density of galaxies κ_g , defined after (142), which is the Fourier transform of $P_{\kappa g}(\ell)$ defined in (141). Hence, a measurement of this correlation, together with a measurement of the correlation function of galaxies, can constrain the values of b and r (Dolag & Bartelmann 1997; Ménard & Bartelmann 2002).

The observed correlation between galaxies and background AGN appears to be significantly larger than can be accounted for by the models presented above. On scales of a few arcmin, Benítez et al. (2001) argued that the observed signal exceeds the theoretical expectations by a factor of a few. This discrepancy can be attributed to either observational effects, or shortcomings of the theoretical modelling. Obviously, selection effects can easily produce spurious correlations, such as patchy dust obscuration or a physical association of AGNs with the galaxies. Furthermore, the weak lensing approximation employed above can break down on small angular scales. Jain et al. (2003, see also Takada & Hamana 2003) argued that the simple biasing model most likely breaks down for the small scales where the discrepancy is seen, and employed the halo model for describing the large-scale distribution of matter and galaxies to predict the expected correlations. For example, the strength of the signal depends sensitively on the redshifts, magnitudes and galaxy type.

At present, the shear method to determine the bias factor and the galaxy-mass correlation has yielded more significant results than the magnification method, owing to the small complete and homogeneous samples of high-redshift AGNs. As pointed out by Ménard & Bartelmann (2002), the SDSS may well change this situation shortly, as this survey will obtain $\sim 10^5$ homogeneously selected spectroscopically verified AGNs. Provided the effects of extinction can be controlled sufficiently well, this data should provide a precision measurement of the QSO-galaxy correlation function.

9 Additional issues in cosmic shear

9.1 Higher-order statistics

On the level of second-order statistics, ‘only’ the power spectrum is probed. If the density field was Gaussian, then the power spectrum would fully characterize it; however, in the course of non-linear structure evolution, non-Gaussian features of the density field are generated, which show up correspondingly in the cosmic shear field and which can be probed by higher-order shear statistics. The usefulness of these higher-order measures for cosmic shear has been pointed out in Bernardeau et al. (1997), Jain & Seljak (1997), Schneider et al. (1998a) and van Waerbeke et al. (1999); in particular, the near-degeneracy between σ_8 and Ω_m as found from using second-order statistics can be broken. However, there are serious problems with higher-order shear statistics, that shall be illustrated below in terms of the third-order statistics.

But first, we can give a simple argument why third-order statistics is able to break the degeneracy between Ω_m and σ_8 . Consider a density field on a scale where the inhomogeneities are just weakly non-linear. One can then employ second-order perturbation theory for the growth of the density contrast δ . Hence, we write $\delta = \delta^{(1)} + \delta^{(2)} + \dots$, where $\delta^{(1)}$ is the density contrast obtained from linear perturbation theory, and $\delta^{(2)}$ is the next-order term. This second-order term is quadratic in the linear density field, $\delta^{(2)} \propto (\delta^{(1)})^2$. The linear density field is proportional to σ_8 , and the projected density $\kappa \propto \Omega_m \sigma_8$. Hence, in the linear regime, $\langle \kappa^2 \rangle \propto \Omega_m^2 \sigma_8^2$, where $\langle \kappa^2 \rangle$ shall denote here any second-order shear estimator. The lowest order contribution to the third-order statistics is of the form

$$\langle \kappa^3 \rangle \propto \left(\delta^{(1)} \right)^2 \delta^{(2)} \propto \Omega_m^3 \sigma_8^4 ,$$

since the term $(\delta^{(1)})^3$ yields no contribution owing to the assumed Gaussianity of the linear density field. Hence, a skewness statistics of the form

$$\langle \kappa^3 \rangle / \langle \kappa^2 \rangle^2 \propto \Omega_m^{-1}$$

will be independent of the normalization σ_8 , at least in this simplified perturbation approach. In more accurate estimates, this is not exactly true; nevertheless, the functional dependencies of the second- and third-order shear statistics on σ_8 and Ω_m are different, so that these parameters can be determined separately.

The shear three-point correlation function. Most of the early studies on three-point statistics concentrated on the third-order moment of the surface mass density κ in a circular aperture, $\langle \kappa(\theta) \rangle$; however, this is not a

directly measurable quantity, and therefore useful only for theoretical considerations. As for second-order statistics, one should consider the correlation functions, which are the quantities that can be obtained best directly from the data and which are independent of holes and gaps in the data field. The three-point correlation function (3PCF) of the shear has three independent variables (e.g. the sides of a triangle) and 8 components; as was shown in Schneider & Lombardi (2003), none of these eight components vanishes owing to parity invariance (as was suspected before – this confusion arises because little intuition is available on the properties of the 3PCF of a polar). This then implies that the covariance matrix has 6 arguments and 64 components! Of course, this is too hard to handle efficiently, therefore one must ask which combinations of the components of the 3PCF are most useful for studying the dark matter distribution. Unfortunately, this is essentially unknown yet. An additional problem is that the predictions from theory are less well established than for the second-order statistics.

A further complication stems from a certain degree of arbitrariness on how to define the 8 components of the 3PCF. For the 2PCF, the vector between any pair of points defines a natural direction with respect to which tangential and cross components of the shear are defined; this is no longer true for three points. On the other hand, the three points of a triangle define a set of centers, such as the ‘center of mass’, or the center of the in- or circum-circle. After choosing one of these centers, one can define the two components of the shear which are then independent of the coordinate frame.

Nevertheless, progress has been achieved. From ray-tracing simulations through a cosmic matter distribution, the 3PCF of the shear can be determined (Takada & Jain 2003a; see also Zaldarriaga & Scoccimarro 2003; furthermore, the three-point cosmic shear statistics can also be determined in the frame of the halo model, see Cooray & Hu 2001; Takada & Jain 2003b), whereas Schneider & Lombardi (2003) have defined the ‘natural components’ of the shear 3PCF which are most easily related to the bispectrum of the underlying matter distribution. Let $\gamma^c(\boldsymbol{\theta}_i) = \gamma_t + i\gamma_x = -\gamma e^{-2i\zeta_i}$ be the complex shear measured in the frame which is rotated by the angle ζ_i relative to the Cartesian frame, so that the real and imaginary parts of γ^c are the tangential and cross components of the shear relative to the chosen center of the triangle (which has to be defined for each triplet of points separately). Then the natural components are defined as

$$\begin{aligned}\Gamma^{(0)} &= \langle \gamma^c(\boldsymbol{\theta}_1) \gamma^c(\boldsymbol{\theta}_2) \gamma^c(\boldsymbol{\theta}_3) \rangle , \\ \Gamma^{(1)} &= \langle \gamma^{c*}(\boldsymbol{\theta}_1) \gamma^c(\boldsymbol{\theta}_2) \gamma^c(\boldsymbol{\theta}_3) \rangle ,\end{aligned}\quad (165)$$

and correspondingly for $\Gamma^{(2)}$ and $\Gamma^{(3)}$. Each of the natural components of the 3PCF constitutes a complex number, which depends just on the three separations between the points. Special care is required for labelling the points, and one should follow the rule that they are labeled in a counter-clock direction around the triangle. If such a unique prescription is not systematically ap-

plied, confusing and wrong conclusions will be obtained about the behaviour of the shear 3PCF with respect to parity transformations (as the author has experienced painfully enough). In Schneider et al. (2004), explicit relations are derived for the natural components of the shear 3PCF in terms of the bispectrum (that is, the generalization of the power spectrum for the three-point statistics) of the underlying mass distribution κ .

Third-order aperture statistics. Alternatively, aperture measures can be defined to measure the third-order statistics. Schneider et al. (1998a) calculated $\langle M_{\text{ap}}^3 \rangle(\theta)$ in the frame of the quasi-linear structure evolution model and showed it to be a strong function of Ω_m . Van Waerbeke et al. (2001) calculated the third-order aperture mass, using a fitting formula of the non-linear evolution of the dark matter bispectrum obtained by Scoccimarro & Couchman (2001) and pointed out the strong sensitivity with respect to cosmological parameters. Indeed, as mentioned before, $\langle M_{\text{ap}}^3 \rangle$ is sensitive only to the E-modes of the shear field. One might be tempted to use $\langle M_{\perp}^3 \rangle(\theta)$ as a measure for third-order B-mode statistics, but indeed, this quantity vanishes owing to parity invariance (Schneider 2003). However, $\langle M_{\perp}^2 M_{\text{ap}} \rangle$ is a measure for the B-modes at the third-order statistical level. Jarvis et al. (2004) have calculated $\langle M_{\text{ap}}^3(\theta) \rangle$ in terms of the shear 3PCF, for the weight function (110) in the definition of M_{ap} . Schneider et al. (2004) have shown that this relation is most easily expressed in terms of the natural components of the shear 3PCF. On the other hand, Jarvis et al. (2004) have expressed $\langle M_{\text{ap}}^3(\theta) \rangle$ in terms of the bispectrum of κ , and as was the case for the aperture dispersion in relation to the power spectrum of κ , the third-order aperture mass is a very localized measure of the bispectrum and is sensitive essentially only to modes with three wavevectors with equal magnitudes. For that reason, Schneider et al. (2004) have generalized the definition of the third-order aperture measures, correlating the aperture mass of three different sizes, $\langle M_{\text{ap}}(\theta_1)M_{\text{ap}}(\theta_2)M_{\text{ap}}(\theta_3) \rangle$. This third-order statistics is again a very localized measure of the bispectrum, but this time with wave vectors of different magnitude $\ell_i \approx \pi/\theta_i$, and therefore, by considering the third-order aperture mass for all combinations of θ_i , one can probe the full bispectrum. Therefore, the third-order aperture mass correlator with three independent arguments (i.e., angular scales) should contain essentially the full third-order statistical information of the κ -field, since in contrast to the two-point statistics, the shear 3PCF does not contain information about long-wavelength modes.

Furthermore, the third-order aperture statistics can be expressed directly in terms of the shear 3PCF through a simple integration, very similar to the relations (125) for the two-point statistics. Finally, the other three third-order aperture statistics (e.g., $\langle M_{\perp}(\theta_1)M_{\text{ap}}(\theta_2)M_{\text{ap}}(\theta_3) \rangle$) can as well be obtained from the natural components of the shear 3PCF. These correlators are expected to vanish if the shear is solely due to lensing, but intrinsic alignments of galaxies can lead to finite correlators which include B-modes.

However, as shown in Schneider (2003), $\langle M_{\text{ap}}(\theta_1)M_{\text{ap}}(\theta_2)M_{\perp}(\theta_3)\rangle$, as well as $\langle M_{\perp}(\theta_1)M_{\perp}(\theta_2)M_{\perp}(\theta_3)\rangle$, are expected to vanish even in the presence of B-modes, since these two correlators are not invariant with respect to a parity transformation. Therefore, non-zero results of these two correlators signify the violation of parity invariance and therefore provide a clean check on the systematics of the data and their analysis.

First detections. Bernardeau et al. (2002) measured for the first time a significant third-order shear from the VIRMOS-DESCART survey, employing a suitably filtered integral over the measured 3PCF (as defined in Bernardeau et al. 2003). Pen et al. (2003) used the aperture statistics to detect a skewness in the same data set. The accuracy of these measurements is not sufficient to derive strong constraints on cosmological parameters, owing to the limited sky area available. However, with the upcoming large cosmic shear surveys, the 3PCF will be measured with high accuracy. Determining the 3PCF from observed galaxy ellipticities cannot be done by straightforwardly considering any triple of galaxies – there are just too many. Jarvis et al. (2004) and Zhang & Pen (2003) have developed algorithms for calculating the 3PCF in an efficient way.

Based on the halo model for the description of the LSS, Takada & Jain (2003b) studied the dependence of the shear 3PCF on cosmological parameters. For relatively large triangles, the 3PCF provides a means to break the degeneracies of cosmological parameters that are left when using the second-order statistics only, as argued above. For small triangles, the 3PCF is dominated by the one-halo term, and therefore primarily probes the mass profiles of halos. Ho & White (2004) show that the 3PCF on small angular scales also contains information on the asphericity of dark matter halos. The full power of third-order statistics is achieved once redshift information on the source galaxies become available, in which case the combination of the 2PCF and 3PCF provides a sensitive probe on the equation-of-state of the dark energy (Takada & Jain 2004).

Beyond third order. One might be tempted to look into the properties of the fourth-order shear statistics (though I’m sure the reader can control herself in doing this – but see Takada & Jain 2002). OK, the four-point correlation function has 16 components and depends on 5 variables, not to mention the corresponding covariance or the redshift dependent fourth-order correlator. One can consider correlating the aperture mass of four different angular sizes, but in contrast to the third-order statistics, this is expected not to contain the full information on the trispectrum (which describes the fourth-order statistical properties of κ). Perhaps a combination of this fourth-order aperture mass with the average of the fourth power of the mean shear in circular apertures will carry most of the information. And how much in-

formation on cosmological parameters does the fourth-order shear statistics contain? And even higher orders?

Already the third-order shear statistic is not accurately predictable from analytic descriptions of the non-linear evolution of the matter inhomogeneities, and the situation worsens with even higher order.¹³ One therefore needs to refer to detailed ray-tracing simulations. Although they are quite time consuming, I do not see a real bottleneck in this aspect: Once a solid and accurate measurement of the three-point correlation function becomes available, certainly considerable effort will be taken to compare this with numerical simulations (in particular, since such a measurement is probably a few years ahead, in which the computer power will increase by significant factors). If we accept this point, then higher-order statistics can be obtained from these simulations, and several can be ‘tried out’ on the numerical data such that they best distinguish between different models. For example, one can consider the full probability distribution $p(M_{\text{ap}}; \theta)$ on a given data set (Kruse & Schneider 2000; Reblinsky et al 1999; Bernardreau & Valageas 2000; Munshi et al. 2004). To obtain this from the observational data, one needs to place apertures on the data field which, as we have argued, is plagued with holes and gaps in the data. However, we can place the same gaps on the simulated data fields and therefore simulate this effect. Similarly, the numerical simulations should be used to find good strategies for combining second- and third-order shear statistics (and potentially higher-order ones) for an optimal distinction between cosmological model parameters, and, in particular, the equation-of-state of Dark Energy. Another issue one needs to consider for third- (and higher-)order cosmic shear measures is that intrinsic clustering of sources, and the correlation between galaxies and the dark matter distribution generating the shear shear field has an influence on the expected signal strength (Bernardeau 1998; Hamana 2001; Hamana et al. 2002). Obviously, there are still a lot of important studies to be done.

Third-order galaxy-mass correlations. We have shown in Sect. 8 how galaxy-galaxy lensing can be used to probe the correlation between galaxies and the underlying matter distribution. With the detection of third-order shear statistics already in currently available data sets, one might expect that also higher-order galaxy-mass correlations can be measured from the same data. Such correlations would then probe, on large angular scales, the higher-order biasing parameters of galaxies, and thereby put additional constraints

¹³ In the limits of small and large angular scales, analytic approximations can be obtained. For small scales, the highly non-linear regime is often described by the hierarchical ansatz and hyperextended perturbation theory (see Munshi & Jain 2001 and references therein), whereas on very large scales second-order perturbation theory can be used. Nevertheless, the range of validity of these perturbation approximations and their accuracy have to be checked with numerical simulations.

on the formation and evolution of galaxies. Ménard et al. (2003) considered the correlation between high-redshift QSOs and pairs of foreground galaxies, thus generalizing the methods of Sect. 8.4 to third-order statistics. The galaxy-galaxy-shear correlation, and the galaxy-shear-shear correlations have been considered by Schneider & Watts (2004). These correlation functions have been related to the underlying bispectrum of the dark matter and the third-order bias and correlation functions, and appropriate aperture statistics have been defined, that are related in a simple way to the bispectra and the correlation functions.

In fact, integrals of these higher-order correlations have probably been measured already. As shown in Fig. 50, galaxies in regions of high galaxy number densities show a stronger, and more extended galaxy-galaxy lensing signal than more isolated galaxies. Hence there is a correlation between the mean mass profile around galaxies and the local number density of galaxies, which is just an integrated galaxy-galaxy-shear correlation. In fact, such a correlation is only first order in the shear and should therefore be much easier to measure than the shear 3PCF. Furthermore, the galaxy-shear-shear correlation seems to be present in the cosmic shear analysis of the COMBO-17 fields by Brown et al. (2003), where they find a stronger-than-average cosmic shear signal in the A901 field, and a weaker cosmic shear signal in the CDFS, which is a field selected because it is rather poor in brighter galaxies.

9.2 Influence of LSS lensing on lensing by clusters and galaxies

The lensing effect of the three-dimensional matter distribution will contaminate the lensing measurements of localized objects, such as galaxies and clusters. Some of the associated effects are mentioned in this section.

Influence of cosmic shear on strong lensing by galaxies. The lensing effect of foreground and background matter in a strong lensing system will affect the image positions and flux ratios. As this 3-D lensing effects are not recognized as such in the lens modelling, a ‘wrong’ lens model will be fitted to the data, in the sense that the mass model for the lensing galaxy will try to include these additional lensing effects not associated with the galaxy itself. In particular, the corresponding predictions for the time delays can be affected through this effect.

Since the image separation of strong lens systems are less than a few arcseconds, the lensing effect of the LSS can be well approximated by a linear mapping across this angular scale. In this case, the effect of the 3-D matter distribution on the lens model can be studied analytically (e.g., Bar-Kana 1996). The lens equation resulting from the main lens (the galaxy) plus the linearized inhomogeneities of the LSS is strictly equivalent to the single-plane gravitational lens equation without these cosmological perturbations, and the mass distribution of the equivalent single-plane lens can be explicitly derived

(Schneider 1997). For example, if the main lens is described by elliptical isopotential curves (i.e., elliptical contours of the deflection potential ψ) plus external shear, the equivalent single-plane lens will be of the same form. The orientation of the ellipticity of the lens, as seen by the observer, will be rotated by the foreground LSS by the same angle as the potential of the equivalent lens, so that no observable misalignment is induced. This equivalence then implies that the determination of the Hubble constant from time-delay measurements is affected by the same mass-sheet degeneracy transformation as for a single plane lens.

LSS effects on the mass determination of clusters. The determination of mass parameters of a cluster from weak lensing is affected by the inhomogeneous foreground and background matter distribution. The effect of local mass associated with a cluster (e.g., filaments extending from the cluster along the line-of-sight) will bias the mass determination of clusters high, since clusters are likely to be located in overdense regions of the LSS, though this effect is considerably smaller than claimed by Metzler et al. (2001), as shown by Clowe et al. (2004a).

Hoekstra (2001, 2003) considered the effect of the LSS on the determination of mass parameters of clusters, using either SIS or NFW models. For the SIS model, the one parameter characterizing this mass profile (σ_v) can be obtained as a linear estimator of the shear. The dispersion of this parameter is then the sum of the dispersion caused by the intrinsic ellipticity of the source galaxies and the cosmic shear dispersion. For the NFW model, the relation between its two parameters (M_{200} , the mass inside the virial radius r_{200} , and the concentration c) and the shear is not linear, but the effect of the LSS can still be estimated from Monte-Carlo simulations in which the cosmic shear is assumed to follow Gaussian statistics with a power spectrum following the Peacock & Dodds (1996) prescription.

For the SIS model, the effect of the LSS on the determination of σ_v is small, provided the cluster is at intermediate redshift (so that most source galaxies are in the background). The noise caused by the finite ellipticity in this case is almost always larger than the effect by the LSS. There is an interesting effect, however, in that the relative contribution of the LSS and shape noise changes as larger aperture fits to the SIS model are considered: The larger the field over which the shear is fitted to an SIS model, the larger becomes the impact of cosmic shear, and this increase compensates for the reduced shape noise. In effect, cosmic shear and shape noise together put an upper limit on the accuracy of the determination of σ_v from shear data. The same is true for the determination of the mass parameters of the NFW model, as shown in Fig. 56. The uncertainties of the mass parameters of NFW profiles are about twice as large as if the effects from the LSS are ignored, whereas the effect is considerably smaller for the one-parameter model of the SIS. One should also note that a decrease of the shape noise, which can be obtained by

using data with a fainter limiting magnitude, yields an increase of the noise from the LSS, since the fainter galaxies are expected to be at higher redshift and therefore carry a larger cosmic shear signal. For low-redshift clusters, these two effects nearly compensate.

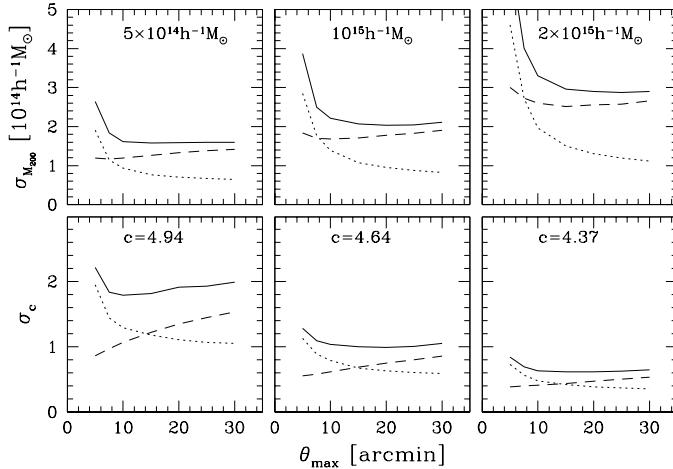


Fig. 56. The dispersion of the determination of the mass and concentration of three NFW halos at redshift $z_d = 0.3$. These parameters were derived by fitting an NFW shear profile to the shear simulated from an NFW halo with parameters indicated in the figure and adding shape noise and noise from cosmic shear. The outer angular scale over which the fit was performed is θ_{\max} . Dotted curves show the effect from shape noise alone, dashed curves show the dispersion from cosmic shear, and the solid curves contain both effects. Surprisingly, the accuracy of the NFW parameters does not increase once $\theta_{\max} \sim 15'$ is reached, as for larger radii, the cosmic shear noise more than compensates for the reduced ellipticity noise. Another way to express that is that the lensing signal at very large distance from the halo center is weaker than the rms cosmic shear and therefore does not increase the signal-to-noise any more (from Hoekstra 2003)

The efficiency and completeness of weak lensing cluster searches. We take up the brief discussion at the end of Sect. 5.8 about the potential of deriving a shear-selected sample of galaxy clusters. The first studies of this question were based on analytical models (e.g., Kruse & Schneider 1999) or numerical models of isolated clusters (Reblinsky & Bartelmann 1999). Those studies can of course not account for the effects of lensing by the LSS. Ray-tracing simulations through N-body generated LSS were carried out by Reblinsky et al. (1999), White et al. (2002), Hamana et al. (2004), Vale & White (2003), Hennawi & Spergel (2004) and others. In these cosmological

simulations, halos were identified based on their 3-D mass distribution. They were then compared to the properties of the lensing results obtained from ray tracing, either by considering the (smoothed) surface mass density κ (that could be obtained from a mass reconstruction from the shear field) or by studying the aperture mass M_{ap} which can be obtained directly from the shear. In both cases, noise due to the finite intrinsic source ellipticity can be added.

The two basic quantities that have been investigated in these studies are *completeness* and *efficiency*. Completeness is the fraction of dark matter halos above some mass threshold M_{\min} that are detected in the weak lensing data, whereas efficiency is the fraction of significant lensing detections that correspond to a real halo. Both of these quantities depend on a number of parameters, like the mass threshold of a halo and the limiting significance ν of a lensing detection [in the case of the aperture mass, this would correspond to (80)], as well as on the choice of the filter function Q . Hennawi & Spergel (2003) have pointed out that even without noise (from observations or intrinsic galaxy ellipticities), the efficiency is limited to about 85% – even under these idealized condition, the selected sample will be contaminated by at least 15% of spurious detections, generated by projection effects of the LSS.

To compare these predictions with observations, the six highest-redshift EMSS clusters were all detected at high significance with a weak lensing analysis (Clowe et al. 2000). Clowe et al. (2004b) have studied 20 high-redshift clusters with weak lensing techniques. These clusters were optically selected and are expected to be somewhat less massive (and potentially more affected by foreground galaxies) than the EMSS clusters. Only eight of these 20 clusters are detected with more than 3σ significance, but for none of them does the SIS fit produce a negative σ_v^2 . Only for four of these clusters are the lensing results compatible with no shear signal.

10 Concluding remarks

Weak lensing has become a standard tool in observational cosmology, as we have learned how to measure the shape of faint galaxy images and to correct them for distortions in the telescope and camera optics and for PSF effects. These technical issues are at the very center of any observational weak lensing research. It appears that at present, the accuracy with which shear can be measured is sufficient for the data available today, in the sense that statistical uncertainties are likely to be larger than potential inaccuracies in the measurement of unbiased shear estimates from faint images. This, however, will change quickly. The upcoming large cosmic shear surveys will greatly reduce statistical uncertainties, and then the accuracy of shear measurements from the data will be the essential limiting factor. Alternatives to KSB have been developed, but they need to undergo thorough testing before becoming

a standard tool for observers. It should also be noted that the KSB method is applied differently by different groups, in particular with regards to the weighting of galaxies and other details. What is urgently needed is a study in which different groups apply their version of KSB to the same data set and compare the results. Furthermore, starting from raw data, the specific data reduction methods will lead to slightly different coadded images, and shear measurements on such differently reduced imaged should be compared. These technical issues will be a central challenge for weak lensing in the upcoming years.

The ongoing and planned wide-field imaging surveys mentioned at the end of Sect. 7.7 will allow us to investigate several central questions of cosmology. The two aspects that I consider most relevant are the investigation of the equation-of-state of the Dark Energy and the relation between galaxies and the underlying dark matter distribution. The former question about the nature of Dark Energy is arguably the central challenge of modern cosmology, and cosmic shear is one of the very few methods how it can be studied empirically. The relation between dark matter and galaxies is central to our understanding of how galaxies form and evolve, and galaxy-galaxy lensing is the only way how this relation can be investigated without a priori assumptions.

Essentially all weak lensing studies today have used faint galaxies as sources, since they form the densest source population currently observable. The uniqueness of faint optical galaxies will not stay forever, with the currently planned future instruments. For example, there is a rich literature of weak lensing of the cosmic microwave background which provides a source of very accurately known redshift. Weak lensing by the large-scale structure enhances the power spectrum of the CMB at small angular scales, and the Planck satellite will be able to measure this effect. In particular, polarization information will be very useful, since lensing can introduce B-modes in the CMB polarization. The James Webb Space Telescope, with its large aperture of 6.5 meters and its low temperature and background will increase the number density of observable faint sources in the near-IR up to $5\mu\text{m}$ to several hundred per square arcminute, many of them at redshifts beyond 3, and will therefore permit much more detailed weak lensing studies, in particular of clusters (see Fig. 57; an observation of this huge number of arcs and multiple images will answer questions about the mass distribution of clusters that we have yet not even dared to ask). The envisioned next generation radio telescope Square Kilometer Array will populate the radio sky with very comparable source density as currently the deepest optical images. Since the beam (that is, the point-spread function) of this radio interferometer will be known very accurately, PSF corrections for this instrument will be more reliable than for optical telescopes. Furthermore, higher-order correlation of the shear field with sources in the field will tell us about non-Gaussian properties

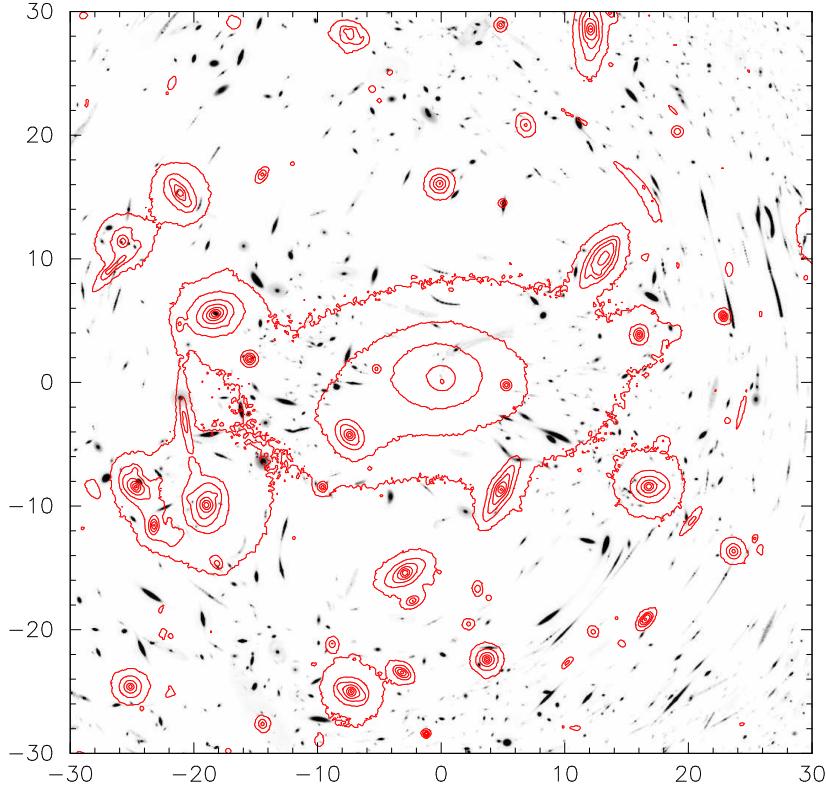


Fig. 57. Simulated image of lensed features in the very central part of the massive cluster A2218, as observed with the future JWST. For these simulations, the mass profile of the cluster as constrained from HST observations and detailed modelling (Kneib et al. 1996) has been used. The number density of (unlensed) sources was assumed to be $4 \times 10^6 \text{ deg}^{-2}$ down to $K=29$. The redshift distribution assumed is broad and extends to redshift $z \sim 10$ with a median value $z_{\text{med}} \sim 3$. The brighter objects (cluster galaxies and brightest arcs) seen by HST are displayed as contours, to make the faint galaxy images visible on this limited dynamic range reproduction. An enormous number of large arcs and arclets are seen; in particular, numerous radial arcs can be easily detected, which will allow us to determine the ‘core size’ of the cluster mass distribution. Due to the broad redshift distribution of the faint galaxies, arcs occur at quite a range of angular separations from the cluster center; this effect will become even stronger for higher-redshift clusters. It should be noted that this 1 arcminute field does not cover the second mass clump seen with HST; an JWST image will cover a much larger area, and more strong lensing features will be found which can then be combined with the weak lensing analysis of such a cluster. For this simulation, a pixel size of $0''.06$ was used; the JWST sampling will be better by a factor of 2 (from Schneider & Kneib 1998)

of galaxy-matter correlations and biasing, and therefore provide important input into models of galaxy formation and evolution.

Acknowledgements. I enjoyed the week of lecturing at Les Diablerets a lot. First, I thank Georges Meylan and his colleagues Philippe Jetzer and Pierre North for organizing this school so efficiently and smoothly; to them also my sincere apologies for not finishing these proceedings much earlier – but sure enough, lecturing is more fun than preparing lectures, and certainly very much more fun than writing them up into something to be published. Second, I thank my fellow lecturers Chris Kochanek and Joe Wambsganss for discussions, great company and good spirits. Third, and foremost, my compliments to the students who patiently sat through these lectures which, I'm sure, were not always easy to follow; nevertheless I hope that they grasped the essential points, and that these lecture notes help to fill in the details. These notes would not have been possible without the many colleagues and students with whom I had the privilege to collaborate over the years on various aspects of weak lensing, among them Matthias Bartelmann, Marusa Bradac, Doug Clowe, Jörg Dietrich, Thomas Erben, Bernhard Geiger, Hannelore Hämerle, Bhuvnesh Jain, Martin Kilbinger, Lindsay King, Martina Kleinheinrich, Guido Kruse, Marco Lombardi, Yannick Mellier, Hans-Walter Rix, Mischa Schirmer, Tim Schrabback, Carolin Seitz, Stella Seitz, Patrick Simon, Ludovic van Waerbeke and Anja von der Linden. This work was supported by the German Ministry for Science and Education (BMBF) through the DLR under the project 50 OR 0106, by the German Ministry for Science and Education (BMBF) through DESY under the project 05AE2PDA/8, and by the Deutsche Forschungsgemeinschaft under the project SCHN 342/3-1.

References

- Abazajian, K., Adelman-McCarthy, J.K., Agüeros, M.A. et al. 2004, AJ 128, 502
 Abazajian, K.N. & Dodelson, S. 2003, PRL 91, 041301
 Abell, G.O. 1958, ApJS 3, 211
 Abell, G.O., Corwin, H.G. & Olowin, R.P. 1989, ApJS 70, 1
 Allen, S.W. 1998, MNRAS 296, 392
 Andersson, K.E. & Madejski, G.M. 2004, ApJ 607, 190
 Arp, H. 1987, *Quasars, redshifts, and controversies*, Interstellar Media, Berkeley
 Athreya, R.M., Mellier, Y., van Waerbeke, L., Pelló, R., Fort, B. & Dantel-Fort, M.
 2002, A&A 384, 743
 Babul, A. & Lee, M.H. 1991, MNRAS 250, 407
 Bacon, D.J., Massey, R.J., Refregier, A.R. & Ellis, R.S. 2003, MNRAS 344, 673
 Bacon, D.J., Refregier, A., Clowe, D. & Ellis, R.S. 2001, MNRAS 325, 1065
 Bacon, D.J., Refregier, A.R. & Ellis, R.S. 2000, MNRAS 318, 625
 Bacon, D.J. & Taylor, A.N. 2003, MNRAS 344, 1307
 Bacon, D.J., Taylor, A.N., Brown, M.L. et al. 2004, astro-ph/0403384
 Barber, A.J. 2002, MNRAS 335, 909

- Barber, A.J., Thomas, P.A., Couchman, H.M.P. & Fluke, C.J. 2000, MNRAS 319, 267
- Bar-Kana, R. 1996, ApJ 468, 17
- Bartelmann, M. 1995, A&A 303, 643
- Bartelmann, M., Huss, A., Colberg, J., Jenkins, A. & Pearce, F. 1998, A&A 330, 1
- Bartelmann, M., King, L.J. & Schneider, P. 2001, A&A 378, 361
- Bartelmann, M. & Meneghetti, M. 2004, A&A 418, 413
- Bartelmann, M., Meneghetti, M., Perrotta, F., Baccigalupi, C. & Moscardini, L. 2003, A&A 409, 449
- Bartelmann, M. & Narayan, R. 1995, ApJ 451, 60
- Bartelmann, M., Narayan, R., Seitz, S. & Schneider, P. 1996, ApJ 464, L115
- Bartelmann, M., Perrotta, F. & Baccigalupi, C. 2002, A&A 396, 21
- Bartelmann, M. & Schneider, P. 1993, A&A 271, 421
- Bartelmann, M. & Schneider, P. 1994, A&A 284, 1
- Bartelmann, M. & Schneider, P. 1999, A&A 345, 17
- Bartelmann, M. & Schneider, P. 2001, Physics Reports 340, 291 (BS01)
- Bartelmann, M. & Steinmetz, M. 1996, MNRAS 283, 431
- Bartelmann, M., Steinmetz, M. & Weiss, A. 1995, A&A 297, 1
- Bartelmann, M. & Weiss, A. 1994, A&A 287, 1
- Bartsch, A., Schneider, P. & Bartelmann, M. 1997, A&A 319, 375
- Beckwith, S., Somerville, R. & Stiavelli, M. 2003, STScI Newsletter 20, Issue 04, p.1
- Benabed, K. & van Waerbeke, L. 2003, astro-ph/0306033
- Benítez, N. 2000, ApJ 536, 571
- Benítez, N. & Martínez-González, E. 1995, ApJ 448, L89
- Benítez, N., Sanz, J.L. & Martínez-González, E. 2001, MNRAS 320, 241
- Bennett, C.L., Halpern, M., Hinshaw, G. et al. 2003, ApJS 148, 1
- Bergmann, A.G. & Petrosian, V. 1993, ApJ 413, 18
- Bernardeau, F. 1998, A&A 338, 375
- Bernardeau, F., Mellier, Y. & van Waerbeke, L. 2002, A&A 389, L28
- Bernardeau, F. & Valageas, P. 2000, A&A 364, 1
- Bernardeau, F., van Waerbeke, L. & Mellier, Y. 1997, A&A 322, 1
- Bernardeau, F., van Waerbeke, L., & Mellier, Y. 2003, A&A 397, 405
- Bernstein, G.M. & Jarvis, M. 2002, AJ 123, 583
- Bertin, E. & Arnouts, S. 1996, A&AS 117, 393
- Blain, A.W., Kneib, J.-P., Ivison, R.J. & Smail, I. 1999, ApJ 512, L87
- Blandford, R. & Narayan, R. 1986, ApJ 310, 568
- Blandford, R.D. & Narayan, R. 1992, ARA&A 30, 311
- Blandford, R.D., Saust, A.B., Brainerd, T.G. & Villumsen, J.V. 1991, MNRAS 251, 600
- Bolzonella, M., Miralles, J.-M. & Pelló, R. 2000, A&A 363, 476
- Bonnet, H. & Mellier, Y. 1995, A&A 303, 331
- Bradac, M., Lombardi, M. & Schneider, P. 2004, A&A 424, 13
- Brainerd, T.G., Blandford, R.D., & Smail, I. 1996, ApJ 466, 623
- Bridle, S.L., Hobson, M.P., Lasenby, A.N. & Saunders, R. 1998, MNRAS 299, 895
- Broadhurst, T.J., Taylor, A.N. & Peacock, J.A. 1995, ApJ 438, 49
- Brown, M.L., Taylor, A.N., Bacon, D.J. et al. 2003, MNRAS 341, 100
- Carlberg, R.G., Yee, H.K.C. & Ellingson, E. 1994, ApJ 437, 63
- Catelan, P., Kamionkowski, M. & Blandford, R.D. 2001, MNRAS 320, L7

- Clowe, D., De Lucia, G. & King, L. 2004a, MNRAS 350, 1038
 Clowe, D., Luppino, G.A., Kaiser, N. & Gioia, I.M. 2000, ApJ 539, 540
 Clowe, D., Luppino, G.A., Kaiser, N., Henry, J.P., & Gioia, I.M. 1998, ApJ 497, L61
 Clowe, D. & Schneider, P. 2001, A&A 379, 384
 Clowe, D. & Schneider, P. 2002, A&A 395, 385
 Clowe, D., Schneider, P., Aragon-Salamanca, A. et al. 2004, A&A, submitted
 Connolly, A.J., Csabai, I., Szalay, A.S., Koo, D.C., Kron, R.G., & Munn, J.A. 1995, AJ 110, 2655
 Contaldi, C.R., Hoekstra, H. & Lewis, A. 2003, PhRvL 90, 221303
 Cooray, A. & Hu, W. 2001, ApJ 548, 7
 Cooray, A. & Hu, W. 2002, ApJ 574, 19
 Cooray, A. & Sheth, R. 2002, Phys. Rep. 372, 1
 Coté, P., McLaughlin, D.E., Cohen, J.G. & Blakeslee, J.P. 2003, ApJ 591, 850
 Crittenden, R.G., Natarajan, P., Pen, U.-L. & Theuns, T. 2001, ApJ 559, 552
 Crittenden, R.G., Natarajan, P., Pen, U.-L. & Theuns, T. 2002, ApJ 568, 20
 Croft, R.A.C. & Metzler, C.A. 2001, ApJ 545, 561
 Czoske, O., Kneib, J.-P., Soucail, G., Bridges, T.J., Mellier, Y. & Cuillandre, J.-C. 2001, A&A 372, 391
 Czoske, O., Moore, B., Kneib, J.-P. Soucail, G. 2002, A&A 386, 31
 Dahle, H., Pedersen, K., Lilje, P.B., Maddox, S.J. & Kaiser, N. 2003, ApJ 591, 662
 Dalal, N., Holder, G. & Hennawi, J.F. 2003, astro-ph/0310306
 de Blok, W.J.G., McGaugh, S.S. & Rubin, V.C. 2001, AJ 122, 2396
 De Filippis, E., Schindler, S. & Castillo-Morales, A. 2003, A&A 404, 63
 Dekel, A. & Lahav, O. 1999, ApJ 520, 24
 Dietrich, J.P., Clowe, D.I. & Soucail, G. 2002, A&A 394, 395
 Dietrich, J.P., Schneider, P., Clowe, D., Romano-Diaz, E. & Kerp, J. 2004, A&A submitted (also astro-ph/0406541)
 Donahue, M., Gaskin, J.A., Patel, S.K., Joy, M., Clowe, D. & Hughes, J.P. 2003, ApJ 598, 190
 Doré, O., Bouchet, F.R., Mellier, Y. & Teyssier, R. 2001, A&A 375, 14
 Dye, S. & Taylor, A. 1998, MNRAS 300, L23
 Dye, S., Taylor, A.N., Greve, T.R. et al. 2002, A&A 386, 12
 Dye, S., Taylor, A.N., Thommes, E.M. et al. 2001, MNRAS 321, 685
 Ebbels, T., Ellis, R., Kneib, J.-P. et al. 1998, MNRAS 295, 75
 Ellis, R.S. 1997, ARA&A 35, 389
 Erben, T., Miralles, J.M., Clowe, D. et al. 2003, A&A 410, 45
 Erben, T., van Waerbeke, L., Mellier, Y., Schneider, P., Cuillandre, J.C., Castander, F.J. & Dantel-Fort, M. 2000, A&A 355, 23
 Erben, T., van Waerbeke, L., Bertin, E., Mellier, Y. & Schneider, P. 2001, A&A 366, 717
 Ettori, S. & Lombardi, M. 2003, A&A 398, L5
 Fabian, A.C. 1994, ARA&A 32, 277
 Fahlman, G., Kaiser, N., Squires, G. & Woods, D. 1994, ApJ 437, 56
 Fischer, P. 1999, AJ 117, 2024
 Fischer, P., McKay, T.A., Sheldon, E. et al. 2000, AJ 120, 1198
 Fluke, C.J., Webster, R.L. & Mortlock, D.J. 2002, MNRAS 331, 180
 Fort, B. & Mellier, Y. 1994, A&AR 5, 239
 Fort, B., Mellier, Y. & Dantel-Fort, M. 1997, A&A 321, 353

- Fort, B., Mellier, Y., Dantel-Fort, M., Bonnet, H. & Kneib, J.-P. 1996, A&A 310, 705
- Fort, B., Prieur, J.L., Mathez, G., Mellier, Y. & Soucail, G. 1988, A&A 200, L17
- Fruchter, A.S. & Hook, R.N. 2002, PASP 114, 144
- Fugmann, W. 1990, A&A 240, 11
- Gavazzi, R., Mellier, Y., Fort, B., Cuillandre, J.-C. & Dantel-Fort, M. 2004, A&A 422, 407
- Gautret, L., Fort, B., & Mellier, Y. 2000, A&A 353, 10
- Geiger, B. & Schneider, P. 1998, MNRAS 295, 497
- Geiger, B. & Schneider, P. 1999, MNRAS 302, 118
- Gerhard, O., Kronawitter, A., Saglia, R.P. & Bender, R. 2001, AJ 121, 1936
- Giavalisco, M. & Mobasher, B. 2004, STScI Newsletter 21, Issue 01, p.1
- Gioia, I.M., Braito, V., Branchesi, M. et al. 2004, A&A 419, 517
- Gladders, M.D., Hoekstra, H., Yee, H.K.C., Hall, P.B. & Barrientos, L.F. 2003, ApJ 593, 48
- Gladders, M.D. & Yee, H.K.C. 2000, AJ 120, 2148
- Golse, G., Kneib, J.-P. & Soucail, G. 2002, A&A 387, 788
- Gray, M.E., Ellis, R.S., Refregier, A. et al. 2000, MNRAS 318, 573
- Gray, M., Ellis, R.S., Lewis, J.R., McMahon, R.G. & Firth, A.E. 2001, MNRAS 325, 111
- Gray, M., Taylor, A.N., Meisenheimer, K., Dye, S., Wolf, C. & Thommes, E. 2002, ApJ 568, 141
- Gunn, J.E. 1967, ApJ 150, 737
- Guzik, J. & Seljak, U. 2001, MNRAS 321, 439
- Hamana, T. 2001, MNRAS 326, 326
- Hamana, T., Colombi, S.T., Thion, A., Devriendt, J.E.G.T., Mellier, Y. & Bernardeau, F. 2002, MNRAS 330, 365
- Hamana, T. & Mellier, Y. 2001, MNRAS 327, 169
- Hamana, T., Miyazaki, S., Shimasaku, K. et al. 2003, ApJ 597, 98
- Hamana, T., Takada, M. & Yoshida, N. 2004, MNRAS 350, 893
- Hämmerle, H., Miralles, J.-M., Schneider, P. et al. 2002, A&A 385, 743
- Heavens, A. 2003, MNRAS 343, 1327
- Heavens, A.F., Refregier, A. & Heymans, C.E.C. 2000, MNRAS 319, 649
- Hennawi, J.F. & Spergel, D.N. 2003, astro-ph/0404349
- Hetterscheidt, M. 2003, Diploma Thesis, Univ. Bonn.
- Heymans, C., Brown, M., Heavens, A., Meisenheimer, K., Taylor, A. & Wolf, C. 2004, MNRAS 347, 895
- Heymans, C. & Heavens, A. 2003, MNRAS 339, 711
- Hirata, C.M., Mandelbaum, R., Seljak, U. et al. 2004, astro-ph/0403255
- Hirata, C.M. & Seljak, U. 2004, astro-ph/0406275
- Ho, S. & White, M. 2004, ApJ 607, 40
- Hoekstra, H. 2001, A&A 370, 743
- Hoekstra, H. 2003, MNRAS 339, 1155
- Hoekstra, H. 2004, MNRAS 347, 1337
- Hoekstra, H., Franx, M. & Kuijken, K. 2000, ApJ 532, 88
- Hoekstra, H., Franx, M., Kuijken, K., et al. 2001, ApJ 548, L5
- Hoekstra, H., Franx, M., Kuijken, K., Carlberg, R.G. & Yee, H.K.C. 2003, MNRAS 340, 609
- Hoekstra, H., Franx, M., Kuijken, K., & Squires, G. 1998, ApJ 504, 636

- Hoekstra, H., Franx, M., Kuijken, K., & van Dokkum, P.G. 2002d, MNRAS 333, 911
- Hoekstra, H., van Waerbeke, L., Gladders, M.D., Mellier, Y. & Yee, H.K.C. 2002c, ApJ 577, 604
- Hoekstra, H., Yee, H.K.C. & Gladders, M.D. 2001, ApJ 558, L11
- Hoekstra, H., Yee, H.K.C. & Gladders, M.D. 2002a, ApJ 577, 595
- Hoekstra, H., Yee, H.K.C. & Gladders, M.D. 2004, ApJ 606, 67
- Hoekstra, H., Yee, H.K.C., Gladders, M.D., Barrientos, L.F., Hall, P.B. & Infante, L. 2002b, ApJ 572, 55
- Hu, W. 1999, ApJ 522, L21
- Hu, W. 2002, PhRvD 66, 083515
- Hu, W. & Keeton, C.R. 2002, PhRvD 66, 063506
- Hu, W. & Jain, B. 2003, astro-ph/0312395
- Hu, W. & Tegmark, M. 1999, ApJ 514, L65
- Hu, W. & White, M. 2001, ApJ 554, 67
- Huterer, D. 2002, PhRvD 64, 063001
- Jain, B., Scranton, R. & Sheth, R.K. 2003, MNRAS 345, 62
- Jain, B. & Seljak, U. 1997, ApJ 484, 560
- Jain, B., Seljak, U. & White, S.D.M. 2000, ApJ 530, 547
- Jain, B. & Taylor, A. 2003, PRL 91, 141302
- Jain, B. & van Waerbeke, L. 2000, ApJ 530, L1
- Jarvis, M., Bernstein, G.M., Fischer, P. et al. 2003, AJ 125, 1014
- Jarvis, M., Bernstein, G.M. & Jain, B. 2004, MNRAS 352, 338
- Jing, Y.P. 1998, ApJ 503, L9
- Jing, Y.P. 2002, MNRAS 335, L89
- Kaiser, N. 1984, ApJ 284, L9
- Kaiser, N. 1987, MNRAS 227, 1
- Kaiser, N. 1992, ApJ 388, 272
- Kaiser, N. 1995, ApJ 439, L1
- Kaiser, N. 1998, ApJ 498, 26
- Kaiser, N. 2000, ApJ 537, 555
- Kaiser, N., Squires, G., 1993, ApJ 404, 441
- Kaiser, N., Squires, G. & Broadhurst, T. 1995, ApJ 449, 460
- Kaiser, N., Squires, G., Fahlman, G. & Woods, D. 1994, in *Clusters of galaxies, proc. XIVth Moriond astrophysics meeting, Méribel, France, 1994*, p. 269
- Kaiser, N., Wilson, G. & Luppino, G. 2000, astro-ph/0003338
- Kaiser, N., Wilson, G., Luppino, G. et al. 1998, astro-ph/9809268
- Kilbinger, M. & Schneider, P. 2004, A&A 413, 465
- King, L.J., Clowe, D.I., Lidman, C., Schneider, P., Erben, T., Kneib, J.-P. & Meylan, G. 2002, A&A 385, L5
- King, L.J., Clowe, D.I. & Schneider, P. 2002, A&A 383, 118
- King, L. & Schneider, P. 2001, A&A 369, 1
- King, L. & Schneider, P. 2002, A&A 396, 411
- King, L. & Schneider, P. 2003, A&A 398, 23
- King, L., Schneider, P. & Springel, V. 2001, A&A 378, 748
- Kleinheinrich, M. 2003, Ph.D. Thesis, University of Bonn
- Kneib, J.-P., Ellis, R.S., Santos, M.R. & Richard, J. 2004, ApJ 607, 697
- Kneib, J.-P., Ellis, R.S., Smail, I., Couch, W.J. & Sharples, R.M. 1996, ApJ 471, 643

- Kneib, J.-P., Hudelot, P., Ellis, R.S. et al. 2003, *ApJ* 598, 804
 Kochanek, C.S. 1990, *MNRAS* 247, 135
 Kruse, G. & Schneider, P. 1999, *MNRAS* 302, 821
 Kruse, G. & Schneider, P. 2000, *MNRAS* 318, 321
 Kuijken, K. 1999, *A&A* 352, 355
 Limber, D.N. 1953, *ApJ* 117, 134
 Link, R. & Pierce, M.J. 1998, *ApJ* 502, 63
 Lombardi, M. & Bertin, G. 1998, *A&A* 335, 1
 Lombardi, M. & Bertin, G. 1999, *A&A* 342, 337
 Lombardi, M. & Schneider, P. 2001, *A&A* 373, 359
 Lombardi, M. & Schneider, P. 2002, *A&A* 392, 1153
 Luppino, G.A., Gioia, I.M., Hammer, F., Le Fèvre, O. & Annis, J.A. 1999, *A&AS* 136, 117
 Luppino, G.A. & Kaiser, N. 1997, *ApJ* 475, 20
 Lynds, R. & Petrosian, V. 1989, *ApJ* 336, 1
 Mackey, J., White, M. & Kamionkowski, M. 2002, *MNRAS* 332, 788
 Maoli, R., van Waerbeke, L., Mellier, Y. et al. 2001, *A&A* 368, 766
 Markevitch, M., Gonzalez, A.H., Clowe, D. et al. 2004, *ApJ* 606, 819
 Marshall, P.J., Hobson, M.P., Gull, S.F. & Bridle, S.L. 2002, *MNRAS* 335, 1037
 Marshall, P.J., Hobson, M.P. & Slosar, A. 2003, *MNRAS* 346, 489
 McKay, T.A., Sheldon, E.S., Racusin, J. et al. 2001, *astro-ph/0108013*
 Mellier, Y. 1999, *ARA&A* 37, 127
 Mellier, Y., Fort, B., Soucail, G., Mathez, G. & Cailloux, M. 1991, *ApJ* 380, 334
 Ménard, B. & Bartelmann, M. 2002, *A&A* 386, 784
 Ménard, B., Bartelmann, M. & Mellier, Y. 2003, *A&A* 409, 411
 Ménard, B. & Péroux, C. 2003, *A&A* 410, 33
 Meneghetti, M., Bartelmann, M., Dolag, K. et al. 2004, *astro-ph/0405070*
 Meneghetti, M., Yoshida, N., Bartelmann, M. et al. 2001, *MNRAS* 325, 435
 Metzler, C.A., White, M. & Loken, C. 2001, *ApJ* 547, 560
 Miralda-Escudé, J. 1991, *ApJ* 380, 1
 Miralles, J.-M., Erben, T., Hämmerle, H. et al. 2002, *A&A* 388, 68
 Miralles, J.-M., Erben, T., Hämmerle, H. et al. 2003, *astro-ph/0310716*
 Miyazaki, S., Hamana, T., Shimasaku, K. et al. 2002, *ApJ* 580, L97
 Mould, J., Blandford, R., Villumsen, J., Brainerd, T., Smail, I., Small, T., & Kells, W. 1994, *MNRAS* 271, 31
 Mulchaey, J.S., Dressler, A. & Oemler, A. 2004, editors, "Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution", Cambridge University Press (in press).
 Munchi, D. & Jain, B. 2001, *MNRAS* 322, 107
 Munshi, D., Valageas, P. & Barber, A.J. 2004, *MNRAS* 350, 77
 Munshi, D. & Wang, Y. 2003, *ApJ* 583, 566
 Narayan, R. & Nityananda, R. 1986, *ARA&A* 24, 127
 Natarajan, P. & Kneib, J.-P. 1997, *MNRAS* 287, 833
 Natarajan, P., Kneib, J.-P. & Smail, I. 2002, *ApJ* 580, L11
 Natarajan, P., Kneib J.-P., Smail, I. & Ellis, R.S. 1998, *Ap* 499, 600
 Natarajan, P. & Refregier, A. 2000, *ApJ* 538, L113
 Norberg, P., Baugh, C.M., Hawkins, E. et al. 2001, *MNRAS* 328, 64
 Oguri, M., Lee, J. & Suto, Y. 2003, *ApJ* 599, 7
 Ota, N., Pointecouteau, E., Hattori, M. & Mitsuda, K. 2004, *ApJ* 601, 120

- Peacock, J.A. & Dodds, S.J. 1996, MNRAS 280, L19
 Pelló, R., Kneib, J.P.; Le Borgne, J.F. et al. 1999, A&A 346, 359
 Pelló, R., Schaeerer, D., Richard, J., Le Borgne, J.-F. & Kneib, J.-P. 2004, A&A 416, L35
 Pen, U.-L., Lu, T., van Waerbeke, L. & Mellier, Y. 2003, MNRAS 346, 994
 Pen, U.-L., Zhang, T., van Waerbeke, L., Mellier, Y., Zhang, P. & Dubinski, J. 2003, ApJ 592, 664
 Pierpaoli, E., Scott, D. & White, M. 2001, MNRAS 325, 77
 Premadi, P., Martel, H., Matzner, R. & Futamase, T. 2001, ApJS 135, 7
 Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. 1986, Numerical Recipes, (Cambridge: University Press)
 Reblinsky, K. 2000, A&A 364, 377
 Reblinsky, K. & Bartelmann, M. 1999, A&A 345, 1
 Reblinsky, K., Kruse, G., Jain, B. & Schneider, P. 1999, A&A 351, 815
 Refregier, A. 2003a, ARA&A 41, 645
 Refregier, A. 2003b, MNRAS 338, 35
 Refregier, A. & Bacon, D. 2003, MNRAS 338, 48
 Refregier, A. Rhodes, J. & Groth, E.J. 2002, ApJ 572, L131
 Refsdal, S. 1970, ApJ 159, 357
 Rhodes, J., Refregier, A., Collins, N.R., Gardner, J.P., Groth, E.J. & Hill, R.S. 2004, ApJ 605, 29
 Richards, E.A., Kellermann, K.I., Fomalont, E.B., Windhorst, R.A. & Partridge, R.B. 1998, AJ 116, 1039
 Rix, H.-W., Barden, M., Beckwith, S.V.W. et al. 2004, ApJS 152, 163
 Rögnvaldsson, Ö.E., Greve, T.R., Hjorth, J. et al. 2001, MNRAS 322, 131
 Rodrigues-Williams, L.L. & Hogan, C.J. 1994, AJ 107, 451
 Sand, D.J., Treu, T. & Ellis, R.S. 2002, ApJ 574, L129
 Sand, D.J., Treu, T., Smith, G.P. & Ellis, R.S. 2004, ApJ 604, 88
 Sarazin, C.L. 1986, RvMP 58, 1
 Schirmer, M. 2004, PhD Dissertation, Univ. Bonn
 Schneider, P. 1995, A&A 302, 639
 Schneider, P. 1996, MNRAS 283, 83
 Schneider, P. 1997, MNRAS 292, 673
 Schneider, P. 1998, ApJ 498, 43
 Schneider, P. 2003, A&A 408, 829
 Schneider, P. 2004, this volume (IN)
 Schneider, P. & Bartelmann, M. 1997, MNRAS 286, 696
 Schneider, P., Ehlers, J. & Falco, E.E. 1992, Gravitational Lenses (New York: Springer) (SEF)
 Schneider, P., Kilbinger, M. & Lombardi, M. 2004a, A&A submitted (also astro-ph/0308328)
 Schneider, P., King, L. & Erben, T. 2000, A&A 353, 41
 Schneider, P. & Kneib, J.-P. 1998, in: *The Next Generation Space Telescope*, Proceedings of the 34th Liège International Astrophysics Colloquium, ESA Sp-429, p.89
 Schneider, P. & Lombardi, M. 2003, A&A 397, 809
 Schneider, P. & Rix, H.-W. 1997, ApJ 474, 25
 Schneider, P. & Seitz, C. 1995, A&A 294, 411
 Schneider, P., van Waerbeke L., Jain B., Kruse G., 1998a, MNRAS, 296, 873

- Schneider, P., van Waerbeke, L. & Mellier, Y. 2002a, A&A 389, 729
 Schneider, P., van Waerbeke, L., Mellier, Y., Jain, B., Seitz, S. & Fort, B. 1998b, A&A 333, 767
 Schneider, P., van Waerbeke, L., Kilbinger, M. & Mellier, Y. 2002b, A&A 396, 1
 Schneider, P. & Watts, P. 2004, A&A submitted
 Schneider, P. & Weiss, A. 1988a, ApJ 327, 526
 Schneider, P. & Weiss, A. 1988b, ApJ 330, 1
 Schrabback, T. 2004, Diploma Thesis, Univ. Bonn
 Schramm, T. & Kayser, R. 1995, A&A 299, 1
 Schuecker, P., Böhringer, H., Collins, C.A. & Guzzo, L. 2003, A&A 398, 867
 Scoccimarro, R. & Couchman, H. 2001, MNRAS 325, 1312
 Seitz, C., Kneib, J.-P., Schneider, P. & Seitz, S. 1996, A&A 314, 707
 Seitz, C. & Schneider, P. 1995a, A&A 297, 287
 Seitz, C. & Schneider, P. 1997, A&A 318, 687
 Seitz, S., Saglia, R., Bender, R., Hopp, U., Belloni, P. & Ziegler, B. 1998, MNRAS 298, 325
 Seitz, S. & Schneider, P. 1995b, A&A 302, 9
 Seitz, S. & Schneider, P. 1996, A&A 305, 383
 Seitz, S. & Schneider, P. 2001, A&A 374, 740
 Seitz, S., Schneider, P. & Bartelmann, M. 1998, A&A 337, 325
 Seitz, S., Schneider, P. & Ehlers, J. 1994, Class. Quantum Grav. 11, 2345
 Seljak, U. 1998, ApJ 506, 64
 Seljak, U. 2002, MNRAS 337, 769
 Seljak, U., Makarov, A., Mandelbaum, R. et al. 2004, astro-ph/0406594
 Sheldon, E.S., Johnston, D.E., Frieman, J.A. et al. 2004, AJ 127, 2544
 Sheth, R.K., Mo, H.J. & Tormen, G. 2001, MNRAS 323, 1
 Simon, P., King, L.J. & Schneider, P. 2004, A&A 417, 873
 Smith, R.E., Peacock, J.A., Jenkins, A. et al. 2003, MNRAS 341, 1311
 Sofue, Y. & Rubin, V. 2001, ARA&A 39, 137
 Song, Y.-S. & Knox, L. 2003, astro-ph/0312175
 Soucail, G., Fort, B., Mellier, Y. & Picat, J.P. 1987, A&A 172, L14
 Soucail, G., Kneib, J.-P. & Golse, G. 2004, A&A 417, L33
 Soucail, G., Ota, N., Böhringer, H., Czoske, O., Hattori, M. & Mellier, Y. 2000, A&A 355, 433
 Spergel, D.N. & Steinhardt, P.J. 2000, PRL 84, 3760
 Spergel, D.N., Verde, L., Peiris, H.V. et al. 2003, ApJS 148, 175
 Squires, G. & Kaiser, N. 1996, ApJ 473, 65
 Takada, M. & Hamana, T. 2003, MNRAS 346, 949
 Takada, M. & Jain, B. 2002, MNRAS 337, 875
 Takada, M. & Jain, B. 2003a, MNRAS 340, 580
 Takada, M. & Jain, B. 2003b, MNRAS 344, 857
 Takada, M. & Jain, B. 2004, MNRAS 348, 897
 Takada, M. & White, M. 2004, ApJ 601, L1
 Taruya, A., Takada, M., Hamana, T., Kayo, I. & Futamase, T. 2002, ApJ 571, 638
 Taylor, A.N. 2001, astro-ph/0111605
 Taylor, A.N., Bacon, D.J., Gray, M.E. et al. 2004, astro-ph/0402095
 Taylor, A.N., Dye, S., Broadhurst, T.J., Benítez, N. & van Kampen, E. 1998, ApJ 501, 539
 Tegmark, M. & Peebles, P.J.E. 1998, ApJ 500, L79

- Tereno, I., Doré, O., van Waerbeke, L. & Mellier, Y. 2004, astro-ph/0404317
- Torri, E., Meneghetti, M., Bartelmann, M., Moscardini, L., Rasia, E. & Tormen, G. 2004, MNRAS 349, 476
- Trager, S.C., Faber, S.M., Dressler, A. & Oemler, A. 1997, ApJ 485, 92
- Trentham, N. & Tully, R.B. 2002, MNRAS 335, 712
- Tyson, J.A. 1988, AJ 96, 1
- Tyson, J.A. & Fischer, P. 1995, ApJ 446, L55
- Tyson, J.A., Valdes, F., Jarvis, J.F. & Mills Jr., A.P. 1984, ApJ 281, L59
- Tyson, J.A., Valdes, F. & Wenk, R.A. 1990, ApJ 349, L1
- Umetsu, K. & Futamase, T. 2000, ApJ 539, L5
- Valdes, F., Jarvis, J.F. & Tyson, J.A. 1983, ApJ 271, 431
- Vale, C. & White, M. 2003, ApJ 592, 699
- van den Bosch, F.C., Abel, T., Croft, R.A.C., Hernquist, L. & White, S.D.M. 2002, ApJ 576, 21
- van Kampen, E. 1998, MNRAS 301, 389
- van Waerbeke, L. 1998, A&A 334, 1
- van Waerbeke, L. 2000, MNRAS 313, 524
- van Waerbeke, L., Bernardeau, F. & Mellier, Y. 1999, A&A 243, 15
- van Waerbeke, L., Hamana, T., Scoccimarro, R., Colombi, S. & Bernardeau, F. 2001, MNRAS 322, 918
- van Waerbeke, L. & Mellier, Y. 2003, astro-ph/0305089
- van Waerbeke, L., Mellier, Y., Erben, T. et al. 2000, A&A 358, 30
- van Waerbeke, L., Mellier, Y. & Hoekstra, H. 2004, astro-ph/0406468
- van Waerbeke, L., Mellier, Y., Pello, R., et al. 2002, A&A 393, 369
- van Waerbeke, L., Mellier, Y., Radovich, M. et al. 2001, A&A 374, 757
- van Waerbeke, L., Mellier, Y., Schneider, P., Fort, B. & Mathez, G. 1997, A&A 317, 303
- von der Linden, A. 2004, Diploma Thesis, Univ. Bonn
- Wambsganss, J., Bode, P. & Ostriker, J.P. 2004, ApJ 606, L93
- Wambsganss, J., Cen, R. & Ostriker, J.P. 1998, ApJ 494, 29
- Weinberg, N.N. & Kamionkowski, M. 2002, MNRAS 337, 1269
- Weinberg, N.N. & Kamionkowski, M. 2003, MNRAS 341, 251
- White, M. & Hu, W. 2000, ApJ 537, 1
- White, M., van Waerbeke, L. & Mackey, J. 2002, ApJ 575, 640
- White, S.D.M., Navarro, J.F., Evrard, A.E. & Frenk, C.S. 1993, Nat 366, 429
- Williams, R.E., Blacker, B., Dickinson, M. et al. 1996, AJ 112, 1335
- Wilson, G., Kaiser, N. and Luppino, G. 2001, ApJ 556, 601
- Wittman, D. 2002, in Lecture Notes in Physics 608, Gravitational Lensing: An Astrophysical Tool, ed. F. Courbin & D. Minniti (Berlin: Springer), 55
- Wittman, D., Margoniner, V.E., Tyson, J.A., Cohen, J.G., Becker, A.C. & Dell'Antonio, I. 2003, ApJ 597, 218
- Wittman, D.M., Tyson, J.A., Kirkman, D., Dell'Antonio, I. & Bernstein, G. 2000, Nat 405, 143
- Wittman, D., Tyson, J.A., Margoniner, V.E., Cohen, J.G. & Dell'Antonio, I. 2001, ApJ 557, L89
- Yang, X.H., Mo, H.J., Kauffmann, G. & Chu, Y.Q. 2003, MNRAS 339, 387
- York, D.G., Adelman, J., Anderson, J.E. et al. 2000, AJ 120, 1579
- Zaldarriaga, M. & Scoccimarro, R. 2003, ApJ 584, 559
- Zaritsky, D. & Gonzalez, A.H. 2003, ApJ 584, 691

- Zaritsky, D., Smith, R., Frenk, C. & White, S.D.M. 1997, ApJ 478, 39
Zaroubi, S., Squires, G., Hoffman, Y. & Silk, J. 1998, ApJ 500, L87
Zaroubi, S., Squires, G., de Gasperis, G., Evrard, A.E., Hoffman, Y. & Silk, J. 2001,
ApJ 561, 600
Zehavi, I., Blanton, M.R., Frieman, J.A. et al. 2002, ApJ 571, 172
Zhang, L.L. & Pen, U.-L. 2003, astro-ph/0305447
Zwicky, F. 1933, Helv. Phys. Acta 6, 110

Hot Gas in Galaxy Clusters: Theory and Simulations

MICHAEL L. NORMAN⁽¹⁾

⁽¹⁾ *Physics Department and CASS, UC San Diego, USA*

Summary. — We review the theory of the formation of galaxy clusters and discuss their role as cosmological probes. We begin with the standard cosmological framework where we discuss the origin of the CDM matter power spectrum and the growth of density fluctuations in the linear regime. We then summarize the spherical top-hat model for the nonlinear growth of fluctuations from which scaling relations and halo statistics are derived. Numerical methods for simulating gas in galaxy clusters are then overviewed with an emphasis on multiscale hydrodynamic simulations of cluster ensembles. Results of hydrodynamic AMR simulations are described which compare cluster internal and statistical properties as a function of their assumed baryonic processes. Finally, we compare various methods of measuring cluster masses using X-ray and the thermal Sunyaev-Zeldovich effect (SZE). We find that SZE offers great promise for precision measurements in raw samples of high-z clusters.

1. – Introduction

The Sunyaev-Zeldovich Effect (SZE) detectable in galaxy clusters has emerged as a powerful new probe of the low to intermediate redshift universe (see articles by Birkingshaw & Rephaeli in this volume, as well the review by Carlstrom, et al. [1]. Within the prevailing theory of cosmological structure formation, galaxy clusters form in rare, massive peaks of the cosmic density field. Because of natural biasing, such regions get a “head start” on structure formation on all scales smaller than the cluster scale. As a consequence, galaxy clusters at the present epoch contain the oldest objects in the universe in an evolutionary sense [2]. This makes galaxy clusters intrinsically interesting as astrophysical objects, worthy of study observationally, theoretically, and computationally.

However, much of the current interest stems from the potential use of galaxy clusters as cosmological probes. As discussed in more detail below, the space density of galaxy clusters as a function of cosmological redshift is sensitive to the RMS mass fluctuations on scales of $10^{14-15} M_{\odot}$, which depends on Ω_m , the mean mass density of the universe, and to a lesser extent, Ω_{de} , the dark energy density of the universe. Attempts to deduce

Ω_m based on X-ray surveys have met with some success [3], but they have been hampered by the fact that at these wavebands cluster samples become sparse at $z>1$ owing to their low surface brightness. Because the SZE is intrinsically redshift independent, one has the possibility of detecting clusters over a wide range of redshifts. Blind surveys with sufficient sensitivity can in principle detect clusters from $z=0$ to their formation redshift $z \leq 1.5$ [1], paving the way for more precise cosmological parameter measurements. Follow-up pointed observations of a large sample of galaxy clusters over a range of redshifts would enable a detailed study of their formation and evolution. Such studies would confirm or modify our theory of structure formation, improve our understanding of galaxy evolution, and reveal a great deal about the complex physical processes operating in the intracluster medium (ICM).

This paper summarizes four lectures the author delivered at the Varenna Summer School entitled “Background Microwave Radiation and Intracaluster Cosmology”, held July 2004 in Varenna, Italy. Originally, the organizers asked me to deliver three lectures covering numerical simulations of galaxy clusters, as well as to review the basics of cosmological structure formation, of which galaxy clusters are just one aspect. The first lecture of the school was to have been given by Dr. Rocky Kolb on the cosmological standard model and the linear growth of density perturbations. When he was unable to attend the school, that responsibility fell to me, increasing my task to four lectures. Fortunately, Dr. Kolb’s lecture slides were made available to me, which I used verbatim. The following Section 2 follows closely the content and organization of Dr. Kolb’s lecture notes, while Sections 3-5 are my own. Section 3 reviews key concepts and results from structure formation theory that provide the vocabulary and framework for interpreting observations and simulations of galaxy clusters. Section 4 discusses the technical challenges associated with simulating gas in galaxy clusters and reviews the numerical methods we have employed. Section 5 presents results of numerical simulations of statistical ensembles of galaxy clusters whose goal is to understand how observables such as X-ray luminosity, emission-weighted temperature, and SZE depend on cluster mass and baryonic physics.

In line with the character of the summer school, I have attempted to be pedagogical, emphasizing the key concepts and results that a student needs to know if he/she wants to understand the current literature or do research in this area. Literature citations are kept to a minimum, except for textbooks, reviews, and research papers that I found to be particularly helpful in preparing this article.

2. – Cosmological framework and perturbation growth in the linear regime

Our modern theory of the structure and evolution of the universe, along with the observational data which support it, is admirably presented in a recent textbook by Dodelson [4]. Remarkable observational progress has been made in the past two decades which has strengthened our confidence in the correctness of the hot, relativistic, expanding universe model (Big Bang), has measured the universe’s present mass-energy contents and kinematics, and lent strong support to the notion of a very early, inflationary phase. Moreover, observations of high redshift supernovae unexpectedly have revealed that the cosmic expansion is accelerating at the present time, implying the existence of a pervasive, dark energy field with negative pressure [5]. This surprising discovery has enlivened observational efforts to accurately measure the cosmological parameters over as large a fraction of the age of the universe as possible, especially over the redshift interval $0 < z < 1.5$ which, according to current estimates, spans the deceleration-acceleration transition.

These efforts include large surveys of galaxy large scale structure, galaxy clusters, weak lensing, the Lyman alpha forest, and high redshift supernovae, all of which span the relevant redshift range. Except for the supernovae, all other techniques rely on measurements of cosmological structure in order to deduce cosmological parameters.

2.1. Cosmological standard model. – The dynamics of the expanding universe is described by the two Friedmann equations derived from Einstein’s theory of general relativity under the assumption of homogeneity and isotropy. The expansion rate at time t is given by

$$(1) \quad H^2(t) \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3} \sum_i \rho_i - \frac{k}{a^2} + \frac{\Lambda}{3}$$

where $H(t)$ is the Hubble parameter and $a(t)$ is the FRW scale factor at time t . The first term on the RHS is proportional to the sum over all energy densities in the universe ρ_i including baryons, photons, neutrinos, dark matter and dark energy. We have explicitly pulled the dark energy term out of the sum and placed it in the third term assuming it is a constant (the cosmological constant). The second term is the curvature term, where $k = 0, \pm 1$ for zero, positive, negative curvature, respectively. Equation (1) can be cast in a form useful for numerical integration if we introduce Ω parameters:

$$(2) \quad \Omega_i \equiv \frac{8\pi G}{3H^2} \rho_i, \quad \Omega_\Lambda \equiv \frac{8\pi G}{3H^2} \rho_\Lambda = \frac{\Lambda}{3H^2}, \quad \Omega_k \equiv \frac{-k}{(aH)^2}$$

Dividing equation (1) by H^2 we get the sum rule $1 = \Omega_m + \Omega_k + \Omega_\Lambda$, which is true at all times, where Ω_m is the sum over all Ω_i excluding dark energy. At the present time $H(t) = H_0$, $a = 1$, and cosmological density parameters become

$$(3) \quad \Omega_i(0) = \frac{8\pi G}{3H_0^2} \rho_i(0), \quad \Omega_\Lambda(0) = \frac{\Lambda}{3H_0^2}, \quad \Omega_k(0) = \frac{-k}{H_0^2}$$

Equation (1) can then be manipulated into the form

$$(4) \quad \dot{a} = H_0 [\Omega_m(0)(a^{-1} - 1) + \Omega_\gamma(0)(a^{-2} - 1) + \Omega_\Lambda(0)(a^2 - 1) + 1]^{1/2}$$

Here we have explicitly introduced a density parameter for the background radiation field Ω_γ and used the fact that matter and radiation densities scale as a^{-3} and a^{-4} , respectively, and we have used the sum rule to eliminate Ω_k . Equation (4) is equation (1) expressed in terms of the *current* values of the density and Hubble parameters, and makes explicit the scale factor dependence of the various contributions to the expansion rate. In particular, it is clear that the expansion rate is dominated first by radiation, then by matter, and finally by the cosmological constant.

Current measurements of the cosmological parameters by different techniques [6] yield the following numbers [(0) notation suppressed]:

$$\begin{aligned} h &\equiv H_0/(100 km/s/Mpc) \approx 0.72 \\ \Omega_{total} &\approx 1, \quad \Omega_\Lambda \approx 0.73, \quad \Omega_m = \Omega_{cdm} + \Omega_b \approx 0.27, \quad \Omega_k \approx 0 \\ \Omega_b &\approx 0.04, \quad \Omega_\nu \approx 0.005, \quad \Omega_\gamma \approx 0.00005 \end{aligned}$$

This set of parameters is referred to as the concordance model [7], and describes a spatially flat, low matter density, high dark energy density universe in which baryons, neutrinos, and photons make a negligible contribution to the large scale dynamics. Most of the matter in the universe is cold dark matter (CDM) whose dynamics is discussed below. As we will also see below, baryons and photons make an important contribution to shaping of the matter power spectrum despite their small contribution to the present-day energy budget. Understanding the evolution of baryons in nonlinear structure formation is essential to interpret X-ray and SZE observations of galaxy clusters.

The second Friedmann equation relates the second time derivative of the scale factor to the cosmic pressure p and energy density ρ

$$(5) \quad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p), \quad \rho = \sum_i \rho_i = \rho_m + \rho_\gamma + \rho_\Lambda$$

p and ρ are related by an equation of state $p_i = w_i \rho_i$, with $w_m=0$, $w_\gamma=1/3$, and $w_\Lambda = -1$. We thus have

$$(6) \quad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho_m + 2\rho_\gamma - 2\rho_\Lambda).$$

Expressed in terms of the current values for the cosmological parameters we have

$$(7) \quad \frac{\ddot{a}}{a} = -\frac{1}{2}H_0^2[\Omega_m(0)a^{-3} + 2\Omega_\gamma(0)a^{-4} - 2\Omega_\Lambda(0)].$$

Evaluating equation 7 using the concordance parameters, we see the universe is currently accelerating $\ddot{a} \approx 0.6H_0^2$. Assuming the dark energy density is a constant, the acceleration began when

$$(8) \quad a \equiv \frac{1}{1+z} = \left(\frac{\Omega_m(0)}{2\Omega_\Lambda(0)} \right)^{1/3} \approx 0.57$$

or $z \sim 0.75$.

2.2. The Linear power spectrum. – Cosmic structure results from the amplification of primordial density fluctuations by gravitational instability. The power spectrum of matter density fluctuations has now been measured with considerable accuracy across roughly four decades in scale. Figure 1 shows the latest results, taken from reference [8]. Combined in this figure are measurements using cosmic microwave background (CMB) anisotropies, galaxy large scale structure, weak lensing of galaxy shapes, and the Lyman alpha forest, in order of decreasing comoving wavelength. In addition, there is a single data point for galaxy clusters, whose current space density measures the amplitude of the power spectrum on $8 \text{ h}^{-1} \text{ Mpc}$ scales [9]. Superimposed on the data is the predicted ΛCDM linear power spectrum at $z=0$ for the concordance model parameters. As one can see, the fit is quite good. In actuality, the concordance model parameters are determined by fitting the data. A rather complex statistical machinery underlies the determination of cosmological parameters, and is discussed in Dodelson (2003, Ch. 11). The fact that modern CMB and LSS data agree over a substantial region of overlap gives us confidence in the correctness of the concordance model. In this section, we define the

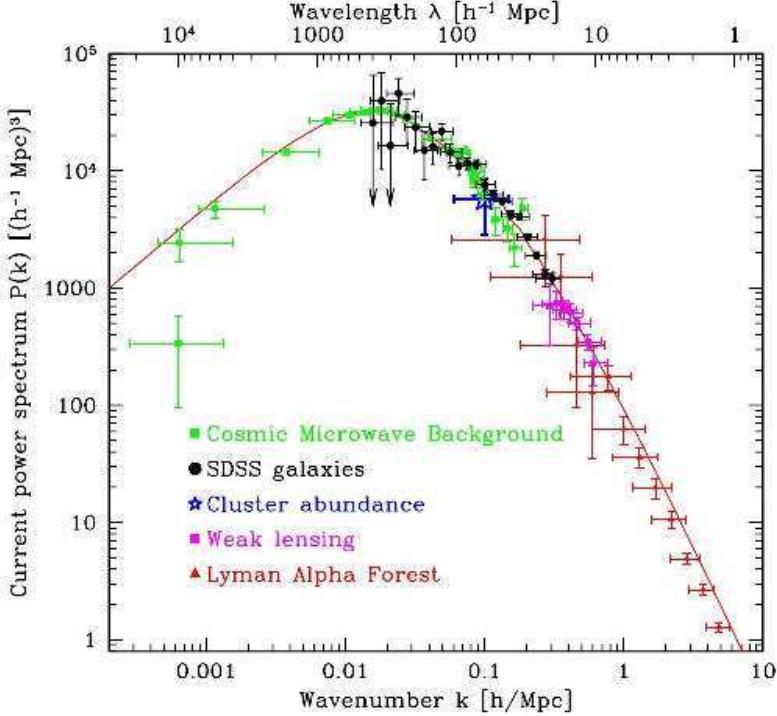


Fig. 1. – Linear matter power spectrum $P(k)$ versus wavenumber extrapolated to $z=0$, from various measurements of cosmological structure. The best fit Λ CDM model is shown as a solid line. From [8].

power spectrum mathematically, and review the basic physics which determines its shape. Readers wishing a more in depth treatment are referred to references [4, 10].

At any epoch t (or a or z) express the matter density in the universe in terms of a mean density and a local fluctuation:

$$(9) \quad \rho(\vec{x}) = \bar{\rho}(1 + \delta(\vec{x}))$$

where $\delta(\vec{x})$ is the density contrast. Expand $\delta(\vec{x})$ in Fourier modes:

$$(10) \quad \delta(\vec{x}) \equiv \frac{\rho(\vec{x}) - \bar{\rho}}{\bar{\rho}} = \int \delta(\vec{k}) \exp(-i\vec{k} \cdot \vec{x}) d^3k.$$

The autocorrelation function of $\delta(\vec{x})$ defines the power spectrum through the relations

$$(11) \quad \langle \delta(\vec{x}) \delta(\vec{x}) \rangle = \int_0^\infty \frac{dk}{k} \frac{k^3 |\delta^2(\vec{k})|}{2\pi^2} = \int_0^\infty \frac{dk}{k} \frac{k^3 P(k)}{2\pi^2} = \int_0^\infty \frac{dk}{k} \Delta^2(k)$$

where we have the definitions

$$(12) \quad P(k) \equiv |\delta^2(\vec{k})|, \text{ and } \Delta^2(k) \equiv \frac{k^3 P(k)}{2\pi^2}.$$

The quantity $\Delta^2(k)$ is called the dimensionless power spectrum and is an important function in the theory of structure formation. $\Delta^2(k)$ measures the contribution of perturbations per unit logarithmic interval at wavenumber k to the variance in the matter density fluctuations. The Λ CDM power spectrum asymptotes to $P(k) \sim k^1$ for small k , and $P(k) \sim k^{-3}$ for large k , with a peak at $k^* \sim 2 \times 10^{-2} h \text{ Mpc}^{-1}$ corresponding to $\lambda^* \sim 350 h^{-1} \text{ Mpc}$. $\Delta^2(k)$ is thus asymptotically flat at high k , but drops off as k^4 at small k . We therefore see that most of the variance in the cosmic density field in the universe at the present epoch is on scales $\lambda < \lambda^*$.

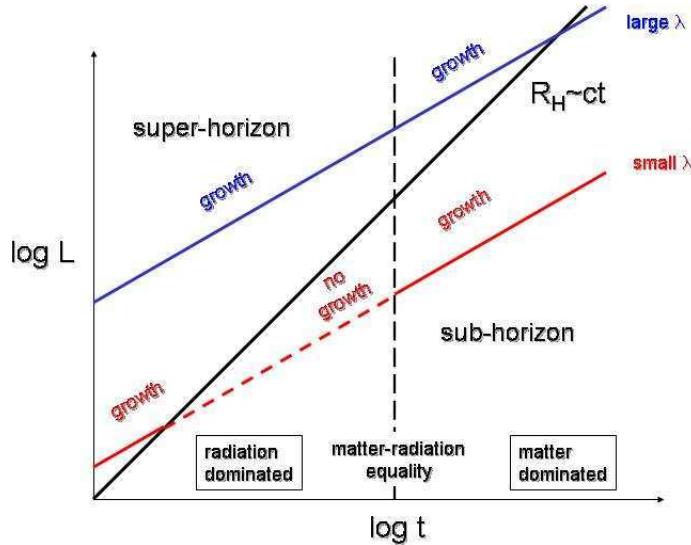


Fig. 2. – The tale of two fluctuations. A fluctuation which is superhorizon scale at matter-radiation equality grows always, while a fluctuation which enters the horizon during the radiation dominated era stops growing in amplitude until the matter dominated era begins.

What is the origin of the power spectrum shape? Here we review the basic ideas. Within the inflationary paradigm, it is believed that quantum mechanical (QM) fluctuations in the very early universe were stretched to macroscopic scales by the large expansion factor the universe underwent during inflation. Since QM fluctuations are random, the primordial density perturbations should be well described as a Gaussian random field. Measurements of the Gaussianity of the CMB anisotropies [11] have confirmed this. The primordial power spectrum is parameterized as a power law $P_p(k) \propto k^n$, with $n = 1$ corresponding to scale-invariant spectrum proposed by Harrison and Zeldovich on the grounds that any other value would imply a preferred mass scale for fluctuations entering the Hubble horizon. Large angular scale CMB anisotropies measure the primordial power spectrum directly since they are superhorizon scale. Observations with the WMAP satellite are consistent with $n = 1$.

To understand the origin of the spectrum, we need to understand how the amplitude of a fluctuation of fixed comoving wavelength λ grows with time. Regardless of its wavelength, the fluctuation will pass through the Hubble horizon as illustrated in Fig. 2. This is because the Hubble radius grows linearly with time, while the proper wavelength $a\lambda$ grows more slowly with time. It is easy to show from Eq. 1 that in the radiation-dominated era, $a \sim t^{1/2}$, and in the matter-dominated era (prior to the onset of cosmic acceleration) $a \sim t^{2/3}$. Thus, inevitably, a fluctuation will transition from superhorizon to subhorizon scale. We are interested in how the amplitude of the fluctuation evolves during these two phases. Here we merely state the results of perturbation theory (e.g., Dodelson 2003, Ch. 7).

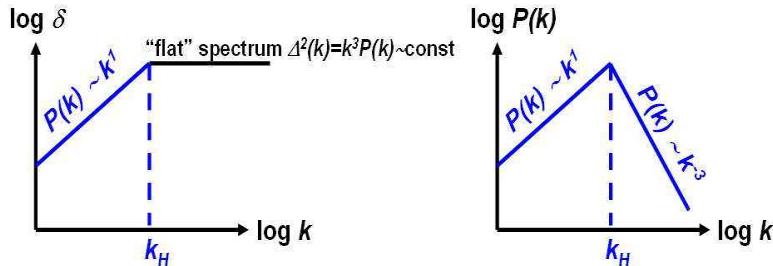
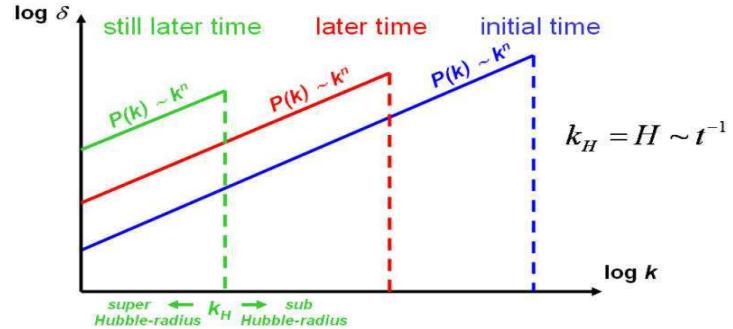


Fig. 3. – a) Evolution of the primordial power spectrum on superhorizon scales during the radiation dominated era. b) Scale-free spectrum produces a constant contribution to the density variance per logarithmic wavenumber interval entering the Hubble horizon (no preferred scale) c) resulting matter power spectrum, super- and sub-horizon. Figures courtesy Rocky Kolb.

2.3. Growth of fluctuations in the linear regime . – To calculate the growth of superhorizon scale fluctuations requires general relativistic perturbation theory, while subhorizon scale perturbations can be analyzed using a Newtonian Jeans analysis. We are interested in scalar density perturbations, because these couple to the stress tensor of the matter-radiation field. Vector perturbations (e.g., fluid turbulence) are not sourced

by the stress-tensor, and decay rapidly due to cosmic expansion. Tensor perturbations are gravity waves, and also do not couple to the stress-tensor. A detailed analysis for the scalar perturbations yields the following results. In the radiation dominated era,

$$\begin{aligned}\delta_+(t) &= \delta_+(t_i)(t/t_i) \text{ superhorizon scales} \\ \delta_+(t) &= \text{constant} \quad \text{subhorizon scales}\end{aligned}$$

while in the matter dominated era,

$$\begin{aligned}\delta_+(t) &= \delta_+(t_i)(t/t_i)^{2/3} \text{ superhorizon scales} \\ \delta_+(t) &= \delta_+(t_i)(t/t_i)^{2/3} \text{ subhorizon scales}\end{aligned}$$

This is summarized in Fig. 2, where we consider two fluctuations of different comoving wavelengths, which we will call large and small. The large wavelength perturbation remains superhorizon through matter-radiation equality (MRE), and enters the horizon in the matter dominated era. Its amplitude will grow as t in the radiation dominated era, and as $t^{2/3}$ in the matter dominated era. It will continue to grow as $t^{2/3}$ after it becomes subhorizon scale. The small wavelength perturbation becomes subhorizon before MRE. Its amplitude will grow as t while it is superhorizon scale, remain constant while it is subhorizon during the radiation dominated era, and then grow as $t^{2/3}$ during the matter-dominated era.

Armed with these results, we can understand what is meant by a scale-free primordial power spectrum (the Harrison-Zeldovich power spectrum.) We are concerned with perturbation growth in the very early universe during the radiation dominated era. Superhorizon scale perturbation amplitudes grow as t , and then cease to grow after they have passed through the Hubble horizon. We can define a Hubble wave number $k_H \equiv 2\pi/R_H \propto t^{-1}$. Fig. 3a shows the primordial power spectrum at three instants in time for $k < k_H$. We see that the fluctuation amplitude at $k = k_H(t)$ depends on primordial power spectrum slope n . The scale-free spectrum is the value of n such that $\Delta^2(k_H(t)) = \text{constant}$ for $k > k_H$. A simple analysis shows that this implies $n=1$. Since $\Delta^2(k) \propto k^3 P(k)$, we then have

$$\begin{aligned}P(k) &\propto k^1, \quad k \leq k_H \\ P(k) &\propto k^{-3}, \quad k > k_H\end{aligned}$$

In actuality, the power spectrum has a smooth maximum, rather than a peak as shown in Fig. 3c. This smoothing is caused by the different rates of growth before and after matter-radiation equality. The transition from radiation to matter-dominated is not instantaneous. Rather, the expansion rate of the universe changes smoothly through equality, as given by Eq. 1, and consequently so do the temporal growth rates. The position of the peak of the power spectrum is sensitive to the when the universe reached matter-radiation equality, and hence is a probe of Ω_γ/Ω_m .

Once a fluctuation becomes sub-horizon, dissipative processes modify the shape of the power spectrum in a scale-dependent way. Collisionless matter will freely stream out of overdense regions and smooth out the inhomogeneities. The faster the particle, the larger its free streaming length. Particles which are relativistic at MRE, such as light neutrinos, are called hot dark matter (HDM). They have a large free-streaming length, and consequently damp the power spectrum over a large range of k . Weakly Interacting Massive Particles (WIMPs) which are nonrelativistic at MRE, are called cold dark matter

(CDM), and modify the power spectrum very little (Fig. 4). Baryons are tightly coupled to the radiation field by electron scattering prior to recombination. During recombination, the photon mean-free path becomes large. As photons stream out of dense regions, they drag baryons along, erasing density fluctuations on small scales. This process is called Silk damping, and results in damped oscillations of the baryon-photon fluid once they become subhorizon scale. The magnitude of this effect is sensitive to the ratio of baryons to collisionless matter, as shown in Fig. 4.

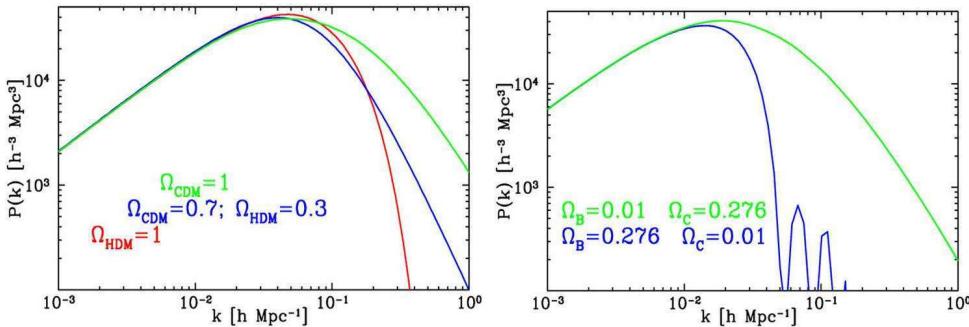


Fig. 4. – Effect of dissipative processes on the evolved power spectrum. Left: Effect of collisionless damping (free streaming) in the dark matter. Right: Effect of collisional damping (Silk damping) in the matter-radiation fluid. Figures courtesy Rocky Kolb.

3. – Analytic models for nonlinear growth, virial scaling relations, and halo statistics

Here we introduce a few concepts and analytic results from the theory of structure formation which underly the use of galaxy clusters as cosmological probes. These provide us with the vocabulary which pervades the literature on analytic and numerical models of galaxy cluster evolution. Material in this section has been derived from three primary sources: Padmanabhan (1993) [12] for the spherical top-hat model for nonlinear collapse, Dodelson (2003) [4] for Press-Schechter theory, and Bryan & Norman (1998) [13] for virial scaling relations.

3.1. Nonlinearity defined. – In the linear regime, both super- and sub-horizon scale perturbations grow as $t^{2/3}$ in the matter-dominated era. This means that after recombination, the linear power spectrum retains its shape while its amplitude grows as $t^{4/3}$ before the onset of cosmic acceleration. When $\Delta^2(k)$ for a given k approaches unity linear theory no longer applies, and some other method must be used to determine the fluctuation's growth. In general, numerical simulations are required to model the nonlinear phase of growth because in the nonlinear regime, the modes do not grow independently. Mode-mode coupling modifies both the shape and amplitude of the power spectrum over the range of wavenumbers that have gone nonlinear.

At any given time, there is a critical wavenumber which we shall call the nonlinear wavenumber k_{nl} which determines which portion of the spectrum has evolved into the nonlinear regime. Modes with $k < k_{nl}$ are said to be linear, while those for which $k > k_{nl}$ are nonlinear. Conventionally, one defines the nonlinear wavenumber such that $\Delta(k_{nl}, z) = 1$.

From this one can derive a nonlinear mass scale $M_{nl}(z) = \frac{4\pi}{3} \bar{\rho}(z) \left(\frac{2\pi}{k_{nl}} \right)^3$. A more

useful and rigorous definition of the nonlinear mass scale comes from evaluating the amplitude of mass fluctuations within spheres of radius R at epoch z . The enclosed mass is $M = \frac{4\pi}{3}\bar{\rho}(z)R^3$. The mean square mass fluctuations (variance) is

$$(13) \quad \langle (\delta M/M)^2 \rangle \equiv \sigma^2(M) = \int d^3k W_T^2(kR) P(k, z),$$

where W is the Fourier transform of the top-hat window function

$$(14) \quad W(\mathbf{x}) = \begin{cases} 3/4\pi R^3, & |\mathbf{x}| < R \\ 0, & |\mathbf{x}| \geq R \end{cases} \rightarrow W_T(kR) = 3 [\sin(kR)/kR - \cos(kR)]/(kR)^2.$$

If we approximate $P(k)$ locally with a power-law $P(k, z) = D^2(z)k^m$, where D is the linear growth factor, then $\sigma^2(M) \propto D^2 R^{-(3+m)} \propto D^2 M^{-(3+m)/3}$. From this we see that the RMS fluctuations are a decreasing function of M . At very small mass scales, $m \rightarrow -3$, and the fluctuations asymptote to a constant value. We now define the nonlinear mass scale by setting $\sigma(M_{nl})=1$. We get that ([17])

$$(15) \quad M_{nl}(z) \propto D(z)^{6/(3+m)} (\propto (1+z)^{-6/(3+m)} \text{ for EdS}).$$

For $m > -3$, the smallest mass scales become nonlinear first. This is the origin of hierarchical (“bottom-up”) structure formation.

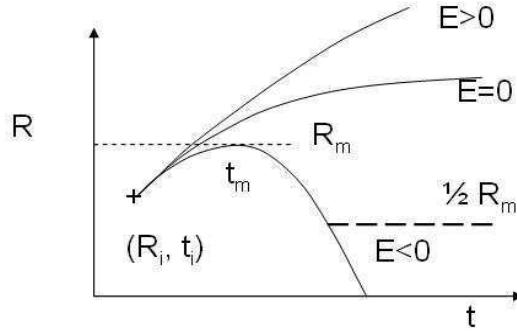


Fig. 5. – Evolution of a top-hat perturbation in an EdS universe. Depending on the E , the first integral of motion, the fluctuation collapses ($E < 0$), continues to expand ($E > 0$), or asymptotically reaches its maximum radius ($E = 0$). Virialization occurs when the fluctuation has collapsed to half its turnaround radius.

3.2. Spherical Top-Hat Model. – We now ask what happens when a spherical volume of mass M and radius R exceeds the nonlinear mass scale. The simplest analytic model of the nonlinear evolution of a discrete perturbation is called the spherical top-hat model. In it, one imagines a spherical perturbation of radius R and some constant overdensity $\bar{\delta} =$

$3M/4\pi R^3$ in an Einstein-de Sitter (EdS) universe. By Birkhoff's theorem the equation of motion for R is

$$(16) \quad \frac{d^2R}{dt^2} = -\frac{GM}{R^2} = -\frac{4\pi G}{3}\bar{\rho}(1+\bar{\delta})R$$

whereas the background universe expands according to Eq. 6

$$(17) \quad \frac{d^2a}{dt^2} = -\frac{4\pi G}{3}\bar{\rho}a.$$

Comparing these two equations, we see that the perturbation evolves like a universe of a different mean density, but with the same initial expansion rate. Integrating Eq. 16 once with respect to time gives us the first integral of motion:

$$(18) \quad \frac{1}{2} \left(\frac{dR}{dt} \right)^2 - \frac{GM}{R} = E,$$

where E is the total energy of the perturbation. If $E < 0$, the perturbation is bound, and obeys

$$(19) \quad \frac{R}{R_m} = \frac{(1 - \cos\theta)}{2}, \quad \frac{t}{t_m} = \frac{(\theta - \sin\theta)}{\pi}$$

where R_m and t_m are the radius and time of “turnaround”. At turnaround (as $\theta \rightarrow \pi$), the fluctuation reaches its maximum proper radius (see Fig. 5). As $t \rightarrow 2t_m$, $R \rightarrow 0$, and we say the fluctuation has collapsed.

A detailed analysis of the evolution of the top-hat perturbation is given in Padmanabhan (1993, Ch. 8) for general Ω_m . Here we merely quote results for an EdS universe. The mean *linear* overdensity at turnaround; i.e., the value one would predict from the linear growth formula $\delta \sim t^{2/3}$, is 1.063. The actual overdensity at turnaround using the nonlinear model is 4.6. This illustrates that nonlinear effects set in well before the amplitude of a linear fluctuation reaches unity. As $R \rightarrow 0$, the nonlinear overdensity becomes infinite. However, the linear overdensity at $t = 2t_m$ is only 1.686. As the fluctuation collapses, other physical processes (pressure, shocks, violent relation) become important which establish a gravitationally bound object in virial equilibrium before infinite density is reached. Within the framework of the spherical top-hat model, we say virialization has occurred when the kinetic and gravitational energies satisfy virial equilibrium: $|U| = 2K$. It is easy to show from conservation of energy that this occurs when $R = R_m/2$; in other words, when the fluctuation has collapsed to half its turnaround radius. The nonlinear overdensity at virialization Δ_c is not infinite since the radius is finite. For an EdS universe, $\Delta_c = 18\pi^2 \approx 180$. Fitting formulae for non-EdS models are provided in the next section.

3.3. Virial Scaling Relations. – The spherical top-hat model can be scaled to perturbations of arbitrary mass. Using virial equilibrium arguments, we can predict various physical properties of the virialized object. The ones that interest us most are those that relate to the observable properties of gas in galaxy clusters, such as temperature, X-ray luminosity, and SZ intensity change. Kaiser [14] first derived virial scaling relations for clusters in an EdS universe. Here we generalize the derivation to non-EdS models of

interest. In order to compute these scaling laws, we must assume some model for the distribution of matter as a function of radius within the virialized object. A top-hat distribution with a density $\rho = \Delta_c \bar{\rho}(z)$ is not useful because it is not in mechanical equilibrium. More appropriate is the isothermal, self-gravitating, equilibrium sphere for the collisionless matter, whose density profile is related to the one-dimensional velocity dispersion [15]

$$(20) \quad \rho(r) = \frac{\sigma^2}{2\pi G r^2}.$$

If we define the virial radius r_{vir} to be the radius of a spherical volume within which the mean density is Δ_c times the critical density at that redshift ($M = 4\pi r_{vir}^3 \rho_{crit} \Delta_c / 3$), then there is a relation between the virial mass M and σ :

$$(21) \quad \sigma = M^{1/3} [H^2(z) \Delta_c G^2 / 16]^{1/6} \approx 476 f_\sigma \left(\frac{M}{10^{15} M_\odot} \right)^{1/3} (h^2 \Delta_c E^2)^{1/6} \text{ km s}^{-1}.$$

Here we have introduced a normalization factor f_σ which will be used to match the normalization from simulations. The redshift dependent Hubble parameter can be written as $H(z) = 100hE(z) \text{ km s}^{-1}$ with the function $E^2(z) = \Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda$, where the Ω 's have been previously defined.

The value of Δ_c is taken from the spherical top-hat model, and is $18\pi^2$ for the critical EdS model, but has a dependence on cosmology through the parameter $\Omega(z) = \Omega_m(1+z)^3/E^2(z)$. Bryan and Norman (1998) provided fitting formulae for Δ_c for the critical for both open universe models and flat, lambda-dominated models

$$(22) \quad \Delta_c = 18\pi^2 + 82x - 39x^2 \text{ for } \Omega_k = 0, \quad \Delta_c = 18\pi^2 + 60x - 32x^2 \text{ for } \Omega_\Lambda = 0$$

where $x = \Omega(z)-1$.

If the distribution of the baryonic gas is also isothermal, we can define a ratio of the “temperature” of the collisionless material ($T_\sigma = \mu m_p \sigma^2 / k$) to the gas temperature:

$$(23) \quad \beta = \frac{\mu m_p \sigma^2}{kT}$$

Given equations (22) and (23), the relation between temperature and mass is then

$$(24) \quad kT = \frac{GM^{2/3} \mu m_p}{2\beta} \left[\frac{H^2(z) \Delta_c}{2G} \right]^{1/3} \approx 1.39 f_T \left(\frac{M}{10^{15} M_\odot} \right)^{2/3} (h^2 \Delta_c E^2)^{1/3} \text{ keV},$$

where in the last expression we have added the normalization factor f_T and set $\beta=1$.

The scaling behavior for the object's X-ray luminosity is easily computed by assuming bolometric bremsstrahlung emission and ignoring the temperature dependence of the Gaunt factor: $L_{bol} \propto \int \rho^2 T^{1/2} dV \propto M_b \rho T^{1/2}$, where M_b is the baryonic mass of the cluster. This is infinite for an isothermal density distribution, since ρ is singular. Observationally and computationally, it is found that the baryon distribution rolls over to a constant density core at small radius. A procedure is described in Bryan and Norman

(1998) which yields a finite luminosity:

$$(25) \quad L_{bol} = 1.3 \times 10^{45} \left(\frac{M}{10^{15} M_\odot} \right)^{4/3} (h^2 \Delta_c E^2)^{7/6} \left(\frac{\Omega_b}{\Omega_m} \right)^2 \text{ erg s}^{-1}.$$

Eliminating M in favor of T in Eq. 25 we get

$$(26) \quad L_{bol} = 6.8 \times 10^{44} \left(\frac{kT/f_T}{1.0 \text{ keV}} \right)^2 (h^2 \Delta_c E^2)^{1/2} \left(\frac{\Omega_b}{\Omega_m} \right)^2 \text{ erg s}^{-1}.$$

The scaling of the SZ “luminosity” is likewise easily computed. If we define L_{SZ} as the integrated SZ intensity change: $L_{SZ} = \int dA \int n_e \sigma_T \left(\frac{kT}{m_e c^2} \right) dl \propto M_b T$, then

$$(27) \quad L_{SZ} = \frac{GM^{5/3} \sigma_T}{2\beta m_e c^2} \left[\frac{H^2(z) \Delta_c}{2G} \right]^{1/3} \left(\frac{\Omega_b}{\Omega_m} \right).$$

We note that cosmology enters these relations only with the combination of parameters $h^2 \Delta_c E^2$, which comes from the relation between the cluster’s mass and the mean density of the universe at redshift z. The redshift variation comes mostly from E(z), which is equal to $(1+z)^{3/2}$ for an EdS universe.

3.4. Statistics of hierarchical clustering: Press-Schechter theory. – Now that we have a simple model for the nonlinear evolution of a spherical density fluctuation and its observable properties as a function of its virial mass, we would like to estimate the number of virialized objects of mass M as a function of redshift given the matter power spectrum. This is the key to using surveys of galaxy clusters as cosmological probes. While large scale numerical simulations can and have been used for this purpose (see below), we review a powerful analytic approach by Press and Schechter [16] which turns out to be remarkably close to numerical results. The basic idea is to imagine smoothing the cosmological density field at any epoch z on a scale R such that the mass scale of virialized objects of interest satisfies $M = \frac{4\pi}{3} \bar{\rho}(z) R^3$. Because the density field (both smoothed and unsmoothed) is a Gaussian random field, the probability that the mean overdensity in spheres of radius R exceeds a critical overdensity δ_c is

$$(28) \quad p(R, z) = \frac{2}{\sqrt{2\pi} \sigma(R, z)} \int_{\delta_c}^{\infty} d\delta \exp \left(-\frac{\delta^2}{2\sigma^2(R, z)} \right)$$

where $\sigma(R, z)$ is the RMS density variation in spheres of radius R as discussed above. Press and Schechter suggested that this probability be identified with the fraction of particles which are part of a nonlinear lump with mass exceeding M if we take $\delta_c = 1.686$, the linear overdensity at virialization. This assumption has been tested against numerical simulations and found to be quite good [9]. The fraction of the volume collapsed into objects with mass between M and $M + dM$ is given by $(dp/dM)dM$. Multiply this by the average number density of such objects ρ_m/M to get the number density of collapsed objects between M and $M + dM$:

$$(29) \quad dn(M, z) = -\frac{\bar{\rho}}{M} \frac{dp(M(R), z)}{dM} dM.$$

The minus sign appears here because p is a decreasing function of M . Carrying out the derivative using the fact that $dM/dR = 3M/R$, we get

$$(30) \quad \frac{dn(M, z)}{dM} = \sqrt{\frac{2}{\pi}} \frac{\bar{\rho}\delta_c}{3M^2\sigma} e^{-\delta_c^2/2\sigma^2} \left[-\frac{d\ln\sigma}{d\ln R} \right].$$

The term in square brackets is related to the logarithmic slope of the power spectrum, which on the mass scale of galaxy clusters is close to unity. Eq. 30 is called the *halo mass function*, and it has the form of a power law multiplied by an exponential. To make this more explicit, approximate the power spectrum on scales of interest as a power law as we have done above. Substituting the scaling relations for σ in Eq. 30 one gets the result [17]

$$(31) \quad \frac{dn}{dM} = \left(\frac{2}{\pi} \right)^{1/2} \frac{\bar{\rho}}{M^2} \left(1 + \frac{m}{3} \right) \left[\frac{M}{M_{nl}(z)} \right]^{\frac{m-3}{6}} \exp \left[- \left(\frac{M}{M_{nl}(z)} \right)^{\frac{3+m}{3}} / 2 \right].$$

Here, $M_{nl}(z)$ is the nonlinear mass scale. To be more consistent with the spherical top-hat model, it satisfies the relation $\sigma(M_{nl}, z) = \delta_c$; i.e., those fluctuations in the smoothed density field that have reached the linear overdensity for which the spherical top-hat model predicts virialization.

3.5. Application to galaxy clusters. – Galaxy clusters correspond to rare ($\sim 3\sigma$) peaks in the density field. Combining the halo mass function as predicted by the PS formalism with the scaling laws derived above, we can predict the evolution of the statistical properties of X-ray and SZ clusters of galaxies. Here we show a few results taken from Eke, Cole & Frenk (1996) [18]. Fig. 6a shows the evolution of the integrated mass function $n(> M)$ for several cosmologies and redshifts. One can see the power-law behavior at lower mass and the exponential cutoff at higher M . One sees strong redshift evolution of the number of massive clusters in the EdS model, but slower evolution on the open and lambda models. This is because of the saturated growth of structure in low density models. This makes number counts of massive clusters a sensitive test of the linear growth factor $D(z)$, which depends on Ω_m and Ω_Λ . Convolving the cluster population with the scaling relations for $T(M)$ and $Y(M)$, one gets distribution functions for $n(> T)$ and $n(> Y)$. Here $Y = L_{SZ}/d_A^2$ is the effective SZE cross section of a cluster, where d_A is its angular diameter distance. These are shown in Figs. 6b and 6c. Another way to present the data is to convolve the mass function with the differential volume element as a function of redshift for the three models. Figs. 6d-f plot the redshift probability of detecting a cluster with M , T , and Y exceeding the fiducial values given in the figure caption. As one can see, the profiles are sharply peaked at low redshift for the EdS model, but substantially broader and peaking at higher redshift for the low density universe models. There is, however, rather little difference between the open and lambda-dominated models as far as the probability distributions for M and Y . Things are somewhat better for T , implying that some combination of X-ray and SZE measurements will be needed for precision cosmological parameter determinations.

4. – Numerical simulations of gas in galaxy clusters

The central task is for a given cosmological model, calculate the formation and evolution of a population of clusters from which synthetic X-ray and SZ catalogs can be

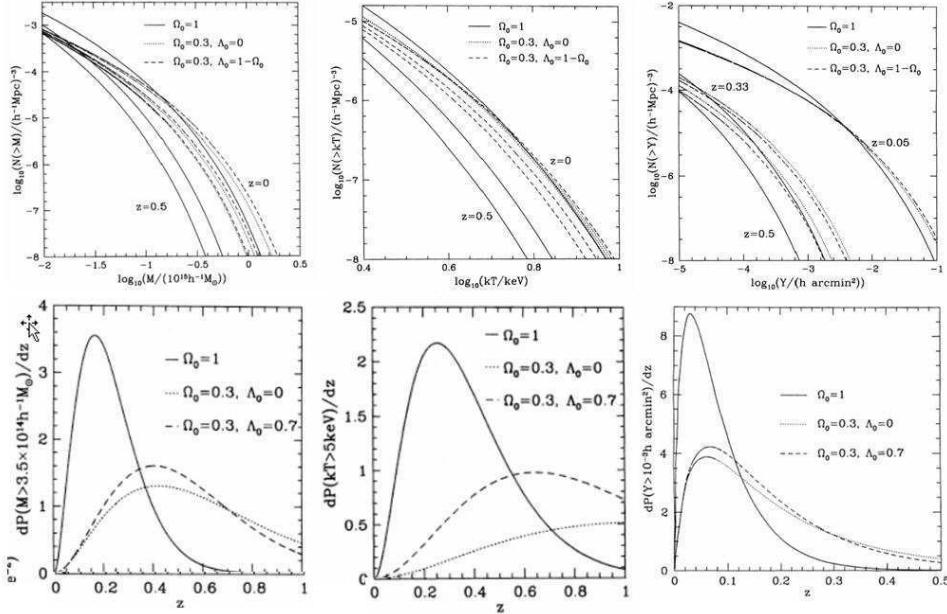


Fig. 6. – Top left to bottom right: a) Integrated cluster mass function for three cosmologies and two redshifts; b) like a), but for integrated temperature function; c) like a) but for integrated SZ cross section; d) redshift distribution of the integrated probability to find a cluster exceeding $M = 3.5 \times 10^{14} h^{-1} M_{\odot}$; e) redshift distribution of the integrated probability to find a cluster exceeding $kT=5$ keV; f) redshift distribution of the integrated probability to find a cluster exceeding $Y=10^{-3}$ h arcmin 2 . From [18].

derived. These can be used to calibrate simpler analytic models, as well as to build synthetic surveys (mock catalogs) which can be used to assess instrumental effects and survey biases. One would like to directly simulate $n(M, z)$, $n(L_x, z)$, $n(T, z)$, $n(Y, z)$ from the governing equations for collisionless and collisional matter in an expanding universe. Clearly, the quality of these statistical predictions relies on the ability to adequately resolve the internal structure and thermodynamical evolution of the ICM.

In Norman (2003) [19] I provided a historical review of the progress that has been made in simulating the evolution of gas in galaxy clusters motivated by X-ray observations. Since X-ray emission and the SZE are both consequences of hot plasma bound in the cluster's gravitational potential well, the requirements to faithfully simulate X-ray clusters and SZ clusters are essentially the same. Numerical progress can be characterized as a quest for higher resolution and essential baryonic physics. In this section I describe the technical challenges involved and the numerical methods that have been developed to overcome them. I then discuss the effects of assumed baryonic physics on ICM structure. Our point of reference is the non-radiative (so-called adiabatic) case, which has been the subject of an extensive code comparison [20]. I review the properties of adiabatic X-ray clusters, and show that they fail to reproduce observed cluster scaling laws. I then show results of numerical hydrodynamic simulations incorporating radiative cooling, star formation, and galaxy feedback and their associated scaling properties.

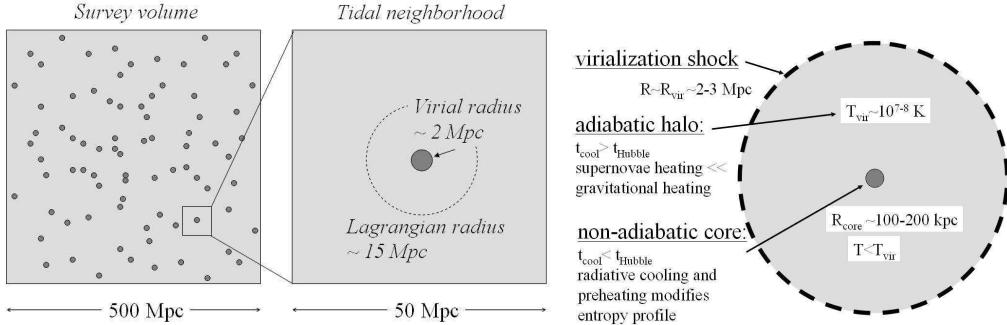


Fig. 7. – Left: A range of length scales of ~ 250 separates the size of a reasonable survey volume and the virial radius of a rich cluster. Right: Simplified structure of the ICM in a massive cluster. A range of length scales of ~ 20 - 30 separates the virial radius and the core radius.

4.1. Dynamic range considerations. – Figure 7 illustrates the dynamic range difficulties encountered with simulating a statistical ensemble of galaxy clusters, while at the same time resolving their internal structure. Massive clusters are rare at any redshift, yet these are the ones most that are most sensitive to cosmology. From the cluster mass function (Fig. 6a), in order to get adequate statistics, one deduces that one must simulate a survey volume many hundreds of megaparsecs on a side (Fig. 7a). A massive cluster has a virial radius of ~ 2 Mpc. It forms via the collapse of material within a comoving Lagrangian volume of ~ 15 Mpc. However, tidal effects from a larger region (50-100 Mpc) are important on the dynamics of cluster formation. The internal structure of cluster's ICM is shown in Fig. 7b. While clusters are not spherical, two important radii are generally used to characterize them: the virial radius, which is the approximate location of the virialization shock wave that thermalizes infalling gas to 10-100 million K, and the core radius, within which the baryon densities plateau and the highest X-ray emissions and SZ intensity changes are measured. A typical radius is ~ 200 kpc. Within the core, radiative cooling and possibly other physical processes are important. Outside the core, cooling times are longer than the Hubble time, and the ICM gas is effectively adiabatic. If we wanted to achieve a spatial resolution of $1/10$ of a core radius everywhere within the survey volume, we would need a spatial dynamic range of $D=500$ Mpc/ 20 kpc = 25,000. The mass dynamic range is more severe. If we want 1 million dark matter particles within the virial radius of a $10^{15} M_\odot$ cluster, then we would need $N_{\text{particle}} = M_{\text{box}}/M_{\text{particle}} = \Omega_m \rho_{\text{crit}} L^3 / 10^9 \approx 10^{11}$ if they were uniformly distributed in the survey volume.

Two solutions to spatial dynamic range problem have been developed: tree codes for gridless N-body methods [21, 22] and adaptive mesh refinement (AMR) for Eulerian particle-mesh/hydrodynamic methods [23, 24, 25, 26]. Both methods increase the spatial resolution automatically in collapsing regions as described below. The solution to the mass dynamic range problem is the use of multi-mass initial conditions in which a hierarchy of particle masses is used, with many low mass particles concentrated in the region of interest. This approach has most recently used by Springel et al. (2000) [27], who simulated the formation of a galaxy cluster dark matter halo with $N = 6.9 \times 10^6$ dark matter particles, resolving the dark matter halos down to the mass scale of the Fornax dwarf spheroidal galaxy. The spatial dynamic range achieved in this simulation was $R = 2 \times 10^5$. Such dynamic ranges have not yet been achieved in galaxy cluster

simulations with gas.

4.2. Simulating cluster formation. – Simulations of cosmological structure formation are done in a cubic domain which is comoving with the expanding universe. Matter density and velocity fluctuations are initialized at the starting redshift chosen such that all modes in the volume are still in the linear regime. Once initialized, these fluctuations are then evolved to $z=0$ by solving the equations for collisionless N-body dynamics for cold dark matter, and the equations of ideal gas dynamics for the baryons in an expanding universe. Making the transformation from proper to comoving coordinates $\vec{r} = a(t)\vec{x}$, Newton's laws for the collisionless dark matter particles become

$$(32) \quad \frac{d\vec{x}_{dm}}{dt} = \vec{v}_{dm}, \quad \frac{d\vec{v}_{dm}}{dt} = -2\frac{\dot{a}}{a}\vec{v}_{dm} - \frac{1}{a^2}\nabla_x\phi$$

where x and v are the particle's comoving position and peculiar velocity, respectively, and ϕ is the comoving gravitational potential that includes baryonic and dark matter contributions. The hydrodynamical equations for mass, momentum, and energy conservation in an expanding universe in comoving coordinates are ([28])

$$(33) \quad \begin{aligned} \frac{\partial\rho_b}{\partial t} + \nabla \cdot (\rho_b\vec{v}_b) + 3\frac{\dot{a}}{a}\rho_b &= 0, \\ \frac{\partial(\rho_b v_{b,i})}{\partial t} + \nabla \cdot [(\rho_b v_{b,i})\vec{v}_b + 5\frac{\dot{a}}{a}\rho_b v_{b,i}] &= -\frac{1}{a^2}\frac{\partial p}{\partial x_i} - \frac{\rho_b}{a^2}\frac{\partial\phi}{\partial x_i}, \\ \frac{\partial e}{\partial t} + \nabla \cdot (e\vec{v}_b) + p\nabla \cdot \vec{v}_b + 3\frac{\dot{a}}{a}e &= \Gamma - \Lambda, \end{aligned}$$

where ρ_b, p and e , are the baryonic density, pressure and internal energy density defined in the proper reference frame, \vec{v}_b is the comoving peculiar baryonic velocity, $a = 1/(1+z)$ is the cosmological scale factor, and Γ and Λ are the microphysical heating and cooling rates. The baryonic and dark matter components are coupled through Poisson's equation for the gravitational potential

$$(34) \quad \nabla^2\phi = 4\pi Ga^2(\rho_b + \rho_{dm} - \bar{\rho}(z))$$

where $\bar{\rho}(z) = 3H_0\Omega_m(0)/8\pi Ga^3$ is the proper background density of the universe.

The cosmological scale factor $a(t)$ is obtained by integrating the Friedmann equation (Eq. 4). To complete the specification of the problem we need the ideal gas equation of state $p = (\gamma - 1)e$, and the gas heating and cooling rates. When simulating the ICM, the simplest approximation is to assume Γ and $\Lambda = 0$; i.e., no heating or cooling of the gas other than by adiabatic processes and shock heating. Such simulations are referred to as adiabatic (despite entropy-creating shock waves), and are a reasonable first approximation to real clusters because except in the cores of clusters, the radiative cooling time is longer than a Hubble time, and gravitational heating is much larger than sources of astrophysical heating. However, as discussed in the paper by Cavaliere in this volume, there is strong evidence that the gas in cores of clusters has evolved non-adiabatically. This is revealed by the entropy profiles observed in clusters [29] which deviate substantially from adiabatic predictions. In the simulations presented below, we consider radiative cooling due to thermal bremsstrahlung, and mechanical heating due to galaxy feedback, details of which are described below.

4.3. Numerical methods overview. – A great deal of literature exists on the gravitational clustering of CDM using N-body simulations. A variety of methods have been employed including the fast grid-based methods particle-mesh (PM), and particle-particle+particle-mesh (P^3M) [30], spatially adaptive methods such as adaptive P^3M [31], adaptive mesh refinement [24], tree codes [32, 33], and hybrid methods such as TreePM [34]. Because of the large dynamic range required, spatially adaptive methods are favored, with Tree and TreePM methods the most widely used today. When gas dynamics is included, only certain combinations of hydrodynamics algorithms and collisionless N-body algorithms are “natural”. Dynamic range considerations have led to two principal approaches: P^3MSPH and TreeSPH, which marries a P^3M or tree code for the dark matter with the Lagrangian smoothed-particle-hydrodynamics (SPH) method [35, 21, 22], and adaptive mesh refinement (AMR), which marries PM with Eulerian finite-volume gas dynamics schemes on a spatially adaptive mesh [23, 26, 25, 36]. Pioneering hydrodynamic simulations using non-adaptive Eulerian grids [37, 38, 13] yielded some important insights about cluster formation and statistics, but generally have inadequate resolution to resolve their internal structure in large survey volumes. In the following we concentrate on our latest results using the AMR code *Enzo* [26]. The reader is also referred to the paper by Borgani et al. [39] which presents recent, high-resolution results from a large TreeSPH simulation.

Enzo is a grid-based hybrid code (hydro + N-body) which uses the block-structured AMR algorithm of Berger & Collela [40] to improve spatial resolution in regions of large gradients, such as in gravitationally collapsing objects. The method is attractive for cosmological applications because it: (1) is spatially- and time-adaptive, (2) uses accurate and well-tested grid-based methods for solving the hydrodynamics equations, and (3) can be well optimized and parallelized. The central idea behind AMR is to solve the evolution equations on a grid, adding finer meshes in regions that require enhanced resolution. Mesh refinement can be continued to an arbitrary level, based on criteria involving any combination of overdensity (dark matter and/or baryon), Jeans length, cooling time, etc., enabling us to tailor the adaptivity to the problem of interest. The code solves the following physics models: collisionless dark matter and star particles, using the particle-mesh N-body technique [41]; gravity, using FFTs on the root grid and multigrid relaxation on the subgrids; cosmic expansion; gas dynamics, using the piecewise parabolic method (PPM)[42]; multispecies nonequilibrium ionization and H₂ chemistry, using backward Euler time differencing [28]; radiative heating and cooling, using subcycled forward Euler time differencing [43]; and a parameterized star formation/feedback recipe [44]. At the present time, magnetic fields and radiation transport are being installed. *Enzo* is publicly available at <http://cosmos.ucsd.edu/enzo>.

4.4. Structure of nonradiative clusters: the Santa Barbara test cluster. – In Frenk et al. [20] 12 groups compared the results of a variety of hydrodynamic cosmological algorithms on a standard test problem. The test problem, called the Santa Barbara cluster, was to simulate the formation of a Coma-like cluster in a standard CDM cosmology ($\Omega_m = 1$) assuming the gas is nonradiative. Groups were provided with uniform initial conditions and were asked to carry out a “best effort” computation, and analyze their results at $z=0.5$ and $z=0$ for a set of specified outputs. These outputs included global integrated quantities, radial profiles, and column-integrated images. The simulations varied substantially in their spatial and mass resolution owing to algorithmic and hardware limitations. Nonetheless, the comparisons brought out which predicted quantities were robust, and which were not yet converged. In Fig. 8 we show a few figures from Frenk et

al. (1999) which highlight areas of agreement (top row) and disagreement (bottom row).

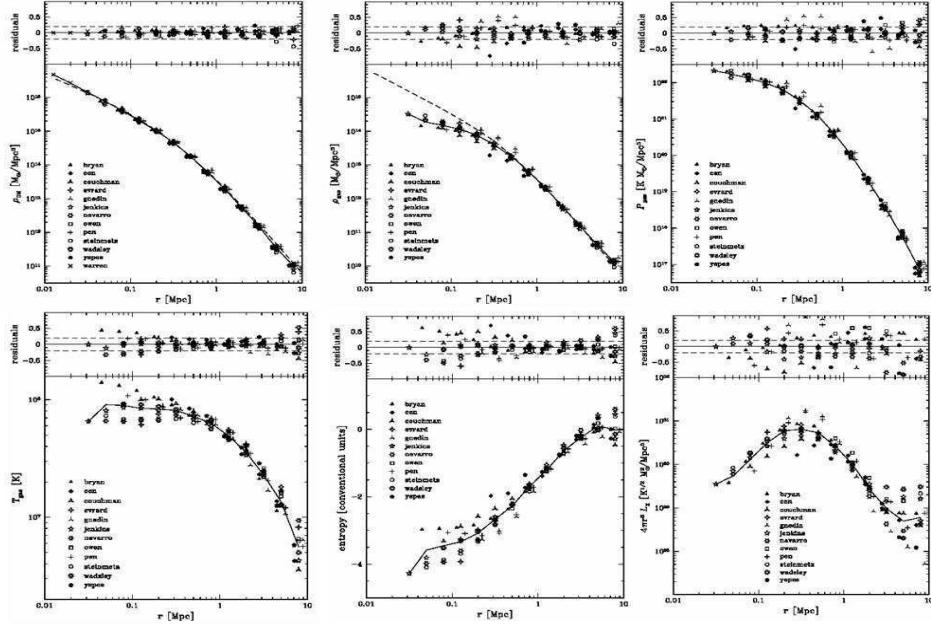


Fig. 8. – The Santa Barbara test cluster. Top row, left to right: profiles of dark matter density, gas density, and gas pressure. Bottom row, left to right: profiles of gas temperature, gas entropy, and X-ray emissivity. Different symbols correspond to different code results. From [20].

The top row shows profile of dark matter density, baryon density, and pressure for the different codes. All are in quite good agreement for the *mechanical structure* of the cluster. The dark matter profile is well described by an NFW profile which has a central cusp [45]. The baryon density profiles show more dispersion, but all codes agree that the profile flattens at small radius, as observed. All codes agree extremely well on the gas pressure profile, which is not surprising, since mechanical equilibrium is easy to achieve for all methods even with limited resolution. This bodes well for the interpretation of SZE observations of clusters, since the Compton y parameter is proportional to the projected pressure distribution. In section 5 we show results from a statistical ensemble of clusters which bear this out.

The bottom row shows the thermodynamic structure of the cluster, as well as the profile of X-ray emissivity. The temperature profiles show a lot of scatter within about one-third the virial radius ($=2.7$ Mpc). Systematically, the SPH codes produce nearly isothermal cores, while the grid codes produce temperature profiles which continue to rise as $r \rightarrow 0$. The origin of this discrepancy has not been resolved, but improved SPH formulations come closer to reproducing the AMR results [51]. This discrepancy is reflected in the entropy profiles. Again, agreement is good in the outer two-thirds of the cluster, but the profiles show a lot of dispersion in the inner one third. Discounting the codes with inadequate resolution, one finds the SPH codes produce an entropy profile which continues to fall as $r \rightarrow 0$, while the grid codes show an entropy core, which is more consistent with observations [29]. The dispersion in the density and temperature profiles are amplified in the X-ray emissivity profile, since $\varepsilon_x \propto \rho_b^2 T^{1/2}$. The different

codes agree on the integrated X-ray luminosity of the cluster only to within a factor of 2. This is primarily because the density profile is quite sensitive to resolution in the core; any underestimate in the core density due to inadequate resolution is amplified by the density squared dependence of the emissivity. This suggests that quite high resolution is needed, as well as a good grasp on non-adiabatic processes operating in cluster cores, before simulations will be able to accurately predict X-ray luminosities.

4.5. A numerical sample of adiabatic clusters: Universal Temperature Profile. – Three questions one can ask about the Santa Barbara cluster results are: 1) is the cluster statistically representative, 2) do the results change substantially for a Λ CDM cosmology (the SB cluster assumed an EdS cosmology), and 3) what is the effect of additional baryonic physics on cluster structure? We address these questions here by summarizing results of *Enzo* simulations of the ICM in a sample of clusters in a concordance Λ CDM model drawn from a survey volume $256h^{-1}$ Mpc on a side. Multimass initial conditions and AMR are used to achieve high spatial and mass resolution within the clusters. More details can be found in [47, 48, 46, 49].

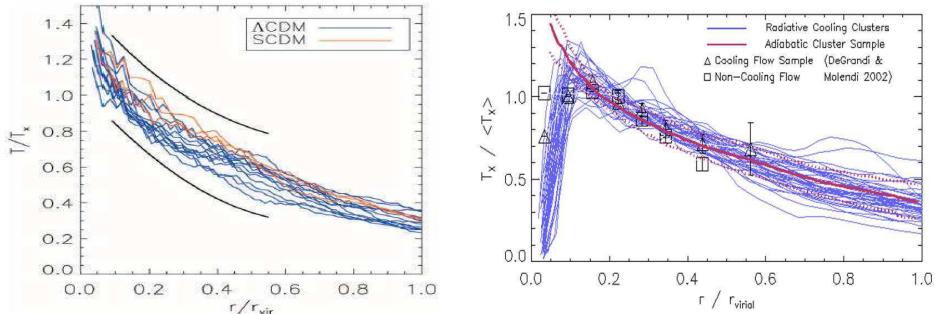


Fig. 9. – Left: Temperature profiles from a sample of adiabatic cluster simulations (from Loken et al. 2002). Black curves bound the 1σ confidence band from Markevitch et al. (1998). Right: Effect of radiative cooling on temperature profiles, compared with adiabatic sample average (red line) and observational data for cooling flow clusters (triangles) and non-cooling flow clusters (squares).

Fig. 9 shows spherically averaged temperature profiles for 13(3) Λ CDM(SCDM) simulated clusters at $z=0$ analyzed by Loken et al. (2002) [47]. These were chosen from a total sample of 22(10) clusters because their 2D projected temperature maps were symmetric; the rejected non-symmetric clusters were in various states of merging. The smooth black curves bound the 1σ confidence band from Markevitch et al. (1998)[62] who analyzed temperature profiles from a sample of 17 symmetric X-ray clusters observed with ASCA. When temperature is normalized by the integrated emission-weighted temperature and the radius by the virial radius, both the observed data and the simulated data collapse to a narrow band, suggesting a universal temperature profile (UTP) outside the core region. The fit to the numerical data is $T \propto (1 + r/\alpha)^{-\delta}$, with $\alpha \sim r_{vir}/1.5$ and $\delta \sim 1.6$. The Λ CDM clusters and SCDM clusters exhibit the same profile, with a suggestion of a slightly higher normalization for clusters in the critically closed model. The fit is in good agreement with observations over the range $0.2 < r/r_{vir} < 0.5$, but diverges at small radius where the effects of non-adiabatic processes appear to be at play [63]. The reality of the UTP was somewhat controversial when early results from Newton/XMM were showing

large isothermal cores. However, the latest Chandra observations of 13 nearby, relaxed clusters have shown that the UTP provides an excellent description for temperature profiles outside $r \sim 0.15r_{vir}$ [50]. Subsequent numerical studies by Ascasibar et al. [51] and Borgani et al. [39] using SPH have found agreement with the AMR results of Loken et al. The general agreement of numerical and observational results suggests that the declining temperature profile is a natural consequence of gravitational heating of the ICM during the process of cluster formation.

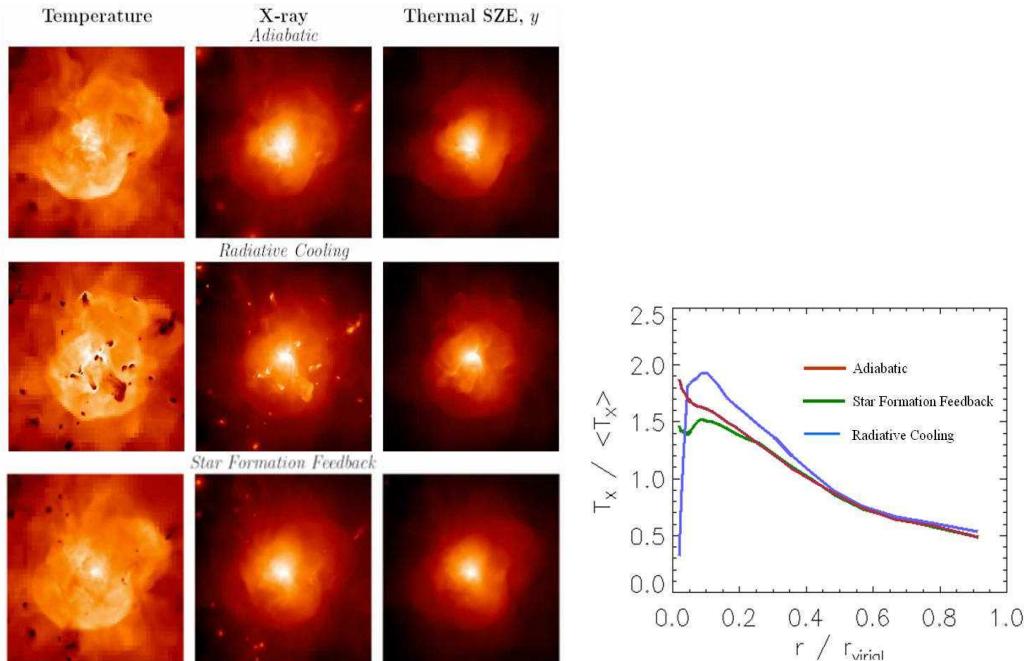


Fig. 10. – Left: Columns show X-ray surface brightness, projected temperature, and Compton y -parameter for a $M = 2 \times 10^{15} M_\odot$ cluster assuming different baryonic physics. Field of view is $5 h^{-1}$ Mpc. Right: Corresponding spherically averaged radial temperature profiles.

4.6. Effect of additional physics. – Within $r=0.15 r_{vir}$, Vikhlinin et al. [50] found large variation in temperature profiles, but in all cases the gas is cooler than the cluster mean. This suggests that radiative cooling is important in cluster cores, and possibly other effects as well. It has been long known that ~ 60 percent of nearby, luminous X-ray clusters have central X-ray excesses, which has been interpreted as evidence for the presence of a cluster-wide cooling flows [64]. More recently, Ponman et al. [29] have used X-ray observations to deduce the entropy profiles in galaxy groups and clusters. They find an entropy floor in the cores of clusters indicative of extra, non-gravitational heating, which they suggest is feedback from galaxy formation. It is easy to imagine cooling and heating both may be important to the thermodynamic evolution of ICM gas.

To explore the effects of additional physics on the ICM, we recomputed the entire sample of clusters changing the assumed baryonic physics, keeping initial conditions the same. Three additional samples of about 100 clusters each were simulated: The “radiative cooling” sample assumes no additional heating, but gas is allowed to cool due to X-ray

line and bremsstrahlung emission in a 0.3 solar metallicity plasma. The “star formation” sample uses the same cooling, but additionally cold gas is turned into collisionless star particles at a rate $\dot{\rho}_{SF} = \varepsilon_{sf} \frac{\rho_b}{\max(\tau_{cool}, \tau_{dyn})}$, where ε_{sf} is the star formation efficiency factor ~ 0.1 , and τ_{cool} and τ_{dyn} are the local cooling time and freefall time, respectively. This locks up cold baryons in a non-X-ray emitting component, which has been shown to have an important effect of the entropy profile of the remaining hot gas [56, 57]. Finally, we have the “star formation feedback” sample, which is similar to the previous sample, except that newly formed stars return a fraction of their rest mass energy as thermal and mechanical energy. The source of this energy is high velocity winds and supernova energy from massive stars. In *Enzo*, we implement this as thermal heating in every cell forming stars: $\Gamma_{sf} = \varepsilon_{SN} \dot{\rho}_{SFC} c^2$. The feedback parameter depends on the assumed stellar IMF the explosion energy of individual supernovae. It is estimated to be in the range $10^{-6} \leq \varepsilon_{SN} \leq 10^{-5}$ [44]. We treat it as a free parameter.

Fig. 10 shows synthetic maps of X-ray surface brightness, temperature, and Compton y-parameter for a $M = 2 \times 10^{15} M_\odot$ cluster at $z=0$ for the three cases indicated. The “star formation” case is omitted because the images are very similar to the “star formation feedback” case (see reference [46].) The adiabatic cluster shows that the X-ray emission is highly concentrated to the cluster core. The projected temperature distribution shows a lot of substructure, which is true for the adiabatic sample as a whole [47]. A complex virialization shock is toward the edge of the frame. The y-parameter is smooth, relatively symmetric, and centrally concentrated. The inclusion of radiative cooling has a strong effect on the temperature and X-ray maps, but relatively little effect on the SZE map. The significance of this is discussed in Section 5. In simulations with radiative cooling only, dense gas in merging subclusters cools to 10^4 K and is brought into the cluster core intact [48]. These cold lumps are visible as dark spots in the temperature map. They appear as X-ray bright features. The inclusion of star formation and energy feedback erases these cold lumps, producing maps in all three quantities that resemble slightly smoothed versions of the adiabatic maps. However, an analysis of the radial temperature profiles (Fig. 10) reveal important differences in the cluster core. The temperature continues to rise toward smaller radii in the adiabatic case, while it plummets to $\sim 10^4$ K for the radiative cooling case. While the temperature profile looks qualitatively similar to observations of so-called cooling flow clusters, our central temperature is too low and the X-ray brightness too high. The star formation feedback case converts the cool gas into stars, and yields a temperature profile which follows the UTP at $r \geq 0.15 r_{vir}$, but flattens out at smaller radii. This is consistent with the high resolution *Chandra* observations of Vikhlinin et al. [50].

5. – Comparisons and predictions for X-ray and SZE surveys

In this section we shall compare the results of numerical hydrodynamical simulations with the analytic scaling laws derived in section 3, and compare with observational data. We will see that the X-ray temperature and the integrated SZE is a robust indicator of cluster mass with relatively little bias, while the X-ray luminosity is not because we cannot reliably simulate the X-ray emission from clusters.

5.1. Analytic and numerical comparisons. – We first ask the question how well do the simple analytic model estimates of cluster statistics agree with the results of numerical hydrodynamic simulations. This question was addressed by Bryan & Norman 1998 [13]. Fig. 11 illustrates how the comparisons are made. For a given cosmological

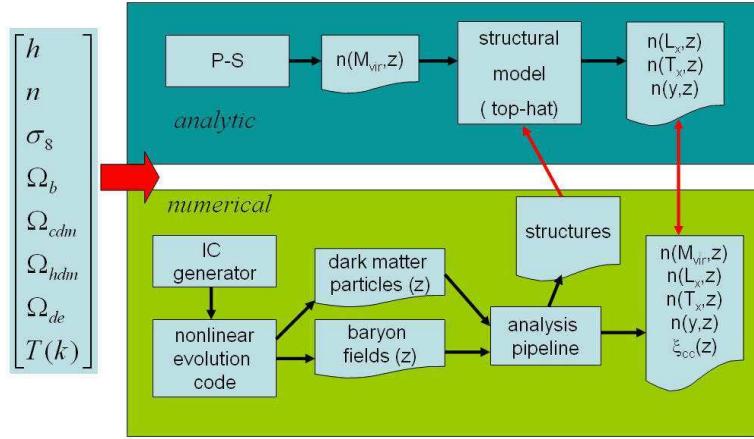


Fig. 11. – Comparing analytic and numerical predictions for cluster statistics.

model Press-Schechter theory is used to calculate the halo mass function versus redshift (top rectangle). The observable quantities $n(T, z)$, $N(L_x, z)$, $n(Y, z)$ are then computed using the scaling relations presented in Section 3 for L_x , T and Y as a function of mass. Somewhat more work is involved deriving these results from numerical simulation (bottom rectangle). Initial conditions for the chosen cosmology are generated which specify dark matter and baryonic perturbations at the starting redshift. These perturbations are evolved use in the methods described in section 4 to $z=0$. The particle and baryonic distributions are output at specified redshifts for analysis. Virialized objects are located using a group-finding algorithm on the dark matter particles list. Two popular techniques are friends-of-friends [52] and HOP [53]. In the friends-of-friends algorithm, two particles are part of the same group if their separation is less than some chosen value; chains of pairs then define groups. In the HOP algorithm, an estimate of the local density is associated with every particle. Each particle is linked to its densest neighbor and on to that particle's densest neighbor until one reaches the particle which is its own densest neighbor. All particles that are traced to the same such particle define the group. Once groups are found, centers of masses for each group are computed. With these centers determined, spherically averaged profiles of dark matter density, baryon density, temperature, etc. are computed by binning the 3D data into spherical shells. For each halo, the virial radius is determined by find the shell inside of which the mean total density (dark matter + baryons) equals the critical overdensity Δ_c (Section 3). Virial mass, X-ray luminosity, and emission weighted temperature are computed by numerical integration over the radial profiles of total density, X-ray emissivity, etc. With these quantities evaluated for each cluster in the sample, distribution functions are then computed.

5.2. Cluster temperatures. – One of the most robust predictions of numerical simulations is the mass-temperature relation. Fig. 12a shows a comparison between analytic scaling relations and simulations for two cosmological models at three epochs. The simulations were carried out on fixed Eulerian grids of size 270^3 and 512^3 assuming the clusters are non-radiative. Good agreement is seen with a slight offset in normalization. Fitting Eq. 24 to the data yields $f_T \approx 0.8$. That the simulations reproduce the analytic

scaling relations despite limited numerical resolution is a consequence of energy conservation, which is maintained to high accuracy by the numerical hydrodynamic method employed. Note that a cluster of a given mass is cooler at lower redshifts.

Fig. 12b shows the temperature distribution function as predicted by simulations (histograms) and Press-Schechter theory (curves) for a critically closed model (SCDM) and a low density model (OCDM). Generally, agreement is good. Simulations underpredict the number of low temperature clusters due to resolution effects. The high temperature clusters are rare, and thus not many are found in our small box. Despite these numerical limitations, one sees that the number of hot clusters evolves rapidly in the flat universe but evolves very little in the open universe.

Fig. 13a shows the predictions of simulations compared with the observational data of Henry & Arnaud (1991)[54]. The SCDM model is ruled out with high confidence, while the CHDM and OCDM models are marginally consistent with data. Eke, Cole & Frenk (1996) [18] showed that with a suitable adjustment of σ_8 , a critically closed, open, and Λ -dominated models could all reproduce the observations (Fig. 13b). This illustrates what is known as the $\Omega_0 - \sigma_8$ degeneracy in cluster abundances [55]. The redshift evolution of cluster abundances can in principle break this degeneracy, however this requires large samples of high redshift clusters with accurately measured temperatures. So far, the samples are small. Temperatures are more difficult to measure than X-ray luminosities. Nonetheless, available data shows mild evolution of the X-ray temperature function, consistent with a low density universe [3].

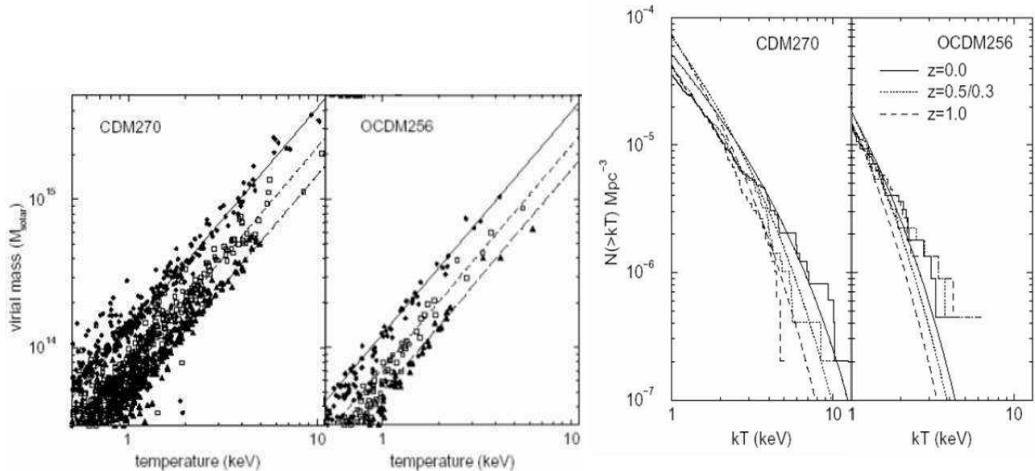


Fig. 12. – Left: M-T scaling in a flat $\Omega_m=1$ universe (left) and an open $\Omega_m=0.34$ universe (right) for $z=0, 0.5$, and 1 (top to bottom). Symbols are measured values hydrodynamic simulations. Lines are the scaling relations from Eq. 24, with $f_T=0.8$ (from [13]). Right: Evolution of cumulative temperature distribution function for the two models shown in Fig 13 as predicted by theory (curves) and hydrodynamic simulations (histograms). The number of hot clusters evolves rapidly in the flat universe but evolves very little in the open universe.

5.3. Cluster X-ray luminosities. – The most easily measured property of an X-ray cluster is its luminosity. However, as we shall see, this is the most difficult quantity to predict using numerical simulations. This is because the integrated X-ray luminos-

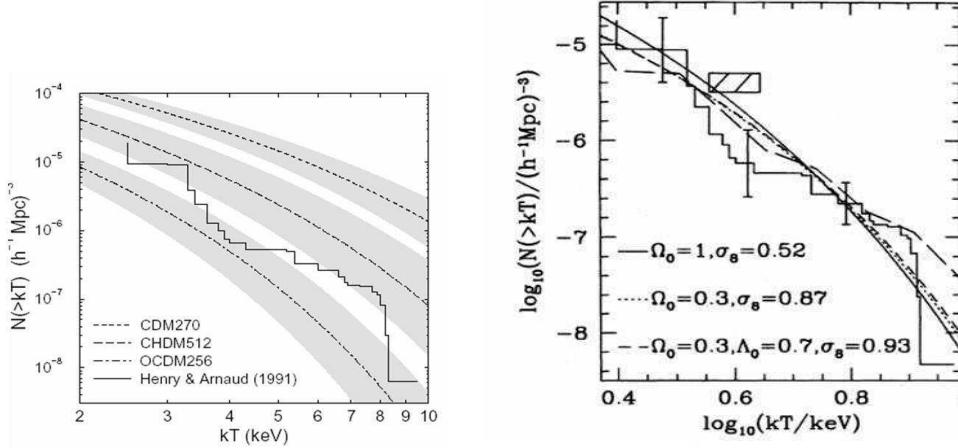


Fig. 13. – Left: Comparison of $z=0$ cluster temperature function from Henry & Arnaud (1991) with hydrodynamic simulations. SCDM model ($\Omega_0=1$, $\sigma_8=1.05$) is ruled out with high confidence, OCDM model ($\Omega_0=0.34$, $\sigma_8=0.75$) is marginally consistent with data. (from Bryan & Norman 1998). Right: Figure 18. Illustration of the $\Omega_0 - \sigma_8$ degeneracy. Good agreement with data is found for flat, open, and Λ -dominated cosmological models with a suitable adjustment of σ_8 . From [18].

ity of a cluster is dominated by emission from the core region, which is challenging to resolve numerically, and it is affected by heating and cooling processes which are as yet not well understood. The advent of multiscale numerical simulation techniques has ameliorated the numerical resolution difficulties. As one can see from Fig. 8f, the X-ray emissivity peaks at about $0.1r_{vir}$ for the adiabatic Santa Barbara cluster. SPH and AMR simulations can now resolve this scale with ten resolution elements or more in large cosmological volumes. Fig. 14 shows the $L_x - M$ and $L_x - T$ scaling relation derived from our large sample of adiabatic galaxy clusters simulated using AMR in a Λ CDM universe. The numerical clusters are in good agreement with the analytic virial scaling relations $L_x \propto M^{4/3}$ and $L_x \propto T^2$ without resort to resolution corrections (cf. Bryan & Norman 1998). However, the adiabatic models are in conflict with the observed scaling relation, which are $L_x \propto M^{1.8}$ and $L_x \propto T^3$ for $T > 2$ keV [3].

The disagreement between the predictions of adiabatic simulations and observations can be taken as strong evidence of the importance of non-adiabatic processes in the cores of galaxy clusters. The effect of radiative cooling is shown by the open diamonds in Fig. 14. Although the $L_x - M$ and $L_x - T$ scaling steepens in the direction of observations, we view these models as unrealistic since every cluster in the sample has too much cold gas in the core, contrary to observations. The scaling relations for the “star formation” and “star formation feedback” samples are show in Fig. 15a. The conversion of cool gas into stars produces clusters whose temperature and X-ray surface brightness profiles are in better agreement with observations, and steepens the $L_x - T$ relation somewhat relative the to adiabatic clusters. The inclusion of supernova heating has a rather minor effect when compared to the magnitude of the change including star formation. This is best illustrated in Fig. 15b, which shows the scatter of central entropy versus central temperature for the adiabatic, star formation, and star formation

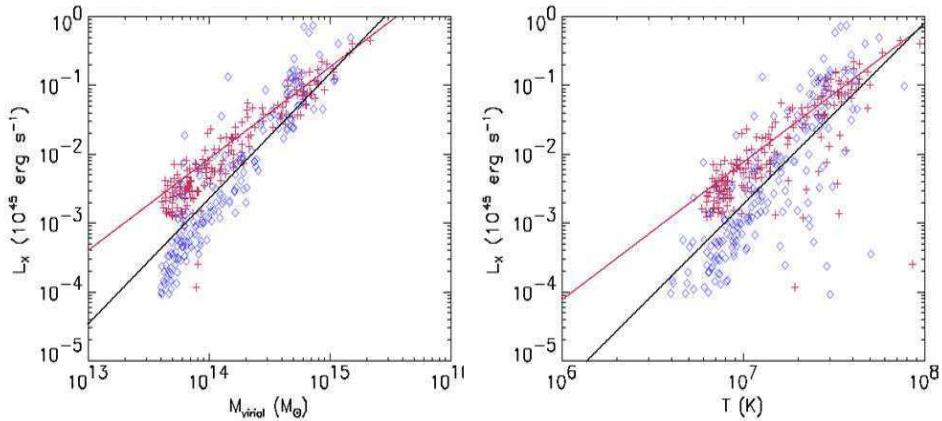


Fig. 14. – High resolution AMR simulations of adiabatic clusters (red crosses) agree with analytic scaling predictions (red lines), but disagree with observations (black lines). Addition of radiative cooling (blue diamonds) improves agreement, but produces too many clusters with cool cores. Figures courtesy P. Motl.

feedback cluster samples. An analysis of a sample of clusters by Ponman et al. (1999) [29] revealed the existence of an “entropy floor”. This feature has been interpreted as evidence of galaxy formation feedback which increases gas entropy. The same data has been explained as the result of radiative cooling [56, 57] which locks up low entropy gas in stars where it does not contribute to X-ray emission. The magnitude of the entropy floor strongly suggests the heating explanation. The failure of star formation feedback simulations to exhibit the entropy floor may be due to limited mass resolution. The galaxy mass function is not well sampled in these simulations; indeed, only the central dominant galaxy and one or two of the most massive galaxies are present in these simulations. Perhaps higher resolution simulations will improve agreement. AGN heating is another source of energy input that may be important, especially in the cores of clusters [58]. Numerical simulations incorporating these effects are in their infancy, and certainly not at the stage where large ensembles can be simulated for statistical analysis.

5.4. Prospects for SZE cluster surveys. – The sensitivity of X-ray luminosity to numerical resolution and baryonic processes motivates us to look for other more robust indicators of a cluster’s mass. Temperature is such an indicator, however this is more difficult to measure than X-ray luminosity even at low redshifts. At high redshifts the task becomes even more difficult because of the severe $(1+z)^{-4}$ surface brightness dimming of the X-ray flux. In this section we explore the thermal SZE effect as a mass indicator based on our four catalogs of simulated galaxy clusters. Based on these models, we find that the integrated SZE y_{500} is a less biased indicator of cluster mass than either the X-ray luminosity or temperature, and shows far less scatter than the central value of the SZE intensity change y_0 . More details can be found in references [46, 49]

As has been discussed elsewhere in this volume (Rephaeli, Birkinshaw), the thermal SZE is an attractive cosmological probe because it is redshift independent. The strength of the SZE is proportional to the Compton parameter, y , which for non-relativistic elec-

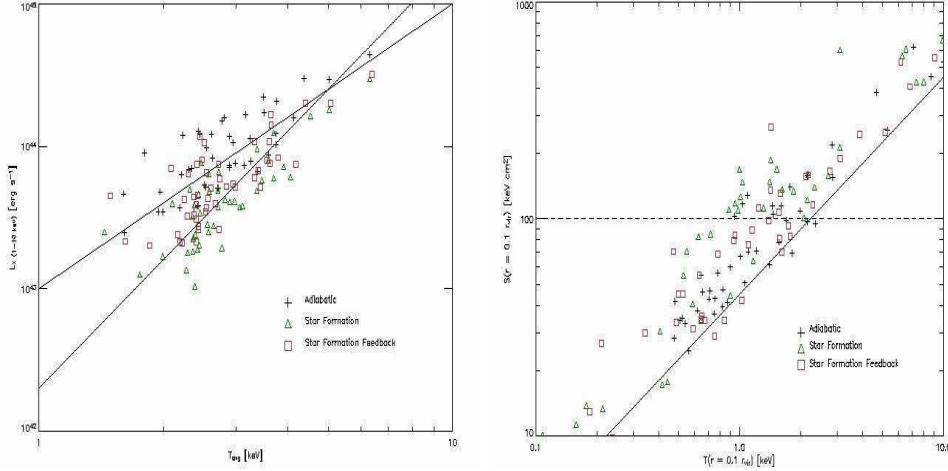


Fig. 15. – Left: Effect of baryonic physics on the L-T relation for three AMR cluster samples: adiabatic (crosses), star formation (triangles), and star formation feedback (squares). Right: Central entropy versus central temperature for the cluster samples in Fig 12. The dashed line is the observed “entropy floor”. Figures courtesy P. Motl.

trons is essentially the integral of the gas pressure through the cluster

$$(35) \quad y = \int \frac{k_B T}{m_e c^2} \sigma_T n_e d\ell \propto \int n T d\ell.$$

The central value of the Compton y parameter we refer to as y_0 . We define the integrated SZE y_{500} as the area integral of the y parameter out to r_{500} , the radius inside of which

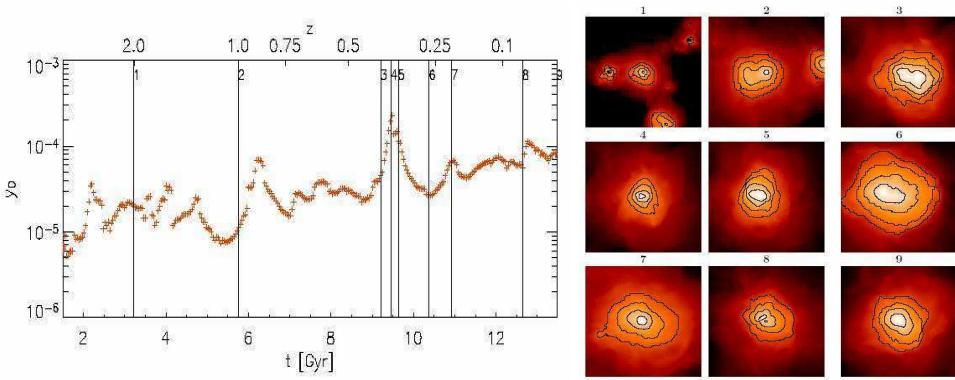


Fig. 16. – Left: The “lightcurve” for the central value of the Compton parameter, y_0 , obtained from tracking one particular halo from a redshift of 4 to the present epoch. Major mergers can boost y_0 by a factor of 10. Right: Projected y parameter distribution of cluster at the epochs marked by vertical lines in the lightcurve. Figures courtesy P. Motl.

the mean density is 500 times the critical density:

$$(36) \quad y_{500} = 2\pi \int_0^{r_{500}} y(r) r dr.$$

The detectability of a cluster is given by its SZ cross section (Section 3), which is essentially $y_{500}/d_A^2 \propto (1+z)^{-2}$. This is far more favorable redshift dependence than X-rays provide.

Fig. 16a shows the redshift evolution of y_0 for the most massive cluster in our sample. As can be seen, y_0 exhibits a secular increase as the cluster potential deepens, but is boosted by up to a factor of $\sim 20(2)$ during major(minor) merger events. The duration of these events is of order the dynamical time $\sim 1-2$ Gyr. The effect of mergers induces considerable scatter into scaling between y_0 and the enclosed mass M_{500} in our sample of clusters at $z=0$ (Fig. 17a). By contrast, y_{500} shows a much tighter correlation (Fig. 17b). The reason for this is illustrated in the lower two panels of Fig. 17 where we plot the central value of the gas pressure p_0 and the volume averaged pressure $p_{500} = \frac{3}{4\pi r_{500}^3} \int_0^{r_{500}} p(\vec{x}) d^3 \vec{x}$. The central pressure exhibits large scatter due to the presence of shock waves induced by mergers. However, the volume averaged pressure exhibits relatively little scatter. This is a consequence of virial equilibrium and tells us that the clusters are approximately in equilibrium within r_{500} .

Fitting the data to a power law of the form

$$(37) \quad y_{500} = A \left[\frac{M_{500}}{10^{14} M_\odot} \right]^\alpha$$

for each of our 4 catalogs, we find $\alpha \sim 1.6, \sigma_\alpha \sim 0.025$ for the adiabatic, star formation, and star formation feedback samples, and $\alpha \sim 1.7, \sigma_\alpha \sim 0.03$ for the radiative cooling sample. The scaling exponent is consistent with the findings of da Silva et al (2004) [59]. Ignoring the radiative cooling only runs as unrealistic, we find that the scaling is relatively insensitive to baryonic physics. This is both reassuring and understandable in that regardless of the thermodynamics of the gas, hydrostatic equilibrium is maintained to a good approximation. By looking back through our catalogs in redshift, we find that the coefficient A is independent of redshift.

5.5. Cluster mass estimates compared. – To assess the systematic biases and relative scatter of various means of estimating cluster masses from X-ray and SZE data, we “observed” our four clusters samples and analyzed the resulting synthetic images in the same way as observations. Our goal was to find both the best cluster mass estimator and best method of analysis. These were defined as the combination which produce the least bias and smallest scatter between inferred cluster mass and actual (simulated) mass. Here we merely summarize our findings; for details the reader is referred to [49].

Cluster masses can be obtained from X-ray and thermal SZE observations in several ways. The most widely used is the isothermal beta model, wherein it is assumed the electron number density is spherically symmetric and follows

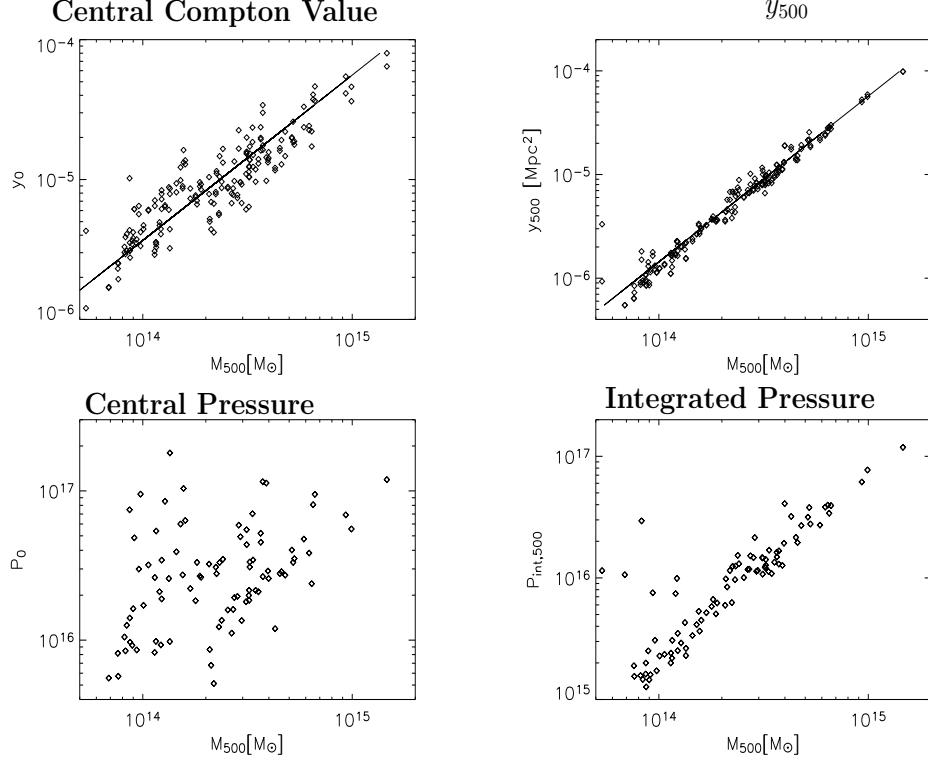


Fig. 17. – Upper: The scaling relations between y_0 and y_{500} and the total cluster mass within the same radius at $z=0$ for the star formation with feedback cluster sample. Two randomly chosen, orthogonal projections for each cluster are plotted as individual points and the catalog contains ~ 100 clusters at this epoch in the mass range $1 \times 10^{14} M_\odot \leq M_{200} \leq 2 \times 10^{15} M_\odot$. The best fit relations are plotted as solid lines. Lower: Central pressure and pressure integrated inside sphere of radius r_{500} plotted against cluster total mass. From [46].

$$(38) \quad n_e(r) = n_{e0} \left[1 + \left(\frac{r}{r_c} \right)^2 \right]^{-3\beta/2},$$

where n_{e0} is the central electron density. Approximating the gas as isothermal with average temperature $\langle T \rangle$ within the fitting radius, then the X-ray surface brightness is

$$(39) \quad S_X(r) = S_{X0} \left[1 + \left(\frac{r}{r_c} \right)^2 \right]^{\frac{1}{2}-3\beta}$$

where $S_{X0} \propto n_{e0}^2 \langle T \rangle^{\frac{1}{2}}$. Similarly for the SZE, a beta model density distribution results

in a projected radial distribution for the Compton y parameter

$$(40) \quad y(r) = y_0 \left[1 + \left(\frac{r}{r_c} \right)^2 \right]^{\frac{1}{2} - \frac{3\beta}{2}}$$

where $y_0 \propto n_{e0} \langle T \rangle$.

By fitting the observed profiles of $S_x(r)$ and $y(r)$ one obtains β and r_c , the core radius. With $\langle T \rangle$ measured observationally, n_{e0} can then be calculated. One then integrates Eq. 38 to find the gas mass within the fitting radius $r_<$. The cluster dynamical mass is then $M_{dyn}(r_<) = M_{gas}(r_<)/f_b(r_<)$, where f_b is the baryon fraction which may in general be different from the cosmic mean Ω_m/Ω_b depending upon the radius. Henceforth we will refer to mass estimates made in this way as X-ray-ISO and SZE-ISO.

Recently it has been shown both in simulations (Loken et al. 2002, Section 4) and in X-ray observations (Vikhlinin et al. 2005) that clusters are not isothermal at large radii, but follow a universal temperature profile (UTP)

$$(41) \quad T(r) = \langle T \rangle_{500} \left[1 + \left(\frac{r}{\alpha r_{500}} \right)^2 \right]^{-\delta}$$

where $\langle T_{500} \rangle$ is the average temperature inside r_{500} , and α and δ are fitting parameters determined from a large sample of clusters. Improved mass estimates can be obtained by geometric deprojection of the X-ray and SZE profiles if one knows the temperature of each radial shell. This is provided by the UTP. For example, the X-ray surface brightness can be deprojected to yield the X-ray emissivity in each spherical shell (e.g., [60]). Knowing the temperature profile, once can obtain the mass in each shell. A similar technique can be applied to the SZE profile. By summing over shells, one obtains the gas mass within the fitting radius. Mass estimates obtained in this way we refer to as X-ray UTP and SZE-UTP.

Fig. 18 shows the ratio of the measured mass to the actual mass for the star formation feedback catalog of simulated clusters for the four methods described above. The triangles are the full sample, whereas the diamonds are for samples which have been cleaned of highly distorted clusters resulting from recent mergers. The error bars enclose the 80% confidence range. As can be seen, cleaning the sample reduces the scatter considerably. Among the different methods, the X-ray measurements yield the smallest scatter, but overestimate the cluster masses by 5-10%. Conversely, the SZE-UTP measurements yield unbiased estimates the cluster mass, with somewhat more scatter. As shown in [49], the scatter in the SZE estimates decreases as the fitting radius is increased to r_{200} , while no improvement is seen in the X-ray estimates. This is to be expected since the X-ray emission is heavily core-weighted, while the SZE samples larger radii.

5.6. Conclusions. – We have seen that galaxy clusters are sensitive cosmological probes provided their masses can be measured with precision. Both analytic estimates and numerical simulations show that the evolution of their comoving number density is sensitive to cosmology. With improvements in X-ray observations and impending large area surveys to detect clusters via the SZE, it is paramount to assess the accuracy to which cluster masses can be obtained observationally. Based on our catalogs of simulated clusters using adaptive mesh refinement, we find that gas masses can be measured to $\sim 10\%$

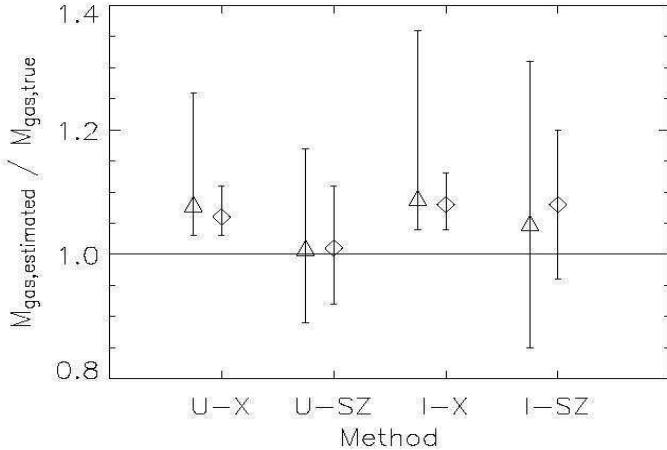


Fig. 18. – Comparison of median values and scatter of gas mass estimates inside r_{500} for full SFF cluster sample (triangles) and cleaned SFF sample (diamonds) at $z=0$ for each of four methods: UTP-X-ray (U-X), UTP-SZE (U-SZ), isothermal X-ray (I-X), and isothermal SZE (I-SZ) as described in the text. From [49].

accuracy with 80% confidence. Our study ignores instrumental or other observational effects. These limits in precision are a direct result of the deviation of the simulated clusters from simple assumptions about their physical and thermodynamic properties, dynamical state, and sphericity. Comparing a variety of methods, we find that SZE methods assuming a UTP produce the smallest scatter when estimating masses from a raw sample of clusters. Cleaning the cluster sample of obvious mergers does not improve the SZE estimates much, but improves the X-ray estimates substantially. As a practical matter, we find SZE methods are superior for mass estimation of large samples of clusters out to high redshift. This is particularly true if the cutoff radius is the virial radius, as this has the effect of smoothing out any boosting effects in the cluster core due to mergers.

Comparing mass estimates from our four catalogs, we find that our conclusions are insensitive to assumed baryonic physics, except for the cooling sample, which yields unrealistic-looking clusters. Mass estimates derived from the cooling sample are systematically high (50–100%) despite excising the overluminous X-ray core. Reasons for this are discussed in detail in reference [49]. We conclude that cool core clusters are poor candidates for precision mass estimation, in disagreement with previous studies [61].

* * *

The author is indebted to his collaborators Greg Bryan, Jack Burns, Eric Hallman, Chris Loken, and Patrick Motl whose results, both published and unpublished, are presented here. Simulations were performed at the National Center for Supercomputing Applications of the University of Illinois, Urbana-Champaign with support from NSF grants ASC-9318185, AST-09803137.

REFERENCES

- [1] CARLSTROM, J., HOLDER, G., & REESE, E., *Ann. Rev. Astron. Astrophys.*, **40** (643) 2002.
- [2] SPRINGEL, V., WHITE, S. ET AL., *Nature*, **435** (629) 2005
- [3] ROSATI, P., BORGANI, S., & NORMAN, C., *Ann. Rev. Astron. Astrophys.*, **40** (539) 2002.
- [4] DODELSON, S., *Modern Cosmology*, (Academic Press, Amsterdam), 2003.
- [5] PERLMUTTER, S., *Physics Today*, **April 2003** (53)
- [6] SPERGEL, D. ET AL., *ApJS*, **148** (175) 2003.
- [7] BAHCALL, N. A.; OSTRIKER, J. P.; PERLMUTTER, S.; STEINHARDT, P. J., *Science*, **284** (1481) 1999
- [8] TEGMARK, M. ET AL., *ApJ*, **606** (702) 2004.
- [9] WHITE, S., EFSTATHIOU, G., & FRENK, C., *MNRAS*, **262** (1023) , 1993.
- [10] KOLB, E. & TURNER, M., *The Early Universe*, (Addison-Wesley, Redwood City, CA), 1990.
- [11] KOMATSU, E. ET AL., *ApJS*, **148** (119) 2003.
- [12] PADMANABHAN, T., *Structure Formation in the Universe*, (Cambridge University Press, Cambridge), 1994.
- [13] BRYAN, G. & NORMAN, M., *ApJ*, **495** (80) 1998.
- [14] KAISER, N., *MNRAS*, **222** (323) 1986.
- [15] BINNEY, J. & TREMAINE, S., *Galactic Dynamics*, (Princeton University Press, Princeton, USA), 1987.
- [16] PRESS, W. & SCHECHTER, S., *ApJ*, **187** (425) 1974.
- [17] WHITE, S. D. M., *Cosmology and Large Scale Structure: Proceedings of Les Houches Summer School*, R. Schaeffer et al., editors, (Elsevier, Amsterdam), 1996.
- [18] EKE, V., COLE, S. & FRENK, C., *MNRAS*, **281** (703)
- [19] NORMAN, M. L., *Matter and Energy in Clusters of Galaxies*, *ASP Conference Series Vol. 301*, S. Boyer & C.-Y. Hwang, eds., (Astronomical Society of the Pacific, San Francisco), p. 419, 2003.
- [20] FRENK, C. ET AL., *ApJ*, **525** (554) 1999
- [21] KATZ, N., WEINBERG, D. & HERNQUIST, L., *ApJS*, **105** (19) 1996
- [22] SPRINGEL, V., YOSHIDA, N., & WHITE, S., *NewA*, **6** (79) 2001
- [23] BRYAN & NORMAN, M., *Computational Astrophysics; 12th Kingston Meeting on Theoretical Astrophysics*, D. A. Clarke and M. Fall, editors, *ASP Conference Series # 123*, 1997.
- [24] KRAVTSOV, A., KLYPIN, A., & KOKHLOV, A., *ApJS*, **111** (73) 1997
- [25] TEYSSIER, R., *Astron. Astrophys.*, **385** (337) 2002
- [26] O'SHEA, B. ET AL., *Adaptive Mesh Refinement—Theory and Applications*, T. Plewa et al., eds., Springer Lecture Notes in Computational Science & Engineering, (Springer, Berlin), 2005.
- [27] SPRINGEL, V. ET AL., *MNRAS*, **328** (726) 2001
- [28] ANNINOS, P. ET AL., *NewA*, **2** (209) 1997
- [29] PONMAN, T., CANNON, D. & NAVARRO, J., *Nature*, **397** (135) 1999
- [30] EFSTATHIOU, G. ET AL., *ApJS*, **57** (241) 1985
- [31] COUCHMAN, H., *ApJL*, **368** (L23) 1991
- [32] BARNES, J. & HUT, P., *Nature*, **324** (446) 1986
- [33] WARREN, M. & SALMON, J., *Comp. Phys. Comm.*, **87** (266) 1995
- [34] XU G., *ApJS*, **98** (355) 1995
- [35] EVRARD, A., *MNRAS*, **235** (911) 1988.
- [36] KRAVTSOV, A., KLYPIN, A. & HOFFMAN, Y., *ApJ*, **571** (563) 2002
- [37] KANG, H. ET AL., *ApJ*, **428** (1) 1994
- [38] BRYAN, G. ET AL., *ApJ*, **428** (405) 1994.
- [39] BORGANI, S. ET AL., *MNRAS*, **348** (1078) 2004
- [40] BERGER, M. & COLELLA, P., *J. Comp. Phys.*, **82** (64) 1989

- [41] R. HOCKNEY AND J. EASTWOOD, *Computer Simulation Using Particles*, (McGraw Hill, New York), 1988.
- [42] P. COLELLA AND P. R. WOODWARD, *J. Comp. Physics*, **54** (174) 1984
- [43] W. Y. ANNINOS & M. L. NORMAN, *ApJ*, **429** (434) 1994
- [44] CEN, R. & OSTRIKER, J., *ApJ*, **417** (404) 1993
- [45] NAVARRO, J., FRENK, C. & WHITE, S., *ApJ*, **462** (563) 1996
- [46] MOTL, P. ET AL., *ApJL*, **623** (L63) 2005
- [47] LOKEN, C. ET AL., *ApJ*, **579** (571) 2002
- [48] MOTL, P. ET AL., *ApJ*, **606** (635) 2004.
- [49] HALLMAN, E. ET AL., *preprint*, astro-ph/0509460
- [50] VIKHLININ, A. ET AL., *ApJ*, **628** (655) 2005
- [51] ASCASIBAR, Y. ET AL., *MNRAS*, **346** (731) 2003
- [52] DAVIS, M. ET AL., *ApJ*, **292** (371) 1985
- [53] EISENSTEIN, D. & HUT, P., *ApJ*, **498** (137) 1998
- [54] HENRY, J. P. & ARNAUD, K., *ApJ*, **372** (410) 1991
- [55] BAHCALL, N., FAN, X. & CEN, R., *ApJ*, **485** (L53) 1997
- [56] BRYAN, G., *ApJ*, **544** (L1) 2000
- [57] VOIT, M. & BRYAN, G., *ApJ*, **551** (L139) 2001
- [58] RUSZKOWSKI, M., BRUGGEN, M. & BEGELMAN, M., *611*, **158** (2004)
- [59] DA SILVA, A. ET AL., *MNRAS*, **348** (1401) 2004
- [60] BUOTE, D. A., *ApJ*, **539** (172) 2000
- [61] ALLEN, S. & FABIAN, A., *MNRAS*, **297** (L57) 1998
- [62] MARKEVITCH, M. ET AL., *ApJ*, **503** (77) 1998
- [63] DE GRANDI, S. & MOLENDI, S., *ApJ*, **567** (163) 2002
- [64] FABIAN, A. C., *ARAA*, **32** (277) 1994