Metallicity Variations and Conduction in the Intracluster Medium

THIS DISSERTATION IS SUBMITTED FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
March 2003

Roger Glenn Morris

INSTITUTE OF ASTRONOMY
&
ROBINSON COLLEGE

UNIVERSITY OF CAMBRIDGE

For J. A. M. and J. K. M.

Declaration

I hereby declare that my thesis entitled *Metallicity Variations and Conduction in the Intracluster Medium* is not substantially the same as any that I have submitted for a degree or diploma or other qualification at any other University.

I further state that no part of my thesis has already been or is being concurrently submitted for any such degree, diploma or other qualification.

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. Those parts of this thesis which have been published or accepted for publication are as follows.

- Sections of Chapters 2 and 4 were published as:
  Morris R. G., Fabian A. C., 2003, MNRAS, 338, 824.
- Some of the issues discussed in Chapter 4 were presented in preliminary form in:
  Morris R. G., Fabian A. C., 2002, in Matteucci and Fusco-Femiano (2002), pp. 85–90.
- Work on thermal conduction (Chapter 5) has benefited from discussions with Lisa Voigt, published as:
  Fabian A. C., Voigt L. M., Morris R. G., 2002b, MNRAS, 335, L71.

Various figures throughout the text are reproduced from the work of other authors, for illustration or discussion. Such figures are always credited in the associated caption.

This thesis contains fewer than 60,000 words.

R. Glenn Morris
Cambridge, 10th March 2003

Summary

This thesis focuses mainly on an investigation of possible explanations for the lack of low-temperature emission lines seen in recent X-ray spectra of cooling flow clusters. I consider two main hypotheses.

Firstly, I examine the idea that the metals in the intracluster medium might be distributed inhomogeneously on small, as yet unresolved scales. I make a theoretical analysis of the effects that such a metal distribution would have on the X-ray spectra of cooling plasmas such as are found in the centres of galaxy clusters. Using the standard isobaric cooling flow model, I quantify the suppression in low-temperature line emission that can be achieved in this way. Different species are found to behave in different fashion, as governed by the temperature dependence of their equivalent widths.

I go on to describe my implementation of a one-dimensional Lagrangian numerical scheme that can be used to model the intracluster medium in either a single-phase or multi-phase mode. Synthetic X-ray observations can be produced and analysed using standard tools. This scheme is used to investigate the spatial and temporal evolution of simulated clusters with small-scale metallicity variations. I stress the effects on the observed spectra, and the deductions that would be made from them. I find that such metal distributions may give rise to interesting effects in observed abundance profiles, in that apparent abundance gradients with central abundance drops and off-centre peaks, similar to those seen in several real clusters, can be produced. The importance of observational testing of this hypothesis is stressed.

Secondly, I study some of the possible effects of efficient thermal conduction in the intracluster plasma. The numerical code is modified to include the effects of thermal conduction. The level of conductivity is controlled by the structure of the cluster magnetic field, and I include a simple parametrization to represent this. The influence of conduction on the evolution of the temperature profile for high- and low-temperature clusters is demonstrated, as well as its ability to delay the onset of cooling in cluster cores over a wide range of cluster parameters. The simulations highlight the difficulty in constructing stable solutions to the cooling flow problem based on thermal conduction, but do show that the properties of clusters are such that conduction is potentially an important process for their evolution.

Acknowledgements*

Or, "This is all your fault, you know"

Well, here we are. To be honest, there were times when I wondered if I'd ever get to this point. That I have done is largely due, I think, to the people listed below. This is poor recompense for putting up with me, but I hope to make it up to you all in the future. It's been a long three-and-some years, and what I've learnt about astronomy is far outweighed by the other things you have taught me. I may go on for too long here, but what the hell, nobody's forcing you to read this. If I were to write fully what I owe to all of you, it would be a second thesis, but one is quite enough.

It's a cliché to say that this thesis wouldn't exist without my supervisor, Andy Fabian, but that doesn't make it any less true. Thanks to him for suggesting many interesting lines of inquiry, and for tolerating my appalling slowness in following them up.

Thanks to Andy Amos for putting up with me during my A-levels, for appreciating the value of sarcasm, and for not complaining too much when I turned my back on biology.

I'd like to thank/blame Alastair Edge for getting me started in the field of X-ray astronomy in the first place, all those years ago. It was him who made me think this seemed like it could be a fun thing to do. Alastair is also responsible for my firm belief that unfeasibly large tea mugs and astronomy should go hand in hand. I'm also grateful for the observing trip to IRAM Granada, to see how the other half lives.

Thanks to Thomas Hove Pedersen for helping me get through my undergraduate years, though I fear our personalities were too similar to be good for either of us. Still, another pair of hands to steady the mortar is always useful. It looks like I haven't learnt your "get out of Cambridge" lesson yet Tom, so you'll have to keep explaining it to me.

Thanks to Felipe Menanteau for being one of those to welcome me to the Hoyle 19 family on my arrival to the IoA, and for being a part of the project to educate me with the *palabra del día*. I'm sorry I turned out to be such a poor pupil.

To Meghan Gray for a lot of sensible advice, many emails, some fun conversations, and generally for taking an interest. As another one of the original senior figures (sorry!) of H19 she contributed enormously to the great (and much missed) atmosphere. Thank you also Meghan for some much-needed support during some difficult times.

To Mateusz "the Disco King" Ruszkowski for his shared appreciation of Desperados beer, and for managing to teach me one word of Polish. Sadly your dancing shoes remain unfilled in the Group.

Thanks to Richard Wilman, who despite having the singular misfortune in life to have been born a Yorkshireman, remains a thoroughly decent bloke.

To Raquel Morales for an awful lot of things, including some very civilised lunches, a lot of fruit tea, some fine walks and many good conversations. Heartfelt thanks for being a pillar of the small but select group that tried to keep me sane, and apologies for making it such a difficult job at times.

Thanks to Andrea Merloni and Alessandra Corchia for tolerating so well my decidedly non-Italian temperament, some good parties, and for making it possible for me to see Rome.

To Enrico Ramirez Ruiz and Kristin Truse I owe more than I can say. Thank you for many fine dinners, for your tolerance and understanding, for Mexico, for putting up with your unsociable flatmate,

*The OED recommends "acknowledgments". What is the world coming to?

for Borges, and above all for your friendship. *El deseo de expresar gratitud aventaja a mi capacidad de enunciarlo, desgraciadamente en cualquier idioma.*

To Jon Gair, Fraser Clarke and Dave Gittins must go collective thanks for many hours of slacking, as the co-inventors of the truly beautiful concepts of Pre-tea Tea, Pre-pre-tea Tea, Post-tea Tea... I'm sorry we never got round to the twenty-four hour tea-a-thon, but we surely put the training in. Thanks guys for making life at the IoA enjoyable, and for some good nights at the St. Radegund. Individually, thanks to Jon for trying to keep me fit with the occasional bout of squash, to Dave for being a games programmer extraordinaire, and to Fraser for inviting me over to watch his TV (several times), and buying a round (once).

Dave gets another mention as a fellow webcomic-devotee, and the most prolific (OK, only) suggester of features for "the script". Thanks Dave, and also Colin Frayn and Fraser, for sharing my obsession (to some degree), and having at least a vague idea of what I was talking about in my references to *Bobbins* et al.

My thanks to Poshak "The Shak" Gandhi for the gifts of his friendship, and his bicycle. The former carries me farther than the latter ever will. I dread to think what impression of British culture we've left you with Poshak, and how badly we must have corrupted your vocabulary.

To Richard Barker, for a lot of things over a lot of years, including many fine walks in the Lakes that provide me with fond memories. I'm looking forward to many more in the future. Thanks also for being a man of letters Richard, and reminding me of the virtues of pen and paper.

To Muon for stress relief, general adorableness, many useful discussions, and occasionally allowing me the use of my coat.

To my filing cabinet, for absorbing a lot of punishment over the years, and not breaking my hand in return.

To *Camellia sinensis* for ritual and refuge.

I (like all the other members of the X-ray group) owe an enormous debt of thanks to Roderick '1337' Johnstone, ably assisted by Jeremy 'm4d 5killz' Sanders, for peerless technical support of the highest order, seven days a week (but more often than not on a Saturday night). Roderick, even your incredibly bad puns are forgiven.

Thanks to Steve Allen and Robert Schmidt for far too many lunchtime conversations. I rarely had a clue what we were talking about, but I enjoyed it all the same. Thanks to Steve also for fun company on the ski-slopes; and to Robert for his enthusiasm, interest and assistance. Collective thanks to Steve, Robert, and Simon Vaughan for some sociable evenings in the curry house and pub.

Thanks to Adrian Turner and Lisa Voigt for putting up with me being a miserable so-and-so in the corner of the office (not that they had much choice).

Thanks to the writers and maintainers of Free Software the world over, for their dedication and commitment to an ideal. They have made the actual process of researching and producing this thesis many times more pleasant than it would otherwise have been. Additionally, I think the standards of openness and co-operation they set could provide an excellent example to many in the scientific community. That, however, is another story. It will be no surprise to some if I single out GNU Emacs in particular for special mention. `(dotimes (i 1000 t) (insert "\nthank you"))`. In a similar vein, thanks to many netizens of various newsgroups. I think there probably *is* an infinite number of monkeys banging on their keyboards out there, but I know there is a finite, yet gladdeningly large, group of helpful and knowledgeable people. Thanks to them for helping me with many computing queries, and for some interesting discussions.

Thanks to the artists of all forms who helped me maintain some kind of equilibrium (if not a particularly stable one). There are far too many to list here, but I must mention: Haruki Murakami, for writing the book I would have written if I could; Raymond Chandler, for a moral compass; Elvis Costello, for proving that the angry guy has all the best tunes; and John Allison for writing the story of my generation — *Bobbins* (it wasn't) — and giving it away for free. I suppose that, a long, long time ago, Asimov, Clarke et al. have an awful lot to do with setting me out on the path I have followed to get here.

On a practical level, thanks to the UK taxpayer for letting me spend their money. Robinson College provided me with a home for three years. Thanks to all the staff of the IoA, academic and support, for helping to make such a great working environment.

Lastly, firstly, always and forever, thanks and love to Mum and Dad. There aren't words.

Finally, my heartfelt apologies to all of you for not writing the thesis you deserve. If I had it all to do over again, I think I could do better. But promise me that if I even think about trying, you'll hit me with something very, very heavy.

RGM
Cambridge, 10th March 2003

I was crazy. I liked it.

The Little Sister
RAYMOND CHANDLER

This thesis was typeset in LaTeX using GNU Emacs.

Contents

Appendices

List of Tables

LIST OF FIGURES

Sometimes I dream of a work of really great breadth, ranging through the whole region of element, object, meaning and style. This, I fear, will remain a dream, but it is a good thing even now to bear the possibility in mind.

The Modern as Ideal
PAUL KLEE

Without warning, she asked me, "Hey Watanabe, can you explain the difference between the English present subjunctive and past subjunctive?"

"I think I can," I said.

"Let me ask you, then, what possible use is stuff like that for everyday life?"

"None at all," I said. "It may not serve any concrete purpose, but it does give you some kind of training to help you grasp things in general more systematically."

Midori gave that a moment's serious thought. "You're amazing," she said. "That never occurred to me before. I always thought of things like the subjunctive case and differential calculus and chemical symbols as totally useless. A pain in the neck. So I've always ignored them. Now I have to wonder if my whole life has been a mistake."

Norwegian Wood
HARUKI MURAKAMI

The known is finite, the unknown infinite; intellectually we stand on an islet in the midst of an illimitable ocean of inexplicability. Our business in every generation is to reclaim a little more land.

THOMAS HENRY HUXLEY

1 Introduction

The birth of observational X-ray astronomy can be dated to 10:00 am (Mountain Standard Time) on Monday 29th September 1949, when a V-2 rocket, launched from the White Sands Missile Range, New Mexico, and carrying Geiger counters made a brief (5 minute) flight to a height of 150 km and recorded the first detection of solar X-rays (Friedman *et al.*, 1951). The first detection of X-rays from a cosmic source (Sco X-1) came the following decade (Giacconi *et al.*, 1962). The field of X-ray astronomy is thus one of the youngest branches of one of the oldest sciences. Indeed, it has only just acquired sufficient age for the first Nobel prize to be awarded in connection with it (Riccardo Giacconi was awarded one-half of the Nobel Prize in Physics in 2002 "for pioneering contributions to astrophysics, which have led to the discovery of cosmic X-ray sources"). Let us not forget that X-rays themselves were only discovered by Röntgen in 1895. Ever since its inception, improvements in detector technology have driven the field of X-ray astronomy forward at a rapid pace.

1.1 X-ray Observatories

Whilst the main focus of this thesis is on theoretical and computational topics, theory does not exist in a vacuum, and therefore I make here some brief discussion of the (in many cases superlative) observatories and instruments that are, or have been, important in this field. The engineering achievements made in the course of X-ray astronomy are immense, and should not be passed over lightly, but there is not nearly enough space available here to do them justice.

Practically every X-ray mission has represented a significant technological advance on its predecessor. NASA's *EINSTEIN* (*HEAO-2*) observatory was the first imaging X-ray satellite. *ROSAT* (Röntgen satellite, a Germany/US/UK collaboration) was the first imaging satellite to carry out an all-sky X-ray and XUV survey. The Japanese *ASCA* (Advanced Satellite for Cosmology and Astrophysics, *ASTRO-D*) was the first to use CCDs (charge-coupled devices) for X-ray astronomy. As another of NASA's Great

Satellite	Lifetime	E (keV)	A_e[a]	PSF[b]	Visibility (hr)
Einstein	12/11/78 – 04/81	0.2–20		7	
ROSAT	01/06/90 – 12/02/99	0.1–2.4	400	3.5	1.3
ASCA	20/02/93 – 02/03/01	0.4–10	350	73	0.9
Chandra	23/07/99 –	0.1–10	800	0.2	50
XMM-Newton	10/12/99 –	0.15–15	4650	6	40

[a]Mirror effective area at 1 keV, cm^2

[b]Mirror point spread function full-width at half maximum (FWHM), arcsec

Table 1.1: A comparison of some X-ray satellites.

Observatories, *Chandra* is in many respects the X-ray equivalent of the *Hubble* telescope. It offers an unparalleled spatial resolution to X-ray observers.

Table 1.1 summarizes some of the main features of various important X-ray missions. It is impossible to fully represent these missions in such a brief format — the values shown are necessarily averages for quantities that change in complex ways as a function of energy, viewing angle, and so on[1]. In addition, most missions have a variety of instruments, each with their own individual characteristics. For example, Figure 1.1 on the facing page shows the energy dependence of the effective area of two of the detectors on the *Chandra* satellite.

Due to the very high energies of X-ray photons, they can only be reflected at grazing angles of incidence. This in turn means that X-ray telescopes necessarily have long focal lengths (e.g. that of *Chandra* is 10 m), since the X-rays can be made to converge only slowly. The designs employed in telescopes are of the so-called Wolter type (originally suggested in the context of microscopy) which require two successive reflections from either a paraboloid/hyperboloid or paraboloid/ellipsoid combination. A review of X-ray optics is provided by Aschenbach (1985).

The latest age of X-ray astronomy began with the launch of the *Chandra* satellite in mid-1999, followed by that of *XMM-Newton* at the end of the same year. One characteristic shared by these missions is that both have been placed in high-Earth orbit, thus allowing for much longer uninterrupted observations as compared to previous missions (Table 1.1).

Chandra[2] has four mirrors constructed from a glass ceramic with a very low (and homogeneous) coefficient of thermal expansion, with an iridium coating for the reflecting surface. Iridium is an efficient reflector of X-rays, although inevitably absorption edges drastically reduce its reflectivity at certain energies (Figure 1.1 on the facing page). The coating is only some 330 Å thick. The largest mirror has a diameter of 1.2 m. The satellite has two major focal plane instruments: the High Resolution Camera (HRC), which is a microchannel plate detector; and the Advanced CCD Imaging Spectrometer (ACIS), which is an array of CCDs. Additionally, there are two transmission gratings, optimized for low (LETG) and high (HETG) energies, formed by gold gratings which can be moved into place behind the mirrors. ACIS itself consists of two separate CCD arrays: the four-chip ACIS-I (for CCD imaging and spectrometry), and the 6-chip ACIS-S (which can also be used with the HETG). ACIS-I is constructed of front-illuminated (FI) chips, whereas ACIS-S has four FI and two back-illuminated (BI) chips. The dif-

[1]Many quantities also appear to be a strong function of which instrument handbook is being consulted.

[2]`http://cxc.harvard.edu/`

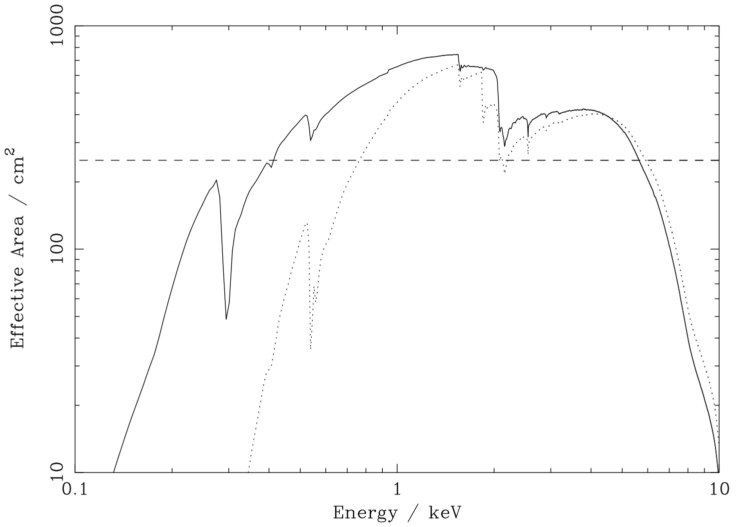


Figure 1.1: Effective area of the *Chandra* High Resolution Mirror Assembly (HRMA) / Advanced CCD Imaging Spectrometer (ACIS) combination. Solid line ACIS-S (spectral), dotted line ACIS-I (imaging). The dashed line is a constant area of $250\,\mathrm{cm}^{-2}$ that forms a reasonable global average for ACIS-S. The drop around 2 keV is due to the iridium M-edge in the mirror coating; the low-energy structure is due to the CCD quantum efficiency. These data were extracted from the AO-2 proposal planning RMF and ARF files for the detector aimpoint position.

ferent forms of chip have different energy responses. The effective area of the ACIS instrument is shown in Figure 1.1.

XMM-Newton[3] has three separate Wolter X-ray telescopes, each with 58 concentric mirrors. As is clear from Table 1.1 on the facing page, these result in a dramatic increase in collecting area over previous missions. Additionally, there is a 30 cm optical/UV telescope with a CCD detector in the focal plane. This allows for simultaneous optical monitoring of X-ray targets. Two of the three X-ray telescopes are equipped with Reflection Grating Arrays (RGAs), *XMM-Newton* being the first X-ray satellite to use such technology. The gratings intercept about half of the incident radiation in each associated telescope, and disperse it into a spectrum across arrays of CCDs. Each combined instrument is known as a Reflection Grating Spectrometer (RGS). The satellite has two different kinds of CCD, which are collectively referred to as the European Photon Imaging Camera (EPIC). The two telescopes with reflection gratings each have a MOS (Metal Oxide Semiconductor) camera of seven CCDs. The cross-dispersion width of these arrays is about 5 arcmin. The RGS instruments allow for high-resolution ($E/\Delta E \sim$ several hundred) spectroscopy in the range 0.3–2.5 keV, a range which includes the L shell (Section 1.4.1) transitions of heavy elements like Fe and Ni, and the K shell transitions of lighter elements like N, O, Ne, Mg, and Si. The third X-ray telescope has a p-n CCD with 6×2 chips. CCDs themselves have a degree of intrinsic energy resolution, which allows for some imaging spectrometry.

For a review of X-ray missions up until the 1990s, see Bradt *et al.* (1992). A readable, general-level account of X-ray astronomy from birth to middle-age is provided by Tucker and Giacconi (1985). The last figure in that work is an artist's impression of the planned Advanced X-ray Astrophysics Facility (*AXAF*). Fifteen years later, *AXAF*, rechristened *Chandra*, was launched. A summary for the general reader of the discoveries (so far) of the *Chandra* and *XMM-Newton* era is provided by Schlegel (2002).

[3] `http://xmm.vilspa.esa.es/`

1.2 X-rays from Clusters of Galaxies

The first observation of extragalactic X-rays was from M87 in the Virgo cluster, which was observed from an Aerobee rocket flown in April 1965 (Byram *et al.*, 1966). An extended X-ray source in the region of the Coma cluster was detected using a combined gas proportional counter and scintillation detector carried on a balloon on 6th December 1965 (Boldt *et al.*, 1966). Initially, of course, it was not possible to be certain that the Coma cluster itself, as opposed to some intervening object, was the source of the X-rays. Operating on the assumption that the Coma cluster was in fact responsible, Felten *et al.* (1966) considered two possibilities for the origin of the X-rays: that they might be produced either by the individual member galaxies; or by a large reservoir of hot ($\sim 10^8$ K) gas in the potential of the cluster. This latter idea was attractive because it might provide an explanation for the observed mass to (visible) light ratios of clusters, which had for some time been known to be unusually high, if member galaxies (with measured velocities) were to be gravitationally bound (e.g. Abell, 1965). After giving brief consideration to two non-thermal mechanisms that might produce the observed X-rays (namely inverse Compton scattering of cosmic photons and synchrotron radiation), Felten *et al.* (1966) concluded that thermal bremsstrahlung radiation (Section 1.4.3) was the most likely source.

After the launch of the *Uhuru*[4] satellite in 1970, it was confirmed that clusters of galaxies are in general sources of extended X-ray emission, with luminosities in the range $10^{43-45}\,\mathrm{erg\,s^{-1}}$. Galaxy clusters are the largest virialized structures in the universe, comprising typically several hundred galaxies that are gravitationally bound and occupy a region several Mpc in size. The total masses of these systems are $\sim 10^{15}\,\mathrm{M_\odot}$. Reviews of galaxy clusters are provided by, for example: Bahcall (1977); Dressler (1984); Biviano (2000).

For a time, the origin of the cluster X-ray emission was uncertain, the possibilities of individual stellar sources (e.g. Katz, 1976), inverse Compton scattering (e.g. Rephaeli, 1977b), and thermal bremsstrahlung all being considered, as suggested by Felten *et al.* (1966). Mitchell *et al.* (1976) reported the first detection of a 7 keV iron line (Section 1.4.3) from observations of the Perseus cluster. This confirmed that the radiation mechanism was thermal bremsstrahlung, and also showed that the intracluster gas must have been enriched somehow (Section 1.3.3).

The source of the X-rays is the intracluster medium (ICM), which is a vast reservoir of hot, diffuse gas, retaining its heat from the initial gravitational collapse which created the cluster (the gas in isolated galaxies cooled much more rapidly). In the X-ray band, the ICM "shines" with the glow of bremsstrahlung emission, providing us with a unique tracer of a major baryonic component of the cluster. A review of the X-ray emission of galaxy clusters is provided by Sarazin (1988).

1.3 Physical Properties of the Intracluster Medium

A suitable order of magnitude estimate for the density of the intracluster medium is $\sim 10^{-3}\,\mathrm{cm^{-3}}$ (see for example Figure 3.8(a) on page 79 for an illustration of how this value changes with radius). As is common in astrophysics, the innocent-looking unit disguises just how far removed from our everyday experience such a value is. At a temperature of 10^8 K, the associated pressure is $10^5\,\mathrm{K\,cm^{-3}}$, or in SI

[4] *Uhuru* was launched from Kenya on 12th December, Kenyan Independence Day. *Uhuru* is the Swahili word for "freedom". *Uhuru* was the first orbiting mission dedicated to X-ray astronomy.

units (see Section B.1), $\sim 10^{-12}\,\mathrm{N\,m^{-2}}$. This is 17 orders of magnitude lower than standard atmospheric pressure. Industrially, pressures are commonly measured in Torr, where 1 Torr corresponds to 1 mm of mercury, and 760 mm of mercury to a pressure of 1 atmosphere. Hence the typical cluster pressure corresponds to around 10^{-14} Torr. This is around one million times lower than an ultra-high vacuum such as is used in a tokamak fusion reactor. For comparison, the density of the solar corona is around $10^9\,\mathrm{cm^{-3}}$, and the temperature $\sim 10^6$ K.

The total mass (e.g. Allen and Fabian, 1994; Ettori *et al.*, 1998) of hot gas inferred for a rich cluster is $\sim 10^{14}\,\mathrm{M_\odot}$, about 20 per cent (e.g. Ettori and Fabian, 1999) of the total mass of the cluster. This is much greater than the mass fraction in visible stars ($\sim$ 5 per cent), but still leaves the majority of the cluster mass in the form of dark matter.

1.3.1 The Definition of a Plasma

An ionized gas may be referred to as a "plasma" if the *Debye shielding length* λ_D is large[5] compared with other lengths of interest.

$$\lambda_\mathrm{D} = \sqrt{\frac{\varepsilon_0 k_\mathrm{B} T}{e^2 n_\mathrm{e}}} \approx 2.2 \times 10^4\,\mathrm{m} \left(\frac{T}{10^8\,\mathrm{K}}\right)^{\frac{1}{2}} \left(\frac{n_\mathrm{e}}{10^{-3}\,\mathrm{cm^{-3}}}\right)^{-\frac{1}{2}} \tag{1.1}$$

λ_D gives a rough measure of the distance over which the electron density may fluctuate significantly, i.e. deviate appreciably from Zn_ion. With $n \sim 10^{-3}\,\mathrm{cm^{-3}}$, the average distance between particles is ~ 10 cm. A plasma has the property that the number of particles contained in a *Debye sphere* of radius λ_D is large, i.e.

$$N_\mathrm{D} = {}^4\!/_3 \pi n_\mathrm{e} \lambda_\mathrm{D}^3 \approx 4.4 \times 10^{16} \left(\frac{T}{10^8\,\mathrm{K}}\right)^{\frac{3}{2}} \left(\frac{n_\mathrm{e}}{10^{-3}\,\mathrm{cm^{-3}}}\right)^{-\frac{1}{2}} \gg 1 \tag{1.2}$$

The ICM clearly fits this definition of a plasma (and is an extremely low-density one, according to most scales).

1.3.2 Magnetic Fields

It has been known for many years that there is an extended radio source with a steep spectrum in the Coma cluster (Large *et al.*, 1959; Willson, 1970). This is the prototype of the so-called "radio haloes", of which some several tens have since been discovered. Radio haloes typically have large extents ($\sim$ 1 Mpc, or a significant fraction of the total cluster size) and centres close to that of the cluster X-ray emission. Indeed, a steep correlation has been found between the X-ray and radio luminosities (e.g. Govoni *et al.*, 2001); such that whilst in an X-ray flux-limited sample of clusters $\lesssim 10$ per cent are detected in the radio, for the brightest X-ray clusters the fraction increases to $\sim$ 35 per cent.

Also found in some clusters are the radio source known as "radio relics" (e.g. Harris *et al.*, 1993). Unlike haloes, these are (by definition) found at the outskirts of the cluster. They tend to be more irregular in shape, and have a higher degree of polarization, than halo sources.

[5] Note that Spitzer Jr (1962) erroneously reverses this definition.

The power-law nature of the cluster radio sources, together with the degree of polarization, are indicative of synchrotron radiation (e.g. Pacholczyk, 1970), emitted by nonthermal relativistic electrons being centripetally accelerated by the presence of a magnetic field. Arguments assuming minimum energy for the total energy due to relativistic particles and magnetic fields lead to field strengths $\sim 1\,\mu\mathrm{G}$ (e.g. Giovannini *et al.*, 1993). In comparison, the strength of the Earth's magnetic field is around $0.5\,\mathrm{G}$.

The phenomenon of *Faraday rotation* also allows an estimate of cluster magnetic field strengths to be made. A magnetized plasma is *birefringent* — this is the phenomenon of "double refraction" exhibited also by some crystals (such as Iceland spar, or calcite) in which light can be split into two rays with different polarizations, as a consequence of an anisotropic refractive index. A magnetic field creates such a refractive index in a plasma by setting a preferential direction for electron motion. Linearly polarized light propagating through a magnetized plasma experiences a progressive rotation of its polarization vector, according to

$$\phi \equiv \mathcal{R}\lambda^2. \tag{1.3}$$

In this expression, ϕ is the change in the position angle of the polarization, and $\mathcal{R}$ is the rotation measure. Since the effect increases with wavelength λ it is most prominent for radio waves. The degree of rotation depends upon the electron density and the path length of the radiation through the magnetic field, such that

$$\mathcal{R} \equiv \frac{e^3}{8\pi^2\varepsilon_0 m_e^2 c^3}\int n_e \boldsymbol{B}\cdot \mathrm{d}\boldsymbol{l} \approx 812\,\mathrm{rad\,m^{-2}} \int \left(\frac{n_e}{\mathrm{cm^{-3}}}\right)\left(\frac{\boldsymbol{B}}{\mu\mathrm{G}}\right)\cdot \mathrm{d}\left(\frac{\boldsymbol{l}}{\mathrm{kpc}}\right). \tag{1.4}$$

The wavelength dependence of the rotation means that observations at multiple frequencies can be used to eliminate the unknown initial polarization direction. An independent measure of the electron density (e.g. from X-ray observations) and path length is required, as well as a radio source with a significant degree of polarization (synchrotron radiation usually has a strong linear polarization), either behind or inside the cluster. Only the component of the magnetic field along the line of sight is measured, and hence a lower limit (or a topology-dependent estimate) of the total field strength can be obtained.

Propagation through the Galactic magnetic field imposes a rotation measure typically of the order of tens of $\mathrm{rad\,m^{-2}}$ on most extragalactic sources. Sources viewed through the intracluster medium (e.g. Cygnus A; Mitton, 1970) can exhibit much higher rotation measures ($\sim 1000\,\mathrm{rad\,m^{-2}}$), varying over relatively small scales. Cooling flow clusters (Section 1.6) exhibit high rotation measures, well-correlated with the X-ray mass deposition rates (Taylor *et al.*, 2002), potentially a very interesting result (Section 5.7). Spatially, the rotation measure distributions tend to have coherence lengths $\sim 10\,\mathrm{kpc}$. Thus, the ICM magnetic field cannot be entirely ordered but is most likely tangled (e.g. Jaffe, 1980) on length scales of a few kpc. See Section 5.5.3 for a discussion of the possible consequences of this.

The intracluster magnetic field may turn out to be important for the evolution of the gas both energetically (see Chapter 5) and dynamically, although the precise effects still remain uncertain. For an up-to-date review of cluster magnetic fields, see Carilli and Taylor (2002).

1.3.3 Chemical Composition

In this thesis, the terms "metals" and "heavy elements" will be used interchangeably to refer to all elements except hydrogen and helium (the standard astrophysical usage). The terms "metallicity", "abundance", and Z will all be used to refer to the number density relative to hydrogen of heavy elements (in almost all cases the helium abundance will be fixed at solar) compared to the value appropriate for a solar plasma.

The solar abundances which have been most commonly used in X-ray studies of the ICM are those of Anders and Grevesse (1989), and so for consistency with previous work that abundance set (listed in Table B.2 on page 161) is employed in this thesis. These values are obtained from observations of the solar photosphere. Solar abundances have also been determined from the corona and from meteorites. For some time there was significant discrepancy between the photospheric results and those obtained from other means, as well as a large variation amongst the photospheric results obtained by different authors. Recent analyses (e.g. Grevesse and Sauval, 1999) have removed this discrepancy. Future observational studies should use the abundance data of Grevesse and Sauval (1998), included in XSPEC (Section 1.4.5) as the `grsa` option, although the data of Anders and Grevesse (1989) are still the default.

The detection of line emission (Section 1.4.3) from multiply ionized iron was a highly significant step on the road towards an understanding of the intracluster medium. The 7 keV Fe Kα line was first detected in the spectrum of the Perseus cluster (Mitchell *et al.*, 1976). Such observations were instrumental not only in confirming that the emission mechanism responsible for the X-rays was thermal bremsstrahlung (Section 1.4.3) from a hot plasma; but also that the ICM could not be primordial, and must have been enriched by processed material from an early generation of stars. Since then, the presence of many other elements in the ICM has been established (e.g. Mushotzky *et al.*, 1996). Measuring the ratio of different element abundances leads to estimates of the relative influences of SNe Ia and SNe II on the ICM (e.g. Finoguenov *et al.*, 2000) and acts as a probe of the enrichment history of the cluster. The average metallicity of the ICM is found to be roughly one-third solar in most cases, both for nearby (Edge and Stewart, 1991) and distant (Mushotzky and Loewenstein, 1997) clusters.

By any reasonable bulk standard (see Section B.3), the heavy elements are almost entirely insignificant components of the ICM. Yet their influence on the spectral properties is profound — see Chapter 2 for more discussion of this issue.

1.3.4 Characteristic Time Scales

There are a number of time-scales that are relevant to a discussion of the ICM. A comparison of the relative magnitudes of these various times is one way to obtain an overview of the bulk behaviour and properties of the intracluster gas.

One of the most important times is the *cooling time*, as discussed in more detail in Section 1.5.3. Using equations 1.26 and 1.19, an approximate value for the (high-temperature) cooling time is

$$t_{\mathrm{cool}} \approx 2.1 \times 10^{10}\,\mathrm{yr} \left(\frac{p}{10^{6}\,\mathrm{K\,cm^{-3}}} \right)^{-1} \left(\frac{T}{10^{8}\,\mathrm{K}} \right)^{3/2}. \tag{1.5}$$

The (adiabatic) sound speed is given by

$$c_{\mathrm{s}}^2 = \frac{\mathrm{d}p}{\mathrm{d}\rho} = \gamma\frac{p}{\rho} = \frac{\gamma}{\mu}\frac{k_{\mathrm{B}}}{m_{\mathrm{H}}}T. \tag{1.6}$$

Therefore the *sound-crossing time* for a region of dimension R (with $\gamma = 5/3$, $\mu = 0.6$) is

$$t_{\mathrm{cross}} = \frac{R}{c_{\mathrm{s}}} \approx 6.5\times10^8\,\mathrm{yr}\left(\frac{R}{\mathrm{Mpc}}\right)\left(\frac{T}{10^8\,\mathrm{K}}\right)^{-\frac{1}{2}}. \tag{1.7}$$

An estimate for the free-fall time from radius r can be obtained under the assumption of a constant gravitational acceleration $\boldsymbol{g}$ (from Figure A.2 on page 155, this is clearly incorrect, but not too far removed from the truth, especially for the NFW profile). An order of magnitude estimate for g follows from the formula for a point mass, $g \sim GM/r^2$. With $M \sim 10^{15}\,\mathrm{M}_{\odot}$ and $r \sim 1\,\mathrm{Mpc}$, $g \sim 10^{-10}\,\mathrm{m\,s^{-2}}$. Hence

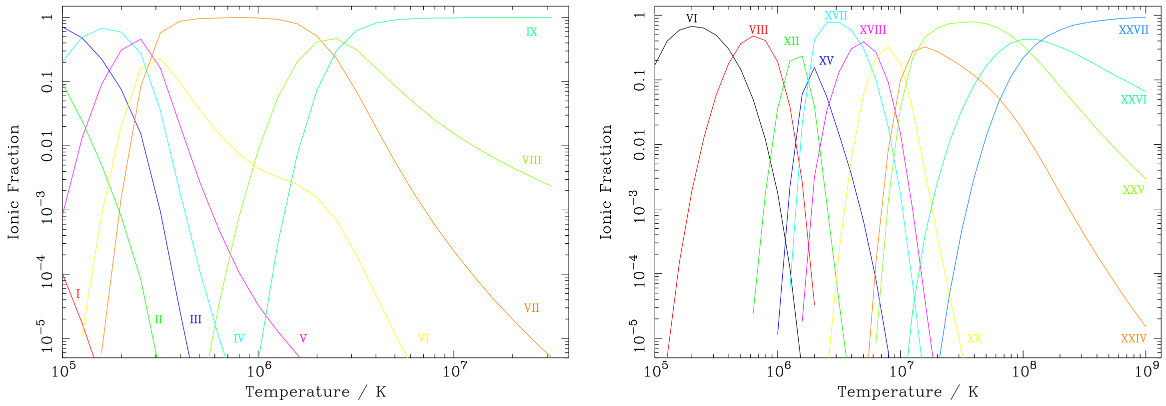
$$t_{\mathrm{ff}} \sim \sqrt{\frac{2r}{g}} \sim 8\times10^8\,\mathrm{yrs}\left(\frac{r}{\mathrm{Mpc}}\right)^{\frac{1}{2}}. \tag{1.8}$$

There are also several atomic time-scales which are relevant. Many of these follow in a relatively straightforward fashion from the consideration of encounters between charged particles — see Spitzer Jr (1962) for an informative summary of the processes at work. It can be shown that the *self-collision time* for a group of interacting charged particles is given by

$$t_{\mathrm{c}} = \frac{2\pi\varepsilon_0^2}{(Ze)^4}\frac{m^{1/2}(3k_{\mathrm{B}}T)^{3/2}}{0.714n\ln\Lambda} \approx \frac{0.22\,\mathrm{Myr}}{Z^4}\left(\frac{m}{m_{\mathrm{e}}}\right)^{\frac{1}{2}}\left(\frac{\ln\Lambda}{37.8}\right)^{-1}\left(\frac{T}{10^8\,\mathrm{K}}\right)^{\frac{3}{2}}\left(\frac{n}{10^{-3}\,\mathrm{cm^{-3}}}\right)^{-1}, \tag{1.9}$$

where the factor of 0.714 arises from the Chandrasekhar (e.g. Chandrasekhar, 1960) correction taking into account the centre-of-mass motion, and $\ln\Lambda$ is the Coulomb logarithm (see Section 5.2). t_{c} is a measure of the time taken for a population of identical charges to achieve equipartition (i.e. thermal equilibrium) amongst themselves. Observe that the time scales as the square root of the particle mass, so that a population of protons will equilibriate some $\sqrt{1836} \approx 43$ times slower than an electron population under the same conditions. The fact that both the electron–electron and proton–proton self-collisions times are short compared to other time-scales of interest means that the electrons and protons in the ICM are both well described by Maxwellian distributions at a single temperature.

Due to the widely differing masses, electron–proton collisions are relatively ineffective in transferring energy. Thus, electron–electron and proton–proton encounters first cause the separate populations to come to equilibrium amongst themselves, but not necessarily at the same temperatures. In order to estimate the time taken for the electrons and protons to come to a mutual thermal equilibrium, consider now the case of an ionized gas consisting of a mixture of "test" (unsubscripted, e.g. electrons) and heavier "field" (subscript f, e.g. protons) particles. If the two populations both initially have Maxwell–Boltzmann velocity distributions, but with different kinetic temperatures T and T_{f} respectively, then the time taken



(a) Oxygen **(b)** Iron

Figure 1.2: The temperature dependence of the ionization state of oxygen and iron (selected states only) in a coronal plasma. Data from Arnaud and Rothenflug (1985).

for equipartition of energy to occur is given (e.g. Spitzer Jr, 1962) by

$$\begin{aligned} t_{\mathrm{eq}} &= 6\pi\sqrt{\frac{\pi}{2}}\frac{mm_{\mathrm{f}}}{n_{\mathrm{f}}\ln\Lambda}\left(\frac{\varepsilon_0}{ZZ_{\mathrm{f}}e^2}\right)^2\left(\frac{k_{\mathrm{B}}T}{m}+\frac{k_{\mathrm{B}}T_{\mathrm{f}}}{m_{\mathrm{f}}}\right)^{\frac{3}{2}} \\ &\approx \frac{0.33\,\mathrm{Myr}}{(ZZ_{\mathrm{f}})^2}\left(\frac{m}{m_{\mathrm{e}}}\right)\left(\frac{m_{\mathrm{f}}}{m_{\mathrm{e}}}\right)\left(\frac{n_{\mathrm{f}}}{10^{-3}\,\mathrm{cm}^{-3}}\right)^{-1}\left(\frac{\ln\Lambda}{37.8}\right)^{-1}\left(\frac{\frac{T}{10^8\,\mathrm{K}}}{\frac{m}{m_{\mathrm{e}}}}+\frac{\frac{T_{\mathrm{f}}}{10^8\,\mathrm{K}}}{\frac{m_{\mathrm{f}}}{m_{\mathrm{e}}}}\right)^{\frac{3}{2}}. \end{aligned} \tag{1.10}$$

For an electron–proton plasma, $t_{\mathrm{eq}}(\mathrm{ep}) \sim m_{\mathrm{p}}m_{\mathrm{e}}^{-1/2}$, whereas $t_{\mathrm{c}}(\mathrm{pp}) \sim m_{\mathrm{p}}^{1/2}$, and $t_{\mathrm{c}}(\mathrm{ee}) \sim m_{\mathrm{e}}^{1/2}$. Therefore at a given temperature and density, $t_{\mathrm{eq}}(\mathrm{ep}) \sim 43t_{\mathrm{c}}(\mathrm{pp}) \sim 43t_{\mathrm{c}}(\mathrm{ee})$, ignoring small numerical factors. The electron population comes into thermal equilibrium most rapidly, then the proton population follows, and finally they reach equilibrium with each other. All these times are short compared to the cooling time, with the net result that the ICM can be described as a collection of particles in a Maxwell–Boltzmann distribution with a single kinetic temperature.

1.3.5 Ionization State

Another set of important time-scales are those for ionization and recombination (e.g. Arnaud and Raymond, 1992). Again, it can be shown that these times are a fraction of the age of the system, or any relevant hydrodynamic time-scale. Consequently, the ICM may be assumed to be in ionization equilibrium. Figure 1.2 shows how the ionization state of two important ICM species, oxygen and iron, changes across the appropriate temperature range. Figure 1.3 on the following page shows similar information, but in a more compact form for more elements, plotting as it does the mean charge of each species as a function of temperature. It is clear that all ICM species are highly ionized. Table 1.2 on the next page lists the approximate temperatures at which each species becomes fully ionized. The corresponding behaviour of the electron density (relative to that of hydrogen) is shown in Figure 1.4 on page 11. Since the bulk of the electrons are contributed by hydrogen and helium (see Section B.3), the electron density

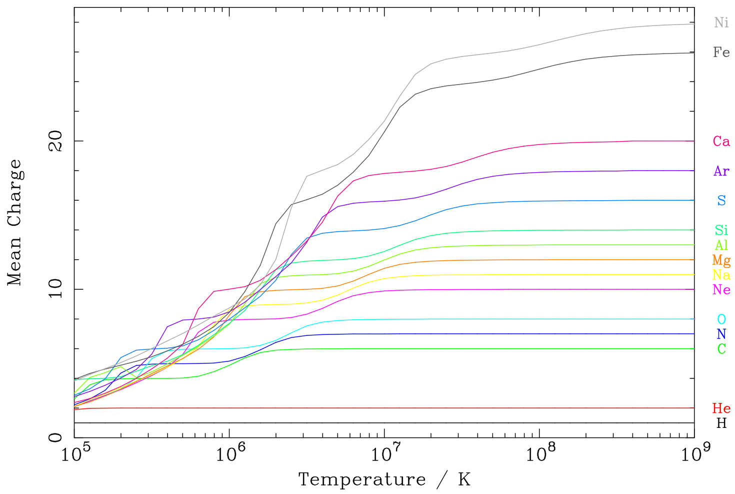


Figure 1.3: The temperature dependence of the mean charge of various species in a coronal plasma. Data from Arnaud and Rothenflug (1985).

Element	T / K
H	2×10^4
He	10^5
C	2×10^6
N	3×10^6
O	4×10^6
Ne	10^7
Na	10^7
Mg	2×10^7
Al	3×10^7
Si	3×10^7
S	4×10^7
Ar	10^8
Ca	10^8
Fe	10^9
Ni	10^9

Table 1.2: Approximate temperatures at which various species in a coronal plasma reach full ionization.

does not change significantly until very low temperatures are reached.

1.3.6 Characteristic Length Scales

From equation (1.9), the mean free path for particle-particle collisions is given by

$$\lambda = t_c\bar{v} = \frac{2\pi}{0.714}\frac{\varepsilon_0^2}{(Ze)^4}\frac{(3k_BT)^2}{n\ln\Lambda} \approx \frac{11\,\mathrm{kpc}}{Z^4}\left(\frac{\ln\Lambda}{37.8}\right)^{-1}\left(\frac{T}{10^8\,\mathrm{K}}\right)^2\left(\frac{n}{10^{-3}\,\mathrm{cm}^{-3}}\right)^{-1}. \tag{1.11}$$

Observe that this is independent of mass. Thus, the mean free path for electron–electron collisions is the same as that for proton–proton collisions (ignoring the slight difference in $\ln\Lambda$ discussed in Section 5.2).

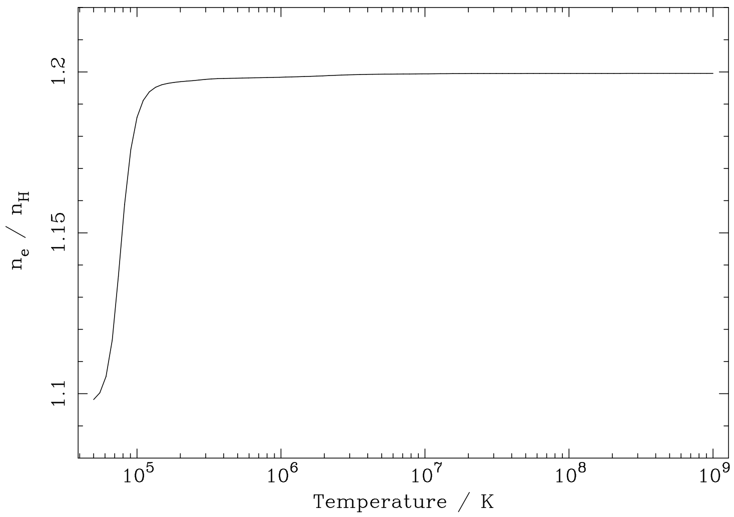


Figure 1.4: Temperature dependence (from MEKAL) of the relative electron density in a 0.3 $Z_\odot$ plasma.

These lengths are much shorter than overall size-scales of clusters (~ 1 Mpc), and therefore the ICM behaves as a collisional, hydrodynamic fluid.

1.4 The X-ray Spectrum of the Intracluster Medium

1.4.1 Spectroscopic Notation

There is a bewildering variety of terminology and notation in use in the fields of spectroscopy and atomic physics, owing to that most dread cause — "historical reasons". Fortunately, we only require a few of the most basic examples. The terms K, L, M, N, O, P, Q shell are commonly used to refer to the states with principal quantum number n = 1–7 respectively. Transitions with their low-energy state in the n =1–6 levels form the Lyman, Balmer, Paschen, Brackett, Pfund, and Humphreys series respectively. For hydrogen, the Lyman series lies in the UV and the Balmer in the visible (for which reason the Balmer series is sometimes known as the Hydrogen series). The lowest-energy transition in each series has the form $n \leftrightarrow n+1$ and is known as the α transition of the series, and so on through β, γ, δ, ε, beyond which the number of the upper level is used. Thus, the first transitions in the Balmer series are denoted Hα, Hβ, Hγ, Hδ, Hε, H8, and H9. The various angular momentum states possible in a given configuration further subdivide the transitions.

The various *ionization states* of a given element are denoted using Roman numerals, with I referring to the neutral species, II to the singly-ionized state, and so on. An element of atomic number Z has $Z+1$ ionization stages. The state with one remaining electron is known as the *hydrogenic* or *hydrogen-like* state, that with two as the *helium-like* state. For example, Fe XXVI is hydrogenic iron, whereas Fe XVII has lost 16 electrons and therefore has 10 remaining (and so is neon-like).

1.4.2 Coronal Plasmas

Those of us engaged in spectral studies of the ICM have a relatively easy task compared to some of our colleagues working in other fields; for example those engaged in some studies of AGN who must contend with the vagaries of time-dependent photoionization. The spectral emission of the ICM is simple, by comparison, largely for two reasons: the low densities ($\sim 0.01\,\mathrm{cm}^{-3}$, say); and high temperatures ($\sim 10^8\,\mathrm{K}$, say). Due to the low density, the gas is *optically thin* (or, the optical depths are low), which means that reprocessing of the emitted radiation field is negligible, and that it is not necessary to consider radiation transport. Actually, this may not always be the case — there are certain circumstances where the optical depth can become significant. This is the phenomenon of "resonant scattering" (e.g. Gil'fanov *et al.*, 1987; Mathews *et al.*, 2001), which can have interesting and possibly important consequences for imaging and spectral analysis (Section 4.7).

In Section 1.3.4 some of the characteristic time-scales of the ICM were discussed. It was shown that the time needed for the ICM electron–proton population to come into thermal equilibrium is relatively short. Therefore, it can be described by a Maxwell–Boltzmann distribution at a single temperature, which controls the rate of ionization and excitation. Additionally, mention was made of the fact that the ionization and recombination times for the ICM plasma are also short, so that the ICM can be regarded as existing in ionization equilibrium.

The main excitation process in the ICM is collisional excitation (of ions by electrons). Photoionization is negligible since the radiation field is not intense enough to have a significant effect. Due to the low density, radiative decay dominates over collisional decay. It is also substantially faster than collisional excitation, with the result that the population of excited states is insignificant and all upward transitions can be assumed to start from the ground state. Three-body processes can be ignored, again due to the low density.

Together, these assumptions describe what is known as the *coronal limit*, since the plasma in the solar corona embodies these conditions. This is convenient, as it means that observations of the solar spectrum can be used to benchmark the plasma codes (e.g. Phillips *et al.*, 1999).

1.4.3 Radiation Mechanisms

At the high temperatures and low densities appropriate for the ICM, the chief emission process is *bremsstrahlung* (braking radiation), the emission of radiation by a charge when it is accelerated in the electrostatic field caused by another charge. The term free–free emission is also used, in the sense that the transitions occur between unbound energy levels of free charges. Expanding the electromagnetic field in terms of multipoles, then taking only the leading, dominant term leads to the so-called dipole approximation. It can be shown that the dipole moment is proportional to the position of the centre of mass. Consequently, the interactions between like particles do not give rise to bremsstrahlung radiation (since the displacement of the centre of mass is a constant of the motion); and so, for the case of the ICM, one must consider the interaction between electrons and ions. In an electron–ion encounter, the electron experiences a much greater acceleration due to its lower mass (the potentially greater charge on the ion compensating only slightly). The net result is that the electrons are the dominant radiators. The much greater mass of the ions means that the problem reduces to one of an electron moving in the fixed field of a stationary ion. At the temperatures of the ICM ($\sim$ keV), the electron (mass $\sim 0.5\,\mathrm{MeV}$) can be treated non-relativistically. Quantum effects, on the other hand, are important, since photons can be produced with energies comparable to those of the electrons.

The derivation (e.g. Rybicki and Lightman, 1979; Longair, 1992) of the bremsstrahlung formula appropriate for the ICM begins with a consideration of the encounter between a single electron and a stationary ion. Coulomb encounters of this kind also control the transport properties of the ICM plasma, as discussed by Spitzer Jr (1962) — see Chapter 5. Integrating over a Maxwellian distribution of electron velocities then leads to the expression for the free–free spectral emissivity,

$$\varepsilon_\nu^{\mathrm{ff}} = \frac{Z^2 e^6}{3\pi^2 \varepsilon_0^3 c^3 m_\mathrm{e}^2} \left(\frac{\pi m_\mathrm{e}}{6 k_\mathrm{B} T}\right)^{1/2} n_\mathrm{e} n_\mathrm{i} g(\nu, T) \exp\left(-\frac{h\nu}{k_\mathrm{B} T}\right) \tag{1.12}$$

$$\begin{aligned}\varepsilon_\mathrm{E}^{\mathrm{ff}} \approx\ & 1.65 \times 10^{-30}\,\mathrm{erg\,s^{-1}\,cm^{-3}\,keV^{-1}} \\ & Z^2 \left(\frac{T}{10^8\,\mathrm{K}}\right)^{-1/2} \left(\frac{n_\mathrm{e}}{10^{-3}\,\mathrm{cm^{-3}}}\right) \left(\frac{n_\mathrm{i}}{10^{-3}\,\mathrm{cm^{-3}}}\right) g(E,T) \exp\left(-\frac{E}{k_\mathrm{B} T}\right).\end{aligned} \tag{1.13}$$

The exponential fall-off at frequencies $h\nu \sim k_\mathrm{B} T$ is a consequence of the Maxwellian electron velocity distribution, and the fact that an electron must have an energy $^1/_2 m v^2 \geqslant h\nu$ to create a photon of frequency ν (the so-called photon discreteness effect). The function $g(\nu, T)$ is the Gaunt factor. Taking $g \equiv 1$ gives the result obtained by a simple, semi-classical treatment of the scattering process, and gives a good approximation to the true result in the range $E \approx k_\mathrm{B} T$. At lower energies, g increases, rising to ~ 2 by the time E has fallen to one-tenth of $k_\mathrm{B} T$, and to ~ 5 by the time $E \sim 10^{-5} k_\mathrm{B} T$ (for an electron–proton plasma at $T \sim 10^8$ K).

In fact, it can be shown that suitable approximations for the high-temperature Gaunt factor are

$$g(E,T) = \begin{cases} \dfrac{\sqrt{3}}{\pi} \ln\left(4\gamma \dfrac{k_\mathrm{B} T}{E}\right) & E \ll k_\mathrm{B} T \\ \left(\dfrac{3}{\pi}\dfrac{k_\mathrm{B} T}{E}\right)^{1/2} & E \gg k_\mathrm{B} T \end{cases} \tag{1.14}$$

where $\gamma = 0.577\ldots$ is the Euler–Mascheroni constant,

$$\gamma \equiv \lim_{n \to \infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \ln n\right), \tag{1.15}$$

which appears as a factor in the solution of various integrals (fascinatingly, it is not yet known if this number is irrational or transcendental). These approximations are illustrated in Figure 1.5 on the next page, which compares the MEKAL (Section 1.4.4) spectrum of a pure hydrogen plasma at 10^8 K with the result of equation (1.13). Gaunt factors applicable to a wide range of cases are presented by Karzas and Latter (1961).

Integrating equation (1.12) over frequency leads to the total bremsstrahlung emissivity,

$$\varepsilon^{\mathrm{ff}} = \frac{Z^2 e^6}{3\pi^2 \varepsilon_0^3 c^3 m_\mathrm{e}^2 h} \left(^1/_6 \pi m_\mathrm{e} k_\mathrm{B} T\right)^{1/2} n_\mathrm{e} n_\mathrm{i} \bar{g}(T) \tag{1.16}$$

$$\approx 1.42 \times 10^{-29}\,\mathrm{erg\,s^{-1}\,cm^{-3}} Z^2 \bar{g}(T) \left(\frac{T}{10^8\,\mathrm{K}}\right)^{1/2} \left(\frac{n_\mathrm{e}}{10^{-3}\,\mathrm{cm^{-3}}}\right) \left(\frac{n_\mathrm{i}}{10^{-3}\,\mathrm{cm^{-3}}}\right), \tag{1.17}$$

where $\bar{g}(T)$ is a frequency (or energy) average of the Gaunt factor, and lies in the range 1.1–1.5, it being usual to take $\bar{g} = 1.2$ for an approximate result.

As well as the free–free continuum, there is also some contribution from free–bound transitions,

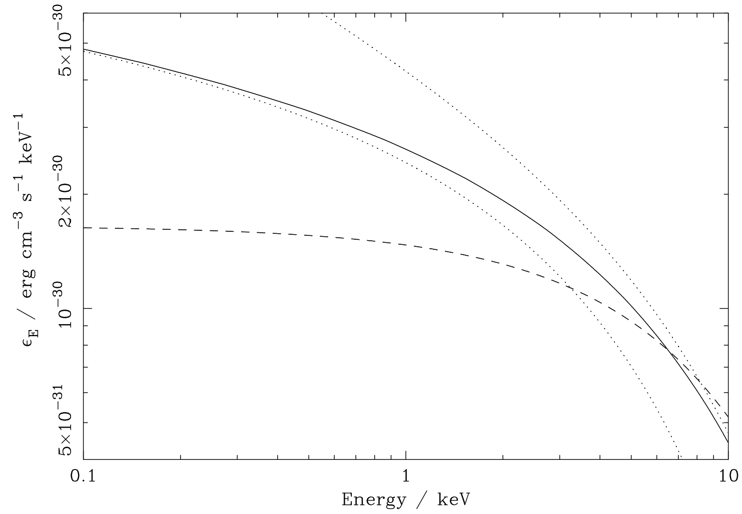
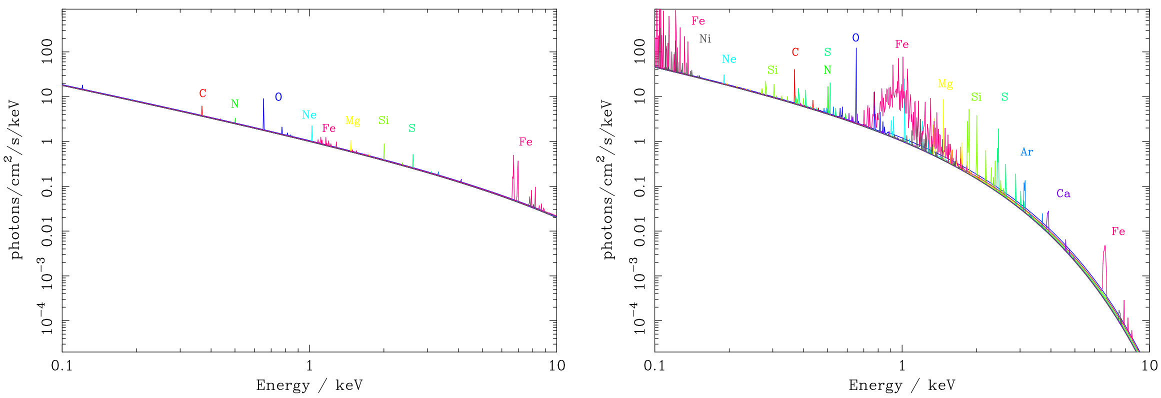


Figure 1.5: The bremsstrahlung spectrum of an electron–proton plasma, with $T = 10^8\,$K, and $n_\mathrm{p} = 10^{-3}\,\mathrm{cm}^{-3}$. The solid curve is from MEKAL. The dashed line is from equation (1.13) with $g \equiv 1$. The dotted lines show the modifications caused by the addition of the high- and low-energy approximations to the Gaunt factor given in equation (1.14). The deviation of the continuum from a simple exponential in the X-ray waveband is non-negligible.



(a) 10 keV

(b) 1 keV

Figure 1.6: Spectra of a solar abundance coronal plasma showing the contributions of individual elements. The fluxes have been scaled to equal 1 at 1 keV. (a) 10 keV plasma, (b) 1 keV plasma. The continuum begins to fall off at energies $\sim k_\mathrm{B}T$. Spectra were produced from MEKAL, with 1000 logarithmically-spaced energy bins.

where an electron in a free continuum state radiatively recombines with an ion.

In addition to the continuum component, bound–bound transitions between the energy states of electrons bound to ions give rise to a host of emission-line features in ICM spectra. Recalling that the energy needed to ionize the hydrogen atom is 13.6 eV, one expects the energy required to remove the last electron from an iron nucleus to be $\sim 26^2 \times 13.6\,\mathrm{eV} \approx 9\,\mathrm{keV}$. This order of magnitude estimate indicates that iron in the ICM will be highly ionized, and that strong emission may be expected near the high-energy end of the 0.1–10.0 keV X-ray waveband. This is indeed the case. Figure 1.6 shows the spectra of 1 and 10 keV coronal plasmas, as calculated by MEKAL. The most prominent features in the spectrum are due to iron. The complex near 7 keV is comprised of K lines from various ionization stages, chiefly Fe XXV and Fe XXVI, and is the most noticeable feature in high-temperature clusters. At lower temperatures, the iron L complex around 1 keV becomes increasingly strong. The continuum is chiefly due to hydrogen and helium, whereas the line intensities obviously depend on the abundance of the appropriate element. Hence taking the ratio of these two components gives a measure of the relative abundances (this is discussed in Section 2.1). The K lines of other elements (e.g. oxygen) are also visible in spectra.

The change in the emission line structure between 1 and 10 keV are large. Good constraints on the temperature of the intracluster gas can be obtained from the iron L lines in the ~ 0.7 to ~ 1.4 keV region of the spectrum. It is clear from Figure 1.6 on the facing page that there is a large number of lines that contribute to this complex. Modelling such a feature is a difficult task. Indeed, observations of cluster spectra in this region (Fabian *et al.*, 1994a) have in the past brought to light inadequacies in the theoretical treatment that have been subsequently corrected (Liedahl *et al.*, 1995). Discrepancies (e.g. Phillips *et al.*, 1999) continue to be found, however, just as theoretical improvements (e.g. Doron and Behar, 2002) continue to be made. Care must be taken with the analysis of the L complex (e.g. Finoguenov *et al.*, 2000). Our understanding of the simple K shell physics is much more complete.

1.4.4 Collisional Ionization Plasma Codes

There is a wide variety of codes available for the modelling of coronal plasmas. For comparative discussions of some of those available, see e.g. Drake (1992) or Brickhouse *et al.* (1995). As a rule, the required inputs are the plasma temperature, density and abundance, the energy range of interest and the required spectral resolution, and some normalization factor. The output is a spectrum in some form (e.g. emissivity or flux), and usually also a relative electron density describing the overall ionization state of the plasma.

The temperature range accessible to most of these coronal codes is generally $\sim 10^4$–10^{10} K. At the high-temperature end of the range, relativistic corrections become important; at the low-temperature limit the plasma approaches neutrality and the calculation of ionization balance becomes more complex due to charge transfer effects. The continuum can be modelled from very low energies up to several hundred keV, whereas the modelling of emission lines is usually restricted to ~ 0.01–10 keV (i.e. ~ 1–1200Å) or so. The applicable density range extends well above that relevant for the ICM, e.g. to $\gg 10^{10}\,\mathrm{cm}^{-3}$.

One of the most widely used has been the Raymond–Smith code (Raymond and Smith, 1977), which has gone through various incarnations since its inception. It models the effects of 13 elements: H, He, C, N, O, Ne, Mg, Si, S, Ar, Ca, Fe, and Ni.

Another commonly-encountered collisional ionization code (and the one which I make use of in the work described in this thesis) is MEKAL (e.g. Mewe *et al.*, 1995), the name deriving from those of its principle authors (Mewe, Kaastra, and Liedahl). It includes the same elements as the Raymond–Smith code, as well as Na and Al. Internally, it uses the solar abundances of Anders and Grevesse (1989), though by scaling the input relative abundances one can of course mimic any desired solar abundance. MEKAL is an update of the earlier code MEKA (e.g. Kaastra and Mewe, 1993), which in turn was developed from work by Mewe and Gronenschild (e.g. Mewe and Gronenschild, 1981).

The main improvement of MEKAL over MEKA is in the treatment of the Fe L complex, through the inclusion of the corrections of Liedahl *et al.* (1995). The Fe L complex in the region of 1 keV is comprised of thousands of lines from ions Fe XVII–XXIV, originating from gas at temperatures $\sim 3\times10^6$–2×10^7 K. Inadequacies in the modelling of this region of the spectrum were highlighted by *ASCA* observations of cooling flow clusters (Fabian *et al.*, 1994a). By way of illustration, MEKA contains some 2400 emission lines in total; whereas MEKAL includes over 2600 Fe L lines in the 6–20 Å region alone (an increase of more than a factor of 10 in that wavelength range).

Codes such as MEKAL have a long history, essentially having their origins in the 1970s, and having been modified, rewritten, and added to many times in the intervening years. This probably does not make for the cleanest or most efficient of implementations. Additionally, experimental spectral data then was of a much lower quantity than that available now. With the development of new laboratory techniques, such as the electron beam ion trap (EBIT) in the late 1980s, the quantity and quality of available spectroscopic data has increased dramatically; and fortunately so, since the ever-increasing spectral resolution available observationally allows (even demands) much more detailed studies of astrophysical plasmas.

There are several new projects in the field of X-ray spectral modelling making use of up-to-date atomic data. Perhaps the two most prominent (or promising) are CHIANTI and ATOMDB. CHIANTI[6] consists of an atomic database together with Interactive Data Language (IDL) routines for the calculation of spectra. The project homepage is `http://wwwsolar.nrl.navy.mil/chianti.html` (Dere *et al.*, 1997; Young *et al.*, 2003). ATOMDB ("The Atomic Database") consists of two components — APED ("The Astrophysical Plasma Emission Database") and the spectral models output from APEC ("The Astrophysical Plasma Emission Code"). The project homepage is `http://cxc.harvard.edu/atomdb/` (Smith *et al.*, 2001). The reasons why I have not investigated these more recent codes in much detail are essentially prosaic. The source code for APEC is not yet available (although at some stage it will be released); and with regards to CHIANTI, IDL is not this author's programming language of choice.

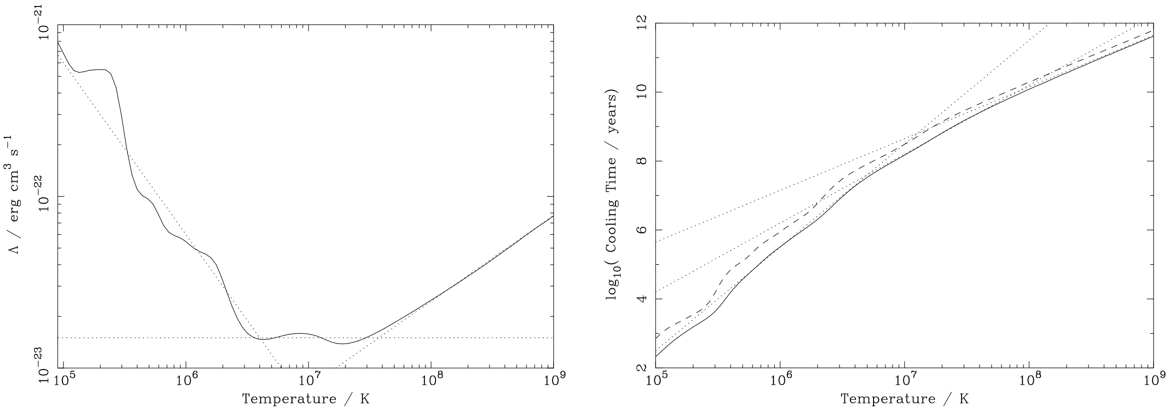
To some extent, the choice of plasma code adopted for this work does not matter too much. This is not to deny that the comparison of spectral codes and databases is in important project — it most certainly is (e.g. Brickhouse *et al.*, 1995). Rather, it is such a large and specialized subject that it is best addressed in isolation. To attempt to look at any such issues here would only muddy the waters. MEKAL is a convenient choice for my purposes for a variety of reasons. It is readily available, written in the same language as the rest of the numerical code (Chapter 3), and can easily be incorporated into the model. Furthermore, it has been widely used in analyses of the X-ray spectra of the ICM. There are, for example, a variety of XSPEC (Section 1.4.5) models that employ it. Using the same spectral synthesis code to fit the spectra as was used to produce them immediately eliminates one possible source of confusion that is present in real-world analyses, namely the possibility that the atomic physics might not be modelled correctly. Again, that is not to say that this is not an important issue, rather it is merely not one I will address in this work.

1.4.5 XSPEC

The software package XSPEC[7] (Arnaud, 1996) has been for many years *the* tool for the analysis of X-ray spectra. It (specifically, Version 11) has been used extensively in the work presented in this thesis, and

[6]As best as I am able to ascertain, the name has no meaning other than to reflect the national origin (and/or imbibitional habits) of some of the project members.

[7]`http://heasarc.gsfc.nasa.gov/docs/xanadu/xspec/`



(a) Cooling function

(b) Cooling time

Figure 1.7: Solid lines — the cooling function and constant-pressure cooling time for an ionized gas of metallicity $0.3\,\mathrm{Z}_\odot$. Dotted lines — simple power-law approximations to the functions, see equations (1.19) and (1.27) for details. The dashed line in (b) shows the approximate cooling time of equation (1.26), based on the instantaneous cooling rate. These figures were produced using numerical integration of MEKAL spectra.

I am indebted to its authors and developers. XSPEC is distributed as part of the XANADU[8] package for X-ray data analysis.

1.5 Radiative Cooling of the Intracluster Medium

1.5.1 The Cooling Function

The cooling function Λ is defined by

$$\text{Radiative power loss per unit volume} = n_\mathrm{e} n_\mathrm{H} \Lambda(T), \tag{1.18}$$

where n_e and n_H are the electron and hydrogen number densities respectively. The choice of $n_\mathrm{e} n_\mathrm{H}$ over n_e^2, n_H^2, or indeed n^2 in this expression is fairly arbitrary. The cooling function is a useful concept since it separates the functional dependence of the radiative cooling into its density and temperature components.

The cooling function of a $0.3\,\mathrm{Z}_\odot$ gas is plotted in Figure 1.7(a). To a reasonable degree of accuracy, the cooling function can be modelled by simple powers laws in various temperature ranges. Shown in Figure 1.7(a) are the approximations

$$\frac{\Lambda(T)}{\mathrm{erg\,cm^3\,s^{-1}}} \approx \begin{cases} 6.0\times10^{-17}\left(\dfrac{T}{\mathrm{K}}\right)^{-1} & 10^5\,\mathrm{K} < T < 3\times10^6\,\mathrm{K} \\ 1.5\times10^{-23} & 3\times10^6\,\mathrm{K} \leqslant T \leqslant 2\times10^7\,\mathrm{K} \\ 2.5\times10^{-27}\left(\dfrac{T}{\mathrm{K}}\right)^{\frac{1}{2}} & 2\times10^7\,\mathrm{K} < T < 10^9\,\mathrm{K} \end{cases} \tag{1.19}$$

More sophisticated representations than this are of course possible (e.g. Sutherland and Dopita, 1993),

[8] `http://heasarc.gsfc.nasa.gov/docs/xanadu/`

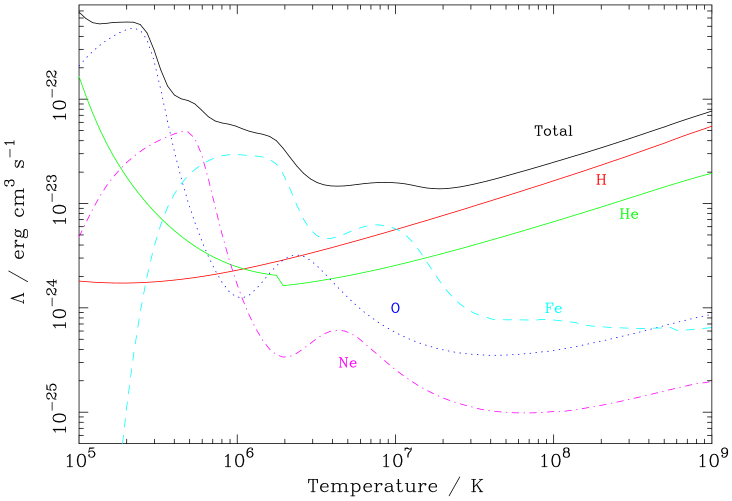


Figure 1.8: Contribution of various elements to the cooling function for a 0.3 $Z_\odot$ plasma (H, He solar). (Solid black) all; (solid red) H; (solid green) He; (dash cyan) Fe; (dot blue) O; (dash-dot magenta) Ne.

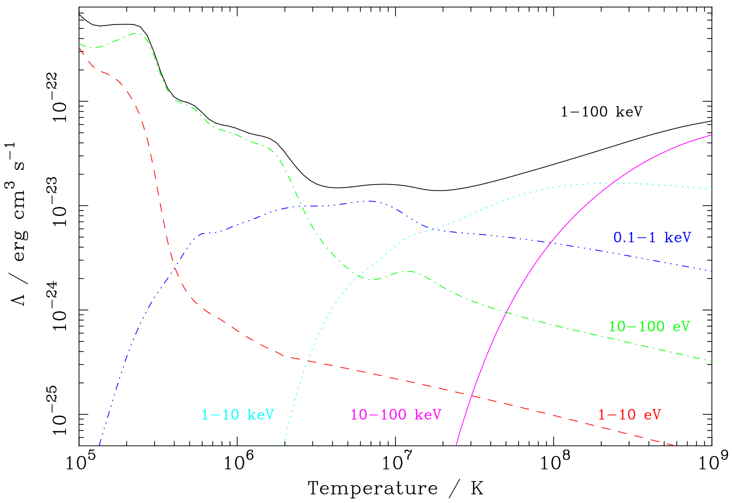


Figure 1.9: Contribution of various spectral energy ranges to the cooling function for a 0.3 $Z_\odot$ plasma (H, He solar). (Solid black) 1 eV–100 keV; (dash red) 1–10 eV; (dash-dot green) 10–100 eV; (dash-dot-dot-dot blue) 0.1–1.0 keV; (dot cyan) 1–10 keV; (solid magenta) 10–100 keV.

this is merely intended to illustrate that the functional form of the Λ profile is not too complex.

Figure 1.8 shows the contributions to the cooling function made by various important ICM species. Different species dominate the cooling function in different temperature ranges. At temperatures $\gtrsim 2 \times 10^7$ K, the cooling is largely due to the bremsstrahlung continuum radiation from hydrogen, and (to a lesser extent) helium. In the range 10^6–10^7 K, iron is the major coolant. At very low temperatures (by ICM standards), oxygen becomes important. The effect of metals on the cooling of the intracluster gas is explored in more detail in Chapter 2.

Figure 1.9 illustrates the contributions made by various spectral energy ranges to the bolometric cooling function. The 0.1–10.0 keV waveband dominates in the range $\sim 2 \times 10^6$ K to $\sim 2 \times 10^8$ K, with the emission shifting to the UV at lower temperatures.

1.5.2 Emission Measure

From the definition of the cooling function, the total power P radiated by a source is

$$P = \int \Lambda(T) n_e n_H \, dV.$$

The *emission measure* (EM) is defined by

$$P \equiv \int \Lambda(T)(\mathrm{EM}) \qquad \Rightarrow \qquad \mathrm{EM} \equiv n_e n_H \, dV. \tag{1.20}$$

Hence (e.g. Buote *et al.*, 1999) the *differential emission measure* (DEM),

$$\mathrm{EM} \equiv \int (\mathrm{DEM}) \, dT \qquad \Rightarrow \qquad \mathrm{DEM} = \xi(T) \equiv n_e n_H \frac{dV}{dT}, \tag{1.21}$$

$$P = \int \Lambda(T) \xi(T) \, dT. \tag{1.22}$$

The integral of the emission measure over the entire source is sometimes referred to as the *emission integral* (EI; see e.g. Sarazin, 1988). The EI sets the normalization of the spectrum, and the DEM weights the contributions of the various temperature ranges to the emission and essentially specifies the temperature structure of the plasma.

1.5.3 The Cooling Time

The radiative cooling time of the ICM may be readily calculated. From the definition of the cooling function, equation (1.18),

$$\rho \frac{dq}{dt} = -n_e n_H \Lambda(T), \tag{1.23}$$

where the specific (per unit mass) energy q obeys

$$dq = d\varepsilon + p \, d\mathcal{V},$$

with ε and $\mathcal{V}$ the specific internal energy and volume respectively. For a constant pressure, ideal gas

$$dq = d\left(\varepsilon + \frac{p}{\rho}\right) = \left(\frac{1}{\gamma - 1} + 1\right) d\left(\frac{p}{\rho}\right) = \frac{\gamma}{\gamma - 1} \frac{k_B}{\mu m_H} dT.$$

Hence from equation (1.23),

$$\frac{\gamma}{\gamma - 1} \frac{\rho k_B}{\mu m_H} \frac{dT}{dt} = -r_e n_H^2 \Lambda(T)$$

$$\frac{\gamma}{\gamma - 1} \frac{k_B n}{(\mu X_H n)^2} \frac{dT}{dt} = -r_e \Lambda(T) \qquad \text{using} \quad n_H = \frac{X_H \rho}{m_H} = \mu X_H n$$

with $r_e \equiv n_e / n_H$ the relative electron density, and $X_H \equiv \rho_H / \rho$ the hydrogen mass fraction. In the following, the temperature dependence of r_e is neglected, since it is almost constant over the temperature range that makes the dominant contribution to the integral (see Figure 1.4 on page 11).

$$\int_0^{t_{\rm cool}} \mathrm{d}t = \frac{-\gamma}{\gamma-1}\left(\frac{k_{\rm B}}{\mu X_{\rm H}}\right)^2 \frac{1}{r_{\rm e}p}\int_T^0 \frac{T\,\mathrm{d}T}{\Lambda(T)} = \frac{\gamma}{\gamma-1}\left(\frac{k_{\rm B}}{\mu X_{\rm H}}\right)^2 \frac{1}{r_{\rm e}}\frac{\mu X_{\rm H}}{n_{\rm H}k_{\rm B}T}\int_0^T \frac{T\,\mathrm{d}T}{\Lambda(T)}$$

$$\therefore \quad t_{\rm cool} = \frac{\gamma}{\gamma-1}\left(\frac{k_{\rm B}}{\mu X_{\rm H}}\right)^2 \frac{1}{r_{\rm e}p}\int_0^T \frac{T\,\mathrm{d}T}{\Lambda(T)}. \tag{1.24}$$

A more approximate value for the constant-pressure cooling time is often calculated as

$$t'_{\rm cool} = \frac{\mathcal{H}}{n_{\rm e}n_{\rm H}\Lambda(T)} \qquad \mathcal{H} \equiv \rho\left(\varepsilon + \frac{p}{\rho}\right) = \frac{\gamma}{\gamma-1}nk_{\rm B}T, \tag{1.25}$$

that is, the time required to radiate enthalpy $\mathcal{H}$ (per unit volume) at temperature T using the instantaneous rate of energy loss at that temperature. This is equivalent to neglecting the integral in equation (1.24), so that

$$t'_{\rm cool} = \frac{\gamma}{\gamma-1}\left(\frac{k_{\rm B}}{\mu X_{\rm H}}\right)^2 \frac{1}{r_{\rm e}p}\frac{T^2}{\Lambda(T)}. \tag{1.26}$$

Both these equations are plotted in Figure 1.7(b) on page 17, along with the power-law approximations

$$\frac{t_{\rm cool}(T)}{\rm yr} \approx \begin{cases} 3.2\times10^{-13}\left(\frac{T}{\rm K}\right)^3 & 10^5\,{\rm K} < T < 3\times10^6\,{\rm K} \\ 1.6\times10^{-6}\left(\frac{T}{\rm K}\right)^2 & 3\times10^6\,{\rm K} \leqslant T \leqslant 2\times10^7\,{\rm K} \\ 1.4\times10^{-2}\left(\frac{T}{\rm K}\right)^{\frac{3}{2}} & 2\times10^7\,{\rm K} < T < 10^9\,{\rm K} \end{cases} \tag{1.27}$$

where the functional form follows directly from equations (1.19) and (1.24). The difference between equation (1.24) and the more approximate equation (1.26) is slight.

Scaling Relations

From equation (1.24), at constant pressure in the high-temperature regime, with $\Lambda \propto T^{1/2}$, $t_{\rm cool} \propto T^{3/2}$. For virially-scaled clusters at a fixed overdensity, $p \propto T$, and consequently $t_{\rm cool} \propto T^{1/2}$.

1.6 Cooling Flows

The various ICM timescales (Section 1.3.4) ensure that, to a good first approximation, the intracluster gas can be regarded as being in hydrostatic equilibrium. This assumption is borne out by the excellent agreement (e.g. Allen *et al.*, 2002, 2001a; Schmidt *et al.*, 2001) that is found to exist between mass profiles derived from X-ray data on this basis, and those obtained from gravitational lensing studies, which are free of any such assumption. In hydrostatic equilibrium, the ICM is naturally densest in the centre of the potential well (Section 3.5.1), and consequently the cooling time is shortest there. The densities in the central 100 kpc or so of many clusters (e.g. RX J137.5-1145, Allen *et al.* 2002;

Abell 1795, Ettori *et al.* 2002; Abell 2199, Johnstone *et al.* 2002; Abell 1835, Schmidt *et al.* 2001; Hydra A, David *et al.* 2001) are sufficiently high that the cooling time is less than 10^{10} yr. This means that, in the absence of any competing factors (but see Section 1.6.1), the gas in the central regions of such clusters can undergo significant energy loss through radiative cooling over the cluster lifetime.

Since the cooling time is a monotonically increasing function of radius, then if the central cooling time is less than the age of the system, there must be some radius at which $t_{\rm cool} = t_{\rm age}$. This is known as the *cooling radius*. To a first approximation, the gas within $r_{\rm cool}$ cools through radiation, whereas outside this radius cooling is negligible. In order to support the weight of the overlying gas, the pressure in the core of the cluster must be maintained, which necessitates a density increase to compensate for the drop in temperature. In the absence of matter sources, gas must flow inward to sustain the pressure. This slow, subsonic inflow through a succession of quasi-hydrostatic equilibrium states is known as a *cooling flow*. Slightly more quantitatively, one may define a cooling flow cluster to be one where the central cooling time is less than a Hubble time. Using this definition, Edge *et al.* (1992) found that at least 70 per cent and possibly as high as 90 per cent of clusters in a flux-limited sample qualified as cooling flow clusters (this is not a selection effect since the flow makes only a minor contribution to the overall luminosity). Changing the definition of the cluster "age" by a factor of two or so does not greatly affect the result. One factor that does affect whether or not a cooling flow is detected is the spatial resolution of the observing instrument. Since the cooling time is an increasing function of radius, taking a radial average from a low-resolution image will always produce an overestimate of the central cooling time. Thus, for example, of the clusters detected by the *Einstein* observatory, less than 50 per cent were inferred to possess cooling flows, while with the higher resolution *EXOSAT* the fraction rose to greater than two-thirds.

Those clusters that do not show signs of a cooling flow often show evidence of recent mergers (e.g. a highly distorted subcluster in the core in the case of Abell 2256). Based on the inferred frequency of mergers, Allen *et al.* (2001b) estimated that merger with a subcluster of mass ~ 30 per cent of the total cluster mass could disrupt a cooling flow.

Observationally, cooling flows manifest themselves in the form of a sharply peaked X-ray surface brightness profile, which betrays the rise of gas density in the core. The contribution of the cooling flow to the total cluster X-ray luminosity ($\sim 10^{43}$–$10^{45}\,{\rm erg\,s^{-1}}$) is usually less than about 10 per cent though. The gas within the cooling radius constitutes only a small fraction of the total. For example, using the Abell 1835 density profile of Schmidt *et al.* (2001), it follows that

$$\frac{M(<100\,{\rm kpc})}{M(<1\,{\rm Mpc})} \sim 3\%.$$

If the ICM were a single-phase medium, in which at any radius there existed only gas with one temperature, then because the coolest gas would inevitably be in the central regions, all mass loss (i.e. cooling to low temperatures and moving out of the X-ray waveband) would occur at the centre. The resulting surface brightness profile would be extremely sharply peaked. Observations, however, suggested that the mass deposition profiles increased roughly linearly with radius (e.g. Stewart *et al.*, 1984), implying distributed mass dropout over a range of radii. This led to the establishment of the multi-phase hypothesis (Chapter 6).

There is an extensive literature of reviews charting the development of the study of cooling flows (e.g. Fabian *et al.*, 1984, 1991; Fabian, 1994, 2002a,b).

1.6.1 The Cooling Flow Problem

In Section 1.6, it was stated that "in the absence of any competing factors" there would be appreciable cooling in the centres of many clusters. Recently, it has become clear that we cannot dismiss these putative "competing factors" so lightly.

The cooling flow hypothesis has always been somewhat controversial, however, since there have been several issues whose explanation has remained unclear. Whilst the picture obtained from X-ray studies in the era before *Chandra* and *XMM-Newton* was essentially self-consistent, there was difficulty with interpreting the results at other wavebands.

The mass deposition rates inferred from X-ray studies ranged from $\sim$ 100–1000 $\mathrm{M_\odot\,yr^{-1}}$ for strong flows. If all this gas seen leaving the X-ray temperature regime were to form stars with a typical IMF, then cluster central galaxies would be both brighter and bluer than is seen. Although an excess of blue light *is* seen in some cases (e.g. Johnstone *et al.*, 1987; Crawford *et al.*, 1999; Bayer-Kim *et al.*, 2002), the rates are substantially less than those implied by the X-ray data. It was therefore suggested (Fabian *et al.*, 1982) that the high (compared to the Galactic interstellar medium) pressures in the cores of cooling flow clusters might reduce the Jeans mass and so produce an IMF biased towards low-mass stars.

The central 10 kpc or so of cooling flow clusters often exhibit optical line nebulosity (Crawford and Fabian, 1992; Donahue and Voit, 1993), at levels which indicate that the gas must have recombined several hundred times. Edge (2001) has detected significant levels of CO emission in the central galaxies of several cluster cooling flows, implying the presence of substantial amounts of warm molecular gas ($\sim 10^9$–10^{11} $\mathrm{M_\odot}$ within the central 50 kpc). Large amounts of dust have also been seen in central cluster galaxies in cooling flow clusters (Edge *et al.*, 1999).

Data from the latest generation of X-ray satellites, *Chandra* and *XMM-Newton*, are forcing us to re-examine some of the basic tenets of the traditional cooling flow model (e.g. Böhringer *et al.*, 2002; Fabian, 2002b). Various authors (e.g. Kaastra *et al.*, 2001; Peterson *et al.*, 2001; Tamura *et al.*, 2001) have drawn attention to the discrepancy between the predictions of the standard cooling flow model (Section 2.5) and observed spectra of cluster central regions. The expected emission lines for several important species (e.g. the Fe XVII 15 and 17 Å lines) do not seem to be present at the levels predicted by simple models. This is a trend that appears to be repeated in many clusters that have traditionally been thought to harbour cooling flows (Peterson *et al.*, 2002). Lines such as these are important because they are strong indicators of low-temperature ($\lesssim$ 1 keV) gas. In a standard cooling flow, in which gas is cooling down to essentially zero, we expect a significant flux in such lines. This issue is discussed in more depth in Chapter 2.

Several ideas (e.g. Fabian *et al.*, 2001a; Peterson *et al.*, 2001) have been put forward to explain this discrepancy. Perhaps the most obvious is to say that, if the spectral signature of cool gas is not seen, maybe this is due to the simple fact that no cool gas is present. For this to be the case, given the known radiative cooling rates, there must be some significant heat source offsetting cooling. Such heating models have a long history (e.g. Bregman and David, 1989). Many clusters contain strong radio sources, which are an obvious candidate for the heat engine. *Chandra* observations, however, have revealed that the ICM surrounding any radio lobes tends to form a bright, *cool* shell (e.g. Fabian *et al.*, 2000; Schmidt *et al.*, 2002). Fabian *et al.* (2002b) have quantified the energy required to offset cooling within the cooling radius over the likely lifetimes of these sources (their figure 3). Assuming the canonical accretion efficiency of around 10 per cent, a central black hole would have to accrete $\sim 10^{10}\,M_\odot$ to offset cooling. Churazov *et al.* (2002) also consider the power requirements of the AGN heating model. The energy must be transferred to the ICM somehow. Various authors have considered the idea that the radio source might heat buoyant plasma bubbles that would then convect through the ICM, distributing energy (e.g. Churazov *et al.*, 2001; Brüggen *et al.*, 2002). Some energy may be lost as sound waves which escape from the cooling flow region. In the case of Abell 2199, Johnstone *et al.* (2002) calculate the way in which heat must be distributed spatially in order to maintain a steady state, concluding that the heating rate per unit volume must be roughly proportional to $r^{1.5}$.

If cooling is occurring at the rates traditionally inferred from X-ray studies, then some way must be found to change the spectral signature of the resulting cool gas. There are two possibilities: either the spectral lines are intrinsically much weaker than expected; or some other process masks the lines from view.

With regards to the former suggestion, even though our understanding of the iron L shell emission from cool gas may be less than perfect, the problem cannot be dismissed in this fashion with just an appeal to atomic physics — the discrepancies are too large and too global, and are not just restricted to certain lines (although examining specific lines in detail is one good way to present the problem). In a similar fashion, the idea of resonant scattering (Gil'fanov *et al.*, 1987), in which the optical depth in certain emission lines becomes non-negligible (i.e. the optically thin assumption breaks down), is not a viable explanation, being inconsistent with the detailed spectral picture. Certain lines of low oscillator strength are weak in observed spectra, whereas others of higher oscillator strengths are relatively unaffected (see e.g. Peterson *et al.*, 2001).

Differential absorption is one possibility suggested by both Peterson *et al.* (2001) and Fabian *et al.* (2001a) to disguise the emission from cool gas. In this hypothesis, absorption by cold gas is selectively applied to the cool X-ray gas alone. One can imagine a self-consistent scenario in which the cold, X-ray absorbing gas is naturally located only in the central regions where the cool gas is to be found. High spatial resolution studies should be able to detect this absorbing screen.

Another possibility is that the cool X-ray gas is somehow cooled non-radiatively to very low temperatures. One way this might happen is through mixing with cold gas. Oegerle *et al.* (2001) report a *FUSE* detection of O VI 1032 Å radiation from the cooling flow cluster Abell 2597. Such emission originates from gas at temperatures $\sim 3\times10^5$ K. Begelman and Fabian (1990) showed that when hot and cool gas are mixed together by turbulence, the final temperature reached is the geometric mean of the starting temperatures. Oegerle *et al.* (2001) point out that mixing hot ($\sim 10^7$ K) and cold ($\sim 10^4$ K, i.e. Hα temperatures) gas results in a temperature close to that appropriate for O VI emission. It is therefore possible that the fact that emission from O VI is observed, whilst that from Fe XVII is not, is indicative of non-radiative cooling due to mixing with cold gas.

Two more hypotheses, namely the ideas of small-scale metallicity variations (first suggested by Fabian *et al.*, 2001a), and efficient thermal conduction (an idea which has a long history in the context of ICM studies), are explored in subsequent Chapters of this thesis. If any one process is to be responsible, a major problem is finding a mechanism that can operate over the wide range of cluster masses and temperatures, in a long-lived, stable fashion.

The problem is all the more bemusing, given that the physics involved seems to be both simple (radiative cooling of hot gas in a potential well), and at the same time so fundamental — very similar processes must be at work during the course of galaxy formation. Fabian *et al.* (2002b) have recently suggested that the efficient thermal conduction (Chapter 5) which has found considerable favour of late in connection with cooling flows, might be the factor responsible for the upper cut-off on the mass distribution of galaxies (e.g. Kauffmann *et al.*, 1999).

During the writing of this thesis, the discrepancy between models and observations has been put on a more quantitative footing by Peterson *et al.* (2003) (see also Kahn *et al.*, 2002). These authors plot the ratio of the observed emission measure originating from gas in a given (fairly broad) temperature range to that predicted from the isobaric cooling flow model against temperature. There is emission originating from gas across the entire X-ray temperature range, but the deficit compared to the standard cooling flow model increases as the temperature decreases. In some sense, this is harder to explain than a complete absence of any cool gas. It is difficult to see how one might keep the situation balanced in this way between the two extremes of no cooling at all, and strong cooling at the traditional rate. It appears that we require a throttling back of the cooling process rather than a complete shutdown. Some form of feedback must come into play between the radiative cooling process and its control mechanism.

Intriguingly, when the temperature abscissa in the emission measure plots is scaled according to the cluster virial temperature, the lines corresponding to various clusters lie fairly close to one another (i.e. there is a degree of similarity between clusters). If this result (based on a sample of 14 clusters) holds generally, then it is, I believe, a very important result, the significance of which is not yet clear.

You will reply that reality hasn't the slightest need to be of interest. And I'll answer you that reality may avoid the obligation to be interesting, but that hypotheses may not.

Death and the Compass
JORGE LUIS BORGES

2 The Effect of Metallicity Inhomogeneities on Cooling Flow Spectra

In this Chapter, I focus on one suggested solution to the deficit of line emission observed in the soft X-ray spectra of cooling flows (Section 1.6.1), namely the idea that the ICM metals might be distributed inhomogeneously on small scales (Fabian *et al.*, 2001a). I investigate the consequences of such a situation within the framework of the cooling flow scenario. Using the standard isobaric cooling flow model, I study the ability of such metallicity variations to suppress preferentially low-temperature line emission, examining in detail the consequences due to the effects of an ICM metallicity which varies on small, unresolved scales (sub-kiloparsec, say). More discussion of the validity of this hypothesis is made in Chapter 4, where I revisit this issue using numerical simulations to examine the temporal and spatial evolution of the ICM and its spectrum when the metals are distributed in such a fashion.

The general basis of this idea (as discussed by Fabian *et al.*, 2001a) is that at any fixed temperature, a mixture of metal-rich and metal-poor gas is indistinguishable from some homogeneous, mean metallicity plasma (Section 2.6.1). Qualitatively, on increasing the metallicity by some factor, the emission line strength (at all temperatures) increases proportionately. Reducing the mass fraction compensates, so that spectra at any fixed temperature are unchanged.

When steady-state cooling is allowed to take place, however, the faster cooling of the metal-rich gas at low temperatures reduces its emission measure and suppresses its spectral contribution (Section 2.6.2). The initial expectation quoted by Fabian *et al.* (2001a) is that concentrating the metals in 10 per cent of the gas, say, would reduce the strength of low-temperature line emission by a factor of ten. I shall put this assumption on a firmer quantitative footing, deriving the reduction in equivalent width that can be expected for several important ICM species as the degree of inhomogeneity is varied.

In published work on this topic (Morris and Fabian, 2002, 2003), I have used the term "bimodal metallicity" when referring to a mixed-metallicity ICM of the sort that will be discussed in this Chapter.

This is not incorrect, but perhaps it is not the best term for the very simple metallicity distributions I consider here, in which the distribution function consists solely of two δ functions at a high and low metallicity. On the other hand, it does serve as a reminder that the models considered are meant to be merely simplified abstractions of a more realistic situation in which there is a continuous spread of metallicity over some range. In this thesis, I shall endeavour to use the term "two-component metallicity", which I feel is more descriptive of the actual models I consider, if not the situation they are intended to represent.

The equivalent width will turn out to be a useful quantity in the analysis, hence I begin this Chapter with its definition and some discussion of its calculation.

2.1 Equivalent Width

The equivalent width (EW) of a spectral line is defined by

$$\mathrm{EW} \equiv \int_{E_0}^{E_1} \frac{I_{\mathrm{L+C}} - I_{\mathrm{C}}}{I_{\mathrm{C}}}\, \mathrm{d}E, \tag{2.1}$$

where $I_{\mathrm{L+C}}$ is the total (line + continuum) intensity at a given energy, I_{C} is the continuum intensity at the same point, and E_0 and E_1 are energies which bracket the line. The precise values of the bounding energies are irrelevant so long as no other lines are included in the integration region. The equivalent width has the units of the integration variable, be they energy, wavelength or frequency. It may be divided by the energy/wavelength/frequency of the line to obtain the normalized equivalent width, a dimensionless quantity. Strictly speaking, the equivalent width is defined to be positive for *absorption* lines and negative for *emission* lines, but I will not adhere to this distinction, since only emission lines are ever discussed in this thesis. Hence I use the definition of equation (2.1). For an absorption line, the equivalent width represents the width of a saturated (i.e. absorbing *all* flux over the relevant energy range) rectangular line that absorbs the same fraction of the total flux as the original line in question (hence the name). For my purposes, it is simply a convenient measure of the strength of a line relative to that of the continuum.

When it is said that there is a lack of low-temperature emission lines in ICM spectra, this lack refers to values relative to the continuum. We have no independent way to determine the expected continuum level — the normalization is a free parameter. In principle, one could calculate an absolute spectrum using equation (2.11), but in practice the mass-flow rate $\dot{M}$ is not known, but rather is itself determined from the fitted normalization. Hence some way of parametrizing the situation in a continuum-independent fashion is required.

The equivalent width is an important quantity in the study of X-ray cluster spectra, although not often referred to directly. The shape of the bremsstrahlung continuum is controlled by the plasma temperature, and its strength by the square of the density. The strength of an emission line also scales as the square of the density, but has an additional dependence on the metallicity. With the temperature determined from the continuum shape, taking the ratio of the line strength to that of the continuum eliminates the density dependence and allows the metallicity to be measured.

Whilst the equivalent width is conceptually simple, in practice it can be difficult to evaluate, with the

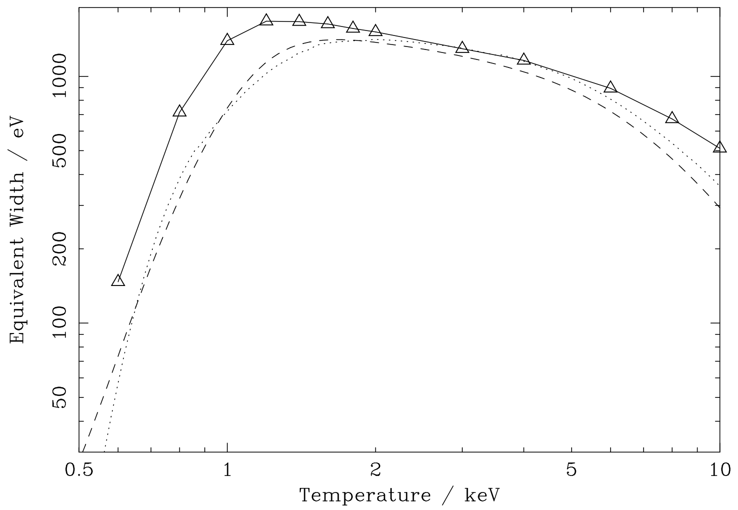


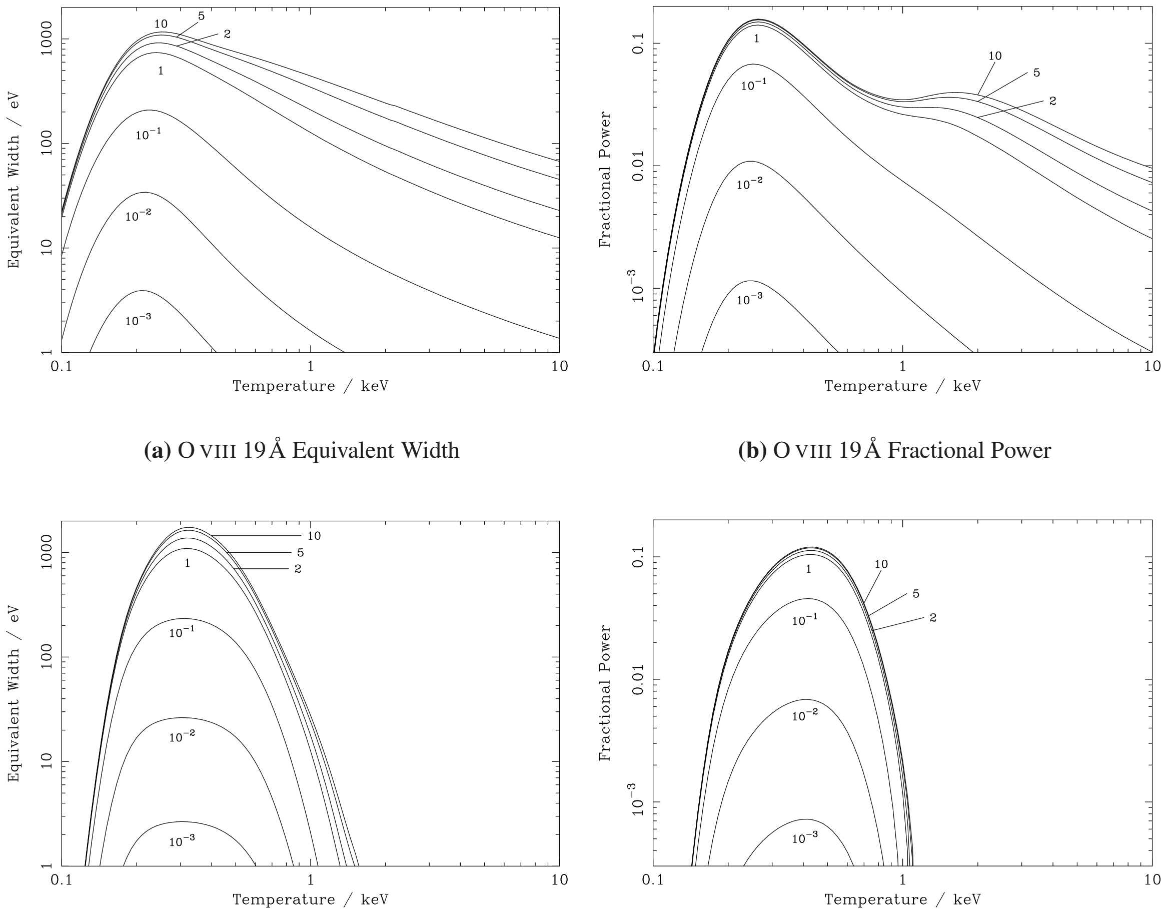
Figure 2.1: Comparison of equivalent width obtained from MEKAL (dashed) with tabulated data (solid) from Arnaud and Rothenflug (1985). Results are for the 6.5–6.7 keV iron complex in a solar abundance plasma. Also shown (dotted) are the results obtained using the Raymond–Smith code.

problem lying in fitting the continuum. Programs such as DIPSO[1], and its successor SPLAT[2] allow for interactive "point-and-click" definition of a continuum. The more difficult problem is that of *automated* continuum fitting, as is essential when there are large numbers of spectra to be handled. Definition of precisely where the continuum lies becomes non-trivial when the spectrum is rich in line features. Fortunately, this is not the sort of problem which we have to put up with[3]. The spectra that will be discussed here all have simple continua that are essentially exponentials (see Section 1.4.3), and are for the most part relatively sparse in lines. Provided one can locate an energy range that encompasses just the line of interest (over the temperature range of interest), equivalent widths can be obtained to a good level of accuracy simply by linearly interpolating (in log-log space) the continuum over the width of the line. This method is much more accurate (and simpler) than, say, trying to fit the whole continuum with a high-order polynomial or some other functional form. I have checked the method is sound by comparing the results with those of DIPSO, for both outlying and intermediate temperatures, say, when calculating over a temperature range. The procedure becomes more difficult at low temperatures due to the increase in both the density of lines and spectral curvature. Figure 2.1 shows a comparison between the equivalent width obtained in this way from MEKAL spectra with data from Arnaud and Rothenflug (1985). Figure 2.2 on the next page shows the temperature dependence of the equivalent width of some prominent ICM spectral lines.

[1]`http://www.ast.man.ac.uk/~dsb/dipso/dipso.html`

[2]Unlike DIPSO, this at least seems to possess *some* justification for its name — SPectraL Analysis Tool. `http://www.starlink.rl.ac.uk/star/docs/sun243.htx/sun243.html`

[3]. . . following Churchill's advice on the matter of prepositions.



(a) O VIII 19 Å Equivalent Width

(b) O VIII 19 Å Fractional Power

(c) Fe XVII 15 Å Equivalent Width

(d) Fe XVII 15 Å Fractional Power

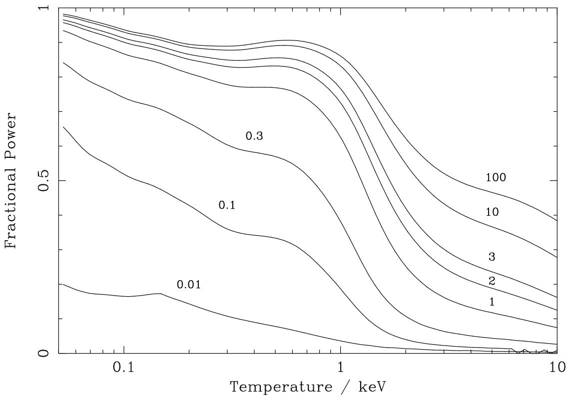
Figure 2.2: Temperature and metallicity dependence of equivalent width (a, c) and fractional power (b, d) for the O VIII 19 Å and Fe XVII 15 Å emission lines. The curve labels refer to metallicity in solar units. Fractional powers are calculated in the 0.1–10.0 keV X-ray waveband.

2.2 Power in Emission Lines

The fraction of the total radiated power in a given waveband being emitted through any particular emission line is given by

$$\mathcal{F} \equiv \frac{\displaystyle\int_{E_0}^{E_1} \left(I_{\mathrm{L+C}} - I_{\mathrm{C}}\right) \mathrm{d}E}{\displaystyle\int_{\mathrm{all}} I_{\mathrm{L+C}} \, \mathrm{d}E}, \tag{2.2}$$

with I the spectral intensity, e.g. erg s^{-1} keV^{-1}, and E_0 and E_1 energies which bracket the line, as before. For a strong emission line, the first term in the numerator is many orders of magnitude greater than the second. Note that whereas the equivalent width is a purely *local* quantity (i.e. it depends only on the spectrum in the vicinity of the line in question), the fractional power depends upon the *global* properties of the spectrum across the entire energy range of interest. The continuum fitting required for the numerator is handled as for the equivalent width, whereas the denominator needs only straightforward



(a) Temperature dependence

(b) Metallicity dependence

Figure 2.3: Temperature and metallicity dependence of the fraction of power radiated in lines for the 0.1–10.0 keV X-ray waveband. In the temperature plot the labels refer to metallicity in solar units; in the metallicity plot to temperature in keV.

quadrature. The XSPEC `flux` command can be used to check the plausibility of some representative results as a test of the method. Observe from Figure 2.2 on the facing page that the fraction of X-ray power being emitted in a single line can be quite high, e.g. peaking around the 10 per cent mark for a solar metallicity for both the O VIII 19 Å line and the Fe XVII 15 Å line. The fraction in a wider energy band encompassing all the thermodynamically relevant energy would of course be lower.

Shown in Figure 2.3 is the fraction of power radiated in *all* emission lines for the 0.1–10 keV X-ray waveband, as a function of temperature and metallicity. Calculation of these curves requires fitting the continuum over the whole energy range of interest, not just at one point. Fourth or fifth order polynomials (in log-log space) were used. These give acceptable results for the X-ray energy range, but not for the wider range including all thermodynamically relevant radiation.

2.3 Metallicity Dependence

In this Section I examine the dependence of various relevant plasma parameters on metallicity.

2.3.1 Intensity and Equivalent Width

It is clear from Figure 2.2 on the facing page that the equivalent width and fractional power of an emission line increase linearly with metallicity at low Z, but eventually saturate at high Z, which is fortunate else eventually more than 100 per cent of the power would be emitted in lines! Recalling that the equivalent width depends on the ratio of the line to the continuum, it is clear that if the line and continuum components both increase by the same factor, there will be no net change in the equivalent width. This is readily verifiable.

The intensity of an emission line obeys $I_{\mathrm{L}} \propto n_{\mathrm{e}} n_{\mathrm{ion}}$. For a metal line, $n_{\mathrm{ion}} \propto Z$. Assuming full ioniza-

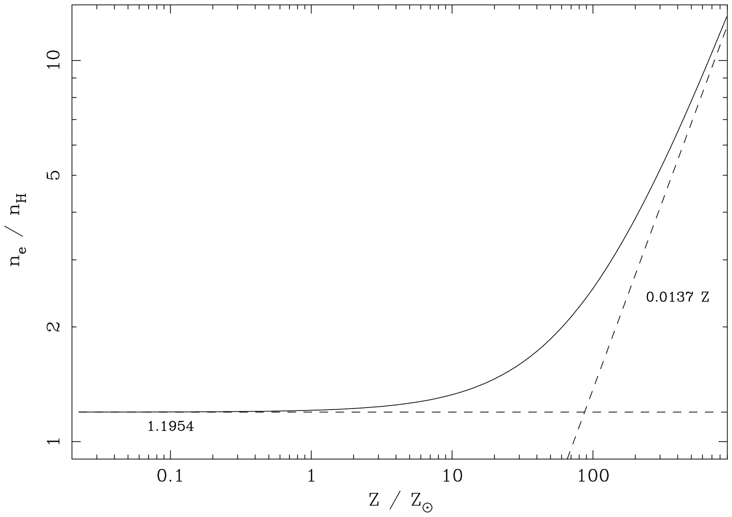


Figure 2.4: Dependence of relative electron density on metallicity for a 1 keV plasma, calculated using MEKAL. Dashed lines show the coefficients of equation (2.3).

tion with fixed H and He density, the electron density is given by

$$n_e = \sum_i Z_i n_i = n_H + 2n_{He} + Z n_H \sum_i Z_i r_i$$
$$\therefore \quad r_e \equiv \frac{n_e}{n_H} = 1 + 2r_{He} + Z \sum_i Z_i r_i = 1.1954 + 0.0137\,Z, \tag{2.3}$$

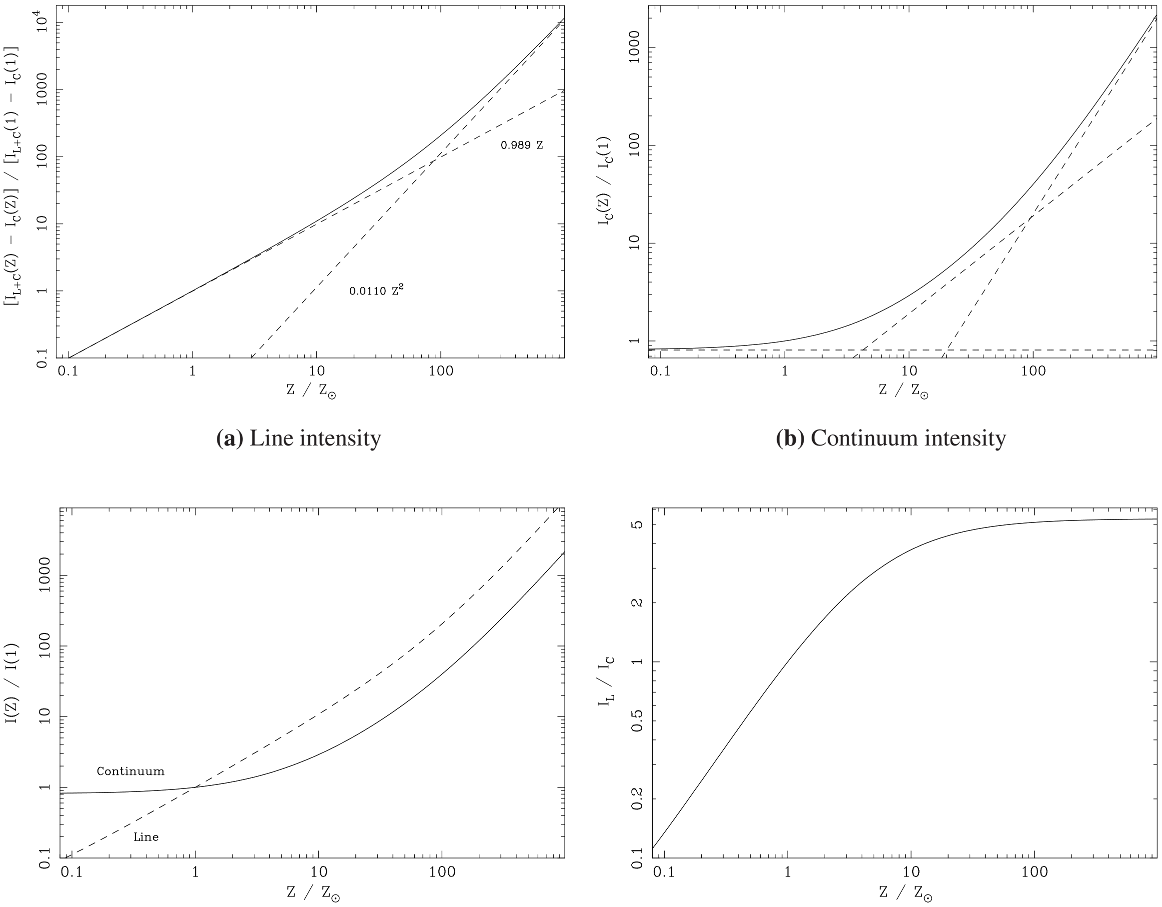
where Z_i and r_i are the atomic number and solar number density relative to hydrogen respectively for species i. Z is the bulk metal abundance, and the solar abundance ratios of Anders and Grevesse (1989) have been used for the final numerical value. Figure 2.4 shows the agreement of the electron density calculated from MEKAL for a 1 keV plasma at various metallicities with this relationship. Using this expression in that for the line intensity leads to

$$\frac{I_L(Z)}{I_L(Z=1)} = \frac{1.1954\,Z + 0.0137\,Z^2}{1.1954 + 0.0137} = 0.9887\,Z + 0.0113\,Z^2, \tag{2.4}$$

where the normalization is taken relative to $Z = 1$ since there is no line component at $Z = 0$ (I_L here refers to the pure line component, i.e. continuum subtracted). Shown in Figure 2.5 on the facing page is the measured change in the intensity of the iron emission line at 0.826 keV with metallicity, showing perfect agreement with equation (2.4).

The bremsstrahlung continuum (Section 1.4.3) due to some species i, on the other hand, behaves according to $I_C \propto Z_i^2 n_i n_e$. Summing the separate continua due to each ion with the common electron density,

$$I_c \propto n_e \sum_i Z_i^2 n_i \propto \left(1 + 2r_{He} + Z \sum_i Z_i r_i\right)\left(1^2 + 2^2 r_{He} + Z \sum_i Z_i^2 r_i\right) \tag{2.5}$$
$$\propto (1.1954 + 0.0137\,Z)(1.3908 + 0.1378\,Z) = 1.663 + (0.1647 + 0.0191)\,Z + 0.0019\,Z^2$$



(a) Line intensity

(b) Continuum intensity

(c) Line and continuum intensity

(d) Line:continuum intensity ratio

Figure 2.5: Metallicity dependence of line and continuum intensity, for a 1 keV plasma. The line intensity plot was produced by subtracting the local continuum to obtain the pure line component. (c) places (a) and (b) on the same graph to facilitate comparison, (d) shows their ratio.

$$\therefore \quad \frac{I_C(Z)}{I_C(Z=1)} = 0.899 + 0.1\,Z + 0.001\,Z^2. \tag{2.6}$$

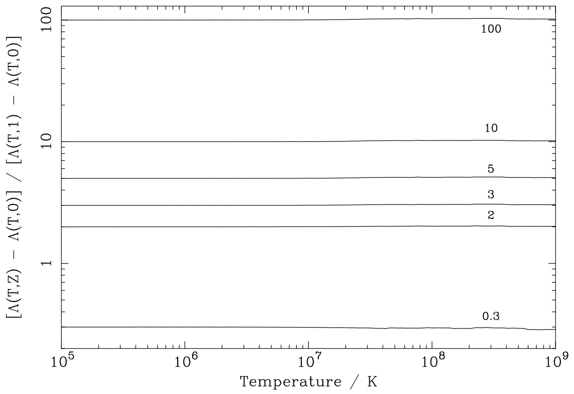
This relationship is also shown in Figure 2.5. The two parts of the Z coefficient are due to electrons from metals and nuclei from metals, respectively.

2.3.2 The Cooling Function

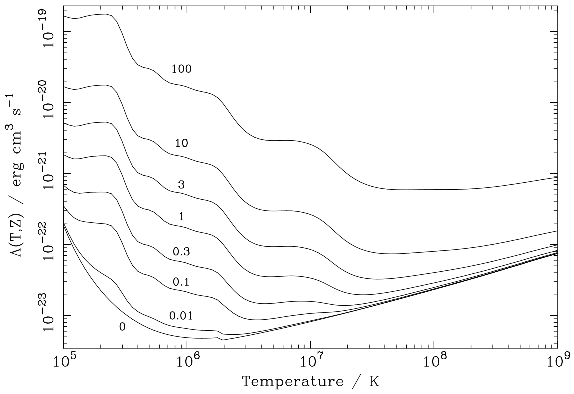
The cooling function can be regarded as having a hydrogen–helium component, and a metal component. The metal component depends linearly on metallicity, so that

$$\Lambda(T,Z) = \Lambda(T,0) + \frac{Z}{\bar{Z}}\left[\Lambda(T,\bar{Z}) - \Lambda(T,0)\right] \tag{2.7}$$

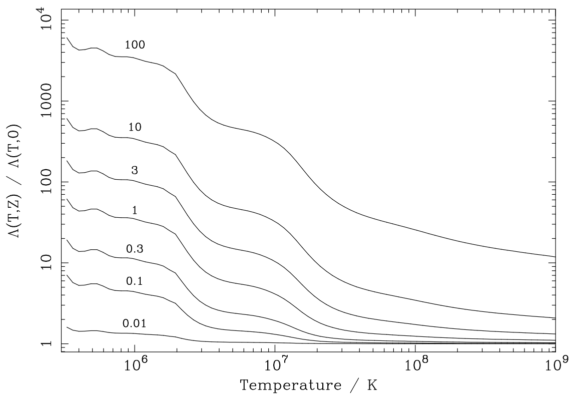
$$\Rightarrow \quad \frac{\Lambda(T,Z) - \Lambda(T,0)}{\Lambda(T,\bar{Z}) - \Lambda(T,0)} = \frac{Z}{\bar{Z}}. \tag{2.8}$$



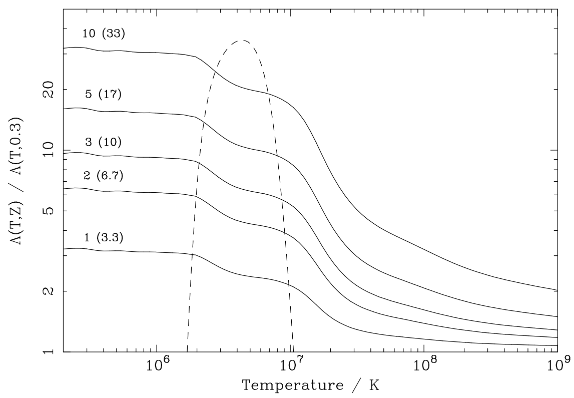
(a) Linear dependence of the metal component on Z



(b) Absolute cooling function



(c) Relative to $Z = 0$



(d) Relative to $Z = 0.3\,\mathrm{Z}_\odot$

Figure 2.6: Various illustrations of the metallicity dependence of the cooling function Λ, calculated from MEKAL spectra. Curve labels are metallicities in solar units. (a) shows how well the linear relationship of equation (2.8) is obeyed in practice. (b) shows the evolution of the absolute cooling function with Z. (c, d) show cooling functions relative to those of 0.0, $0.3\,\mathrm{Z}_\odot$ plasmas, respectively. (d) also shows (dashed curve) the equivalent width of the Fe XVII 15 Å line in arbitrary units. The numbers in parentheses in (d) are the metallicities as multiples of $0.3\,\mathrm{Z}_\odot$. At low temperatures, the curves asymptote to these values.

In this expression, $\bar{Z}$ is any reference metallicity, e.g. $1.0\,\mathrm{Z}_\odot$. Figure 2.6(a), using MEKAL-calculated cooling functions, shows how well this relationship is obeyed in practice.

2.4 Intracluster Medium Thermometers

From Figure 2.2 on page 28 it is clear that the equivalent width curve of the Fe XVII 15 Å line is sharply peaked around 0.3 keV ($\sim 3 \times 10^6$ K), so this emission line is a sensitive indicator of gas at such temperatures. Furthermore, it also accounts for a sizable fraction of the power being radiated in X-rays at such temperatures. Very similar results apply for the Fe XVII 17 Å line as well. The Fe XX lines around 12.8 Å also become strong below ~ 2 keV. All these lines are highly important coolants at temperatures $\lesssim 1$ keV, and therefore their presence (or absence) functions as a thermometer for the ICM in this range (e.g. Tamura *et al.*, 2001). In the standard cooling flow model (in which gas cools from a temperature of several keV to essentially zero), we therefore expect to see a strong flux in these lines.

In contrast, Figure 2.2 on page 28 also shows that the O VIII 19 Å line, for example, is not a good temperature indicator, as it originates from gas at a wide range of temperatures and changes its properties relatively slowly. The 16 Å iron line is also a poor temperature indicator. Although it increases below 2 keV, the presence of an oxygen line in the same region at higher temperatures makes interpretation difficult.

2.5 The Isobaric Cooling Flow Model

I briefly review here the derivation of the standard isobaric cooling flow model (e.g. Johnstone *et al.*, 1992). An isobaric treatment is appropriate because the "flow" in a cooling flow is a slow, subsonic process that passes through a series of near-equilibrium states. Thus, at any instant, the system is in quasi-hydrostatic equilibrium, in which the pressure is determined by the weight of overlying uncooled gas, which is essentially constant. The energy released per unit mass of gas on cooling by $\mathrm{d}T$ is

$$\begin{aligned} \mathrm{d}q &= \mathrm{d}\varepsilon + p\,\mathrm{d}\left(\frac{1}{\rho}\right) = \mathrm{d}\left(\varepsilon + \frac{p}{\rho}\right) && \text{(constant } p\text{)},\\ &= \gamma\,\mathrm{d}\varepsilon = \frac{\gamma}{\gamma-1}\frac{k_\mathrm{B}}{\mu m_\mathrm{H}}\,\mathrm{d}T = \frac{5}{2}\frac{k_\mathrm{B}}{\mu m_\mathrm{H}}\,\mathrm{d}T && \text{(monotonic ideal gas)}, \end{aligned}$$

where use has been made of equations (A.21) and (A.25).

Suppose that gas is flowing through some region, cooling as it does so from T_max to T_min. If the flow is steady, with a constant mass-flow rate $\dot{M}$, then the power released on cooling through $\mathrm{d}T$ is given by

$$\mathrm{d}L = \frac{5}{2}\frac{k_\mathrm{B}}{\mu m_\mathrm{H}}\dot{M}\,\mathrm{d}T. \tag{2.9}$$

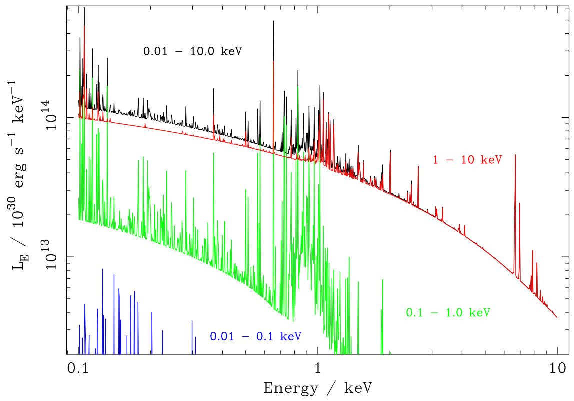
From the definition of the cooling function, equation (1.18), we also have the relations

$$\begin{aligned} \mathrm{d}L &= n_\mathrm{e} n_\mathrm{H} \Lambda(T)\,\mathrm{d}V,\\ \mathrm{d}L_\nu &= n_\mathrm{e} n_\mathrm{H} \Lambda_\nu(T)\,\mathrm{d}V = \frac{\Lambda_\nu(T)}{\Lambda(T)}\,\mathrm{d}L. \end{aligned} \tag{2.10}$$

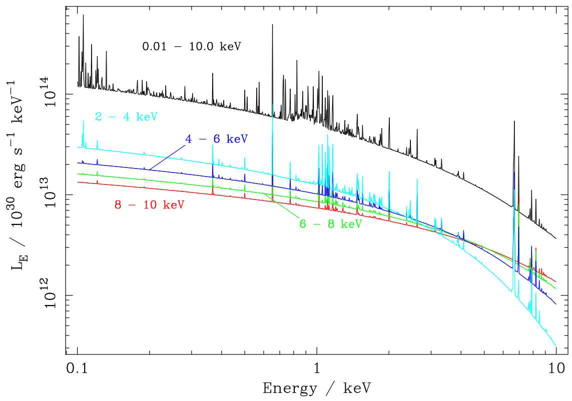
Eliminating $\mathrm{d}L$ and integrating over temperature leads to the canonical expression for the spectral power for a steady-state flow cooling from T_max to T_min,

$$L_\nu = \frac{5}{2}\frac{k_\mathrm{B}}{\mu m_\mathrm{H}}\dot{M}\int_{T_\mathrm{min}}^{T_\mathrm{max}} \frac{\Lambda_\nu(T)}{\Lambda(T)}\,\mathrm{d}T. \tag{2.11}$$

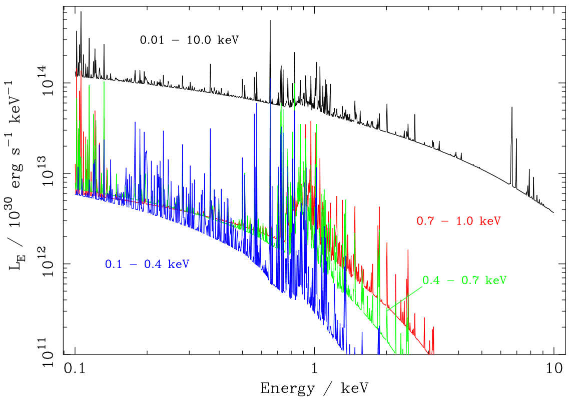
Thus, we have the simple result that the emission measure of each temperature component in the flow is inversely proportional to the cooling function $\Lambda(T)$ at that temperature. The differential emission measure (Section 1.5.2) is shown in Figure 2.8 on page 35. The contributions of various temperature ranges to the overall spectrum are shown in Figure 2.7 on the following page.



(a) Temperature decades



(b) High-temperature gas



(c) Low-temperature gas

Figure 2.7: Contributions of various temperature decades to an `mkcflow`-type cooling flow spectrum. $Z = 0.3\,\mathrm{Z}_\odot$, $\dot{M} = 100\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$, curve labels give T_{min} and T_{max}. The bulk of the continuum emission originates from gas above 1 keV, although cooler gas does make a contribution at low energies.

2.5.1 Practical Implementation

This model is implemented by the XSPEC package as the `mkcflow` and `vmcflow` models, the only difference being that the latter allows specification of the abundances of individual elements (indeed internally the former simply calls the latter). There is also the `cflow` model, which uses Raymond–Smith spectra rather than those of MEKAL, and adds a weighting-factor to the emission measure, after Mushotzky and Szymkowiak (1988). In this model,

$$L_\nu = \frac{5}{2}\frac{k_{\mathrm{B}}}{\mu m_{\mathrm{H}}}\dot{M}\int_{T_{\mathrm{min}}}^{T_{\mathrm{max}}}\left(\frac{T}{T_0}\right)^{\omega}\frac{\Lambda_\nu(T)}{\Lambda(T)}\,\mathrm{d}T, \tag{2.12}$$

with the normalization temperature $T_0 = 10^7\,\mathrm{K}$. The effect of the ω parameter on the emission measure is shown in Figure 2.8 on the facing page.

I draw attention to the following issues in connection with the XSPEC implementation of the cooling flow models. In particular the second point can be significant when working with high metallicities, as shown in Figure 2.9 on page 36.

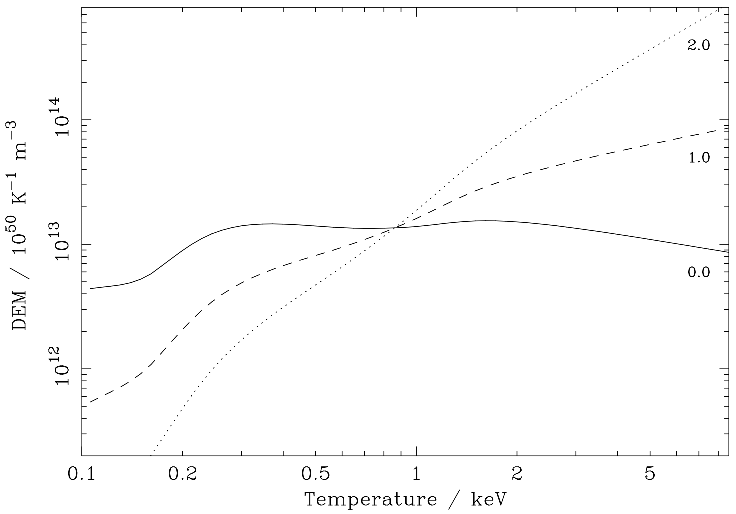


Figure 2.8: Differential emission measure for `cflow`-type models. With ω defined as per equation (2.12), the solid curve has $\omega = 0.0$ (equivalent to an `mkcflow` model), whereas the dashed and dotted curve have $\omega = 1.0, 2.0$ respectively. Other parameters: $\dot{M} = 1.0\,\mathrm{M}_{\odot}$, $T_{\mathrm{min}} = 10^5\,\mathrm{K}$, $T_{\mathrm{max}} = 10^8\,\mathrm{K}$.

- Since equation (2.11) returns a luminosity, and XSPEC uses flux, the normalization factor (defined in `xsvcfl.f`) requires the luminosity distance, and hence a (non-zero) redshift. In older versions of XSPEC, the formula used for the luminosity distance was an approximation for the $q_0 = 0$ case, i.e. it did not take into account the cosmology specified by the `cosmo` command. This was fixed in version 11.2.0b.

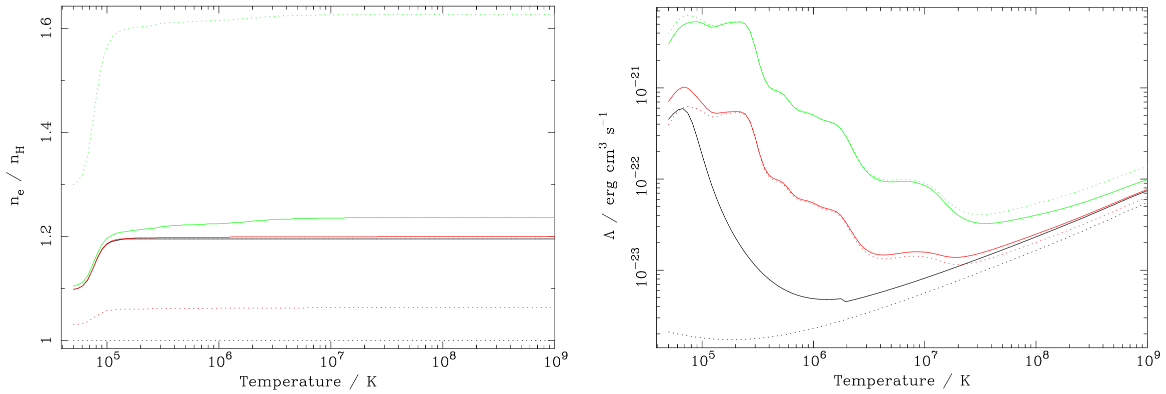
- In some versions of the `mkcflow` model, altering the plasma bulk metallicity varies the helium abundance. This is inconsistent with the other collisional plasma models which fix He to solar (unless all element abundances are free parameters). This can have a non-negligible effect on the calculation of equivalent widths when high metallicities are involved (see Figure 2.9 on the next page). A workaround is to use the `vmcflow` model and specify abundances on an element-by-element basis. This was fixed in version 11.1.0ab.

- At one point, the default upper temperature limits for the `mkcflow` and `vmcflow` models were too high, which, together with some coding errors, could cause segmentation faults when fitting. As a workaround, the upper limit could be manually set to $\sim 60\,\mathrm{keV}$. This was fixed in version 11.1.0k.

- The XSPEC help description for the `mkcflow` and `cflow` models was incorrect at one point, stating that the emission measure was inversely proportional to the cooling *time*, when cooling *function* is the correct description. This was corrected in version 11.2.

The tabulated MEKAL spectra used by XSPEC are stored in the file `mekal.mod`. The energy range runs from 0.005–12 keV in 2400 bins of width 0.005 keV, then from 12–52 keV in 400 bins of width 0.1 keV. The contributions due to each of the 15 elements are stored separately. The spectra are evaluated at 40 logarithmically-spaced temperature points from 0.008–80 keV (10^5–10^9 keV). With single-precision 4-byte reals, this amounts to $\sim 7\,\mathrm{MB}$ in total.

Rather than using the XSPEC implementations, I found it more convenient to write my own, calling MEKAL directly, which allows for greater flexibility. Normally, cooling functions are calculated by in-



(a) Electron density

(b) Cooling function

Figure 2.9: The effect of varying the helium abundance on some bulk plasma properties. Black $Z_{\mathrm{met}} = 0.0\,\mathrm{Z}_\odot$; red $Z_{\mathrm{met}} = 0.3\,\mathrm{Z}_\odot$; green $Z_{\mathrm{met}} = 3.0\,\mathrm{Z}_\odot$. In each case, the solid line has helium at solar abundance and the dotted line has it at the associated metal abundance.

tegrating spectra evaluated at 1000 logarithmically spaced points from 0.005–200 keV. The temperature integral typically involves 100 logarithmically-spaced temperature points and uses the trapezium rule.

2.6 The Spectra of Plasmas with Two Metallicities

The previous sections of this chapter have all implicitly assumed that the radiating plasma is chemically homogeneous. I will now relax this assumption.

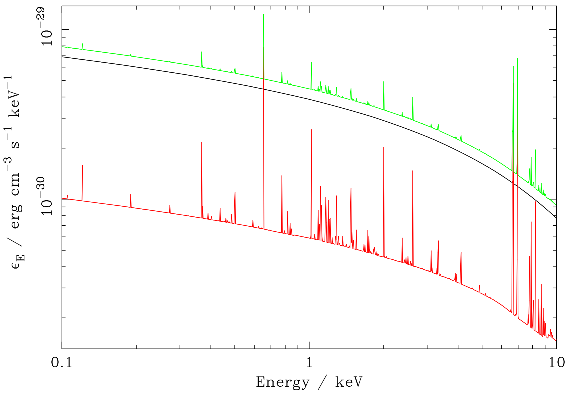
2.6.1 Fixed-temperature Plasmas

First, consider the case of a two-component plasma consisting of metal-rich and metal-poor ‘phases’, at a single, fixed temperature. Note that the usage of the term ‘phase’ here is not necessarily strictly the same as its more traditional application in the context of multi-phase cooling flows (see Chapter 6), because here there is no requirement for the phases to be spatially coincident (see Section 4.1.2), in pressure equilibrium, etc.

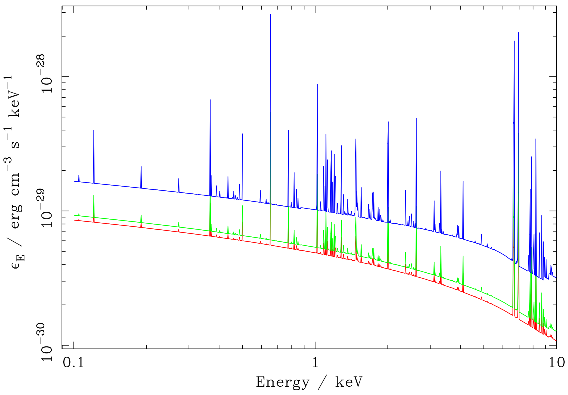
As shown in Figure 2.10(a) on the facing page, such a two-component plasma is spectroscopically indistinguishable from a homogeneous plasma of some mean metallicity $\bar{Z}$,

$$\bar{Z} = f_{\mathrm{hi}} Z_{\mathrm{hi}} + (1 - f_{\mathrm{hi}}) Z_{\mathrm{lo}}; \tag{2.13}$$

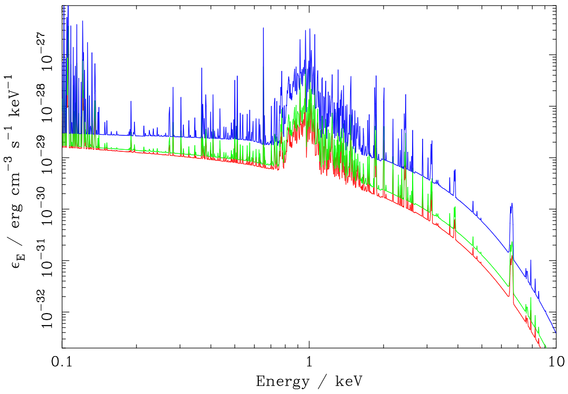
where the metal-rich phase has a metallicity Z_{hi} and accounts for a mass fraction f_{hi}, and the metal-poor phase has a metallicity Z_{lo} and a corresponding mass fraction $f_{\mathrm{lo}} = 1 - f_{\mathrm{hi}}$. For simplicity, but without loss of generality, I will take $Z_{\mathrm{lo}} \equiv 0$ from now on, unless otherwise stated. The value $f_{\mathrm{hi}} = 1$ corresponds to a uniform plasma, and decreasing values of f_{hi} to increasingly segregated plasmas. It is to be understood that the helium abundance is fixed at the solar value, with the heavy elements varying as specified by Z. The indistinguishability of the spectra in these two cases is a simple consequence of the fact that the strength of the continuum radiation is independent of Z, whilst that of the emission lines



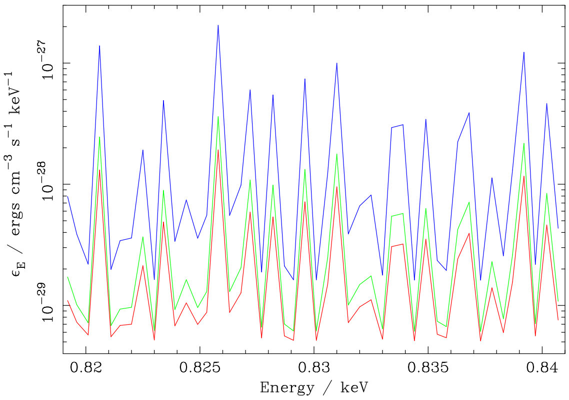
(a) $T = 10\,\mathrm{keV}$, $f_{\mathrm{hi}} = 0.1$, high and low Z components



(b) $T = 10\,\mathrm{keV}$, various f_{hi}



(c) $T = 1\,\mathrm{keV}$, various f_{hi}

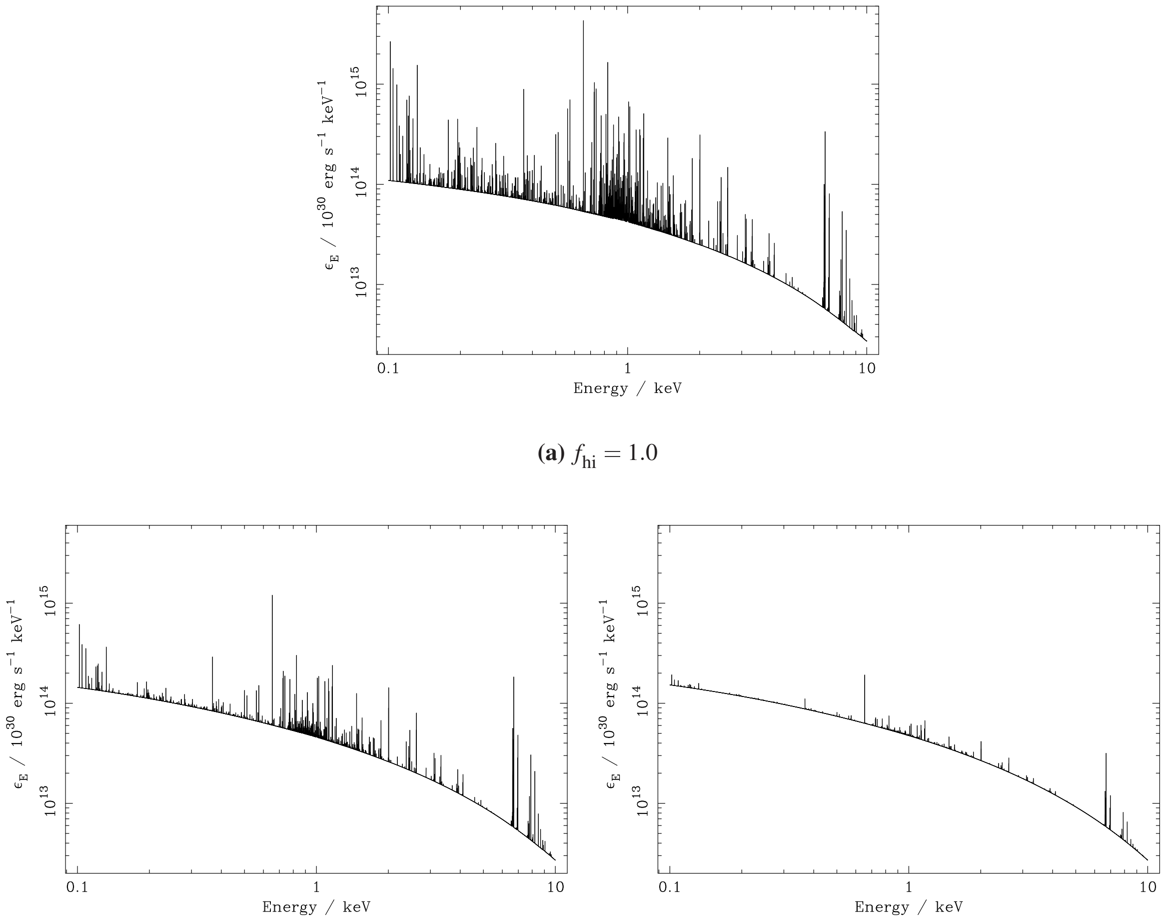


(d) $T = 1\,\mathrm{keV}$ zoom, various f_{hi}

Figure 2.10: Spectra of single-temperature, two-metallicity plasmas. (a) black $Z = 0$, normalization 0.9; red $Z = 3.0\,Z_\odot$, normalization 0.1; green their sum (and coincident with $Z = 0.3\,Z_\odot$, normalization = 1.0). (b, c, d) Two-metallicity plasmas with f_{hi}: 0.1 (red), 0.01 (green), 0.001 (blue). $f_{\mathrm{hi}} = 0.1$ is visually indistinguishable from $f_{\mathrm{hi}} = 1.0$ (i.e. a homogeneous plasma) in each case.

is directly proportional to Z. The reduction in the mass fraction of the enriched phase is therefore offset by the increased line strength. Alternatively, in the coronal limit, the local concentration of the heavy elements makes no difference to their radiation: a given number of heavy ions will radiate in the same way whether uniformly dispersed throughout the emitting volume or concentrated in a specific region.

The simple relationship of equation (2.13) relies on the line-strength increasing linearly with Z (and the continuum being independent of Z), so that making the normalization $\propto Z^{-1}$ leaves the resultant spectrum unchanged. It has already been shown in various ways (e.g. Figure 2.5 on page 31, Figure 2.4 on page 30) that this breaks down at very high metallicities, when the free electrons contributed by the metal ions become significant. This concept is illustrated more explicitly in Figures 2.10(b) to 2.10(d) on the current page. These show the form of spectra constructed according to equation (2.13) for $f_{\mathrm{hi}} = 10^{-1}$, 10^{-2}, 10^{-3}. The $f_{\mathrm{hi}} = 10^{-1}$ case is coincident with that for $f_{\mathrm{hi}} = 1.0$ (i.e. a uniform plasma, which is therefore not shown), but those for $f_{\mathrm{hi}} = 10^{-2}$ and especially $f_{\mathrm{hi}} = 10^{-3}$ are noticeably different. This does not invalidate the idea that a mixture of a metal-rich and a metal-poor plasma can look identical to a uniform-metallicity plasma — it would merely change the dependence of the normalization of each



(a) $f_{\rm hi} = 1.0$

(b) $f_{\rm hi} = 0.1$

(c) $f_{\rm hi} = 0.01$

Figure 2.11: Spectra of two-metallicity, isobaric cooling flows. Note the slight increase in the low-energy continuum as $f_{\rm hi}$ is reduced. Common parameters: $T_{\rm min} = 10^6$ K, $T_{\rm max} = 10^8$ K, 100 logarithmically-spaced temperature points; 10000 logarithmic spectral bins; $\dot{M} = 100\,{\rm M}_\odot\,{\rm yr}^{-1}$.

component on Z from the simple linear relationship of equation (2.13) to something more complicated. Such ultra-high metallicities are of no relevance here though, since in practice we shall not make use of $f_{\rm hi} \lesssim 0.1$, anything less than this probably representing too extreme a situation to be relevant to the ICM.

2.6.2 Cooling Plasmas

The situation becomes more complicated (and interesting) when a two-component metallicity plasma is allowed to cool. Figure 2.11 illustrates how the spectrum of a two-component cooling flow, in which one component has a metallicity $Z_{\rm hi}$ and accounts for a fraction $f_{\rm hi}$ of the total mass-flow rate, and the other component has zero metals, varies with the inhomogeneity parameter $f_{\rm hi}$. The obvious effect of decreasing $f_{\rm hi}$ is to suppress the emission lines, more so at the low-energy end of the spectrum. It is also possible to note a slight increase in the low-energy continuum.

I now make some quantitative analysis of the changes in the spectrum, using the equivalent width, equation (2.1), as a convenient tool to measure the relative strength of emission lines, independent of the continuum. For fixed, narrow line profiles, the integral over energy can be neglected in favour of simply comparing intensity ratios at the line energy. Making use of the isobaric cooling flow equation (2.11), the equivalent width, EW(H), for a given spectral line (with a narrow energy profile centred on some energy E) from a cooling flow of some homogeneous metallicity $\bar{Z}$ can be written as

$$\mathrm{EW(H)} = \frac{\displaystyle\int_{T_{\min}}^{T_{\max}} \frac{\Lambda_{\mathrm{L}}(T,E,\bar{Z})}{\Lambda(T,\bar{Z})}\,\mathrm{d}T}{\displaystyle\int_{T_{\min}}^{T_{\max}} \frac{\Lambda_{\mathrm{C}}(T,E,\bar{Z})}{\Lambda(T,\bar{Z})}\,\mathrm{d}T}. \tag{2.14}$$

In this expression, I explicitly include the metallicity dependence of the cooling function, $\Lambda = \Lambda(T,Z)$. Furthermore,

$$\Lambda(T,Z) = \int \Lambda(T,E,Z)\,\mathrm{d}E. \tag{2.15}$$

$\Lambda_{\mathrm{L}}(T,E,Z)n_{\mathrm{e}}n_{\mathrm{H}}$ would be the spectral power per unit volume per unit energy radiated by the line component at a given energy, with Λ_{C} having an analogous meaning for the continuum component. The total cooling function is the sum of the line and continuum components, $\Lambda = \Lambda_{\mathrm{L}} + \Lambda_{\mathrm{C}}$.

Defining I as power per unit volume, then at a fixed hydrogen density n_{H}, $\Lambda \propto I/n_{\mathrm{e}}$, from the definition of the cooling function, equation (1.18). For the line component, $I_{\mathrm{L}} \propto n_{\mathrm{e}}n_{\mathrm{ion}}$, as discussed in Section 2.3, so that $\Lambda_{\mathrm{L}} \propto Z$. To a first approximation, the continuum component Λ_{C} is independent of Z, provided very high metallicities are not encountered (see Figure 2.5 on page 31).

Consider now the two-component cooling flow. For the prototypical narrow spectral line centred at energy E, the line component is obviously only contributed by the metal-rich component, whereas both components contribute to the continuum. The corresponding expression for the equivalent width, EW(B), of the line is thus

$$\mathrm{EW(B)} = \frac{f_{\mathrm{hi}}\displaystyle\int_{T_{\min}}^{T_{\max}} \frac{\Lambda_{\mathrm{L}}(T,E,Z_{\mathrm{hi}})}{\Lambda(T,Z_{\mathrm{hi}})}\,\mathrm{d}T}{f_{\mathrm{hi}}\displaystyle\int_{T_{\min}}^{T_{\max}} \frac{\Lambda_{\mathrm{C}}(T,E,Z_{\mathrm{hi}})}{\Lambda(T,Z_{\mathrm{hi}})}\,\mathrm{d}T + (1-f_{\mathrm{hi}})\displaystyle\int_{T_{\min}}^{T_{\max}} \frac{\Lambda_{\mathrm{C}}(T,E,0)}{\Lambda(T,0)}\,\mathrm{d}T}. \tag{2.16}$$

With $f_{\mathrm{hi}} \to 1$, $Z_{\mathrm{hi}} \to \bar{Z}$, this obviously reduces to equation (2.14).

Our interest lies in the ratio of the equivalent widths Δ,

$$\begin{aligned}\Delta &\equiv \frac{\mathrm{EW(B)}}{\mathrm{EW(H)}} \\ &= f_{\mathrm{hi}}\frac{\displaystyle\int_{T_{\min}}^{T_{\max}} \frac{\Lambda_{\mathrm{L}}(T,E,Z_{\mathrm{hi}})}{\Lambda(T,Z_{\mathrm{hi}})}\,\mathrm{d}T}{\displaystyle\int_{T_{\min}}^{T_{\max}} \frac{\Lambda_{\mathrm{L}}(T,E,\bar{Z})}{\Lambda(T,\bar{Z})}\,\mathrm{d}T}\;\frac{\displaystyle\int_{T_{\min}}^{T_{\max}} \frac{\Lambda_{\mathrm{C}}(T,E,\bar{Z})}{\Lambda(T,\bar{Z})}\,\mathrm{d}T}{f_{\mathrm{hi}}\displaystyle\int_{T_{\min}}^{T_{\max}} \frac{\Lambda_{\mathrm{C}}(T,E,Z_{\mathrm{hi}})}{\Lambda(T,Z_{\mathrm{hi}})}\,\mathrm{d}T + (1-f_{\mathrm{hi}})\displaystyle\int_{T_{\min}}^{T_{\max}} \frac{\Lambda_{\mathrm{C}}(T,E,0)}{\Lambda(T,0)}\,\mathrm{d}T},\end{aligned} \tag{2.17}$$

which we hope to suppress through a partition of the ICM metals into two phases. One possible way to make progress here would be to assume some (approximate) analytic form for the line profile $\Lambda_{\mathrm{L}}(E)$ and the cooling function $\Lambda(T)$. Instead, I will assume that the emission line of interest is one originating from gas over only a narrow temperature range (from Figure 2.2 on page 28, this is not too unreasonable an assumption for the temperature-diagnostic lines). If so, we can make the simplifying approximation

$$\int \frac{\Lambda_{\rm L}(T)}{\Lambda(T)}\,{\rm d}T \rightarrow \frac{\Lambda_{\rm L}(\bar{T})}{\Lambda(\bar{T})}\delta T. \tag{2.18}$$

From the relationship $\Lambda_{\rm L} \propto Z$,

$$\frac{\Lambda_{\rm L}(\bar{T},E,Z_{\rm hi})}{\Lambda_{\rm L}(\bar{T},E,\bar{Z})} = \frac{Z_{\rm hi}}{\bar{Z}} = \frac{1}{f_{\rm hi}}, \tag{2.19}$$

also using equation (2.13) with $Z_{\rm lo} = 0$.

With these two results, equation (2.17) reduces to

$$\Delta \approx \frac{\Lambda(T,\bar{Z})}{\Lambda(T,Z_{\rm hi})}\left[\frac{\displaystyle\int_{T_{\rm min}}^{T_{\rm max}} \frac{\Lambda_{\rm C}(T,E,\bar{Z})}{\Lambda(T,\bar{Z})}\,{\rm d}T}{\displaystyle f_{\rm hi}\int_{T_{\rm min}}^{T_{\rm max}} \frac{\Lambda_{\rm C}(T,E,Z_{\rm hi})}{\Lambda(T,Z_{\rm hi})}\,{\rm d}T + (1-f_{\rm hi})\int_{T_{\rm min}}^{T_{\rm max}} \frac{\Lambda_{\rm C}(T,E,0)}{\Lambda(T,0)}\,{\rm d}T}\right]. \tag{2.20}$$

The $[\cdots]$ term in this equation represents the ratio of the continuum in the homogeneous case to that in the two-component case. By inspection of cooling flow spectra (Figure 2.11 on page 38), this ratio is very close to one. Two-metallicity plasmas tend to have fractionally stronger continua at low ($\lesssim 1\,$keV) energies. Figure 2.7 on page 34 shows that for $E \gtrsim 1\,$keV, the continuum is almost entirely due to gas with $T \gtrsim 1\,$keV. In this regime, the cooling function has only a weak dependence on Z, as shown in Figure 2.6(b) on page 32. Consequently, the denominator in the $[\cdots]$ term is almost identical to the numerator, and the continuum is unchanged. One has to move to a very high value of $Z_{\rm hi}$ before Z has a significant effect on Λ in this temperature range. By then, of course, $f_{\rm hi}$ is small, so that almost all the continuum is coming from the metal-poor gas anyway.

Below energies of 1 keV or so, there is a small contribution to the continuum from gas below 1 keV (Figure 2.7 on page 34). The same argument as before shows that the continuum component from hotter gas is still unchanged in this energy range. For the gas below 1 keV, the $f_{\rm hi}/\Lambda(T,Z_{\rm hi})$ term (i.e. the continuum contribution from metal-rich gas) can be neglected, because $\Lambda(T,Z_{\rm hi}) \gg \Lambda(T,0)$ in this range. Alternatively, if the metals dominate the cooling then

$$\Lambda(T,Z_{\rm hi}) \sim \frac{Z_{\rm hi}}{\bar{Z}}\Lambda(T,\bar{Z}) = \frac{1}{f_{\rm hi}}\Lambda(T,\bar{Z}), \tag{2.21}$$

so that the metal-rich continuum term is roughly proportional to $f_{\rm hi}^2$ (this also justifies ignoring the term proportional to Z in the expression for the continuum intensity).

Similarly, the metal-poor gas at low temperatures has an enhanced continuum relative to the homogeneous case, because $\Lambda(T,0) < \Lambda(T,\bar{Z})$. The amount by which this is true depends on $\bar{Z}$, not $f_{\rm hi}$, and cannot be predicted in advance. From Figure 2.6(c) on page 32, $\Lambda(T,0.3) \sim 10\Lambda(T,0)$ at low temperatures. In summary, the metal-rich gas makes little contribution to the low-temperature continuum because its enhanced cooling function reduces its emission measure in this range. The metal-poor gas, on the other hand, has a reduced cooling function at low temperatures, and this boosts its emission measure relative to the mean metallicity case.

To a good approximation, any changes in the continuum for the two-metallicity case can therefore be

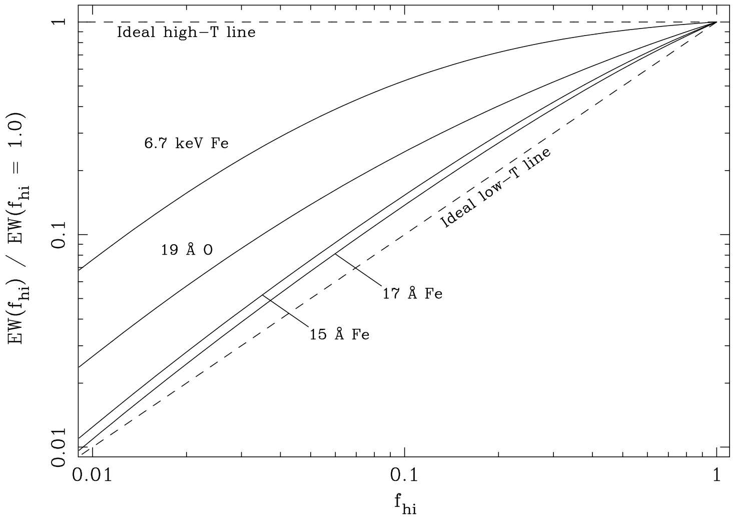


Figure 2.12: Equivalent widths (relative to the homogeneous case $f_{\mathrm{hi}} = 1.0$) of some important ICM emission lines as a function of the mass-fraction of the enriched phase, f_{hi}, for a two-component isobaric cooling flow. Parameters of the cooling flow model: $T_{\mathrm{min}} = 10^6$ K, $T_{\mathrm{max}} = 10^9$ K, 100 logarithmically-spaced temperature points; $\bar{Z} = 0.3$. The two dashed lines show the theoretical behaviour derived in Section 2.6.2 for the limit of a high-temperature ($y = 1$) and low-temperature ($y = x$) line. Approximate values at $f_{\mathrm{hi}} = 0.1$ are: 0.53 (6.7 keV Fe); 0.25 (19 Å O); 0.15 (15 Å Fe); 0.13 (17 Å Fe).

ignored. The expression for the equivalent width ratio Δ for a narrow emission line originating from gas over a small temperature range therefore reduces to the simple form

$$\Delta \approx \frac{\Lambda(T,\bar{Z})}{\Lambda(T,Z_{\mathrm{hi}})}.$$

We can consider two limiting cases:

- A high-temperature emission line. In this range, the cooling is approximately independent of metallicity, so that $\Lambda(T,Z_{\mathrm{hi}}) \sim \Lambda(T,\bar{Z}) \sim \Lambda(T,0)$. Therefore $\Delta \to 1$, i.e. there is no change in the equivalent width of a high-temperature line.

- A low-temperature emission line. In the regime where the cooling is dominated by metals, equation (2.21) implies that $\Delta \to f_{\mathrm{hi}}$.

In summary:

$$\Delta \approx \frac{\Lambda(T,\bar{Z})}{\Lambda(T,Z_{\mathrm{hi}})} \sim \begin{cases} 1 & \text{high } T \\ \dfrac{\bar{Z}}{Z_{\mathrm{hi}}} = f_{\mathrm{hi}} & \text{low } T \end{cases} \qquad (2.22)$$

2.7 Discussion

The value $f_{\mathrm{hi}} = 0.1$ corresponds to $Z_{\mathrm{hi}} \approx 3\,\mathrm{Z}_\odot$ for $\bar{Z} = 0.3$. This may seem like an excessively high metallicity, but consider that supernovae can be viewed as sources of essentially infinite metallicity. The *Chandra* observation (Iwasawa *et al.*, 2001) of 4C 55.15 reveals a metallicity $\approx 2\,\mathrm{Z}_\odot$ within the central

50 kpc, as well as a metal-rich 'plume' extending over ~ 25 kpc with $Z \approx 8\,Z_\odot$. Whilst this is clearly an extreme case, if such high-metallicity features can be produced on these scales, there is no reason they cannot occur on the much smaller scales I consider.

In practice, for lines existing at intermediate temperatures, or over a non-negligible range of temperatures, we expect behaviour intermediate between the limits derived at the end of Section 2.6.2. Figure 2.12 on the preceding page shows how the equivalent widths of some important ICM emission lines vary for a two-component plasma as a function of the degree of inhomogeneity f_{hi}. The actual curves (solid) are indeed contained between the theoretical limits (dashed).

The iron K line, for example, deriving as it does from gas over a wide, high temperature range (see Figure 4.7 on page 96), experiences a relatively low degree of suppression: when $f_{\mathrm{hi}} = 0.1$, the equivalent width is reduced to around half that of a homogeneous plasma. In contrast, the 15 and 17 Å lines of Fe XVII undergo a greater degree of suppression, approaching (but not reaching) the ideal of a narrow, low-temperature emission line. At $f_{\mathrm{hi}} = 0.1$, these equivalent widths are reduced to around 0.15 and 0.13 (respectively) of the homogeneous values.

This behaviour is consistent with the results of Section 2.6.2. From Figure 2.6(d) on page 32, we see that in the temperature range where the 15 Å line is strong, $\Lambda(T, 3.0) \approx 7\Lambda(T, 0.3)$. Equation (2.22) would therefore predict an equivalent width ratio of $1/7 \approx 0.14$ for the case $Z_{\mathrm{hi}} = 3.0$, $\bar{Z} = 0.3$, i.e. $f_{\mathrm{hi}} = 0.1$.

Due to its strength and the fact that the atomic physics of the K shell is well understood, the iron K line has been used as the basis of many measurements of the ICM "metallicity". Suppose that the ICM metals *are* actually inhomogeneous on small scales, and that a two-component plasma with $f_{\mathrm{hi}} = 0.1$ is a suitable description. The above results imply that analyses not taking this into account would have underestimated the mean ICM metallicity by a factor of two.

Note that the *relative* suppression at $f_{\mathrm{hi}} = 0.1$ of the Fe XVII 15 Å line (say) compared to that of the Fe K line, is only a factor of $0.15 \div 0.53 \approx 0.3$. When presented with an observed spectrum, the actual metallicity is of course unknown and must be fitted for. If the iron K complex dominates the fit, then for a two-component plasma with $f_{\mathrm{hi}} = 0.1$, then the observed Fe XVII 15 Å line will appear to be around one third (not one tenth) of the value that a homogeneous model would predict. This is of course appreciably less than the simple 'factor of ten' reduction that might have been expected.

In other words, because the prominent emission lines of the ICM do not divide neatly up into "high" and "low" temperature lines (where high indicates an origin in a temperature range where cooling is independent of metallicity, and low an origin in the range where metals dominate the cooling), the actual degree of suppression that might be obtained through small-scale metallicity variations is not as great as might be expected at first sight. Indeed, if Z_{lo} and Z_{hi} are taken to represent the extremes of a continuous metallicity distribution, then the reduction taking into account the intermediate abundances would be even milder. See Section 4.6 for discussion of another factor that would mitigate the actual suppression even more.

In the next Chapter, I describe a numerical scheme that can be applied to simulations of the ICM in general, before going on in Chapter 4 to apply it specifically to the case of small-scale metallicity variations.

On two occasions I have been asked [by members of Parliament], "Pray, Mr Babbage, if you put into the machine wrong figures, will the right answers come out?" I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.

CHARLES BABBAGE

3 Numerical Models of the Intracluster Medium

Although the ICM is in some respects undoubtedly a complex entity, one can make a great deal of progress towards an understanding of it with comparatively straightforward methods. I describe in this Chapter a simple numerical scheme that can be used to model the ICM, its behaviour, evolution, and X-ray emission. The fact that the model itself is not too complicated makes it relatively easily to focus on the physical processes at work (as opposed to the details of the computational implementation), and to obtain output that can easily be compared with genuine observational data. I begin with some brief discussion of the equations of fluid dynamics that govern the ICM, and mention some of the key ideas of the numerical methodology as it applies to such equations. Then the actual numerical scheme is presented, in terms of the steps necessary to follow the evolution of the physical variables concerned. A process by which simulated observational data can be obtained is described. In this Chapter, I confine myself to a discussion of the workings of the scheme — examples of the outputs will be presented in later Chapters. Discussion of some possible ways in which the model might be extended and improved is made in Chapter 7.

3.1 The Equations of Gasdynamics

The equations of *gasdynamics* (compressible, inviscid fluid flow), like those of electromagnetism, allow for a rich expression of behaviours, yet have a concise mathematical formulation (e.g. Landau and Lifshitz, 1987):

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho\nabla\cdot\boldsymbol{u} = 0 \qquad \text{continuity equation} \qquad (3.1)$$

$$\rho\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\nabla p + \boldsymbol{F} \qquad \text{motion equation} \qquad (3.2)$$

$$\frac{\mathrm{D}\varepsilon}{\mathrm{D}t} + p\frac{\mathrm{D}\mathcal{V}}{\mathrm{D}t} = \mathcal{Q} \qquad \text{energy equation} \tag{3.3}$$

with the *total derivative* $\frac{\mathrm{D}}{\mathrm{D}t}$, defined by

$$\frac{\mathrm{D}}{\mathrm{D}t} \equiv \frac{\partial}{\partial t} + \boldsymbol{u} \cdot \nabla, \tag{3.4}$$

expressing the total rate of change of a fluid property. This has two components, one due to purely temporal evolution, the other due to the motion of fluid particles through space in accordance with the fluid velocity field $\boldsymbol{u}$. The origin of these equations is outlined in Appendix A.5.

These five equations (two scalar, one vector) represent the conservation of mass, momentum, and energy of fluid elements throughout the course of a flow. ρ, p, ε, and $\mathcal{V}$ are the fluid mass density, pressure, and specific (per unit mass) internal energy and volume respectively. Body forces (such as a gravity field) are represented by $\boldsymbol{F}$ (per unit volume), and sources or sinks of heat by $\mathcal{Q}$ (per unit mass). Viscous dissipation and thermal conduction are neglected for the present, though the latter issue is visited in Chapter 5. Viscosity is mentioned in the context of the pseudoviscosity in Section 3.3.4.

With any heat sources $\mathcal{Q}$ and body forces $\boldsymbol{F}$ known, the five equations (3.1), (3.2), and (3.3) express the behaviour of the five fluid variables ρ, $u_{1,2,3}$, and p as functions of time and position. The thermodynamic properties of the fluid are assumed to be known:

$$\varepsilon = \varepsilon(p, \rho). \tag{3.5}$$

The internal energy is often expressed in terms of the density and temperature, meaning that the fluid *equation of state* (relating the pressure, density, and temperature) is also required. For example, the ideal gas (see Appendix A.6) equation of state

$$p = k_{\mathrm{B}} n T, \tag{3.6}$$

with n the fluid number density.

3.1.1 Eulerian and Lagrangian Formulations

For mathematical purposes, it is often convenient to speak of fluid “particles”, or fluid packets. Two conceptual frameworks are useful when dealing with fluid flow. In the *Eulerian* picture, the motion of the fluid is considered with respect to a fixed external reference frame, and the relevant independent variables are position $\boldsymbol{r}$ and time t. In the *Lagrangian* picture, a frame of reference which moves with the fluid particles is adopted. Each fluid particle can then be specified uniquely by a *Lagrangian co-ordinate*; for example, the position of that particle at the start of the time of interest. The relevant independent variables are time and the Lagrangian co-ordinate. The spatial position of the fluid particles must be solved for. In Lagrangian co-ordinates the time derivative is equivalent to the total derivative.

The two treatments are of course mathematically equivalent, but very different in terms of implementation when solving numerically. The Lagrangian approach provides extra information, in that it is possible to say where each fluid particle originated. This is especially useful when dealing with flows containing more than one fluid. The Lagrangian approach is less well suited in general to problems in two- or three-dimensions that the Eulerian method. Firstly, the transformed fluid equations become significantly more complex. Secondly, a fixed reference frame is more easily handled in the two- and three-dimensional case, as opposed to a Lagrangian one which deforms with the fluid flow.

3.2 Basic Numerical Methods for Differential Equations

Problems involving the solution of differential equations can be classified into two broad categories. In an *initial-value problem*, the state of the system is known completely at the initial time, and then must be solved for at future times. This is to be contrasted with a *boundary-value problem*, in which a static solution subject to specified boundary conditions is to be sought.

A common technique for the numerical solution of differential equations is the method of *finite differences*, in which the independent variables are discretized on a grid of some form. Derivatives are then approximated by differences calculated using neighbouring grid points. There are other methods for the solution of such equations, for example Monte Carlo methods, but the finite-difference approach is the one adopted here (see e.g. Press *et al.*, 1990; Ames, 1992, for more details).

The simplest family of finite-difference solutions for differential equations are the Euler methods, of which there are three basic varieties — Forward, Backward, and Centred. For the purposes of illustration, consider the equation

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t,y). \tag{3.7}$$

The finite-difference solution is obtained by discretizing the independent variable t to $t_0, t_0+\Delta t, t_0+2\Delta t, \cdots$. The accuracy of each approximation can be obtained using the Taylor series expansion of the points in the vicinity of some general point t_n.

$$\begin{aligned} y_{n+1} &= y_n + \Delta t y'_n + {}^1\!/_2 \Delta t^2 y''_n + \mathrm{O}\left(\Delta t^3\right) \\ y_n &= y_n \\ y_{n-1} &= y_n - \Delta t y'_n + {}^1\!/_2 \Delta t^2 y''_n + \mathrm{O}\left(\Delta t^3\right) \end{aligned}$$

Forward Euler

$$y'_n = \frac{y_{n+1} - y_n}{\Delta t} \tag{3.8}$$

$$y_{n+1} = y_n + \Delta t f(t_n, y_n) \tag{3.9}$$

By inspection of the Taylor series expansions, the error in the approximation for y'_n is of order Δt^2. Hence this method is first order accurate. It is also *explicit*, in that y_{n+1} is calculated from quantities known at time t_n.

Backward Euler

$$y'_n = \frac{y_n - y_{n-1}}{\Delta t} \tag{3.10}$$

$$y_{n+1} = y_n + \Delta t f(t_{n+1}, y_{n+1}) \tag{3.11}$$

Since the right-hand side of the evolution equation involves quantities evaluated at time t_{n+1}, it is not in general possible to give an explicit equation for y_{n+1} in terms of known quantities at time t_n, although it may be possible in some specific cases. Hence this is an *implicit* method. The accuracy is again first order.

Centred Euler

$$y'_n = \frac{y_{n+1} - y_{n-1}}{2\Delta t} \tag{3.12}$$

To avoid the inconvenient factor of two, this is often used in the form

$$y'_n = \frac{y_{n+\frac{1}{2}} - y_{n-\frac{1}{2}}}{\Delta t}. \tag{3.13}$$

From the Taylor series expansions about y_{n+1} and y_{n-1}, it is clear that subtracting these two series leads to cancellation of the Δt^2 term. Thus the centred Euler (or Crank–Nicholson) method is a second order one. There is a wide range of finite-difference methods (e.g. Press *et al.*, 1990) of varying complexities: Runge–Kutta; leapfrog; etc.

3.2.1 Stability of Finite-difference Schemes

Consider again the prototypical differential equation to be solved through the use of a finite-difference scheme, equation (3.7). Inevitably, the discretization of the solution will introduce small errors. This is acceptable, provided that these errors do not grow to overwhelm the true solution; else the numerical results will rapidly become meaningless. These small disturbances (z, say) will obey the approximate relationship

$$z' = -\Gamma z, \qquad \text{where} \quad \Gamma \equiv -\frac{\partial f(t,y)}{\partial y}.$$

For the purpose of investigating the stability of small disturbances, Γ can be treated as a constant. Consider the case $\Gamma > 0$, so that the true solution is stable (i.e. bounded) as $t \to \infty$. The numerical scheme adopted to solve the equation may not preserve this stability.

The forward Euler scheme for this equation has

$$z_{n+1} = z_n + \Delta t z'_n = (1 - \Gamma\Delta t) z_n = \cdots = (1 - \Gamma\Delta t)^{n+1} z_0.$$

Stability in the limit $n \to \infty$ thus requires $|1 - \Gamma\Delta t| < 1$. For $\Gamma > 0$, this implies $\Gamma\Delta t < 2$. This is usually quoted in the form

$$\Delta t < \frac{2}{\Gamma},$$

which makes it apparent that there is a maximum size of time step that can be taken with the forward Euler method if the solution is not to become unstable.

The backward Euler solution for this case has the form

$$z_{n+1} = z_n + \Delta t z'_{n+1} = z_n - \Gamma \Delta t z_{n+1},$$
$$\therefore \quad z_{n+1} = \frac{z_n}{1+\Gamma\Delta t} = \cdots = \frac{z_0}{(1+\Gamma\Delta t)^{n+1}}.$$

In this instance, the stability requirement is $|1+\Gamma\Delta t| > 1$. For $\Gamma > 0$, this is always satisfied, so the backward Euler method is unconditionally stable.

This illustrates the general point that implicit schemes are on the whole more stable than their explicit counterparts, so that it is possible to take larger time steps. The cost of computing each time step, though, is usually greater in the implicit case.

It can be shown (e.g. Press *et al.*, 1990) that in the case of a wave equation

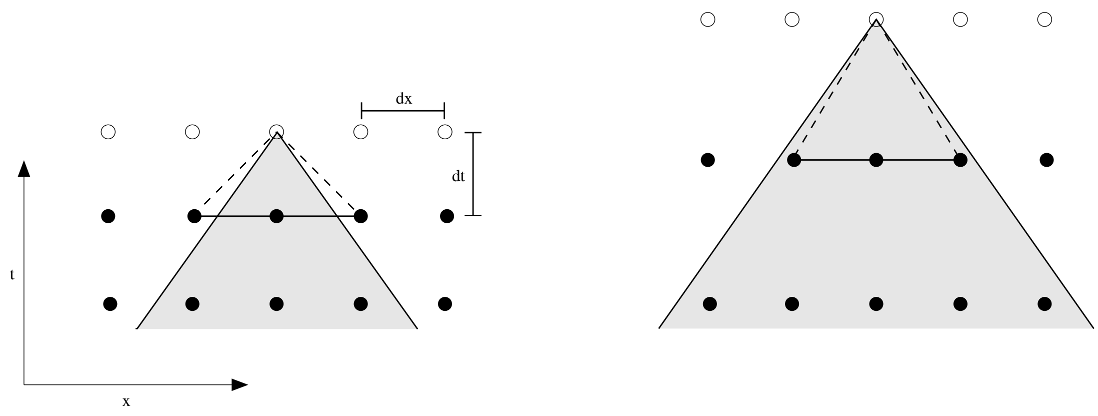
$$\frac{\partial y}{\partial t} = -c_{\mathrm{w}} \frac{\partial y}{\partial x}, \tag{3.14}$$

the stability criterion for an explicit scheme can be written in the form

$$\mathcal{C} \equiv c_{\mathrm{w}} \frac{\Delta t}{\Delta x} < 1, \quad \text{i.e.} \tag{3.15}$$
$$\Delta t < \frac{\Delta x}{c_{\mathrm{w}}} = t_{\mathrm{cross}}, \tag{3.16}$$

where c_{w} is the wave speed, and t_{cross} is the time for the wave to cross a region of width Δx. $\mathcal{C}$ is the *Courant number*, and this is the Courant–Friedrichs–Lewy stability criterion, usually referred to as the *Courant condition*. It states that the size of the time step must be less than the time it takes sound to cross the spatial zones. The sound speed represents the speed at which information can propagate through the fluid. Thus the physical conditions at any point in the fluid can be influenced only by those points which lie within the "sound cone" of that point, i.e. from which there has been time for information to arrive. The larger the time step is made, the larger the region of the fluid which can affect the updated conditions. Any differencing scheme, however, only uses a finite number of grid points to calculate the updated properties of each fluid particle. If the region encompassed by these points includes the sound cone as a subset, then the numerical system will be stable. If not, then the system will be unstable, because not all relevant physical information is captured by the numerical scheme. These ideas are illustrated in Figure 3.1 on the next page.

Qualitatively, the stability time prevents one from being too demanding of a numerical simulation. It shows that it is not possible to increase the spatial resolution (i.e. reduce Δx) without also increasing the time resolution (reducing Δt). Finite-difference schemes operate in violation of the principles of capitalism, in that the more one tries to buy (i.e. the greater the spatial resolution one tries to obtain), the more one has to pay (the greater the number of time steps that must be taken to evolve through a fixed



(a) Stable **(b)** Unstable

Figure 3.1: Representation of the Courant condition for the stability of a finite-difference scheme. Filled circles represent known grid points; open circles are future grid points to be solved for through advancing the numerical solutions. Filled circles connected by a solid line are those used in computing (dashed line) the quantity at the next time step. The shaded region represents the "sound cone" of the fluid, that is, the region bounded by lines of slope $\Delta x/\Delta t = c$. If the sound cone is contained within the region of numerical dependency, the scheme is stable. If not, as is the case if the time step is made too large, then the scheme is unstable, as some relevant information is omitted from the difference equations. Adapted from Press *et al.* 1990, figure 17.1.3.

time). The total number of computations that must be performed to model a system at a spatial resolution of Δx thus scales (in the absence of other factors) as $(\Delta x)^2$.

The larger time-step sizes possible with implicit schemes are not always of use if factors other than numerical stability are relevant; for example if the system itself changes significantly in time $\Delta t = \Delta x/c$ anyway. They are beneficial if the system evolves slowly through a set of near-equilibrium states, or if some parts of the system have a much higher sound speed than others (and hence are always close to equilibrium).

3.3 A One-dimensional, Single-phase Model of the Intracluster Medium

I describe here a simple one-dimensional model of the ICM that assumes spherical symmetry. The numerical scheme used for the difference equations is that of Thomas (1987, 1988). The ICM is represented as a single-phase (the multi-phase case is discussed in Chapter 6) ideal gas in the potential well of a dark matter halo. The gas is allowed to radiate at the rate appropriate to its temperature and chemical composition, and moves in response to gravity and pressure. The emitted radiation is accurately modelled in the X-ray waveband, so that synthetic spectra can be produced at any time in the evolution of the system and analysed in the same way as genuine observational data.

3.3.1 Simplified Form of the Gasdynamics Equations

A Lagrangian frame of reference is adopted, and the system of fluid equations to be solved reduces to the form

$$\frac{\mathrm{d}r}{\mathrm{d}t} = u \tag{3.17}$$

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -\left(g + \frac{1}{\rho}\frac{\mathrm{d}p}{\mathrm{d}r}\right) \tag{3.18}$$

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}t} = -p\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{\rho}\right) - \mathcal{L}(\rho,T)\ , \tag{3.19}$$

with the velocity u positive for outflow, g the magnitude of the gravitational acceleration, and $\mathcal{L}$ the power radiated per unit mass. If g is defined as the gravitational acceleration due solely to the dark matter, then in order to add the effects of the gas self gravity, $g \to g + \frac{GM}{r^2}$, where M is the gas mass enclosed within radius r. In addition, there are auxiliary equations for the gas equation of state (assumed ideal), and the radiative power loss:

$$p = \frac{k_\mathrm{B}}{\mu m_\mathrm{H}}\rho T = (\gamma - 1)\varepsilon\rho \tag{3.20}$$

$$\mathcal{L}(\rho,T) = \frac{n_\mathrm{e} n_\mathrm{H}\Lambda(T)}{\rho} = \frac{\Lambda(T)}{\rho}\left(\frac{X_\mathrm{H}\rho}{m_\mathrm{H}}\right)^2 r_\mathrm{e} = \rho\Lambda(T)\left(\frac{X_\mathrm{H}}{m_\mathrm{H}}\right)^2 r_\mathrm{e}, \tag{3.21}$$

with γ the adiabatic index, and μ the mean molecular weight of the gas. $\Lambda(T)$ is the cooling function, as defined in Section 1.5.1. Strictly speaking, the relative electron density r_e is a function of temperature, but as shown in Figure 1.4 on page 11, the actual change over the temperature range of interest is slight (this effect is included, though).

3.3.2 The Finite-difference Scheme

After Richtmyer and Morton (1967), the differential equations can be modelled by the following Lagrangian finite-difference system:

$$\frac{r_j^{n+1} - r_j^n}{\Delta t} = u_j^{n+\frac{1}{2}} \tag{3.22}$$

$$\frac{u_j^{n+\frac{1}{2}} - u_j^{n-\frac{1}{2}}}{\Delta t} = -g_j^n - \frac{1}{{}^1\!/_2\left(\rho^n_{j+\frac{1}{2}} + \rho^n_{j-\frac{1}{2}}\right)}\,\frac{\left(p^n_{j+\frac{1}{2}} - p^n_{j-\frac{1}{2}}\right)}{{}^1\!/_2\left(r^n_{j+1} - r^n_{j-1}\right)} \tag{3.23}$$

$$\frac{\varepsilon^{n+1}_{j+\frac{1}{2}} - \varepsilon^n_{j+\frac{1}{2}}}{\Delta t} = -p^n_{j+\frac{1}{2}}\frac{1}{\Delta t}\left(\frac{1}{\rho^{n+1}_{j+\frac{1}{2}}} - \frac{1}{\rho^n_{j+\frac{1}{2}}}\right) - \mathcal{L}^n_{j+\frac{1}{2}} \tag{3.24}$$

$$\rho^{n+1}_{j+\frac{1}{2}}\left[(r^{n+1}_{j+1})^3 - (r^{n+1}_j)^3\right] = \rho^n_{j+\frac{1}{2}}\left[(r^n_{j+1})^3 - (r^n_j)^3\right], \tag{3.25}$$

where subscripts refer to spatial, and superscripts to temporal, grid points. Conceptually, the gas is to be viewed as being divided into *zones* of radius Δr. The term *shell* will be used to refer to that single radius which separates any two zones. The labelling scheme used is one in which shell $j = 0$ corresponds to

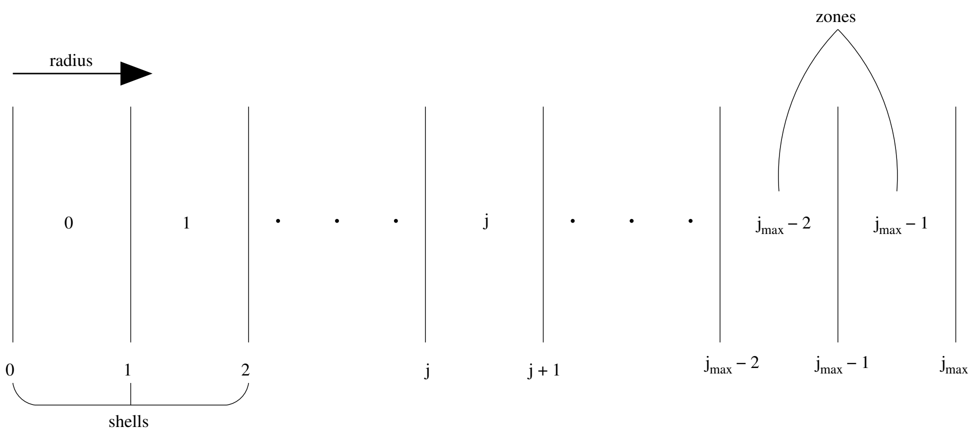


Figure 3.2: Illustrating the shells and zones as used by the finite-difference scheme for the fluid equations.

the innermost radius of the simulation, and shell $j = j_{\max}$ to the outermost radius. Between these two radii lie $j_{\max} - 1$ zones, numbered from zone 0 in the centre to zone $j_{\max} - 1$ at the outside. This is illustrated in Figure 3.2. Note that some quantities (ρ, p, ε, T, $\mathcal{L}$) are evaluated on half-integer mesh points (i.e. at zone centres), and some (u) on integer mesh points (i.e. on zone boundaries, or shells). This is purely a matter of notational convenience, to avoid numerous factors of two appearing in the difference equations. In the same way, the speed u (and later q — see Section 3.3.4) is evaluated half a time step out of synchronization with the other quantities. It is to be understood that the auxiliary quantity g_j^n, for example, represents $g(r_j^n)$.

The continuity equation (3.25) has a particularly simple form thanks to the adoption of a reference frame which moves with the fluid, expressing the conservation of mass in any given zone as the system evolves. This issue is made a little more complex if mass loss is allowed to occur — see Section 3.6.

Observe that most of the difference terms are *centred*, and hence second-order accurate, with the exception of the pressure p and power-loss $\mathcal{L}$ terms in the energy equation. Leaving this issue aside for the moment, the equations can be re-written to solve for the time-advanced quantities as follows:

$$u_j^{n+\frac{1}{2}} = u_j^{n-\frac{1}{2}} - g_j^n \Delta t - \frac{4\Delta t}{\left(\rho_{j+\frac{1}{2}}^n + \rho_{j-\frac{1}{2}}^n\right)} \frac{\left(p_{j+\frac{1}{2}}^n - p_{j-\frac{1}{2}}^n\right)}{\left(r_{j+1}^n - r_{j-1}^n\right)} \tag{3.26}$$

$$r_j^{n+1} = r_j^n + u_j^{n+\frac{1}{2}} \Delta t \tag{3.27}$$

$$\rho_{j+\frac{1}{2}}^{n+1} = \left[\frac{(r_{j+1}^{n+1})^3 - (r_j^{n+1})^3}{(r_{j+1}^n)^3 - (r_j^n)^3}\right] \rho_{j+\frac{1}{2}}^n \tag{3.28}$$

$$\varepsilon_{j+\frac{1}{2}}^{n+1} = \varepsilon_{j+\frac{1}{2}}^n - p_{j+\frac{1}{2}}^n \left(\frac{1}{\rho_{j+\frac{1}{2}}^{n+1}} - \frac{1}{\rho_{j+\frac{1}{2}}^n}\right) - \mathcal{L}_{j+\frac{1}{2}}^n \Delta t \tag{3.29}$$

$$p_{j+\frac{1}{2}}^{n+1} = (\gamma - 1)\varepsilon_{j+\frac{1}{2}}^{n+1} \rho_{j+\frac{1}{2}}^{n+1} \tag{3.30}$$

$$T_{j+\frac{1}{2}}^{n+1} = \frac{\mu m_{\mathrm{H}}}{k_{\mathrm{B}}} \frac{p_{j+\frac{1}{2}}^{n+1}}{\rho_{j+\frac{1}{2}}^{n+1}} \tag{3.31}$$

$$\mathcal{L}^{n+1}_{j+\frac{1}{2}} = \rho^{n+1}_{j+\frac{1}{2}} \Lambda \left(T^{n+1}_{j+\frac{1}{2}}\right) \left(\frac{X_{\rm H}}{m_{\rm H}}\right)^2 r_{\rm e} \tag{3.32}$$

As it stands, this scheme is fully explicit when solved in the stated order.

Centring of the Pressure Term in the Energy Equation

As was commented above, the original difference representation for the energy equation is not centred with respect to the pressure term. It can be made so by the change $p^n_{j+\frac{1}{2}} \to {}^1/_2(p^{n+1}_{j+\frac{1}{2}} + p^n_{j+\frac{1}{2}})$. This is at the expense of making the equations for the updated energy and pressure implicit, since the unknown $p^{n+1}_{j+\frac{1}{2}}$ then appears in both. It is, however, straightforward to solve these two equations simultaneously and obtain a revised energy equation.

$$\varepsilon^{n+1}_{j+\frac{1}{2}} = \varepsilon^n_{j+\frac{1}{2}} - \frac{1}{2}\left[(\gamma-1)\varepsilon^{n+1}_{j+\frac{1}{2}}\rho^{n+1}_{j+\frac{1}{2}} + p^n_{j+\frac{1}{2}}\right]\left(\frac{1}{\rho^{n+1}_{j+\frac{1}{2}}} - \frac{1}{\rho^n_{j+\frac{1}{2}}}\right) - \mathcal{L}^n_{j+\frac{1}{2}}\Delta t$$

$$\varepsilon^{n+1}_{j+\frac{1}{2}}\left[1 + \frac{1}{2}(\gamma-1)\rho^{n+1}_{j+\frac{1}{2}}\left(\frac{1}{\rho^{n+1}_{j+\frac{1}{2}}} - \frac{1}{\rho^n_{j+\frac{1}{2}}}\right)\right] = \varepsilon^n_{j+\frac{1}{2}} - \frac{1}{2}p^n_{j+\frac{1}{2}}\left(\frac{1}{\rho^{n+1}_{j+\frac{1}{2}}} - \frac{1}{\rho^n_{j+\frac{1}{2}}}\right) - \mathcal{L}^n_{j+\frac{1}{2}}\Delta t$$

$$\therefore \quad \frac{1}{2}\varepsilon^{n+1}_{j+\frac{1}{2}}\left[(\gamma+1) - (\gamma-1)\frac{\rho^{n+1}_{j+\frac{1}{2}}}{\rho^n_{j+\frac{1}{2}}}\right] = \varepsilon^n_{j+\frac{1}{2}} - \frac{1}{2}p^n_{j+\frac{1}{2}}\left(\frac{1}{\rho^{n+1}_{j+\frac{1}{2}}} - \frac{1}{\rho^n_{j+\frac{1}{2}}}\right) - \mathcal{L}^n_{j+\frac{1}{2}}\Delta t \tag{3.33}$$

Centring of the Luminosity Term in the Energy Equation

The energy equation is also not centred in terms of the radiative power loss $\mathcal{L}$. A similar transformation $\mathcal{L}^n_{j+\frac{1}{2}} \to {}^1/_2(\mathcal{L}^n_{j+\frac{1}{2}} + \mathcal{L}^{n+1}_{j+\frac{1}{2}})$ can in principle be applied to correct this. For example, if $\mathcal{L}$ has the simple analytic form (appropriate to high-temperature bremsstrahlung radiation for example)

$$\mathcal{L} = \alpha_{\mathcal{L}}\rho\sqrt{T}. \tag{3.34}$$

$$T = \frac{\mu m_{\rm H}}{k_{\rm B}}\frac{p}{\rho} = \frac{\mu m_{\rm H}}{k_{\rm B}}(\gamma-1)\varepsilon \qquad \Rightarrow \qquad \mathcal{L} = \alpha_{\mathcal{L}}\left[\frac{\mu m_{\rm H}}{k_{\rm B}}(\gamma-1)\right]^{\frac{1}{2}}\rho\sqrt{\varepsilon}$$

Centring the $\mathcal{L}$ term in equation (3.33) then leads to

$$c_1\varepsilon^{n+1}_{j+\frac{1}{2}} = c_2 - c_3\sqrt{\varepsilon^{n+1}_{j+\frac{1}{2}}}, \qquad \text{where}$$

$$c_1 \equiv (\gamma+1) - (\gamma-1)\frac{\rho^{n+1}_{j+\frac{1}{2}}}{\rho^n_{j+\frac{1}{2}}}$$

$$c_2 \equiv 2\varepsilon^n_{j+\frac{1}{2}} - p^n_{j+\frac{1}{2}}\left(\frac{1}{\rho^{n+1}_{j+\frac{1}{2}}} - \frac{1}{\rho^n_{j+\frac{1}{2}}}\right) - \mathcal{L}^n_{j+\frac{1}{2}}\Delta t$$

$$c_3 \equiv \alpha_{\mathcal{L}}\left[\frac{\mu m_H}{k_B}(\gamma-1)\right]^{\frac{1}{2}}\rho^{n+1}_{j+\frac{1}{2}}\Delta t.$$

Hence

$$
\begin{aligned}
c_3^2 \varepsilon_{j+\frac{1}{2}}^{n+1} &= (c_2 - c_1 \varepsilon_{j+\frac{1}{2}}^{n+1})^2 \\
c_1^2 \left(\varepsilon_{j+\frac{1}{2}}^{n+1}\right)^2 - \left(2c_1c_2 + c_3^2\right) \varepsilon_{j+\frac{1}{2}}^{n+1} + c_2^2 &= 0 \\
\varepsilon_{j+\frac{1}{2}}^{n+1} &= \frac{1}{2c_1^2}\left(2c_1c_2 + c_3^2 - c_3\sqrt{4c_1c_2 + c_3^2}\right),
\end{aligned}
$$

where the sign of the root is chosen for the correct behaviour in the limit of large Δt. The functional form used for Λ in practice, however, is numerical rather than analytic (see Section 3.7.1), and so this transformation is of purely academic interest. It is presented here only to show proof of concept, and is not used in the actual numerical implementation. An iterative scheme would be necessary to correctly centre the $\mathcal{L}$ term in the general case. Since Λ is not a strong function of temperature over the majority of the temperature range of interest, the lack of centring is not too worrisome. The $\mathcal{L}$ term also, however, has a linear density dependence. In a roughly isobaric situation, ρT is approximately constant; and consequently in the regions where Λ is an increasing function of T, the changes in ρ and Λ with cooling act antagonistically to reduce the net change in $\mathcal{L}$.

Final form of the Energy Equation

Since the specific internal energy ε is not a particularly useful quantity per se, it can be eliminated in favour of the pressure p. This will also facilitate comparison with the multi-phase equations (see Section 6.2), where such a transformation of the equations is a necessary step on the road to a solution. Removing ε from the pressure-centred equation (3.33) using the ideal gas equation of state leads to

$$
\frac{p_{j+\frac{1}{2}}^{n+1}}{\rho_{j+\frac{1}{2}}^{n+1}}\left[(\gamma+1) - (\gamma-1)\frac{\rho_{j+\frac{1}{2}}^{n+1}}{\rho_{j+\frac{1}{2}}^{n}}\right] = 2\frac{p_{j+\frac{1}{2}}^{n}}{\rho_{j+\frac{1}{2}}^{n}} - (\gamma-1)p_{j+\frac{1}{2}}^{n}\left(\frac{1}{\rho_{j+\frac{1}{2}}^{n+1}} - \frac{1}{\rho_{j+\frac{1}{2}}^{n}}\right) - 2(\gamma-1)\mathcal{L}_{j+\frac{1}{2}}^{n}\Delta t
$$

$$
\therefore \quad p_{j+\frac{1}{2}}^{n+1}\left(\frac{\gamma+1}{\rho_{j+\frac{1}{2}}^{n+1}} - \frac{\gamma-1}{\rho_{j+\frac{1}{2}}^{n}}\right) = p_{j+\frac{1}{2}}^{n}\left(\frac{\gamma+1}{\rho_{j+\frac{1}{2}}^{n}} - \frac{\gamma-1}{\rho_{j+\frac{1}{2}}^{n+1}}\right) - 2(\gamma-1)\mathcal{L}_{j+\frac{1}{2}}^{n}\Delta t \tag{3.35}
$$

3.3.3 Form of the Courant Condition

It can be shown (by considering small perturbations about a steady state, e.g. Richtmyer and Morton 1967) that in systems such as this, the Courant condition for numerical stability takes the form

$$
\Delta t < t_{\mathrm{stab}} \equiv \min(t_{\mathrm{stab,j+\frac{1}{2}}}) \qquad\qquad t_{\mathrm{stab,j+\frac{1}{2}}} \equiv \frac{r_{j+1}^{n} - r_{j}^{n}}{c_{\mathrm{s}\,j+\frac{1}{2}}}, \tag{3.36}
$$

where $c_{\mathrm{s}\,j+\frac{1}{2}}$ is the value in zone j of the sound speed,

$$
c_{\mathrm{s}}^2 = \frac{\mathrm{d}p}{\mathrm{d}\rho} = \gamma\frac{p}{\rho} \qquad \text{(adiabatic).} \tag{3.37}
$$

Thus, the stability time t_{stab} is simply the minimum zone sound-crossing time. With equally-spaced zones, it is thus the hottest zone which acts as the limiting factor on the stability time. In terms of the Courant number, $\mathcal{C} < 1$ is required. The form of the stability criterion used in practice is discussed in Section 3.3.6.

3.3.4 Pseudoviscosity

The study of gasdynamics often involves the treatment of discontinuities, where some or all of the fluid properties change abruptly in a non-smooth manner, with differing limits obtained depending on which direction the discontinuity is approached from. Examples includes interfaces between differing fluids, contact discontinuities (p and u are continuous but ρ and ε are not), shocks (p, u, ρ, ε all discontinuous), and the heads of rarefaction waves (some derivatives are discontinuous). Mathematically, in order to solve uniquely for the fluid motion jump conditions (acting as internal boundary conditions) must be added to the differential equations.

Standard difference schemes, like the one outlined here, are able to handle interfaces, contact discontinuities, and rarefaction waves without modification. It turns out that this is not, however, the case for shocks. As it stands, there is nothing in the difference scheme to represent the increase in entropy generated by a shock. For shock modelling, the differential equations are supplemented by the so-called Rankine–Hugoniot relations, which express the conservation of mass, momentum, and energy across the shock front.

Using the Rankine–Hugoniot relations directly in numerical schemes is possible, but difficult, due to the fact that the shock front (where the conditions are to be applied) is in motion through the fluid. The way in which the shock moves is not predetermined, but must be solved for. The motion is itself a function of the differential equations and jump conditions, and as a consequence any numerical method adopted must be highly implicit. The process is known as "shock fitting". Whilst it performs reasonably well for the special case of an external shock entering undisturbed fluid, it is less successful for more complex problems, or the general case where shocks may develop spontaneously at any point in the fluid.

An approximate (but accurate) numerical scheme that is able to deal with shocks automatically was developed by von Neumann and Richtmyer in the 1940s[1]. It is based more upon the real-world nature of shocks rather than their existence as mathematical abstractions. In practice, shocks are not step functions where the fluid properties change over an infinitesimal distance, but rather are smoothed (to some extent) by dissipative process such as viscosity and thermal conduction. This results in a rapid, but not instantaneous, change in the fluid properties over some small but finite distance. It can be shown that the size of the transition region is proportional to the coefficient of the dissipative process, so that in the limit of zero dissipation the abstraction of a truly discontinuous shock is regained.

The method of von Neumann and Richtmyer introduces an artificial dissipation mechanism in order to smooth out numerical shocks over a small number of spatial grid points. Finite-difference equations including this dissipation obviate the need for shock fitting and its associated problems, and model shocks in a manner close to the exact representation. An artificial viscosity is found to be more useful than an artificial conductivity. With the latter, strong shocks are still partly discontinuous, whereas with the former shocks of all strengths are smoothed (Becker, 1922). The simplest form of viscosity would be a

[1]Much pioneering work on fluid dynamics originated in Los Alamos during the Second World War.

linear one with stress proportional to rate of shear. This, however, results in a shock thickness which is a function of shock strength, tending to zero for a strong shock and infinity for a weak one. The use of a quadratic term, on the other hand, can be shown to give a shock extent that is independent of shock strength, as desired. This represents a viscosity coefficient that is very weak in the regions of smooth flow, and increases rapidly in strength in the presence of a shock (to many orders of magnitude greater than the real-world viscosity of fluids). The method is also constructed so that the shock thickness does not depend on the properties of the ambient medium, but remains fixed at a few grid spacings.

The form of the pseudoviscosity adopted by Richtmyer and Morton (1967) involves the addition of a pseudoviscous pressure q to the differential equations. For the one-dimensional Cartesian case,

$$q = \begin{cases} \rho l^2 \left(\frac{\partial u}{\partial x}\right)^2 & \frac{\partial u}{\partial x} < 0 \\ 0 & \frac{\partial u}{\partial x} \geqslant 0 \end{cases} \tag{3.38}$$

with l being a constant with the dimensions of length. The thickness of the numerical shocks is proportional to l. Note that q is zero when the fluid is expanding.

The motion and energy equations are modified by $p \to p+q$, taking the forms

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -\left[g + \frac{1}{\rho}\frac{\mathrm{d}}{\mathrm{d}r}(p+q)\right] \tag{3.39}$$

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}t} = -(p+q)\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{\rho}\right) - \mathcal{L}(\rho,T). \tag{3.40}$$

The other equations, including the equation of state, remain unchanged. The pseudoviscous pressure q is a dynamic, rather than a thermodynamic, quantity. Richtmyer and Morton (1967) show that such a scheme reproduces the correct shock motion and automatically satisfies the Rankine–Hugoniot relations, without the need for shock fitting. In finite-difference schemes, the length l is taken as $q_c \Delta x$, where the constant $q_c \sim 1$; i.e. l (and hence the shock thickness) is a few times the grid spacing.

If q_c is small, shocks are sharply resolved, but there are pressure oscillations in the volume trailing any shock. Making q_c larger smooths the post-shock profile, but increases the thickness of the shock transition region. The appropriate value is a compromise between these two extremes. If the pseudoviscosity is omitted altogether, there are large pressure oscillations and the shocks travel at the wrong speed. The oscillations are not instabilities — their amplitude remains roughly constant with time. Finite-difference schemes can be interpreted (in one dimension) as modelling a fluid in terms of many thin parallel plates where the mass is concentrated, separated by weightless, compressible fluid. The passage of a shock sets up oscillations in these plates, analogous to thermal excitation of fluid particles. The pseudoviscosity converts the oscillation energy into true internal energy, representing the dissipation of energy in a shock, in accordance with the Rankine–Hugoniot relations. This dissipation in the numerical scheme does not correspond to a violation of energy conservation, it merely redistributes kinetic energy from the oscillations into internal fluid energy. For the numerical scheme used here, q has the form

$$q^{n+\frac{1}{2}}_{j+\frac{1}{2}} = \begin{cases} q_c^2 \left[\frac{1}{2}\left(\frac{1}{\rho^{n+1}_{j+\frac{1}{2}}} + \frac{1}{\rho^{n}_{j+\frac{1}{2}}}\right)\right]^{-1} \left(u^{n+\frac{1}{2}}_{j+1} - u^{n+\frac{1}{2}}_{j}\right)^2 & u^{n+\frac{1}{2}}_{j+1} - u^{n+\frac{1}{2}}_{j} < 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.41}$$

It is to be noted that q, like u, is evaluated half a time step out of synchronization with the other quantities, again purely as a matter of convenience.

The modified form of the motion equation (3.26), in accordance with equation (3.39), is obtained via

$$\left(p^n_{j+\frac{1}{2}} - p^n_{j-\frac{1}{2}}\right) \to \left(p^n_{j+\frac{1}{2}} - p^n_{j-\frac{1}{2}}\right) + \left(q^{n-\frac{1}{2}}_{j+\frac{1}{2}} - q^{n-\frac{1}{2}}_{j-\frac{1}{2}}\right).$$

The energy equation (3.29) is adapted in accordance with equation (3.40), with

$$p^n_{j+\frac{1}{2}} \to {}^1\!/_2\left(p^{n+1}_{j+\frac{1}{2}} + p^n_{j+\frac{1}{2}}\right) \to {}^1\!/_2\left(p^{n+1}_{j+\frac{1}{2}} + p^n_{j+\frac{1}{2}}\right) + q^{n+\frac{1}{2}}_{j+\frac{1}{2}}, \tag{3.42}$$

where the first change is the correct centring of the true pressure as discussed previously, and is unrelated to the addition of the pseudoviscosity. Making this change in equation (3.29) is algebraically equivalent to

$$\mathcal{L}^n_{j+\frac{1}{2}}\Delta t \to \mathcal{L}^n_{j+\frac{1}{2}}\Delta t + q^{n+1}_{j+\frac{1}{2}}\left(\frac{1}{\rho^{n+1}_{j+\frac{1}{2}}} - \frac{1}{\rho^{n}_{j+\frac{1}{2}}}\right).$$

Therefore, the version of equation (3.35) incorporating the pseudoviscosity is

$$p^{n+1}_{j+\frac{1}{2}}\left(\frac{\gamma+1}{\rho^{n+1}_{j+\frac{1}{2}}} - \frac{\gamma-1}{\rho^{n}_{j+\frac{1}{2}}}\right) = p^n_{j+\frac{1}{2}}\left(\frac{\gamma+1}{\rho^{n}_{j+\frac{1}{2}}} - \frac{\gamma-1}{\rho^{n+1}_{j+\frac{1}{2}}}\right) - 2(\gamma-1)\left[\mathcal{L}^n_{j+\frac{1}{2}}\Delta t + q^{n+1}_{j+\frac{1}{2}}\left(\frac{1}{\rho^{n+1}_{j+\frac{1}{2}}} - \frac{1}{\rho^{n}_{j+\frac{1}{2}}}\right)\right].$$

Note that the pseudoviscous pressure term added to the energy equation is correctly centred (at time $t^{n+\frac{1}{2}}$), whereas that added to the motion equation is not (for correct centring one would need a time average of $q^{n+\frac{1}{2}}$ and $q^{n-\frac{1}{2}}$). Richtmyer and Morton (1967) comment on the fact that it is not found worthwhile to make the effort to correctly centre this term (and indeed it may cause instability to try), the lack of accuracy being justified on the grounds of the limited physical relevance of q.

Strictly speaking, the addition of a pseudoviscosity is only necessary for the modelling of shocks. I find it promotes stability of the equations, though, so I include it; even though $q \ll p$ in the absence of shocks.

Effect of Pseudoviscosity on Stability

It is shown by Richtmyer and Morton (1967) that the addition of pseudoviscosity does not alter the stability condition from the Courant one, equation (3.15), in regions where no shocks are present. In the vicinity of a shock, the condition for stability is somewhat modified, becoming

$$\mathcal{C} < \mathcal{C}_\eta \equiv \frac{\sqrt{\eta\left[\eta - \dfrac{\gamma-1}{\gamma+1}\right]}}{2q_\mathrm{c}(\eta-1),} \tag{3.43}$$

where η is the ratio of the densities either side of the shock. For a weak shock, $\eta \to 1$, and the standard Courant condition $\mathcal{C} < 1$ dominates. In strong shocks, however, equation (3.43) is the controlling factor. It can be shown from the Rankine–Hugoniot relations that in the limit of an infinitely strong shock,

$$\eta \to \frac{\gamma+1}{\gamma-1}; \tag{3.44}$$

in which case

$$\mathcal{C}_\eta = \frac{\left[\frac{\gamma+1}{\gamma-1}\left(\frac{\gamma+1}{\gamma-1} - \frac{\gamma-1}{\gamma+1}\right)\right]^{\frac{1}{2}}}{2q_\mathrm{c}\left(\frac{\gamma+1}{\gamma-1} - 1\right)} = \frac{\frac{1}{\gamma-1}}{\frac{4q_\mathrm{c}}{\gamma-1}}\left[(\gamma+1)^2 - (\gamma-1)^2\right]^{\frac{1}{2}} = \frac{\sqrt{\gamma}}{2q_\mathrm{c}} \approx 0.32 \quad (\gamma = {}^5/_3,\ q_\mathrm{c} = 2). \tag{3.45}$$

As Richtmyer and Morton (1967) point out, however, the derivation of equation (3.43) is not rigorous. In practice, the restriction on $\mathcal{C}$ is found to be not so severe for strong shocks, but somewhat more limiting for weak shocks.

3.3.5 Summary of the Single-phase Numerical Scheme

The final single-phase numerical scheme incorporating pseudoviscosity is presented here for reference.

$$u_j^{n+\frac{1}{2}} = u_j^{n-\frac{1}{2}} - g_j^n \Delta t - \frac{4\Delta t}{\left(\rho^n_{j+\frac{1}{2}} + \rho^n_{j-\frac{1}{2}}\right)} \frac{\left(p^n_{j+\frac{1}{2}} - p^n_{j-\frac{1}{2}} + q^{n-\frac{1}{2}}_{j+\frac{1}{2}} - q^{n-\frac{1}{2}}_{j-\frac{1}{2}}\right)}{\left(r^n_{j+1} - r^n_{j-1}\right)} \tag{3.46}$$

$$r_j^{n+1} = r_j^n + u_j^{n+\frac{1}{2}}\Delta t \tag{3.47}$$

$$\rho^{n+1}_{j+\frac{1}{2}} = \left[\frac{(r^{n+1}_{j+1})^3 - (r^{n+1}_j)^3}{(r^n_{j+1})^3 - (r^n_j)^3}\right]\rho^n_{j+\frac{1}{2}} \tag{3.48}$$

$$q^{n+\frac{1}{2}}_{j+\frac{1}{2}} = \begin{cases} q_c^2\left[\dfrac{1}{2}\left(\dfrac{1}{\rho^{n+1}_{j+\frac{1}{2}}} + \dfrac{1}{\rho^n_{j+\frac{1}{2}}}\right)\right]^{-1}\left(u^{n+\frac{1}{2}}_{j+1} - u^{n+\frac{1}{2}}_j\right)^2 & u^{n+\frac{1}{2}}_{j+1} - u^{n+\frac{1}{2}}_j < 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.49}$$

$$p^{n+1}_{j+\frac{1}{2}}\left(\frac{\gamma+1}{\rho^{n+1}_{j+\frac{1}{2}}} - \frac{\gamma-1}{\rho^n_{j+\frac{1}{2}}}\right) = p^n_{j+\frac{1}{2}}\left(\frac{\gamma+1}{\rho^n_{j+\frac{1}{2}}} - \frac{\gamma-1}{\rho^{n+1}_{j+\frac{1}{2}}}\right) - 2(\gamma-1)\left[\mathcal{L}^n_{j+\frac{1}{2}}\Delta t + q^{n+1}_{j+\frac{1}{2}}\left(\frac{1}{\rho^{n+1}_{j+\frac{1}{2}}} - \frac{1}{\rho^n_{j+\frac{1}{2}}}\right)\right] \tag{3.50}$$

$$T^{n+1}_{j+\frac{1}{2}} = \frac{\mu m_\mathrm{H}}{k_\mathrm{B}}\frac{p^{n+1}_{j+\frac{1}{2}}}{\rho^{n+1}_{j+\frac{1}{2}}} \tag{3.51}$$

$$\mathcal{L}^{n+1}_{j+\frac{1}{2}} = \rho^{n+1}_{j+\frac{1}{2}} \Lambda \left(T^{n+1}_{j+\frac{1}{2}} \right) \left(\frac{X_{\mathrm{H}}}{m_{\mathrm{H}}} \right)^2 r_{\mathrm{e}} \left(T^{n+1}_{j+\frac{1}{2}} \right) \tag{3.52}$$

When solved in the order presented, this scheme is fully explicit. It is also *single-step*, in that it is only necessary to store the values from one prior time in order to be able to calculate the updated conditions. This is obviously beneficial from the point of view of memory requirements and speed of execution.

3.3.6 Control of the Time Step

In practice, I make use of an adaptive step size, so that the actual time step taken in any iteration is some fraction of the stability time appropriate for that iteration,

$$\Delta t = f_{\mathrm{stab}}\, t_{\mathrm{stab}} \qquad\qquad f_{\mathrm{stab}} < 1\,, \tag{3.53}$$

with t_{stab} defined as in equation (3.36). Thus, the time step at each iteration is the largest compatible with stability. f_{stab} (equivalent to the Courant number $\mathcal{C}$) is not equal to one, due to the influence of pseudoviscosity, and the somewhat vague nature of the stability criterion in practice — see equation (3.45) and associated comments. Empirically, $f_{\mathrm{stab}} \approx 0.4$ is found to give stable evolution. It is straightforward to detect if instability is occurring due to the use of too large a time step, since the numerical results rapidly (and obviously) become meaningless.

As well as the concept of a stability time, common to explicit difference schemes, there is another factor that limits the time step in this particular case, namely the radiative cooling of the gas. Clearly, time steps cannot be so large that any zone cools below zero, and indeed it is preferable if the fractional cooling in any iteration is not too large. A second limit is therefore imposed on the time step,

$$\Delta t = f_{\mathrm{cool}}\, t_{\mathrm{cool}} \qquad\qquad f_{\mathrm{cool}} < 1 \tag{3.54}$$

$$t_{\mathrm{cool}} \equiv \min(t_{\mathrm{cool,j+\frac{1}{2}}}) \qquad\qquad t_{\mathrm{cool,j+\frac{1}{2}}} = t_{\mathrm{cool}} \left(n^n_{\mathrm{e}_{\mathrm{j+\frac{1}{2}}}}, T^n_{j+\frac{1}{2}} \right), \tag{3.55}$$

with t_{cool} being evaluated according to equation (1.24). In practice, I use $f_{\mathrm{cool}} = 0.1$. Whilst t_{stab} forms the most severe limit on the time step for high-temperature zones, t_{cool} is the limiting factor for cool zones. Since t_{stab} is a function of the spatial resolution (i.e. value of j_{max}) employed, so is the crossover temperature.

The addition of thermal conduction to the system adds a third constraint on the time step — see Section 5.8.2.

3.3.7 Boundary Conditions

Observe from equation (3.46) that u_{j} is a function of (amongst other things) $\rho_{j+\frac{1}{2}}$, $p_{j+\frac{1}{2}}$, and r_{j+1}. This means that it is not possible to calculate a value for the outer boundary speed, $u_{j_{\mathrm{max}}}$, from this equation, there being no such grid points as $\rho_{j_{\mathrm{max}}+\frac{1}{2}}$, $p_{j_{\mathrm{max}}+\frac{1}{2}}$, and $r_{j_{\mathrm{max}}+1}$ (these would lie outside the simulation volume). Similar considerations apply to the inner boundary at u_0. It is of course necessary to impose boundary conditions of some form in order to obtain unique solutions of the differential equations. At the inner boundary, the constraint $u_0 \equiv 0$ is imposed.

Since $j = 0$ corresponds to $r = 0$, the velocity must go to zero there in a one-dimensional, spherically-symmetric system. In order to calculate the speed at the outer boundary, $u_{j_{\max}}$, a "dummy mesh point" is used. This is a notional extra zone that lies beyond the outer boundary of the simulation, with pressure $p_{j_{\max}}$, density $\rho_{j_{\max}}$, and pseudoviscous pressure $q_{j_{\max}}$ (quantities that would not otherwise exist). q for this dummy zone is fixed at 0, the pseudoviscous pressure being unimportant at the outer edge of the simulation where the conditions are almost unchanging. p and ρ remain fixed at the values appropriate to this point in the initial conditions. Thus, the dummy zone is a static one that does not evolve. It is thus necessary to ensure that the integration proceeds outward far enough for this to be a valid assumption. This is certainly the case if integrating to the virial radius. Additionally, it is necessary to check that the simulation does not proceed for so long that the constant-pressure outer boundary condition is invalidated. This never occurs in practice.

The adoption of such a boundary condition is inconvenient in some ways, since it requires integrating out to a large radius. Obviously if one could reduce the radial range, higher spatial resolution would be obtained. Integrating out to the virial radius is also convenient, though, for the normalization of the gas mass. The gas mass is fixed to be some fraction of the total mass at the virial radius. If instead the outer radius of the simulation were used, this would mean that changing the radial extent of the simulation would change the conditions being modelled, which is not usually what is wanted. In order to calculate an accurate surface brightness (Section 3.8), it is also important not to truncate the radial extent of the cluster too soon. This is not a very strong constraint, though, since the surface brightness declines strongly with radius, so distant zones make only a small contribution to the more central regions.

3.4 The Dark Matter Profile

The model simulates a cluster as a two-component system, comprising hot gas in the potential well of a dark matter halo. Initial conditions must be specified for both these constituents. The dark matter profile is treated as static, so the initial conditions apply throughout the simulation.

There are a variety of functional forms for the density profile of a dark matter halo that have been or are being used; for example the traditional King profile

$$\rho(x) = \frac{\rho_{\mathrm{K}}}{(1+x^2)^{3/2}}, \tag{3.56}$$

and the now widely-used NFW profile (Navarro *et al.*, 1997)

$$\rho(x) = \frac{\rho_{\mathrm{N}}}{x(1+x)^2}. \tag{3.57}$$

In both cases, $x = r/r_{\mathrm{scale}}$ is a dimensionless radius. The masses, accelerations, and potentials corresponding to these profiles are derived in Appendix A.8. The King profile is an approximation to the non-singular isothermal sphere (e.g. Binney and Tremaine, 1987). The NFW profile is advertised as being a universal density profile, applicable to dark matter haloes over a wide range of mass scales, independent of cosmology. It is derived from the output of numerical simulations of cluster formation. There is an ongoing debate as to which of the modern density profiles provides the best description of actual haloes (e.g. Power *et al.*, 2003).

From the point of view of the software implementation of the numerical scheme, the dark matter profile is specified by a set of external functions for the enclosed mass, potential, and gravitational acceleration. It is therefore easy to switch between parametrizations or introduce new ones. I have mainly made use of the NFW profile.

The NFW profile has two free parameters, a scale density ρ_N and a scale radius r_N. These are not the most physically relevant parameters for my purposes, however, and can be replaced by making using of calibrated virial scaling relations.

3.4.1 The Virial Scaling Relations

Before deriving the virial scaling relations, it is necessary to introduce the concept of *overdensity*. At any given radius, the overdensity (or density contrast) of a cosmological mass concentration is simply the mean density enclosed within that radius measured in units of the universal critical density for closure (at the appropriate redshift),

$$\Delta(r) \equiv \frac{\bar{\rho}(r)}{\rho_\mathrm{c}(z)}. \tag{3.58}$$

Since the overdensity is a monotonic function of radius (increasing towards the centre of a mass concentration), it can be used as an alternative measure of the radial co-ordinate.

The mass enclosed within an overdensity Δ at redshift z is, by definition,

$$M_\Delta = {}^4\!/_3\pi r_\Delta^3 \Delta \rho_\mathrm{c}(z), \quad \text{where} \tag{3.59}$$

$$\rho_\mathrm{c}(z) = \rho_\mathrm{c}(0) E^2(z), \quad \text{and} \tag{3.60}$$

$$H(z) \equiv E(z) H_0. \tag{3.61}$$

$$\therefore \quad M_\Delta \propto E^2(z) r_\Delta^3. \tag{3.62}$$

This is true for any value of Δ. During the process of cosmological structure formation, high-density fluctuations collapse in on themselves under the influence of gravity. The central regions of these condensations are the first sites to achieve thermal (virial) equilibrium in the collapse. Simple spherical collapse theory (for $\Omega_\mathrm{M} = 1$, although the result is often used for other cosmologies) shows that matter within $\Delta_\mathrm{V} \approx 178$ (often approximated as $\Delta_\mathrm{V} = 200$) will have virialized at any time, while matter outside this radius will still be infalling (e.g. Cole and Lacey, 1996).

In a relaxed cluster, the ICM within Δ_V is hydrostatic and in virial equilibrium, so that its thermal energy reflects the gravitational potential energy of the cluster. Thus (e.g. Mathiesen and Evrard, 2001),

$$k_\mathrm{B} T \propto \frac{G M_\mathrm{V}}{r_\mathrm{V}}, \tag{3.63}$$

where r_V is the virial radius, and M_V the mass enclosed within it. Assuming that the constant of proportionality in this relationship is independent of redshift (self-similar evolution) and making use of equation (3.62) leads to $T \propto E^2(z) r_\mathrm{V}^2$, and hence

$$r_\mathrm{V} \propto E^{-1}(z) T^{\frac{1}{2}} \qquad M_\mathrm{V} \propto E^{-1}(z) T^{\frac{3}{2}}. \tag{3.64}$$

Cosmology	h_{50}	A	C
SCDM	0.5	7.0×10^{14}	9.9×10^{5}
ΛCDM	0.7	5.4×10^{14}	7.3×10^{5}

Table 3.1: Coefficients of the calibrated virial scaling relations.

These are the scaling relations for the virial radius, r_{V}, and the mass within it, M_{V}, as functions of the cluster virial temperature.

Evrard *et al.* (1996) found that the scaling relations were well obeyed in their simulations of structure formation, and obtained values for the associated constants of proportionality.

$$\begin{aligned} M_{200}(T) &= \frac{2.9\times10^{15}\,\mathrm{M}_\odot}{h_{50}E(z)} \left(\frac{T}{10\,\mathrm{keV}}\right)^{\frac{3}{2}} \\ r_{200}(T) &= \frac{3.7\,\mathrm{Mpc}}{h_{50}E(z)} \left(\frac{T}{10\,\mathrm{keV}}\right)^{\frac{1}{2}}, \end{aligned} \tag{3.65}$$

where r_{200} is the radius at which $\Delta=200$, i.e. the virial radius.

Calibrated scaling relations were also obtained by Allen *et al.* (2001b), this time from *Chandra* observations of relaxed clusters. These authors quote results for $\Delta=2500$, this being roughly the greatest radius at which reliable *Chandra* temperature measurements can be made for the selected clusters.

$$h_{50}E(z)\left(\frac{M_{2500}}{\mathrm{M}_\odot}\right) = A\left(\frac{k_{\mathrm{B}}T_{2500}}{10\,\mathrm{keV}}\right)^{\alpha} \tag{3.66}$$

$$h_{50}E(z)\left(\frac{r_{2500}}{\mathrm{pc}}\right) = C\left(\frac{k_{\mathrm{B}}T_{2500}}{10\,\mathrm{keV}}\right)^{\gamma} \tag{3.67}$$

Allen *et al.* do not calculate the radius–temperature relation (focusing on the luminosity–temperature relation instead), so I have made my own fits to their tabulated data. They find the mass scaling relation to be consistent with $\alpha=1.5$, and I find the radius scaling to be consistent with $\gamma=0.5$. Accordingly, values for the constants A and C are found with the power indices fixed at these values. The results obtained are listed in Table 3.1 (A as quoted by Allen *et al.*, C from my fits to their data, excluding 3C295 as they do).

For my purposes, I require the scaling relations at $\Delta=200$, rather than 2500. For comparison with the results of Evrard *et al.* (1996), Allen *et al.* convert their results for the mass relation at $\Delta=2500$ according to

$$M_{2500} = M_\Delta\left(\frac{2500}{\Delta}\right)^{-0.5}\left(\frac{T_{2500}}{T_\Delta}\right)^{1.5}. \tag{3.68}$$

This relies on assuming the density profile in the relevant range has the form $\rho\propto r^{-2}$ (i.e. the mid-range of the NFW profile), so that $r\propto\Delta^{-0.5}$. $M\propto\int r^2\rho(r)\,\mathrm{d}r\propto r$ in this case. Hence $M\propto\Delta^{-0.5}$. After a transformation of this kind, Allen *et al.* (2001b) find that the normalization of Evrard *et al.* (1996) lies approximately 40 per cent above the observed value. Similar results were found by other authors in the analysis of *ASCA* and *ROSAT* data (e.g. Horner *et al.*, 1999; Nevalainen *et al.*, 2000).

Conversion of the scaling relations from one overdensity to another requires assumption of a functional form for the density profile. One may as well make use of the NFW profile itself, rather than any approximation to it, since the added complication is slight. Begin by transforming the dependent variable x in the NFW profile according to

$$x \equiv \frac{r}{r_\mathrm{N}} = \frac{r}{r_\mathrm{V}}\frac{r_\mathrm{V}}{r_\mathrm{N}} = cy,$$

where $y \equiv r/r_\mathrm{V}$ is the radius expressed in units of the virial radius, and c is the *concentration parameter*.

$$c \equiv \frac{r_\mathrm{V}}{r_\mathrm{N}}, \tag{3.69}$$

and expresses the virial radius in units of the NFW scale radius.

Hence equation (3.57) becomes

$$\rho(y) = \frac{\rho_\mathrm{N}}{cy(1+cy)^2}.$$
$$\rho(r = r_\mathrm{V}) = \rho(y=1) \equiv \Delta_\mathrm{V}\rho_\mathrm{c} \qquad (\Delta_\mathrm{V} = 200)$$
$$\therefore \quad \rho_\mathrm{N} = \Delta_\mathrm{V}\rho_\mathrm{c}c(1+c)^2.$$

Therefore the NFW profile can be written in the form

$$\Delta(y) = \frac{\Delta_\mathrm{V}}{y}\left(\frac{1+c}{1+cy}\right)^2. \tag{3.70}$$

This is a cubic equation for $y(\Delta)$. Rather than inverting it analytically, it is straightforward to do so numerically. Re-arranging the equation as

$$y(1+cy)^2\frac{\Delta}{\Delta_\mathrm{V}(1+c)^2} - 1 = 0,$$

it is easily solved by (for example) the Newton–Raphson method for iterative solution of the equation $f(y) = 0$,

$$y_{n+1} = y_n - \frac{f(y_n)}{f'(y_n)}. \tag{3.71}$$

In this specific case,

$$y_{n+1} = y_n - \frac{\left[y(1+cy)^2\dfrac{\Delta}{\Delta_\mathrm{V}(1+c)^2} - 1\right]}{(1+cy)(1+3cy)\dfrac{\Delta}{\Delta_\mathrm{V}(1+c)^2}}. \tag{3.72}$$

In practice, of course, Δ_V here need not refer to the virial radius per se, but rather any desired reference radius.

It can be shown (see Appendix A.8) that the mass within a given radius for the NFW profile is given

	$c=3$	$c=5$	$c=8$	Approximate
y_{2500}	0.327	0.363	0.387	-
M_{2500}/M_{200}	0.300	0.407	0.500	0.283
y_{500}	0.685	0.704	0.716	-
M_{500}/M_{200}	0.698	0.762	0.806	0.632
M_{2500}/M_{500}	0.423	0.535	0.620	0.447
Simulation:observation excess	33%	68%	95%	41%

Table 3.2: Conversion between different overdensities.

by

$$M_{\rm N}(x) = 4\pi\rho_{\rm N}{r_{\rm N}}^3\left[\ln(1+x) - \frac{x}{1+x}\right]. \tag{3.73}$$

Therefore

$$\frac{M_\Delta}{M_{\Delta_{\rm v}}} = \frac{\ln(1+cy) - \dfrac{cy}{1+cy}}{\ln(1+c) - \dfrac{c}{1+c}}.$$

Since the shape of the profile has been assumed, the final result (i.e. the conversion from one reference radius to another) depends on c. Table 3.2 lists some results for three representative concentrations that span most of the relevant range. The approximate values are obtained via the relationship $M_{\Delta_1}/M_{\Delta_2} \approx (\Delta_1/\Delta_2)^{-0.5}$.

The Allen *et al.* (2001b) normalization for the mass-temperature relation in a ΛCDM cosmology is 0.7×10^{15} at $\Delta = 2500$, whilst that of Evrard *et al.* (1996) at $\Delta = 500$ is 2.2×10^{15}. The fractional excess of the Evrard *et al.* (1996) normalization above that of Allen *et al.* (2001b) is thus given by

$$\left(\frac{M_{2500}}{M_{500}}\times\frac{2.2}{0.7}\right) - 1.$$

This is tabulated in the final row of Table 3.2. The discrepancy between simulation and observation is even worse for high-concentration clusters. The figure of 40 per cent quoted by Allen *et al.* (2001b) is consistent with the fact that the clusters in their sample are biased towards low concentrations.

To specify the simulation mass distribution for a cluster of some specified virial temperature, r_{200} and the mass M_{200} enclosed within it are obtained from the virial scaling relations. Given that there are two free parameters in the density profile, this single data point does not suffice to specify a unique profile. The specification is completed with a value for the concentration parameter, c. Less massive systems tend to have higher concentrations, since they formed at earlier epochs when the average cosmological density was higher (e.g. Wechsler *et al.*, 2002). A concentration of 5 can be considered as representative of the cluster regime (e.g. Bullock *et al.*, 2001), although there is considerable range (3–8 not excessive, as shown by the sample of Allen *et al.*, 2001b).

The concentration and virial temperature are the more physically interesting parameters which replace the scale density and radius in my specification of an NFW profile. The temperature and radius scaling relation give a value for the virial radius, and dividing by the required concentration leads to the scale radius. Finally, the scale density is obtained by normalizing the mass profile so that the correct mass is enclosed within r_{200}, using equation (3.73). I initially used the scaling relations of Evrard *et al.* (1996), and then later those of Allen *et al.* (2001b). With the latter, once the concentration is specified, the above method is used to solve for y_{2500} and then M_{2500}/M_{200}. The $\Delta = 2500$ normalization of Allen *et al.* (2001b) is thus converted to the required $\Delta = 200$.

3.5 The Gas Initial Conditions

I have assumed that the gas is initially in isothermal, hydrostatic equilibrium. Whilst there are other initial conditions that could be applied, to do so would I feel shift the focus of investigation away from the evolution of the ICM, and more towards the process of cluster assembly, say. This is a very broad topic that is best addressed by other means.

3.5.1 Isothermal, Hydrostatic Equilibrium

If a fluid is an rest in a gravitational field, the equation of motion, equation (3.2), reduces to

$$-\nabla p + \rho \boldsymbol{g} = 0. \tag{3.74}$$

For a spherically symmetric system, this becomes

$$\frac{\mathrm{d}p}{\mathrm{d}r} + \rho g = 0, \tag{3.75}$$

where $g = \mathrm{d}\Phi/\,\mathrm{d}r$ is the magnitude of the gravitational acceleration. For the specific case of an isothermal ideal gas

$$\frac{k_{\mathrm{B}}T}{\mu m_{\mathrm{H}}}\frac{\mathrm{d}\rho}{\mathrm{d}r} + \rho\frac{\mathrm{d}\Phi}{\mathrm{d}r} = 0,$$

and hence

$$\rho(r) \propto \exp\left[-\frac{\mu m_{\mathrm{H}}}{k_{\mathrm{B}}T}\Phi(r)\right]. \tag{3.76}$$

This expresses the density profile of an isothermal, ideal gas in hydrostatic equilibrium in a gravitational field as a function of the potential.

In practice, I begin by specifying the gas pressure in terms of a smooth, isothermal profile

$$p(r) = p(r_{200})\exp\left\{\frac{\rho(r_{200})}{p(r_{200})}\left[\Phi(r_{200}) - \Phi(r)\right]\right\}. \tag{3.77}$$

This expression is applied at half-integer mesh points to generate the initial discretized pressure profile, $p^0_{j+\frac{1}{2}}$ for $j = 0$ through to $j_{\max} - 1$. Additionally, the profile is extrapolated to a dummy mesh point beyond the outer edge, for use as a boundary condition as discussed in Section 3.3.7. The scale pressure $p(r_{200})$ is initially arbitrary. The density at the virial radius, $\rho(r_{200})$, is calculated from $p(r_{200})$, $T(r_{200})$ and the equation of state, with $T(r_{200})$ usually taken to be the virial temperature of the cluster (although any value can be chosen). In the case of the NFW profile, it can be shown (see Appendix A.8) that the gravitational potential is given by

$$\Phi = -4\pi G r_{\mathrm{N}}^{2} \rho_{\mathrm{N}} \frac{\ln(1+x)}{x} \qquad \text{(normalized to } \Phi \to 0 \text{ as } x \to \infty\text{)}. \tag{3.78}$$

The gas is assigned a density profile which is a scaled version of the pressure profile,

$$\rho(r) = \frac{\rho(r_{200})}{p(r_{200})} p(r). \tag{3.79}$$

Since the pressure profile is smooth, so too is the density profile. This makes normalization of the total gas mass straightforward (note that the initial temperature profile is not always smooth — see Section 4.1.1). The total mass of gas is then calculated through numerical integration of the density profile. The trapezium rule is used, there being no advantage in applying any more sophisticated quadrature methods when the density is only tabulated on discrete points to begin with.

Since the initial value for $p(r_{200})$ is arbitrary, so is the initial gas mass. I normalize the gas mass so that it corresponds to some fraction f_{gas} of the total mass.

$$f_{\mathrm{gas}} \equiv \frac{M_{\mathrm{gas}}}{M_{\mathrm{gas}} + M_{\mathrm{DM}}}, \tag{3.80}$$

$$\Rightarrow \quad M_{\mathrm{gas}} = \left(\frac{f_{\mathrm{gas}}}{1 - f_{\mathrm{gas}}} \right) M_{\mathrm{DM}}. \tag{3.81}$$

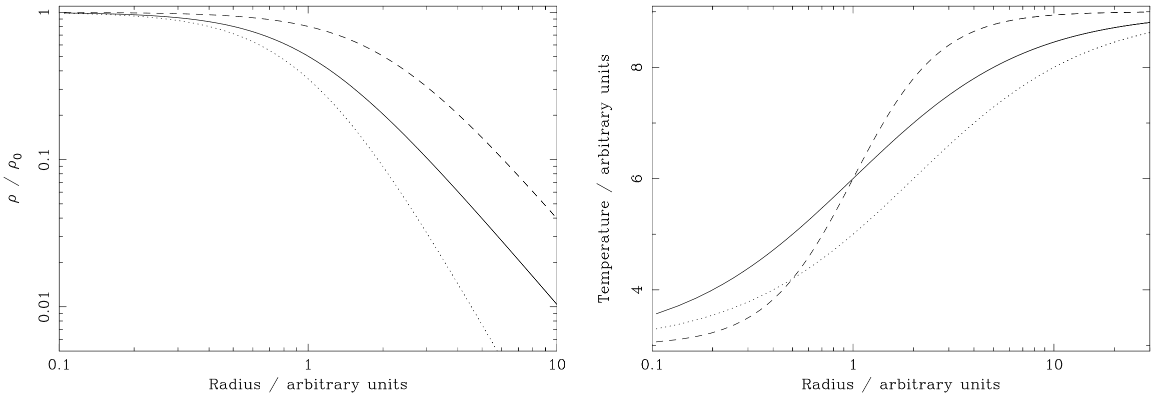
The total mass of gas is linearly proportional to the outer boundary pressure, $p(r_{200})$, hence it is trivial to adjust the outer boundary pressure to obtain the correct gas mass. A reasonable value for the gas mass fraction is $f_{\mathrm{gas}} = 0.17$ (e.g. Ettori and Fabian, 1999). The initial temperature profile is then assigned using the equation of state, and auxiliary quantities such as $\mathcal{L}$ are calculated.

3.5.2 Real-world Initial Conditions

Although specifying a hydrostatic initial condition is useful, it is also interesting to be able to base the initial states on observed profiles. In this way, the effect of differing internal properties on the future evolution of specific objects can be investigated. If just the initial outputs are used, then this provides a way to compute the X-ray emission profiles and spectra that would result from the output of other simulations (e.g. SPH ones). Also, by producing spectra corresponding to known profiles, one can use standard observational techniques to attempt to reproduce the input physical conditions, thus giving a useful check for any systematic errors in the analysis procedure. For example, with Roderick Johnstone I have been carrying out some tests of the XSPEC `projct` model in this way.

In such cases, the density and pressure profiles (and optionally the radial grid-points) are read into the main body of the program from external files. If the radial grid is not specified, then the maximum and minimum radii and number of zones must be given, and the grid is then calculated.

Irrespective of whether the radial grid is supplied or calculated, the observed profiles must be regridded so that they are evaluated at an appropriate number of points. The only sensible way to do this in practice is to fit the observed profiles with functional forms, then evaluate these forms at suitable points.



(a) Density profile

(b) Temperature profile

Figure 3.3: Examples of simple functional forms for some cluster profiles. (a) β density profile, equation (3.82). Solid line $\beta = 0.66$, $r_\beta = 1.0$; dashed line $\beta = 0.66$, $r_\beta = 2.0$; dotted line $\beta = 1.0$, $r_\beta = 1.0$. (b) η temperature profile, equation (3.83). $T_0 = 3$, $T_1 = 6$ in all cases. Solid line $\eta = 1.0$, $r_{T\eta} = 1.0$; dashed line $\eta = 2.0$, $r_\eta = 1.0$; dotted line $\eta = 1.0$, $r_\eta = 2.0$.

Common forms are

$$\rho(r) = \rho_0 \left[1 + \left(\frac{r}{r_\beta}\right)^2\right]^{-3\beta/2} \tag{3.82}$$

$$T(r) = T_0 + T_1 \left[1 + \left(\frac{r}{r_\eta}\right)^{-\eta}\right]^{-1}, \tag{3.83}$$

that is, the β density profile (Cavaliere and Fusco-Femiano, 1978); and what will hereinafter be referred to as the η temperature profile (e.g. Allen *et al.*, 2001c, figure 1). The former has a constant density core of radius $\sim r_\beta$, outside which the density asymptotes to $r^{-3\beta}$. The η profile ($\eta > 0$) has a constant limit both as $r \to 0$ and $r \to \infty$, of T_0 and $T_0 + T_1$ respectively. At intermediate radii, in the region of $r \sim r_\eta$, there is a transition region whose sharpness is controlled by η, with larger η values giving more abrupt transitions. Some examples are given in Figure 3.3.

Normally, observed temperature and density (in one form or another, e.g. n_e) profiles are available. As described in Section 3.5.1, though, in the numerical implementation the pressure and density profiles are first assigned, then the temperature profile is calculated. Whilst one could change this when external profiles are used, the approach I adopt is to externally calculate a pressure profile from the observed temperature and density profiles, using the ideal gas equation of state. This requires a value of μ, which strictly speaking is metallicity-dependent, although in practice not very sensitively (see Table B.4 on page 162). The same is true for the calculation of r_e if n_e is supplied rather than ρ. The pressure and density profiles are then input to the main program. The calculated temperature profile can be compared with the observed one as a sanity check.

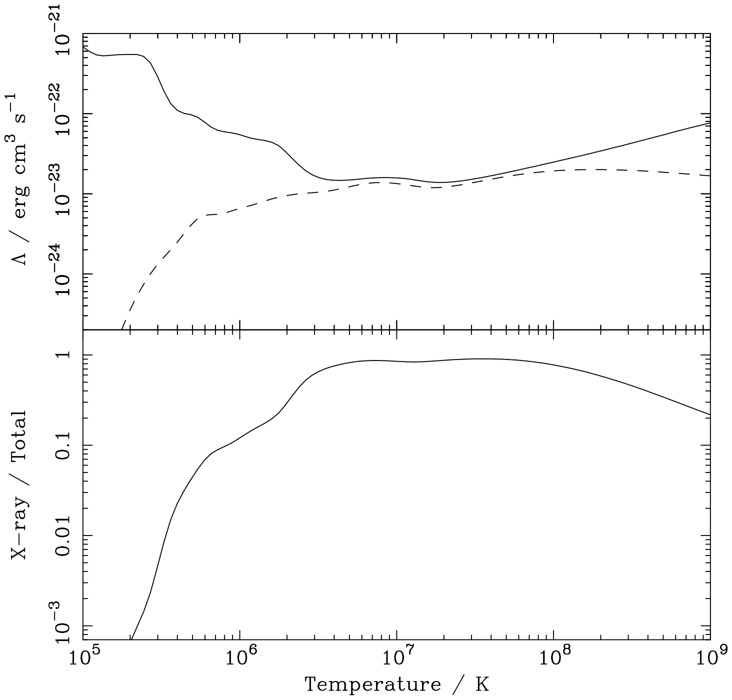


Figure 3.4: A comparison of the 0.1 – 10.0 keV (dashed line in top panel) and total (solid line in top panel) cooling functions for a 0.3 $Z_\odot$ plasma. The lower panel shows the ratio of the two.

3.6 Mass Loss

The implementation of the numerical scheme is made considerably more complicated due to the fact that mass loss can (and frequently does) occur. The ICM is a cooling plasma, but it is not possible to follow the cooling down to arbitrarily low temperatures where, for example, many complex molecular processes would eventually occur as the plasma recombined. In practice, therefore, the gas cooling is only followed down to some finite temperature. Although this is to some extent an action taken out of necessity, it is in fact desirable, for a number of reasons. As can be seen from Figure 3.4, the fraction of the radiated power being emitted in the "X-ray" waveband (hereinafter 0.1–10.0 keV, unless otherwise stated) declines precipitously below around 10^6 keV (by which point the fraction has fallen to 10 per cent or so for a 0.3 $Z_\odot$ plasma). For our purposes, there is therefore little incentive to follow the gas cooling below such temperatures (and indeed it would not be possible with the plasma radiation code employed — see Section 3.7). Also, as shown in Figure 1.7(b) on page 17, the cooling time is extremely short once such temperatures are reached. Satisfying the requirement of equation (3.54) on the time step would necessitate a large number of iterations to simulate a relatively short time scale. Furthermore, if the cooling is roughly isobaric then the gas occupies only a fraction of its initial volume.

For all these reasons, the gas cooling is only followed down to some temperature $T_{\rm min}$. In practice, I have used 10^5 K for the cut-off, though this could probably be increased somewhat without affecting the results noticeably. Conceptually, is it being assumed that once the zones reach this temperature, they go on to rapidly collapse down into condensed objects of negligible size. Note that I neglect the possibility of any energy feedback (e.g. due to star formation) from the gas that has cooled out.

There then arises the question of what to do with a zone of cooling gas when it reaches the cut-off temperature T_{min}. It must be removed from the simulation in some fashion, but clearly it is not acceptable to simply leave an empty space behind.

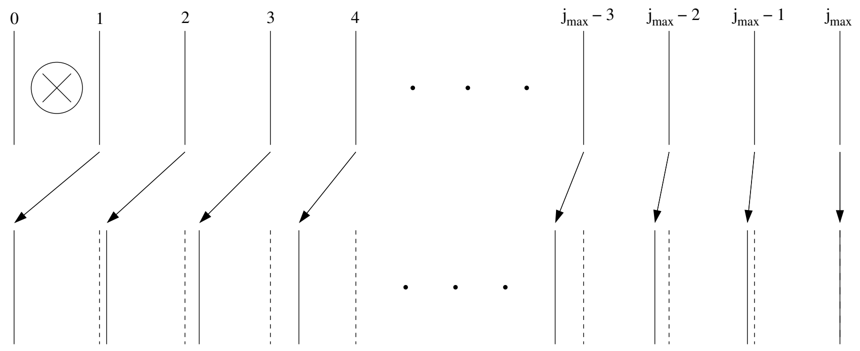
First consider the situation in which it is the innermost zone which is lost. This is almost always the case with a single-phase medium, because the temperature is a monotonically increasing function of radius. The simplest way to deal with the loss of the central zone would be to expand the next zone to fill the available space. This is somewhat artificial though, roughly akin to kicking away the bottom box in a pile and only having the one immediately above it react. Such an approach tends to excite a rarefaction wave which then propagates outward through the medium as it adapts to the adjusted conditions over a period of time.

An improved approach allows *all* the remaining zones to expand to some degree. Those closest to the loss site undergo the greatest degree of expansion, whereas distant zones hardly react at all. Thus, the differential change in properties between one zone and its neighbours is slight, which is obviously more stable. This is illustrated in Figure 3.5(a) on the next page. Algebraically, $r_{\text{j}} \to r_{\text{j}} - c_{\text{j}}(r_{\text{j}} - r_{\text{j}-1})$. The question then becomes one of how to assign the coefficients c_{j}. If only the first zone responds to the loss of the zeroth zone, this is equivalent to $c_{\text{j}} = \delta_{1j}$; with the Kronecker delta function, δ_{1j}, equal to one for $j = 1$, and zero otherwise. c_{j} = constant represents a translation in which all the remaining zones move inwards. Neither of these is particularly realistic. If the zones closest to the loss site are to respond the most, then a decreasing series of some kind is required.

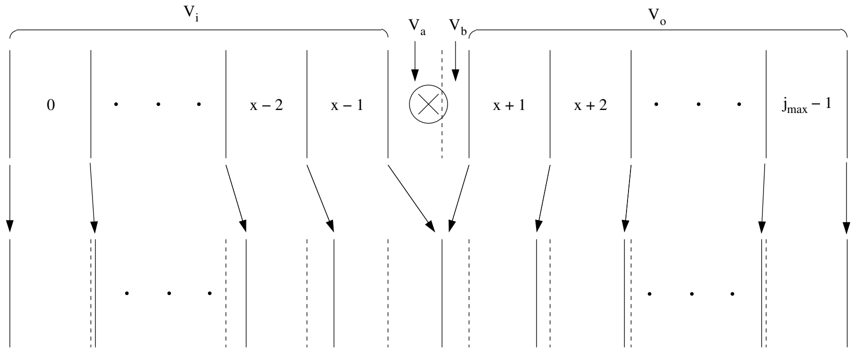
In fact, it is probably preferable to work in terms of volume, rather than radius, given the nature of the volume element in spherical polars (i.e. a given radius change corresponds to a greater volume change at large radii than it does at small radii). If zone zero is lost, there is a volume V_0 to be filled. Assigning all this volume to the next zone, or distributing it equally between all remaining zones, are both unphysical. Some abstraction of the collapse of zone zero down to zero volume is sought, in which the change in volume of the remaining zones decreases with radius. A geometric series is a convenient choice (an exponential decay would be another). A scheme using a geometric series has the form:

$$\begin{array}{llll}
V_1 & \to V_1 & + b & V_0 \\
V_2 & \to V_2 & + br & V_0 \\
V_3 & \to V_3 & + br^2 & V_0 \\
\vdots & & & \\
V_j & \to V_j & + br^{j-1} & V_0 \\
\vdots & & & \\
V_{j_{\max}-1} & \to V_{j_{\max}-1} & + br^{j_{\max}-1} & V_0,
\end{array}$$

where V_0 is the volume of the central zone that has been lost, and b and $r < 1$ are parameters of the geometric series. I regard r as a fixed parameter for each simulation, that controls how fast the volume change decays, whilst the normalization constant b is to be determined for each instance of zone loss. A value of $r \ll 1$ corresponds to a highly decaying response in which the first zone undergoes the greatest change, and $r \to 1$ to a more evenly distributed change affecting all zones nearly equally. I have used



(a) Central zone loss



(b) Off-centre zone loss

Figure 3.5: Illustration of how zone loss is treated in the numerical implementation. Upon loss of the marked zone, the remaining zones expand to fill the vacated volume. Shells move as indicated by the arrows, with the dashed lines marking original shell positions before the zone loss occurred. Zones closest to the loss site undergo the greatest amount of expansion, and those furthest away the least. (a) Loss of central zone. Numeric labels are shell numbers. (b) Loss of off-centre zone. Numeric labels are zone numbers. The vacated volume is partitioned into two. V_a is expanded into by the zones lying inward of the loss site, and V_b by those lying outward of the loss site. The volume is apportioned such that $V_a : V_b = V_i : V_o$.

$r = 0.9$. Through conservation of volume, b is determined:

$$\sum_{j=1}^{j_{\max}-1} br^{j-1} = \sum_{j=0}^{j_{\max}-2} br^{j} = 1 = \frac{b(1-r^{j_{\max}-1})}{1-r}$$
$$\therefore \quad b = \frac{1-r}{1-r^{j_{\max}-1}}, \tag{3.84}$$

using equation (A.2) for the sum of a geometric series.

Off-centre zone loss can occur even in the single-phase case — see Chapter 4 for some examples. A technique for handling the more general case of non-central zone loss, as illustrated in Figure 3.5(b), is therefore required. Although expansion into the freed space might be restricted only to zones lying outward of the loss site, the more general treatment allows the inward zones to expand also. If zone x is lost, freeing a volume $V_{\rm x}$, then this volume is partitioned according to

$$V_{\rm x} = V_{\rm a} + V_{\rm b}, \qquad \text{where} \tag{3.85}$$
$$\frac{V_{\rm a}}{V_{\rm b}} \equiv \frac{V_{\rm i}}{V_{\rm o}} \tag{3.86}$$
$$V_{\rm i} \equiv \sum_{j=0}^{x-1} V_j \qquad\qquad V_{\rm o} \equiv \sum_{j=x+1}^{j_{\max}-1} V_j \tag{3.87}$$

From equations (3.85) and (3.86),

$$V_{\rm a} = V_{\rm x}\left(1+\frac{V_{\rm o}}{V_{\rm i}}\right)^{-1}, \qquad\qquad V_{\rm b} = V_{\rm x}\left(1+\frac{V_{\rm i}}{V_{\rm o}}\right)^{-1}. \tag{3.88}$$

Zones inward of the loss site expand into the volume $V_{\rm a}$, and those outward of the loss site into the volume $V_{\rm b}$. Setting $V_{\rm a} \equiv 0$ allows only exterior zones to expand. Algebraically, individual zones expand as follows.

Interior zones:

$$\begin{aligned}
V_{\rm x-1} &\rightarrow V_{\rm x-1} + a\, V_{\rm a} \\
V_{\rm x-2} &\rightarrow V_{\rm x-2} + ar\, V_{\rm a} \\
&\vdots \\
V_{\rm j} &\rightarrow V_{\rm j} + ar^{x-j-1} V_{\rm a} \\
&\vdots \\
V_0 &\rightarrow V_0 + ar^{x-1}\, V_{\rm a}
\end{aligned} \tag{3.89}$$

Exterior zones:

$$\begin{aligned}
V_{\rm x+1} &\rightarrow V'_{\rm x} = V_{\rm x+1} + b\, V_{\rm b} \\
V_{\rm x+2} &\rightarrow V'_{\rm x+1} = V_{\rm x+2} + br\, V_{\rm b} \\
&\vdots \\
V_{\rm j} &\rightarrow V'_{\rm j-1} = V_{\rm j} + br^{j-x-1}\, V_{\rm b} \\
&\vdots \\
V_{\rm j_{max}-1} &\rightarrow V'_{\rm j_{max}-2} = V_{\rm j_{max}-1} + br^{j_{\max}-2-x} V_{\rm b}
\end{aligned} \tag{3.90}$$

Zones that lie outward of the loss site also have their indices shifted (primed quantities) so that there is no gap in the number sequence. For example, after zone x is lost, the old zone $x+1$ is renumbered as the new zone x, and so on. It is also necessary to reduce $j_{\max}$ to take account of the updated conditions, $j_{\max} \rightarrow j'_{\max} = j_{\max} - 1$.

Note that the geometric ratio parameter r is common to both series, but the normalizations a and b differ. Since there are x zones inward of the loss site,

$$S_{\rm x} = \frac{a(1-r^x)}{1-r} \qquad\qquad \Rightarrow \quad a = \frac{1-r}{1-r^x}. \tag{3.91}$$

There are $j_{\max} - x - 1$ zones exterior to the loss site (giving a total of $j_{\max} - 1$ remaining zones as required), and hence

$$S_{\rm j_{max}-x-1} = \frac{b(1-r^{j_{\max}-x-1})}{1-r} \qquad\qquad \Rightarrow \quad b = \frac{1-r}{1-r^{j_{\max}-x-1}}. \tag{3.92}$$

If the central zone is lost, $x = 0$ and equation (3.92) reduces to equation (3.84) as it should.

Equations (3.88) to (3.92) together specify the expansion of the zone volumes after the loss of zone x. Mass conservation gives the updated density in each zone. The expansion is assumed to be adiabatic, and hence the new zone pressures can be calculated from constancy of pV^{γ}. T then follows from the equation of state. The new zone radii are calculated from the volumes, working from the inside to the outer edge. If more than one zone cools out in any iteration, multiple passes through the loss algorithm are made.

3.6.1 Calculation of the Cooling Radius

Initially, there is no defined r_{cool}. If at any time t there exists a zone j such that $t_{\mathrm{cool,j}} < t < t_{\mathrm{cool,j+1}}$, then I linearly interpolate to find the cooling time at all points between,

$$t_{\mathrm{cool}}(r) = t_{-} + \frac{r - r_{-}}{r_{+} - r_{-}}(t_{+} - t_{-}),$$

where r_{-} and r_{+} are the central radii of zones j and $j+1$ respectively, with associates cooling times t_{-} and t_{+}. By definition, the cooling radius is that point where $t_{\mathrm{cool}} = t$, and therefore

$$r_{\mathrm{cool}}(t) = r_{-} + \frac{t - t_{-}}{t_{+} - t_{-}}(r_{+} - r_{-}). \tag{3.93}$$

Interpolating to find the cooling radius gives a smoothly evolving profile. A warning is issued if r_{cool} approaches r_{max}, as this would invalidate the constant pressure outer boundary condition. In the multiphase case (Chapter 6), or the inhomogeneous metallicity case (Chapter 4), each phase or abundance set has its own cooling radius.

3.7 The Gas Radiation

As shown in Section 1.5.1, the form of the cooling function Λ can be approximated to a reasonable degree of accuracy with simple analytic forms. I have not made use of such an approach — instead I have chosen to calculate the appropriate value of Λ in each simulation using the MEKAL plasma code, as discussed in Section 1.4.4. This allows for greater flexibility (e.g. one can vary any of the fifteen MEKAL elements without having to worry about the validity of any functional approximations to the cooling function), and the associated computational cost is small (given that MEKAL is required anyway for spectral output).

3.7.1 Practicalities of Using MEKAL

The source code for MEKAL is available as a part of the XSPEC (Section 1.4.5) package. The only source files required are `fmekal.f` and its associated include file `mekal.inc`. In addition, the spectral data files `mekaln.dat` (where n = 1–6) are needed. The data files are in FITS[2] (Flexible Image Transport System) format, therefore it is necessary to link with the CFITSIO[3] library routines. These form the only other (non-standard) external dependency. Actually, the unmodified `fmekal.f` needs to be linked with the XANADU library `libxanlib` as well. This is only in order to expand an environment variable and locate the spectral data files, using the routine `xsftop`. If the data files are copied to a fixed location and `fmekal.f` modified accordingly, `libxanlib` is no longer required.

[2] http://heasarc.gsfc.nasa.gov/docs/heasarc/fits.html
[3] http://heasarc.gsfc.nasa.gov/docs/software/fitsio/

`fmekal.f` has a simple input–output format. Defined as a subroutine, it takes as input: energy arrays with the lower and upper boundaries of each energy bin in keV; the number of energy bins; the hydrogen density in $\mathrm{cm^{-3}}$; the temperature in K; the abundances of each of the 15 included elements with respect to solar; and a normalization parameter. Internally, the solar abundances of Anders and Grevesse (1989) are used. To use a different solar abundance, the *relative* abundances input to `fmekal.f` must be adjusted accordingly. The outputs are the relative electron density, and an array containing the flux ($\mathrm{ph\,cm^{-2}\,s^{-1}\,keV^{-1}}$). Since it is a *flux* that is output, the input normalization is proportional to the emitting volume divided by the square of the source distance.

As an aside, I note at this point that an incarnation of the Raymond–Smith code is also distributed with XSPEC. This includes both source code and tabulated spectra, although the former is never actually used by XSPEC itself (i.e. unlike with MEKAL, one can only use pretabulated Raymond–Smith spectra in XSPEC). Unlike the MEKAL implementation, there are numerous (~ 30) Fortran source files for the Raymond–Smith code[4]. These can be compiled into a single library for subsequent use. The top-level routine is the subroutine `rayspec_m`. As well as temperature, electron density, abundance and energy binning (unlike `fmekal.f` the energy bins must be linearly-spaced, and there is also a compiled-in limitation on the maximum number that can be used), there are two input switches that can be set. These control whether charge transfer and density-dependent dielectric recombination (this has little influence at cluster densities) effects are active. The solar abundances are the same as those used by `fmekal.f` (expect the helium abundance is slightly different, and Na and Al are not included). The output spectrum is in emissivity units, $10^{-23}\, n_\mathrm{e}\, n_\mathrm{H}\, \mathrm{erg\,cm^3\,s^{-1}\,bin^{-1}}$. It can provide a useful comparison to have a second spectral code available.

Calculation of Emissivities and Cooling Functions

As we seek to calculate an emissivity ε per unit volume, the MEKAL input normalization factor is irrelevant as it cancels out of the final result:

$$\begin{aligned}\text{flux}\ \left(\mathrm{ph\,cm^{-2}\,s^{-1}\,keV^{-1}}\right) \quad &\times \quad 4\pi\left(\frac{d}{\mathrm{cm}}\right)^2\left(\frac{V}{\mathrm{cm}^3}\right)^{-1}\left(\frac{E}{\mathrm{erg\,ph^{-1}}}\right)\\ &\rightarrow \quad \text{spectral emissivity } \varepsilon_\nu\left(\mathrm{erg\,s^{-1}\,keV^{-1}\,cm^{-3}}\right)\end{aligned}$$

Alternatively, setting $V/d^2 = 4\pi\,\mathrm{cm}$ in the input normalization gives the (photon) emissivity directly. Integrating the spectral emissivity over energy (or frequency) gives the total emissivity (power per unit volume) ε. By the definition of the cooling function, equation (1.18), this is equal to $n_\mathrm{e} n_\mathrm{H}\Lambda = n_\mathrm{H}^2 r_\mathrm{e}\Lambda$. Since the relative electron density r_e is also returned by `fmekal.f`, the cooling function Λ may thus be calculated. The final result for Λ is independent of the density n_H as it should be, since the flux is of course proportional to n_H^2.

In this way, the cooling function $\Lambda(T)$ and the relative electron density $r_\mathrm{e}(T)$ (although as shown in Figure 1.4 on page 11 the magnitude of any change in r_e with temperature is slight) appropriate for the metallicity in question are calculated at the start of each simulation and stored on a grid of 100 logarithmically spaced temperature points over the range of interest ($\sim 10^5$–10^9 K). At the same

[4] An equivalent version not split into multiple files can be downloaded from `ftp://heasarc.gsfc.nasa.gov/software/plasma_codes/`.

time, the quantity $\int T\,\mathrm{d}T/\Lambda(T)$ which appears in the expression for the cooling time, equation (1.24) is also calculated and stored. Also calculated are an "X-ray cooling function" (obtained by limiting the range over which the spectral emissivity is integrated to 0.1–10 keV), and a "photon cooling function" (obtained by omitting the multiplication by photon energy). The former is used in calculating the total X-ray luminosity at any instant, and the latter for photon surface brightnesses (which cannot be derived from the *energy* surface brightnesses).

During the course of a simulation, the values of the cooling function, cooling function integral, and relative electron density appropriate for a zone of a given temperature are obtained by linearly interpolating amongst the tabulated values. The interpolation is made more efficient by exploiting the fact that the tabulated temperatures are known quantities. If they were not, it would be necessary to step through the tabulated arrays testing the stored temperatures against the input temperature each time — a slow process. Since the gridded temperatures are known, one can just step straight to the correct point. For example, if quantity y is evaluated and stored at n_x logarithmically-spaced points between $x_{\min}$ and $x_{\max}$, then the jth point is given by

$$y_j = x_{\min}\left(\frac{x_{\max}}{x_{\min}}\right)^{\frac{\mathrm{real}(j)}{n_x}} \qquad \Rightarrow \qquad j = \mathrm{int}\left[n_x \frac{\log\left(\frac{x}{x_{\min}}\right)}{\log\left(\frac{x_{\max}}{x_{\min}}\right)}\right]. \tag{3.94}$$

Since the `int` operator truncates downward, this returns the index j of the tabulated value y_j lying just below x. One can then linearly interpolate between j and $j+1$:

$$y(x) \approx y_j + \frac{x \quad - x_j}{x_{j+1} - x_j}(y_{j+1} - y_j). \tag{3.95}$$

There is a gain in speed through the use of this process, but a loss of flexibility over the general interpolation case, since the nature of the j grid used to store quantity y must be known at compilation time.

Calculation of Spectra

The `fmekal.f` routine and its output are all in single-precision (4 byte reals), therefore there is little point working to any greater degree of accuracy with spectral quantities and those derived from them.

Calling MEKAL to calculate spectra can account for a significant fraction of the program run-time. When a large number of zones have the same physical conditions, one is performing a slow calculation many times over in order to produce the same result each time. This is obviously inefficient, so I added an option to precalculate spectra at the start of each simulation. Spectra are calculated on a grid of $n_{T,\mathrm{spec}}$ temperature points (say 1000 logarithmically-spaced points from $\sim 10^5$ K to $\sim 10^9$ K) for the appropriate abundance, and for a density of $1\,\mathrm{cm}^{-3}$. Each individual spectrum is evaluated at $n_{\mathrm{bin,spec}}$ logarithmically-spaced energy points between 0.1 keV and 10.0 keV. Over the range of densities relevant to the ICM, there is no deviation from a density squared dependence, which is obviously extremely beneficial from the point of view of cutting down the required computations. During the course of a simulation, the spectrum appropriate for any particular zone is obtained from the stored grid by linearly interpolating between the stored temperature points, and scaling according to the square of the hydrogen density. This is obviously faster than calling MEKAL to calculate spectra each time they are needed, but more memory-intensive. With 4 bytes single-precision reals, the required memory is $4 \times n_{\text{bin, spec}} \times n_{T,\text{spec}} \approx 3.8\,\text{MB}$ for the default values (this is per abundance set — see Chapter 4). Any loss in spectral accuracy due to the interpolation is insignificant.

Note that there might in principle be some advantages to calculating the spectral contributions for each element separately, then summing the contributions in the required ratios. This would be of no real use for our purposes though, where all the abundance values are known in advance. In a similar fashion, there might be some advantage in storing the grid of precalculated spectra in files to be read in as needed. The saving in run-time would only be slight, though, since it is not the calculation of the spectral grid that dominates the run-time in practice. Obviously, there would also be reduced flexibility.

3.8 Surface Brightness

Consider projecting a line of sight (in the form of a cone expanding from the observer) towards and through an object, as shown in Figure 3.6(a) on the following page. If the distance to the object is much greater than its dimensions, then the line of sight will be cylindrical, to a good approximation, when it passes through the object.

The surface brightness S is defined such that the power emitted from the annular ring (spherical symmetry is assumed throughout) seen projected on the sky between radii b and $b + \mathrm{d}b$ is given by $2\pi b S(b)\,\mathrm{d}b$, i.e. the product of the surface brightness and the projected area on the sky.

If the object is viewed as being comprised of shells, then each shell with a radius $r > b$ contributes to $S(b)$. Consider the contribution made to $S(b)$ by radii between r and $\mathrm{d}r$. From Figure 3.6(b) on the next page, the element of length along the line of sight due to radii in this range is

$$\begin{aligned} \mathrm{d}t &= \sqrt{(r+\mathrm{d}r)^2 - b^2} - \sqrt{r^2 - b^2} \approx \sqrt{r^2 - b^2}\left[\sqrt{1 + \frac{2r\,\mathrm{d}r}{r^2 - b^2}} - 1\right] \\ &\approx \frac{r\,\mathrm{d}r}{\sqrt{r^2 - b^2}} \qquad \text{to first order.} \end{aligned} \tag{3.96}$$

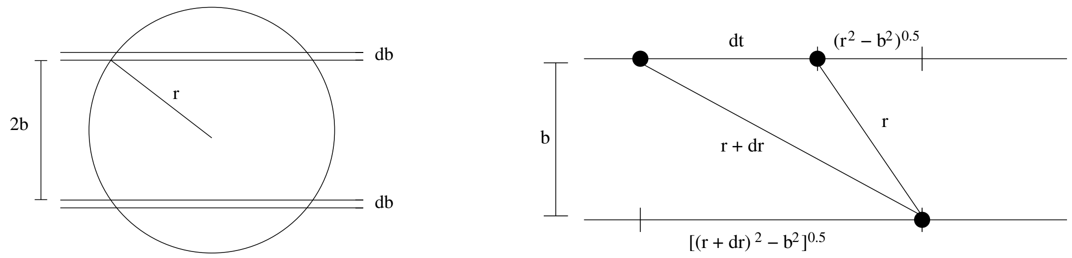
If $\mathrm{d}S$ is that part of $S(b)$ due to the shells between r and $r + \mathrm{d}r$, then

$$2\pi b\,\mathrm{d}b\,\mathrm{d}S = \varepsilon(r)\,\mathrm{d}V = 2\pi b\,\mathrm{d}b\,2\,\mathrm{d}t\,\varepsilon(r),$$

where $\varepsilon(r)$ is the emissivity (power volume^{-1}) of the volume element $\mathrm{d}V$ at radius r. The factor of $2\,\mathrm{d}t$ accounts for the fact that the line of sight intersects the shell of radius r at two points, in both a forward and backward direction. Again, if the dimensions of the object are negligible compared to the distance to it, then each of these contributions is the same. Hence

$$\begin{aligned} \mathrm{d}S &= 2\varepsilon(r)\,\mathrm{d}t \\ \therefore \quad S(b) &= \int_b^\infty \frac{\varepsilon(r) 2r\,\mathrm{d}r}{\sqrt{r^2 - b^2}}. \end{aligned} \tag{3.97}$$

Replacing the total emissivity ε by a spectral emissivity ε_ν leads to the "spectral surface brightness" S_ν.



(a) Line of sight **(b)** Element of thickness

Figure 3.6: The geometry of the surface brightness integral. A cylindrical line of sight of radius b, thickness $\mathrm{d}b$ is projected through a sphere. $\mathrm{d}t$ is the element of thickness along the line of sight due to spherical shells lying between radii r and $r + \mathrm{d}r$.

3.8.1 Inverting the Surface Brightness

In principle, it is possible to invert the surface brightness integral and recover the emissivity profile. It is shown in Appendix A.9 that

$$S(b) = \int_{b^2}^{\infty} \frac{\varepsilon(r)\,\mathrm{d}r^2}{\sqrt{r^2 - b^2}} \qquad \Rightarrow \qquad \varepsilon(r) = -\frac{1}{2\pi r}\frac{\mathrm{d}}{\mathrm{d}r}\int_{r^2}^{\infty} \frac{S(b)\,\mathrm{d}b^2}{\sqrt{b^2 - r^2}}. \tag{3.98}$$

This is a specific form of the *generalized Abel equation*[5],

$$f(x) = \int_{x}^{\infty} \frac{g(t)\,\mathrm{d}t}{(t-x)^{\alpha}} \qquad 0 < \alpha < 1 \qquad \Leftrightarrow \qquad g(t) = -\frac{\sin(\pi\alpha)}{\pi}\frac{\mathrm{d}}{\mathrm{d}t}\int_{t}^{\infty} \frac{f(x)\,\mathrm{d}x}{(x-t)^{1-\alpha}}. \tag{3.99}$$

3.8.2 Numerical Issues Concerning the Surface Brightness

There is an integrable singularity on the lower limit of the integral in equation (3.97) where $r = b$. When evaluating the integral numerically, this can cause problems. Since we essentially have the freedom to define $S(b)$ as being associated with either integer or half-integer radial mesh points, it might be tempting to sweep the issue under the rug; by, for example, defining $S(b)$ on half-integer points and taking contributions to the integral from integer points. This sleight of hand is not successful though — there is then a non-negligible discrepancy between the two alternative formulations for the integrated surface brightness presented in Section 3.8.3, with the double integral method of equation (3.101) giving significantly lower results. The notion of integer versus half-integer mesh points is one of convenience only, and it should not be used in an attempt to avoid mathematical difficulties.

Instead, one should transform the integral to remove the singularity (e.g. Press *et al.*, 1990, "improper integrals"). In the case of a simulated cluster extending out to radius $r_{\max}$, the approach is as follows.

$$S(b) = \int_{b}^{r_{\max}} \frac{2\varepsilon(r)\,\mathrm{d}r}{\sqrt{r^2 - b^2}}.$$

Let $y^2 = r^2 - b^2 \qquad \Rightarrow \qquad 2y\,\mathrm{d}y = 2r\,\mathrm{d}r$

[5]The "Abel" in question being Niels Henrik Abel (1802–1829), a Norwegian mathematician who also worked on quintic equations. The "Abell" more usually associated with clusters is the American astronomer, George Ogden Abell (1927–1983).

$$\therefore \quad S(b) = \int_0^{\sqrt{r_{\max}^2 - b^2}} \varepsilon(\sqrt{y^2+b^2})\,\frac{2y\,\mathrm{d}y}{y}.$$

So, numerically one would have (by the trapezium rule)

$$S(b) = 2\left[{}^1\!/\!{}_2(\varepsilon_0+\varepsilon_1)(y_1-y_0) + {}^1\!/\!{}_2(\varepsilon_1+\varepsilon_2)(y_2-y_1) + \cdots\right],$$

where $y_j = \sqrt{r_j^2 - b^2}$, $b < r_j < r_{j\max}$, and $\varepsilon_j = \varepsilon(\sqrt{y_j^2+b^2}) = \varepsilon(r_j)$.

This approach associates the contributions to the integral, and hence $S(b)$, with integer mesh points. Since ε is stored on half-integer mesh points, it must be interpolated (or extrapolated at the outer edge).

3.8.3 Integrating the Surface Brightness

It is of great interest to be able to integrate the spectral surface brightness profile between any two specified radii, b_1 and b_2 say. In this way, one obtains the spectrum due to an *annulus* of the source. Breaking an image down into annuli is often used as a method to analyse real data (e.g. Schmidt *et al.*, 2001, 2002; Sanders and Fabian, 2002). The total power P received from material lying between projected radii b_1 and b_2 is obtained by integrating the surface brightness over area

$$P(b_1 \leqslant b \leqslant b_2) \equiv \int_{b_1}^{b_2} 2\pi b S(b)\,\mathrm{d}b \tag{3.100}$$

$$= \int_{b_1}^{b_2} \mathrm{d}b\, 2\pi b \int_b^\infty \mathrm{d}r\, \frac{\varepsilon(r)2r}{\sqrt{r^2-b^2}}. \tag{3.101}$$

With reference to Figure 3.7 on the following page, the order of the integrations can be interchanged as follows:

$$\int_{b_1}^{b_2} \mathrm{d}b \int_b^\infty \mathrm{d}r = \int_{b_1}^{b_2} \mathrm{d}r \int_{b_1}^{r} \mathrm{d}b + \int_{b_2}^{\infty} \mathrm{d}r \int_{b_1}^{b_2} \mathrm{d}b. \tag{3.102}$$

Hence

$$\begin{aligned}
P(b_1 \leqslant b \leqslant b_2) &= 4\pi \int_{r=b_1}^{b_2} \mathrm{d}r\, \varepsilon(r) r \int_{b=b_1}^{r} \mathrm{d}b\, \frac{b}{\sqrt{r^2-b^2}} + 4\pi \int_{r=b_2}^{\infty} \mathrm{d}r\, \varepsilon(r) r \int_{b=b_1}^{b_2} \mathrm{d}b\, \frac{b}{\sqrt{r^2-b^2}} \\
&= 4\pi \int_{r=b_1}^{b_2} \mathrm{d}r\, \varepsilon(r) r \left\{ -\left[\sqrt{r^2-b^2}\right]_{b=b_1}^{r} \right\} + 4\pi \int_{r=b_2}^{\infty} \mathrm{d}r\, \varepsilon(r) r \left\{ -\left[\sqrt{r^2-b^2}\right]_{b=b_1}^{b_2} \right\} \\
&= 4\pi \int_{r=b_1}^{b_2} \mathrm{d}r\, \varepsilon(r) r \sqrt{r^2-b_1^2} + 4\pi \int_{r=b_2}^{\infty} \mathrm{d}r\, \varepsilon(r) r \left[\sqrt{r^2-b_1^2} - \sqrt{r^2-b_2^2}\right]
\end{aligned}$$

$$\therefore \quad P(b_1 \leqslant b \leqslant b_2) = 4\pi[\mathcal{F}(b_1) - \mathcal{F}(b_2)], \qquad \text{where} \tag{3.103}$$

$$\mathcal{F}(b) \equiv \int_b^\infty \mathrm{d}r\, \varepsilon(r) r \sqrt{r^2-b^2}. \tag{3.104}$$

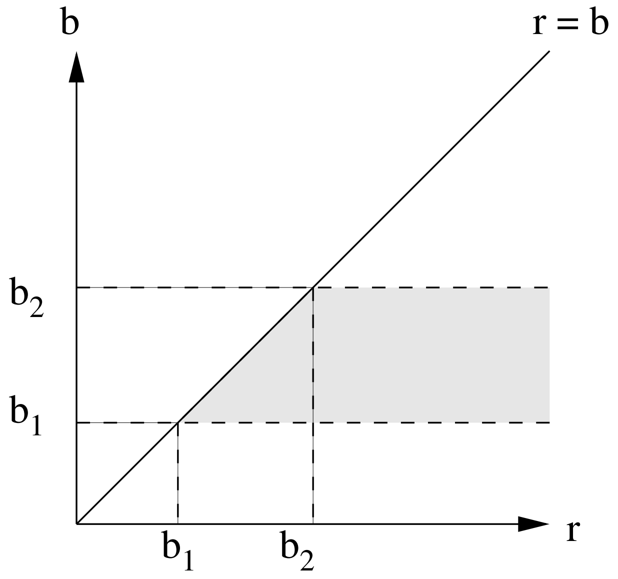


Figure 3.7: Interchanging the order of the surface brightness integrations.

For the special case of an isothermal, isodensity cluster of radius R,

$$P(0 \leqslant b \leqslant \infty) = 4\pi\mathcal{F}(0) = 4\pi \int_0^\infty r^2 \varepsilon(r)\,\mathrm{d}r = {}^4\!/_3\pi R^3 \varepsilon.$$

This result is useful as a gross test of the numerics (though it is of course impossible to generalize from the special case of constant density and temperature to a more typical setup).

A Comparison of the Two Integration Methods

The transformation of equation (3.101) into equation (3.103) is of more than academic interest. From a computational point of view, using equation (3.103) is more appealing. Reversing the order of the integrations allows one of the integrals to be evaluated analytically, so that there is less work to be done numerically. It also has the advantage that the potentially troublesome integrable singularity, as discussed in Section 3.8.2, is removed. If equation (3.101) is used, then the quantity $\varepsilon(r)$, for $r > b$, is needed for every different value of b in the outer integral. One must either accept the necessity of recalculating ε many times, or attempt to store the value of ε at each value of r. Taking this latter approach can be quite memory-intensive, because what is actually required here (in order to calculate a spectrum, rather than just the total power) is ε_ν, i.e. the *spectral* surface brightness, which in practice has ~ 1000 frequency components at each radial point. One could implement some kind of mixed scheme, whereby values of ε_ν were cached at as many points as practical (it would make more sense to cache those points at large radius, since these are used for more values of b), but that is probably an unnecessary level of complication.

The disadvantage of the method represented by equation (3.103) is a lack of flexibility. The radii of the spectral annuli must be specified in advance. With the method of equation (3.101), one can output a *spectral surface brightness profile*, essentially leaving the second integral to be done by an external program. This enables one to re-bin the spectral annuli as required; for example, to obtain a given number of counts in each annulus (under some specified observing conditions). This can involve writing large files, though. Actually, what is written out is not the spectral surface brightness per se, but rather for each zone the product of this quantity with its projected area; that is, the contribution that this zone would make to a trapezium-rule style integral over surface brightness to form a spectral annulus. This multiplication always has to be done (there is no use for the spectral surface brightness itself), so it may as well be done in the main program where greater accuracy is available.

In summary, equation (3.103) is the more efficient approach, but equation (3.101) is more flexible. As is often the case, one must compromise according to requirements. When the issues of Section 3.8.2 are taken into account, the agreement between the two different methods is excellent (even in the case of fairly low spatial resolution), which is a reassuring test of the numerics.

3.8.4 Conversion to Observed Surface Brightness

The quantity $S(b)$ is a *theoretical* surface brightness. It has units of the form $\mathrm{erg\,s^{-1}\,cm^{-2}}$, and refers to the luminosity emitted per unit projected area *at the source*. That is (again for the spherically symmetric case), $\Delta I = 2\pi b S(b)\Delta b$ is the luminosity emitted by an annulus of radius b, width Δb.

A *measured* surface brightness S', on the other hand, is most commonly expressed per solid angle, e.g. $\mathrm{ct\,s^{-1}\,arcmin^{-2}}$, and refers to the luminosity as received by an *observer*. In order to properly compare simulation and observation, these two forms of the surface brightness must be converted.

If a source of photon luminosity ΔI_{ph} is placed at a luminosity distance [see equation (A.76)] D_{L}, then by definition

$$\Delta F_{\mathrm{ph}} = \frac{\Delta I_{\mathrm{ph}}}{4\pi D_{\mathrm{L}}^2}$$

is the photon flux received at earth. If observed with a detector of effective area A_{eff}, $A_{\mathrm{eff}}\Delta F_{\mathrm{ph}}$ $\mathrm{ct\,s^{-1}}$ will be detected. In reality, A_{eff} is a function of energy of course (Figure 1.1 on page 3), but one can make use of an average value.

The observed photon surface brightness per unit solid angle is then given by

$$S'_{\mathrm{ph}} = \frac{A_{\mathrm{eff}}\Delta F_{\mathrm{ph}}}{\Delta\Omega},$$

where $\Delta\Omega$ is the solid angle subtended by the emitting area at the detector. By the definition of solid angle, this is related to the physical area of the source by

$$\Delta\Omega = \frac{\Delta A}{D_{\mathrm{A}}^2},$$

where D_{A} is the angular diameter distance, as defined in equation (A.74). Hence

$$S'_{\mathrm{ph}} = D_{\mathrm{A}}^2 \frac{A_{\mathrm{eff}}\Delta F_{\mathrm{ph}}}{\Delta A} = D_{\mathrm{A}}^2 \frac{A_{\mathrm{eff}}\Delta I_{\mathrm{ph}}}{4\pi D_{\mathrm{L}}^2 \Delta A} = D_{\mathrm{A}}^2 \frac{A_{\mathrm{eff}} S_{\mathrm{ph}}}{4\pi D_{\mathrm{L}}^2}$$

$$\therefore \quad S'_{\mathrm{ph}} = \left(\frac{D_{\mathrm{A}}}{D_{\mathrm{L}}}\right)^2 \frac{A_{\mathrm{eff}} S_{\mathrm{ph}}}{4\pi} = \frac{A_{\mathrm{eff}} S_{\mathrm{ph}}}{4\pi(1+z)^4}, \tag{3.105}$$

where use has been made of equation (A.77). This relates the observed surface brightness S' (measured per rad^2) to the calculated surface brightness S at the source (per unit area). Note that since the result depends only on the ratio $D_{\mathrm{A}}/D_{\mathrm{L}}$ it is independent of cosmology. Using equation (B.5),

$$\frac{S'_{\mathrm{ph}}}{\mathrm{arcmin}^2} = \left(\frac{\pi}{60 \times 180}\right)^2 \frac{A_{\mathrm{eff}} S_{\mathrm{ph}}}{4\pi(1+z)^4}. \tag{3.106}$$

Note that it is not possible to extract the *photon* surface brightness S_{ph} ($\mathrm{ph\,s^{-1}\,cm^{-2}}$) from the *energy* surface brightness S ($\mathrm{erg\,s^{-1}\,cm^{-2}}$), since the spectral information has been lost. That is, it is not possible to know if a given flux of energy is caused by a few high-energy photons or many low-energy photons. So the photon surface brightness $S_{\mathrm{ph}}(b)$ must be calculated separately.

3.8.5 Verification of the Surface Brightness

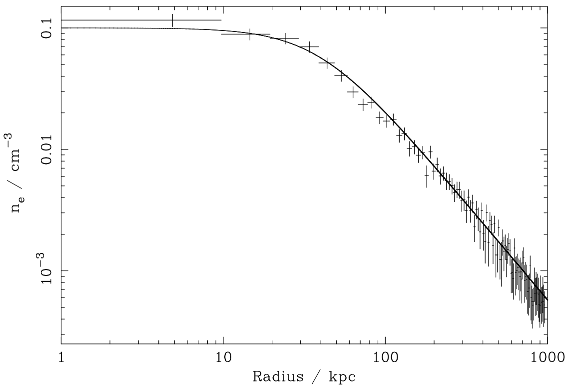
Equation (3.106) provides a method for verifying the calculation of the surface brightness profile, since it allows for direct comparison with observational results. Observational data for Abell 1835 were kindly supplied by Robert Schmidt — namely the electron density profile and the (deprojected) temperature profiles, figures 13 and 8 of Schmidt *et al.* 2001. As described in Section 3.5.2, I fitted the observed density and temperature profiles with a β and η model respectively, the results of which are shown in Figure 3.8 on the facing page. Assuming a constant metallicity of 0.3 $\mathrm{Z}_{\odot}$ (figure 7 of Schmidt *et al.* 2001), and using $j_{\mathrm{max}} = 10000$, these profiles were then used to form the initial conditions. The resulting surface brightness profile as it would be observed was calculated assuming an energy-independent effective area of $A_{\mathrm{eff}} = 250\,\mathrm{cm}^2$ as an appropriate average for the *Chandra* ACIS-S — see Figure 1.1 on page 3. The correspondence between the calculated and the observed surface brightness profiles is gratifying, and strongly suggests that both the setting up of the initial conditions and the evaluation of the surface brightness integral are operating correctly.

Note that the situation here is the reverse of the observational one, where the surface brightness profile is the measured quantity, and the temperature and density are derived quantities.

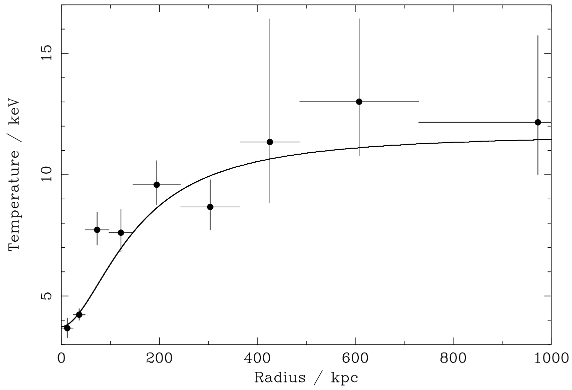
3.9 Simulated Observations

The spectral annuli files output by the simulation are used to produce synthetic observations. The spectral files ($\mathrm{erg\,s^{-1}\,keV^{-1}}$) are first converted to a flux ($\mathrm{cm^{-2}}$) appropriate for a given redshift, making use of the luminosity distance as given in equation (A.76). These ASCII files are then converted to XSPEC table model format (Arnaud, 1999) FITS files. The various library routines in the CFITSIO package make the format conversion fairly straightforward. The approach adopted is the simplest one, in which each single spectrum is converted to a single additive table model, with no free parameters other than normalization.

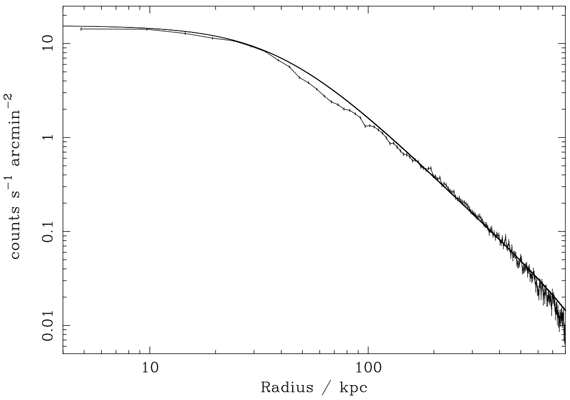
The resulting FITS files can be read into XSPEC using the `model atable` command, and the `fakeit` command then used to produce simulated observations incorporating the effects of random Poisson noise (in principle it is also possible to add a background, but I have not done this). One can add further model components before producing the fake observations; for example a `phabs` component to represent the effects of (Galactic) photoelectric absorption, say. The faking process requires specification of



(a) Electron density



(b) Temperature



(c) Surface brightness

Figure 3.8: Observed and modelled profiles of Abell 1835. Observational data provided by Robert Schmidt. (a, b, c) correspond to figures 13, 8, and 3 respectively of Schmidt *et al.* (2001). Observational error bars are 1σ. (a, b) show fitted β and η profiles; (c) shows the resulting calculated surface brightness, assuming $Z = 0.3\,Z_{\odot}$, $A_{\mathrm{eff}} = 250\,\mathrm{cm}^2$.

the observation time, and provision of redistribution matrix (RMF) and ancillary response (ARF) files[6]. For response files I have made use of both *Chandra* (specifically, the AO-2 proposal planning files[7]) and *XMM-Newton* (specifically, the AO-1 proposal planning files[8]) files. The utility `grppha` from the FTOOLS collection is used to group the PHA[9] (Pulse Height Analyser) files that are the result of the faking process to a minimum of 20 counts per bin to ensure applicability of the χ^2 statistic when fitting models to these data.

A common observational procedure is to divide an image into annuli such that each contains an equal number of counts, say 10000 to give a number useful for spectral analysis. It is possible to achieve the same effect with the simulated observational data, though the method is not entirely straightforward and hence is outlined here. It is immediately obvious that this is a major disadvantage of the method of

[6] The format of RMF and ARF files is described in George *et al.* (1998) and its addendum George and Arnaud (1998a).

[7] `http://asc.harvard.edu/caldb/Aeff/`

[8] `http://xmmssc-www.star.le.ac.uk/ukos/ao1tools.html`

[9] The format of PHA files is described in Arnaud *et al.* (1998) and its addendum George and Arnaud (1998b).

calculating spectral annuli files represented by equation (3.103). In that approach, the radii of the annuli are fixed in advance. It is not therefore possible to achieve a binning in which the size of each annulus is controlled by the number of counts it contains. Even if it were somehow possible to calculate these radii in advance, this would have the disadvantage that the simulation output would then no longer be independent of the detector used in making the fake data files (since this obviously affects the number of counts received) — it is more aesthetic to keep the two stages of simulated evolution and simulated observation separate. Furthermore, the required radial positions would evolve with the simulated system in a way that could not be predicted in advance. This is the major reason why it is preferable to write a spectral surface brightness file that may be binned into annuli separately as desired. The task of binning into equal count radii is handled as follows.

At the same time as the spectral surface brightness file (containing the contribution of each zone to an annulus integral) is written, a count-rate file is produced, giving the number of photons per second emitted by each zone; i.e.

$$\text{count-rate}(j) = \sum_{E} \frac{S_{\mathrm{E}}(j)\Delta A(j)\Delta E}{E}, \tag{3.107}$$

where $S_{\mathrm{E}}(j)$ is the spectral surface brightness of zone j, $\Delta A(j)$ is its projected area, and ΔE is the width of the energy bin centred at energy E. Given such a file, one can easily sum the count-rate to find the total for the region of interest, and hence calculate the radial positions appropriate for a division into equal count annuli. Normally, the integration time constrains the maximum number of counts available, but we have the luxury of selecting a fake integration time to give the required number of counts per annulus. Using a mean, energy-independent effective area appropriate for the detector in question leads to the required counts per unit area at the detector. Multiplying by $4\pi D_{\mathrm{L}}^2$ gives the required counts at the source. Dividing this by the count-rate at the source results in the necessary integration time.

This method is algorithmically simple and therefore easy to script, but it suffers from a lack of flexibility. It fails if the detector effective area exhibits a significant energy dependence, since then a simple bulk count rate is not an adequate way to calculate integration times. It does not allow for an arbitrary distribution of counts between annuli. Achieving both these goals would require a procedure that binned some number of zones into an annulus, faked the data in XSPEC and derived the number of counts in the resulting observation, then added or subtracted zones from the annulus until convergence on the desired number of counts was obtained.

3.10 Summary of the Computational Scheme

The following flowchart[10] presents a summary of the code. The sections relating to metallicity variations (Chapter 4), thermal conduction (Chapter 5), and multi-phase issues (Chapter 6) are included for completeness, even though these issues have not been discussed yet.

[10] This is felt to be mildly more palatable than the presentation of pages of Fortran 90 comparable in length to this entire thesis.

START

Parameters
Λ: $T_{\text{min}} = 5 \times 10^4\,\text{K}$, $T_{\text{max}} = 10^9\,\text{K}$, $E_{\text{min}} = 5\,\text{eV}$, $E_{\text{max}} = 200\,\text{keV}$.
$\gamma = {}^5\!/_3$, $r_{\text{min}} = 0.1\,\text{pc}$, $f_{\text{stab}} = 0.4$, $f_{\text{cool}} = 0.1$, $T_{\text{cut}} = 10^5\,\text{K}$, $q_{\text{c}} = 2.0$, $\Delta t_{\text{min}} = 1.0\,\text{yr}$,
mass-loss geometric series ratio $r = 0.9$.

USER INPUT

Cluster mass profile
King, NFW, or empty (i.e. no gravity, useful for checks).
For NFW profile, choose either EMN or ASF scaling relations (with choice of cosmology and virial overdensity). Specify T_{virial} and c, obtain r_{virial} and M_{virial}. Alternatively, specify "by hand" a reference radius and the mass contained within it, and the profile scale radius.

Numerical parameters
r_{max}, j_{max}, n_{ph}
Number of outputs for profiles, and global properties (e.g. luminosity).
Total number of iterations, or simulated time to reach.

Gas properties
If empty profile, specify pressure. Otherwise:
f_{gas} (default 0.17), T_{gas} (default T_{virial}).

Conductivity parameters
(See Chapter 5)

Spectral output (optional)
Precalculate spectra on a grid? Number of spectral outputs; write spectral surface brightness out to some radius; or specify radii for spectral annuli (read from file, or assign with linear or log spacing).

VFDD output (optional)
Radii at which to calculate (read from file, or assign with linear or log spacing).

Abundances
Solar abundance (default MEKAL) read from file.
Mean relative abundances (H, He usually solar).
Gradient? If yes, specify central abundance, number of steps, index for gradient function, maximum radius of gradient region.
Bimodal profile? If yes, specify high and low metallicity as multiples of mean metallicity.

INITIALIZATION

Allocate and initialize array storage

Abundance calculations
If a gradient is in use, calculate radii where steps occur. With a bimodal profile, set the appropriate fraction of zones high and low. Set up abundance mask recording which abundance is appropriate for each phase in each zone. For each abundance set, calculate mass fractions, relative electron density and mean molecular weights (for full ionization).

Initial MEKAL calculations
Calculate spectral grid (if required), Λ, Λ_{X}, Λ_{phot}, $\int T\,\mathrm{d}T/\Lambda(T)$, r_{e}.

Initial conditions
$j_{\max}+1$ radial grid points between $r_{\min}$ and $r_{\max}$. Calculate zone volumes. Multi-phase case: assign phase volumes as per vfdd. Pressure profile: read from file; or isobaric (empty profile); or hydrostatic. Density profile: read from file; or scaled version of pressure profile. Adjust outer boundary pressure to normalize gas mass to correct f_{gas}. Establish dummy mesh point for outer boundary conditions. u, q initialized to zero.
Auxiliaries: T from equation of state; interpolate Λ etc. from stored values.
Or, load in save file of previous configuration if present.

ITERATION LOOP

Evolve physical variables
$t \rightarrow t+\Delta t$. Update u, r, V, $\bar{\rho}$, q, p, ρ_i, V_i, T, and other auxiliaries.

Safety checks
Positive pressure, excess cooling, volume conservation, energy conservation.

Conduction
Calculate heat flux across shells, and heat power from divergence.

New time step
Calculate next Δt based on stability, radiative cooling, and conductive times for each zone. Warn if several consecutive very small time steps.

Quantities only needed on output
r_{cool}, bulk X-ray luminosity, $\dot{M}$, surface brightness (energy and photon), gas fraction profile.
Spectra: interpolate from stored, or calculate, as appropriate. Integrate for spectral annuli, or write spectral surface brightness profile.
Conduction: sum of luminosity from within radius over flux across boundary shell.
Output bulk of data: u, ρ, p, T, $\mathcal{L}$, t_{cool} profiles.

Mass loss
If any phase cooled below T_{cut} record details of mass lost. Expand remaining phases and zones. If zone(s) lost, shift indices of remaining zones.

End of iteration loop
Check for signal file indicating output buffers should be flushed. Exit iteration loop if stop file encountered, or end-time reached. Cycle new quantities to old, empty new, and return to top of iteration loop.

CONCLUSION

Final issues
Write save file to enable simulation to be restarted from this point. Details of mass loss, zones lost, time reached, time acceleration factor. Close output files.

STOP

If you don't fear God, fear him through the metals.

One Hundred Years of Solitude
GABRIEL GARCÍA MÁRQUEZ

4 The Evolution of Cooling Flows with Metallicity Variations

It was shown in Chapter 2 that interesting spectral effects can arise through the partition of a plasma into components of differing metallicity. The treatment in that Chapter was a static one, in that there was no consideration of the time evolution of such a plasma. The numerical method described in Chapter 3 provides a framework that allows us to revisit this issue in more detail. In this Chapter, I investigate what can happen when a mixed-metallicity plasma is allowed to evolve in the context of the cooling flow paradigm.

4.1 Adding Chemical Inhomogeneities to the Numerical Scheme

I describe here modifications to the numerical scheme of Chapter 3 that allow for the modelling of an ICM with a spatially varying abundance. Each zone (and each phase within a zone in the multi-phase case, see Chapter 6) is allowed to have a different abundance of any of the elements included in MEKAL. In practice, one almost always has the helium abundance fixed at solar, and varies the other elements *en masse* (as controlled by the metallicity Z), but the full flexibility is there, at no extra cost.

4.1.1 Initial Conditions

The specification of the initial pressure and density profile is largely unchanged. The only difference is, when the metallicity is allowed to vary spatially, then the molecular mass ratio μ does so too. The magnitude of the effect is not too large, as shown in Table B.4 on page 162, but for the sake of generality if nothing else I take account of it. Recall from Section 3.5.1 that the first step in the initialization is to specify a pressure (arbitrary) and gas temperature (usually the cluster virial temperature) at the outer boundary, then calculate the density using the equation of state. This requires a value of μ to be chosen. In practice, in the inhomogeneous metallicity case I use a "mean" μ, so that, as in the multi-phase case (see Section 6.5), the "mean" profile is in hydrostatic equilibrium. The mean μ is calculated as the value appropriate for some mean abundance set, which must be specified (usually just the standard $0.3\,\mathrm{Z}_\odot$ abundance set). It is not really possible to *calculate* a suitable mean value of μ without a mean Z, since this would depend on the density profiles, which are themselves functions of μ of course. Once the density profiles have been set up, a true mean μ can be calculated according to

$$\bar{\mu} \equiv \frac{M_{\mathrm{total}}}{m_{\mathrm{H}} N_{\mathrm{total}}} \tag{4.1}$$

$$\bar{\mu}_j = \frac{\sum_i \rho_{i,j} V_{i,j}}{\sum_i \frac{\rho_{i,j} V_{i,j}}{\mu_{i,j}}} \qquad \text{(for zone } j\text{)} \qquad\qquad \bar{\mu} = \frac{\sum_j \sum_i \rho_{i,j} V_{i,j}}{\sum_j \sum_i \frac{\rho_{i,j} V_{i,j}}{\mu_{i,j}}} \qquad \text{(global)} \tag{4.2}$$

This can then be compared with the value that was used to set up the density profile, to check any discrepancy is not significant.

As described in Section 3.5.1, a smooth pressure profile is first calculated, and the density profile is then assigned as a scaled version of this (thus neglecting the possibility that the metal-rich gas might be intrinsically denser). Consequently, if μ varies spatially, then the initial temperature profile will not be isothermal. In accordance with the equation of state, equation (3.20), it is the quantity T/μ which must be independent of position. Therefore, regions of high μ (i.e. metal-rich regions) are initially hotter (see for example the initial conditions in Figure 4.3 on page 91). As discussed in Section 4.3, this is unimportant.

4.1.2 Assigning Two-component Metallicities

We wish to model a plasma along the lines discussed in Chapter 2, where there are to be two components, one metal rich, one metal poor. Suppose that 10 per cent of the gas is to be made metal rich. Obviously, this can be achieved by making every tenth radial zone metal rich. More generally, if a fraction f_{hi} of zones are to be metal rich, then this is achieved by enriching every Δjth zone, where $\Delta j = \mathtt{int}\left(f_{\mathrm{hi}}^{-1} + 0.5\right)$, the $+0.5$ term being used since the `int` operator rounds down. The actual fraction of zones which are metal rich is then $1/\Delta j$. Since Δj is constrained to be an integer, this is not necessarily exactly the same as f_{hi}. I tend to avoid making the central zone metal-rich.

Note also that f_{hi} as defined in equation (2.13) is a *mass* fraction. Making a fraction f_{hi} of *zones* metal rich is not strictly the same thing (it is closer to a *volume* fraction), although in practice the distinction is unimportant.

4.1.3 The Numerical Equations

There are no conceptual modifications to be made to the difference equations of Chapter 3. The implementation is made slightly more complex because it is necessary to keep track of the metallicity Z appropriate for each phase, and hence the associated value of Λ, μ, and so on.

For each abundance set in use, values of $\Lambda(T)$ etc. are calculated at the start of a simulation and stored

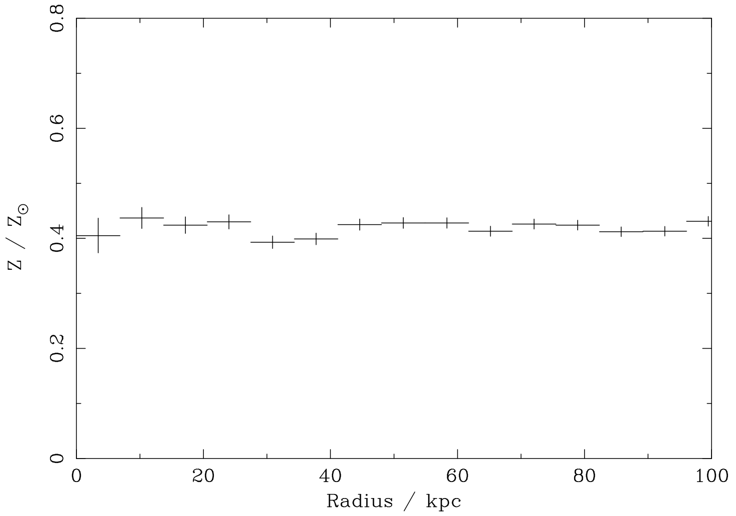


Figure 4.1: Fitting the metallicity of an isothermal ICM with spatially varying abundance. Simulation parameters: $f_{\rm hi} = 0.1$, $\bar{Z} = 0.5$, $Z_{\rm lo} = 0$, 10^8 K cluster. Observed data faked and fitted using the ACIS-S matrices, fitting to the 0.3–7.0 keV energy range. No absorption component was added in this case. Vertical error bars are 1σ.

on grids. A "mask" array is constructed which records the abundance set appropriate for each phase in each zone. It is obviously more efficient to do this (a poor man's hash table) rather than store the actual abundances, Λ, etc., in every case; something that would clearly involve a lot of unnecessary duplication of information.

With hindsight, it perhaps would have been more efficient to exploit the fact that the cooling function can be separated into a hydrogen–helium part and a metal part, as described in Section 2.3.2. Then rather than storing a grid of $\Lambda(T)$ values for discrete Z sets, the linear dependence of the metal part of Λ on Z, equation (2.8), could have been exploited to calculate $\Lambda(Z)$ as needed. This would probably be slower, though, and certainly less flexible, in that one could not vary *individual* element abundances, only the bulk metallicity.

There are various other adaptations that are more cosmetic than anything else: keeping track of how much mass has cooled out for each abundance set; computing separate cooling radii for each set at any instance (which turns out to be a useful diagnostic — see Figure 4.11 on page 101 for example); and so on.

4.2 Fitting to the Initial Conditions

As a verification of the basic premise, I show in Figure 4.1 the results of XSPEC fits to the initial conditions of one particular simulation. The details of the simulated cluster are not important at this point, since I am merely demonstrating proof of concept. Simulated observations were produced from the initial conditions according to the method described in Section 3.9. The resulting fake observations were then fitted in XSPEC, using a single-temperature `mekal` model for each annulus; with the temperature, abundance and normalization as free parameters (the redshift was fixed at the correct value), and using the 0.3–7 keV energy range of the ACIS-S detector.

Two facts are immediately apparent from Figure 4.1. First, the fitted abundance profile is flat (within

Z	r_e	X_H	V	K
0	1.195	0.719	0.9	0.556
5	1.264	0.654	0.1	0.054

Table 4.1: The properties of metal-rich and metal-poor zones in Figure 4.1 on the previous page.

the error bars), with no high-metallicity spikes. This confirms the basic premise that a simulated isothermal plasma with a spatially-varying abundance can produce results that are observationally indistinguishable from the homogeneous case.

Second, it is clear that the initial conditions are fit by a uniform abundance of around $0.4\,\mathrm{Z}_\odot$. This is significantly lower than the value of $0.5\,\mathrm{Z}_\odot$ that would probably be expected for a combination of $f_\mathrm{hi} = 0.1$, $5.0\,\mathrm{Z}_\odot$ gas together with the zero-metallicity gas that constitutes the bulk of the ICM in this case. There are two factors contributing to this apparent discrepancy.

Firstly, the XSPEC normalization for the `mekal` model is defined as

$$K = \frac{10^{-14}}{4\pi D_\mathrm{L}^2} \int n_\mathrm{e} n_\mathrm{H}\,\mathrm{d}V, \tag{4.3}$$

where D_L is the luminosity distance. By definition of the hydrogen mass fraction (Section A.7), $n_\mathrm{H} = \rho X_\mathrm{H}/m_\mathrm{H}$. Using this, together with $n_\mathrm{e} \equiv r_\mathrm{e} n_\mathrm{H}$, and assuming a constant mass density (which is valid for neighbouring metallicity regions in the case of a smooth density profile), leads to

$$K \propto r_\mathrm{e} X_\mathrm{H}^2 V \tag{4.4}$$

as the normalization appropriate for some uniform volume V. Thus, there are extra factors besides the straightforward volume ratio that should be taken into account. In the case of Figure 4.1 on the previous page, the metal-rich and metal-poor zones have properties as listed in Table 4.1.

Thus the expected mean abundance from such a mixture would be

$$\bar{Z} = \frac{K_1 Z_1 + K_2 Z_2}{K_1 + K_2} = \frac{(0.556 \times 0) + (0.054 \times 5)}{0.556 + 0.054} \approx 0.44. \tag{4.5}$$

Qualitatively, the metal-rich zones obviously have a higher relative electron density; but their equivalent hydrogen number density is lower (for a fixed mass density) due to their higher mean molecular weight.

This explains only a part of the reduction of the mean abundance. The remainder is explained by recalling that the metal-rich zones are initially somewhat hotter (see the initial conditions in Figure 4.3 on page 91), and consequently in the regime where line emission is less important. This issue (which is discussed in more detail in Section 4.3.1) explains the rest of the decrease. Indeed, combining two `mekal` models in XSPEC with the appropriate metallicities, temperatures, and normalizations, faking and fitting with a single `mekal` model leads to an abundance of $0.42\,\mathrm{Z}_\odot$. We can therefore have confidence in the modelling and fitting of the initial conditions in the simulations at least.

4.3 Results for the Evolution of the Intracluster Medium

I present here the results for one particular cluster with an 8.6 keV (10^8 K) virial temperature[1]. The cosmological parameters were $h_{50} = 1$, $\Omega_{\mathrm{M}} = 1.0$, $\Omega_{\Lambda} = 0.0$, and $z = 0.017$ ($D_{\mathrm{L}} = 100\,\mathrm{Mpc}^1$). From equation (3.65), the virial radius for such a halo is 3.4 Mpc, and the halo mass is $2.3 \times 10^{15}\ \mathrm{M}_{\odot}$. The radial range (out to the virial radius) was modelled with 5000 zones, corresponding to a resolution of 700 pc, or around 1.5 arcsec for the redshift in question (where 1 arcsec = 0.48 kpc). Using a concentration of 5 (e.g. Bullock *et al.*, 2001), the NFW scale radius is 700 kpc. From equation (A.49), the cluster velocity dispersion is $1300\,\mathrm{km\,s^{-1}}$. The gas fraction was 0.17 (e.g. Ettori and Fabian, 1999). The metals are distributed such that every zone contains hydrogen and helium in the solar abundance ratio, and in addition every tenth zone has a three times solar abundance of the heavy elements. When producing the fake observations, the effect of Galactic absorption was added via a `phabs` component, with $N_{\mathrm{H}} = 10^{21}\,\mathrm{cm}^{-2}$. Noise was added via counting statistics, but no background file was included. The resulting fake data were grouped to a minimum of 20 counts per channel to ensure applicability of the χ^2 statistic.

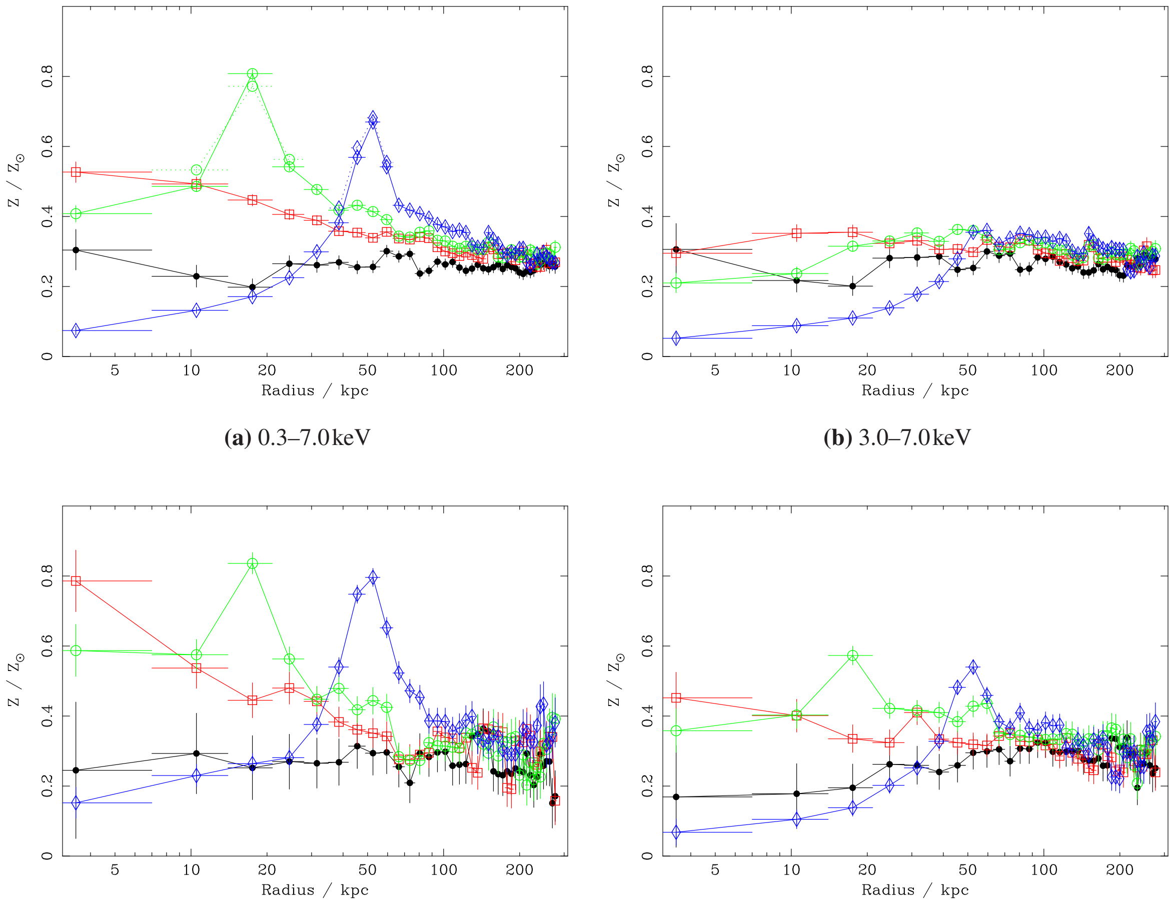
Figure 4.2 on the next page illustrates the derived abundance profiles as obtained from various XSPEC fits to the simulated observations produced for this system. Only the central regions where cooling has progressed significantly in the elapsed time are shown. Each spectral annulus was fit with a `mekal` model, multiplied by a `phabs` component. The redshift z and column density N_{H} were fixed at the correct values, and the temperature, abundance and normalization were left free.

Shown in Figure 4.3 on page 91 is the time-evolution of the true temperature profile. As discussed in Section 4.1.1, with the system adopted, the metal-rich zones are hotter at the outset. It is not unreasonable to suppose that metal-rich zones are initially hotter, since whatever process injects metals into the ICM presumably also injects energy as well. Regardless of this, these early high temperatures are in any case of no consequence for the subsequent evolution of the system.

As the system evolves, it is obviously apparent that the metal-rich regions cool more swiftly than the metal-poor regions, leading to the rapid inversion of the "spikes" in the temperature profile shown in Figure 4.3 on page 91. This is an inevitable consequence of the shorter cooling time of the high abundance gas, which in turn is due to its enhanced radiation — see Figure 4.4 on page 91. The cooling is naturally most extreme in the central regions where the gas density is highest, and the two-body bremsstrahlung radiation is most intense. The disappearance of the innermost spikes in the temperature profile at late times is due to the removal of the metal-rich gas from the simulation as it cools below 10^5 K. Note that this is an example of non-central mass loss (Section 3.6) in what is conceptually a single-phase medium (according to the usual definition, Section 6.1).

Figure 4.2(a) on the next page shows the results of fits to the spectral range 0.3–7 keV. As discussed in Section 4.2, the initial fitted abundance profile is flat, within the error bars. The differential cooling of the metal-rich and metal-poor phases has marked effects upon the measured abundance profile. As time passes, there is a gradual increase in the derived abundances, particularly in the central regions (for example, after 1.0 Gyr the measured value for the central abundance has risen to $0.5\ \mathrm{Z}_{\odot}$). With further evolution, the central abundance begins to decrease again, leading to the situation where there is a peak in the abundance profile at an off-centre position. This peak declines in magnitude and moves outward with subsequent evolution.

[1] The author has an unfashionable attachment to SI units.



(a) 0.3–7.0 keV **(b)** 3.0–7.0 keV

(c) Silicon **(d)** Oxygen

Figure 4.2: Time evolution of the observed abundance profile for a two-metallicity ICM. Vertical error bars are 1σ, horizontal bars show the extent of the spectral annuli. The different curves represent different times: black filled circles 0.0 Gyr; red open squares 1.0 Gyr; green open circles 1.2 Gyr; blue open diamonds 1.6 Gyr. Unless otherwise stated, all results are for single temperature `mekal` models fitted to the 0.3–7.0 keV ACIS-S spectral range. (a) dotted lines are the results of two temperature fits, shown where the reduced χ^2 for the single temperature fits exceeded 2.0. (b) effect of excluding the iron L complex. (c, d) `vmekal` fits for the silicon and oxygen abundances respectively. The simulated observation time was 25 ks for (a, b), and 50 ks for (c, d). See Section 4.3 for properties of the model cluster. The corresponding true temperature profiles are shown in Figure 4.3 on the facing page.

4.3.1 Discussion

This behaviour may be explained as follows. In the simulations, the metal-rich gas (contributing the emission lines) cools relatively quickly and enters the regime where line emission is more important, whereas the metal-poor gas (contributing the bulk of the continuum) cools much more slowly. The combined consequence of these two processes is an increase in the strength of the lines relative to the continuum, that is, an increase in the equivalent width of the lines (an example of the changing appearance of the spectrum from one of the annuli is shown in Figure 4.5 on page 92). It is from the equivalent width that the fitting procedure obtains the plasma abundance, as discussed in Section 2.1. One could view this as a form of emission-weighting — as the metal-rich regions cool, the intensity of their emission increases, so they increasingly dominate the spectral fits. The decrease in the central abundance at late times occurs

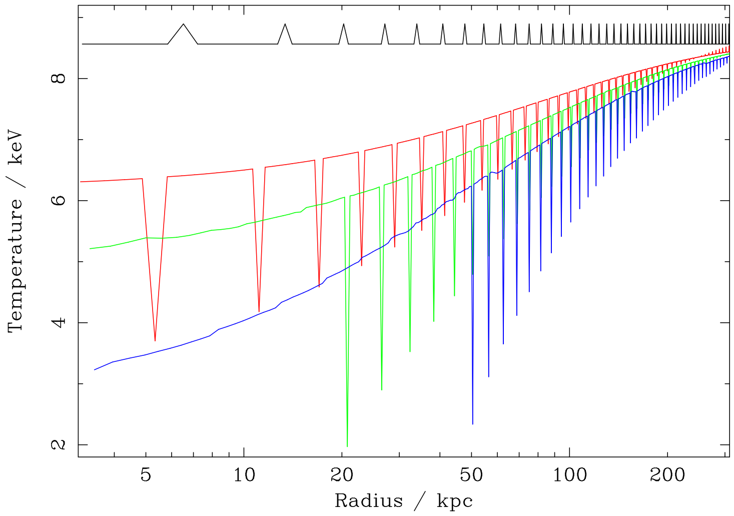


Figure 4.3: Time evolution of the true temperature profile in a two-metallicity ICM. The different curves show the temperature at different times: black 0.0 Gyr; red 1.0 Gyr; green 1.2 Gyr; blue 1.6 Gyr. See Section 4.3 for properties of the model cluster. The fitted abundance profiles at the corresponding times are shown in Figure 4.2 on the facing page.

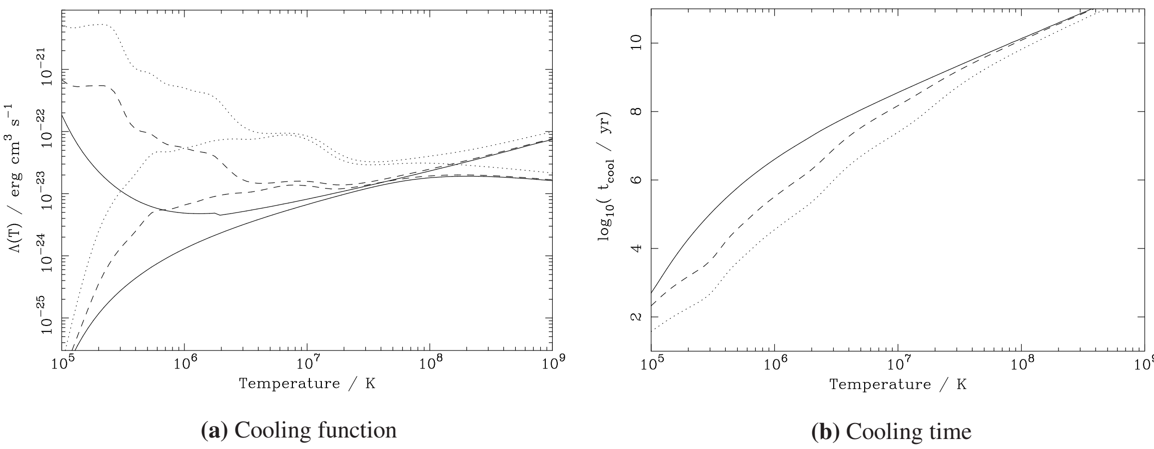


Figure 4.4: Cooling functions (a) and cooling times (b) relevant to the simulation of Figure 4.2 on the facing page. The various curves are for different metallicities: solid 0; dashed 0.3 $Z_\odot$; dotted 3.0 $Z_\odot$. (a) the upper curve in each set is for all thermodynamically relevant radiation (in practice 5 eV–200 keV); the lower curve is for the X-ray wave-band 0.1–10 keV. (b) representative constant pressure (10^6 K cm^{-3}) cooling times.

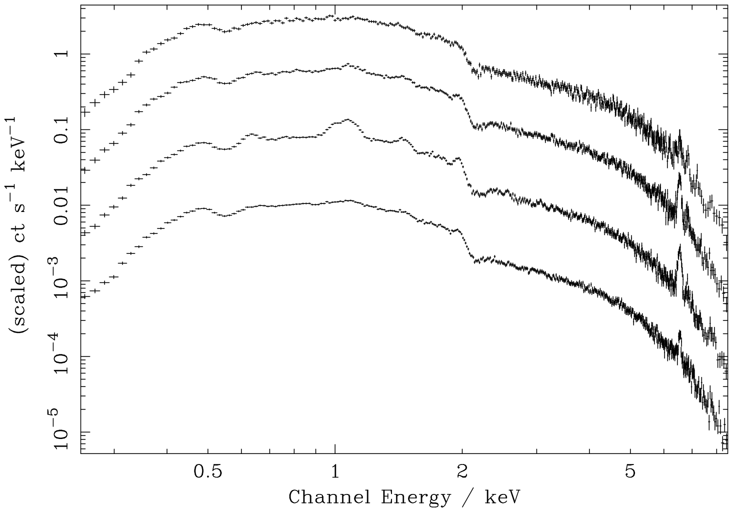


Figure 4.5: Spectra for the third annulus from the centre in Figure 4.2(a) on page 90, as observed in 25 ks with ACIS-S at the times plotted in that figure. From the top, the spectra correspond to: 0.0, 1.0, 1.2, 1.6 Gyr; and have been scaled by factors of 1.0, 10^{-1}, 10^{-2} and 10^{-3} respectively for clarity. Note the changes in the relative strengths of the Fe K ($\sim$ 7 keV) and L ($\sim$ 1 keV) regions, bearing in mind that the spectrum for 1.2 Gyr corresponds to the peak in apparent abundance for this annulus.

as metal-rich gas there cools out of the X-ray regime and is lost from sight.

The maximum amplitude of the abundance peak is obviously related to the underlying spread in metallicity, but it is not obvious how one might use the former in practice to reconstruct the latter. From the theoretical point of view there is a lower limit to the extent of the variations that must exist if they are to be of any use in resolving the cooling flow problem (i.e. the lack of low-energy spectral lines). The upper limit to the spread depends on what one considers reasonable for the injection process to produce, assuming there is no active segregation of metals. Cooling time considerations also become important with extreme abundance values, if one does not wish the entire central region to become devoid of metals in the high-temperature phases.

Figure 4.2(a) on page 90 also displays the results of two-temperature `mekal` fits to the data (dotted lines). These are shown only in those cases where the reduced χ^2 for the one-temperature fits was greater than 2.0. It was necessary to restrict the abundances of the two `mekal` components to be equal, there not being enough information in the spectra to obtain meaningful constraints on two individually varying abundances. Even taking this step, the values obtained for the two temperatures were more often than not very poor limits. The resulting abundances, however, agree very well with the one-temperature results.

To check that these effects are a genuine consequence of the non-uniform metal distribution, I have carried out identical simulations where all the gas has a uniform abundance of 0.3 $Z_{\odot}$. In this case, all four abundance profiles analogous to those plotted in Figure 4.2 on page 90 remain resolutely flat within the errors at all times as the system evolves.

The results for lower temperature clusters (simulations were also carried out for 3 and 5 keV clusters) are qualitatively similar. Obviously, with lower temperature systems there is a greater need for two-temperature models to fit the correspondingly richer spectra. Also, while for the 8 keV cluster the maximum height of the abundance peak ($\sim$ 0.9 $Z_{\odot}$) is some three times the average abundance, for the 3 keV system the peak has a maximum amplitude roughly double the average abundance. There is less opportunity for differential cooling if the bulk of the gas begins life at lower temperatures. The evolution of lower temperature systems also takes place more rapidly (as mentioned in Section 1.5.3, $t_{\rm cool} \propto T^{1/2}$ for virially-scaled clusters). The general trends of behaviour are, however, the same. The small-scale metallicity variation scenario would therefore predict that the off-centre abundance peaks would be proportionately stronger in higher temperature systems.

4.3.2 Matters of Resolution

The size of the spectral annuli is controlled by two factors. First, it is obviously necessary to ensure that they are large enough to collect sufficient photons in a reasonable integration time to produce a meaningful signal-to-noise ratio. Of course, this is only an issue if striving for verisimilitude (we have the ability to award ourselves an arbitrary amount of virtual observing time if we so desire). Second, there is also an issue of numerical resolution. In this case, the size of the annuli was selected such that each encompasses ten radial zones from inner to outer edge (although of course as a result of projection effects each annulus receives emission from every zone whose radius is greater than that of the inner edge of the annulus). It might be thought that this is the minimum number necessary to avoid artificial oscillations in the fitted abundance profile, given that the fraction of metal-rich zones is 10 per cent. With a fraction $f_{\rm hi}$ of zones being metal rich, spectral annuli must encompass at least $f_{\rm hi}^{-1}$ radial zones. This point is discussed below.

Alternatively, if the spectral annuli were any smaller then we would be able to *resolve* the discrete nature of the metal distribution. Given that each annulus spans a projected radius of 7 kpc, whilst the size of the metal-rich zones is 0.7 kpc, it might seem that I am claiming it would be possible to resolve the inhomogeneities with a resolution length greater than their size. This is not the case, however, because the spatial extent of the metal-*poor* zones is nine times that of the metal-rich zones. When the resolution length is less than or equal to the greater of those extents, the distribution can in principle be resolved.

As is standard practice with numerical simulations, I have doubled the numerical resolution to check that the results are unaffected, and indeed they are not. This is important in this case for another reason. As well as confirming the numerical result, it illustrates just how difficult it will be to probe such metallicity distributions in real-world observations. Looked at another way, then, going from the high-resolution simulation to the default-resolution simulation doubles the physical size-scale of the metallicity variations, so that the largest relevant length-scale (that of the metal-poor regions in the model) is just less than the resolution length of the spectral observations. Yet it does not do anything towards hinting that there are in fact extreme enrichment variations on just slightly smaller scales. In other words, detection of such metal variations will be an all-or-nothing affair — without adequate spatial resolution, there is no hope.

Recall as well that this is for a highly simplistic, regular, one-dimensional distribution. Given an irregular, two- or three-dimensional distribution pattern, with metal-rich ‘clouds’ drifting in and out of various lines of sight, then the problem becomes even more difficult.

In reality, the problem is even worse than all-or-nothing, and perhaps might better be described as ‘something-or-nothing’. If I do reduce the size of the spectral annuli below the level used in Figure 4.2(a) on page 90, then I do not immediately begin to resolve the correct (i.e. 0.0–3.0 $Z_\odot$) metal distribution. Instead, one begins to detect only slight variations around the mean level. These will not be significant unless one has a good quality spectrum with adequate signal-to-noise ratio. The smaller one makes the spatial resolution element (in order to probe finer and finer regions), the larger the collecting area required in order to obtain good-quality spectra in reasonable time-scales. Good spectral resolution will also be necessary to separate the effects of the varying enrichments and allow one to constrain two-temperature models with freely varying abundances for each component. This will depend on just how extreme the metal variations are. Given that this is all for a regular, one-dimensional system, with no background file added, one can begin to appreciate the difficulties that genuine observations present in this regard.

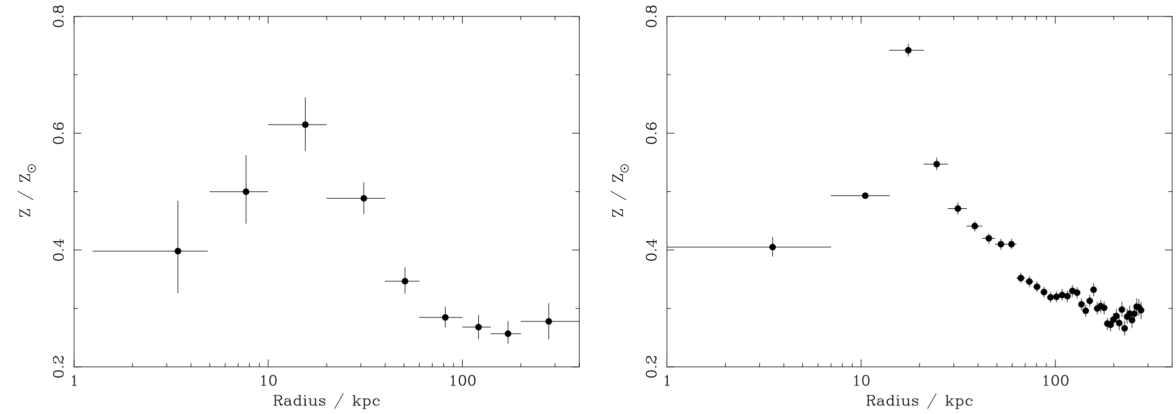
Accepting that *direct* detection of small-scale ICM metallicity deviations will be so difficult, we are forced to consider what forms of *circumstantial* evidence it may prove possible to employ.

4.3.3 Abundance Gradients

Figure 4.2(a) on page 90 illustrates that a bimodal abundance distribution within a cooling flow scenario has the ability to produce an *apparent* metallicity gradient where one does not in reality exist (the real mean abundance in the simulations remains unchanged, at least until metal-rich gas begins to drop out in the inner regions). For example, consider the results after 1.0 Gyr of evolution, when the central abundance has risen to around $0.5\,Z_\odot$. At later times the gradient effect would be even more pronounced if one were not able to resolve the central abundance drop.

Large-scale abundance gradients were detected in several clusters with *ASCA* and *ROSAT*, for example Centaurus (Fukazawa *et al.*, 1994; Allen and Fabian, 1994), Virgo (Matsumoto *et al.*, 1996), AWM 7 (Ezawa *et al.*, 1997), and Abell 496 (Dupke and White III, 2000). The presence of an abundance gradient appears to be correlated with the existence of a cooling flow (e.g. De Grandi and Molendi, 2001), although this may not be a true correlation (cooling flows generate gradients), but merely due to the mergers which disrupt cooling flows also erasing any abundance gradient. Here we have demonstrated one mechanism by which cooling flows may actually give rise to the appearance of such abundance gradients. The interplay between cooling flows and abundance gradients has been studied previously by several authors. For example, Allen and Fabian (1998) argued that it was the presence of abundance gradients in cooling flow systems that leads to the higher emission-weighted metallicities in these sources as compared to non-cooling flow clusters. Reisenegger *et al.* (1996) investigated the ability of cooling flows to create metallicity gradients by transport of metals.

Recently, *Chandra* studies have confirmed the presence of large-scale abundance gradients in many clusters. The unprecedented spatial resolution of *Chandra* has revealed more detail in the profiles. In several cases, clusters are found to exhibit peaked abundance profiles (i.e. a positive radial abundance gradient in the innermost regions, coupled with a negative gradient further out), very similar to those produced in our simulations. Some examples are: Centaurus (Sanders and Fabian, 2002), where the negative abundance gradient peaks at 1.3–1.8 $Z_\odot$ at a radius ~ 15 kpc before falling back to $0.4\,Z_\odot$ at the centre; Abell 2199 (Johnstone *et al.*, 2002), where the metallicity rises from $\sim 0.3\,Z_\odot$ at 200 kpc to $\sim 0.7\,Z_\odot$ at 30 kpc before dropping back to $0.3\,Z_\odot$ within the central 5 kpc; and possibly Perseus (Schmidt *et al.*, 2002), where there may be a high-metallicity ring of around $0.6\,Z_\odot$ at a radius of 60 kpc. It is of course intriguing when a hypothesis designed to answer one issue (lack of low-temperature line emission) ends up providing a potential explanation for another (abundance gradients with off-centre



(a) Observed

(b) Simulated

Figure 4.6: Comparison of observed and simulated abundance gradients. (a) *Chandra* abundance profile of A2199. Data provided by Roderick Johnstone. Corresponds to figure 5 of Johnstone *et al.* (2002). (b) abundance profile obtained from simulated observations as in Figure 4.2 on page 90.

peaks); though of course there is no shortage of alternative explanations for both these phenomena. Figure 4.6 compares the observed abundance profile of Abell 2199 (Johnstone *et al.*, 2002) with one particular simulation output.

Extrapolating the results of Figure 4.2 on page 90 for several Gyr, one would predict that, over not too long a timescale, the central region would become devoid of metals, with a ring of high metallicity at large (hence easily observable) radii that would have been detected before the *Chandra* era. This is clearly unphysical, and is easily explained away by my simple treatment of a complex problem. I start with fully formed metallicity variations rather than allowing them to develop; the variation is extreme — rather than allowing for several ‘phases’ of differing abundances I use only two, one with no metals and one with a high abundance; and I do not account for replenishment of metals through continued enrichment, which would offset the extremes of behaviour that these simple models predict.

The metallicity distribution in Perseus appears to show no correlation with the galaxies. Intriguingly, Schmidt *et al.* (2002) also report what may be the first signs of small-scale metallicity variations in the Perseus ICM, although the scale of the effect is at the limit of detectability with current exposure times and is not statistically significant.

A similar off-centre abundance peak is seen in the *Chandra* observation of the merging cluster Abell 3266 (Henriksen and Tittley, 2002). These authors suggest an alternative explanation for this feature involving the merging subcluster depositing enriched gas close to the centre of the cluster. Shock heating is then called upon to preferentially devoid the enriched region of the lightest, most mobile elements, H and He, so that the metallicity may be increased without an associated raising of the density, for which there is no observational evidence. This explanation requires efficient motion of ions, something which it is not clear can happen even if thermal conduction (Chapter 5) is relatively unimpeded. Moreover, these efficient transport properties are required precisely in the regions where subcluster merging is taking place. Cold fronts (Section 5.7) are plausible evidence that transport properties may be highly suppressed in such volumes, presumably due to separate magnetic structures existing in the merging components. It is intriguing that off-centre abundance peaks appear to be found both in clusters with

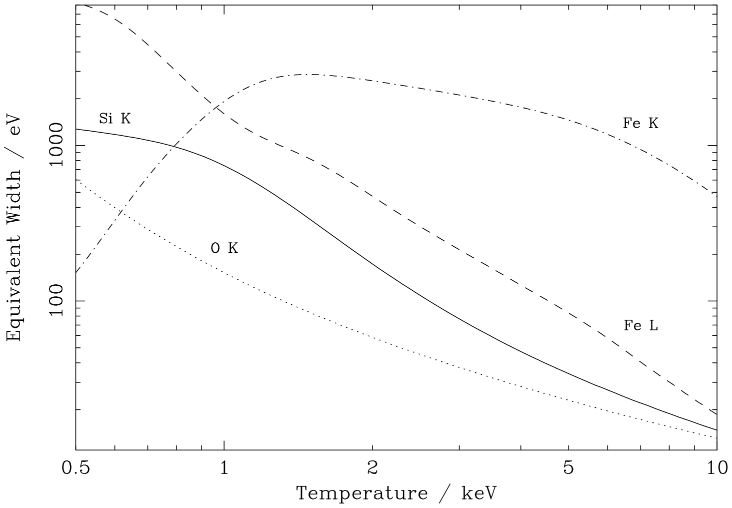


Figure 4.7: The temperature dependence of the equivalent widths for various ICM plasma spectral lines, for a solar metallicity plasma. Dot-dashed iron K; dashed iron L; dotted oxygen K; solid silicon K.

and without cooling flows. If we are looking at the same phenomenon in both cases (this of course by no means clear), and if a common mechanism is responsible, this would suggest that it is something associated with neither the cooling flow process nor the merger process.

4.3.4 Equivalent Width Effects

The original *ASCA* abundance gradients were clearly detected in the equivalent width of the iron K line, e.g. in Centaurus where Fukazawa *et al.* (1994) report an increase by a factor of three on moving to the central regions of the cluster. Excluding the iron L complex from the *Chandra* data for Centaurus by simply fitting to the high-energy end of the spectrum (3–7 keV), whilst obviously increasing the noise, does not affect the qualitative nature of the conclusions regarding the abundance profile (Sanders and Fabian, 2002). Fitting my model spectra just using the data in this region produces different results, however, as illustrated in Figure 4.2(b) on page 90. Obviously, the noise level has increased somewhat due to the omission of part of the spectrum from the fit, but the qualitative behaviour has also altered. There is still something of an initial increase in the central abundance at early times, leading once again to the generation of an apparent gradient, albeit much weaker than before. At later times, however, the highly peaked abundance profile that was present when fitting to the entire spectrum does not appear. Instead the central abundance merely dies away. Excluding the iron K lines, however, by fitting only to the spectra between 0.3 and 5 keV produces much the same results as using the entire spectrum.

Thus there is a qualitative difference in behaviour when fitting to the iron K lines as compared to the iron L lines that must be explained. In Figure 4.7 I illustrate the temperature dependence of the equivalent width for various important ICM spectral lines, calculated from MEKAL spectra. Consider just the results for the iron. Upon cooling from high temperatures, the equivalent width of the iron K lines increases somewhat, then rapidly dies away below 1 keV. The width of the L complex, on the other hand, increases strongly and monotonically as the temperature decreases. Consequently, with a bimodal metallicity, the strength of the iron K lines relative to the continuum increases to a relatively small extent on cooling, producing the initial slight increase in the central abundance, before dying away as cooling

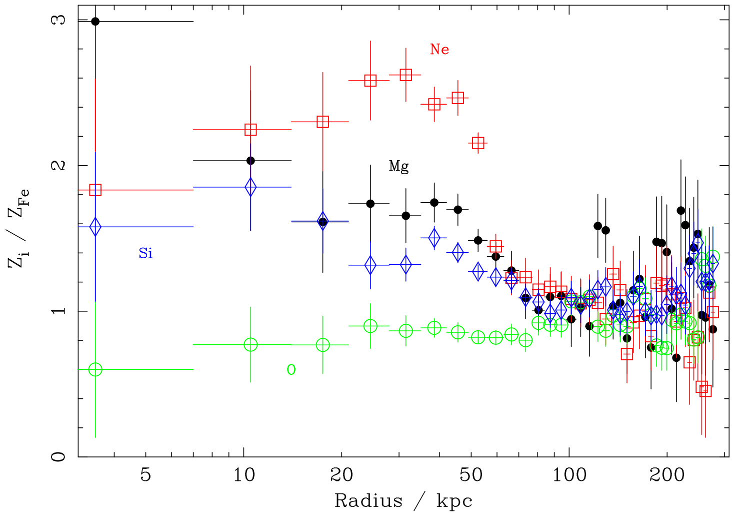


Figure 4.8: Observed ratios of element abundance to that of iron, for the last of the times plotted in Figure 4.2 on page 90. Black filled circles magnesium; red open squares neon; green open circles oxygen; blue open diamonds silicon.

progresses. The strong monotonic increase in the iron L width with cooling produces a much greater change relative to the continuum.

To illustrate this point further, the equivalent widths for the silicon and oxygen K lines are also plotted in Figure 4.7 on the facing page. Both increase monotonically as the temperature is reduced, but the rate of change for oxygen is relatively small, whereas silicon shows very similar behaviour to iron L. The arguments of the previous paragraph are borne out by the results of `vmekal` fits to the model spectra. Figures 4.2(c) to 4.2(d) on page 90 show the fits for silicon and oxygen respectively. The qualitative behaviour of both profiles is the same, but the oxygen profile shows a less pronounced peak.

Shown in Figure 4.8 are the ratios of the abundances of various elements to that of iron, for the last of the times illustrated in Figure 4.2 on page 90. Neon and magnesium (and to a lesser extent silicon) appear overabundant, whilst oxygen appears slightly underabundant.

4.4 Genuine Abundance Gradients

Given that flat metallicity profiles coupled with bimodal distributions are unable to fully reproduce the observed abundance trends, I have considered also clusters with genuine gradients in their abundance profiles.

4.4.1 Functional Form

It is necessary to parametrize the abundance profile in some way. The functional form I have chosen is basically a transformation of $Z \propto r^n$,

$$Z(r) = \begin{cases} \bar{Z} + (Z_c - \bar{Z}) \left(1 - \dfrac{r}{r_Z}\right)^{n_Z} & r < r_Z \\ \bar{Z} & r \geqslant r_Z \end{cases} \tag{4.6}$$

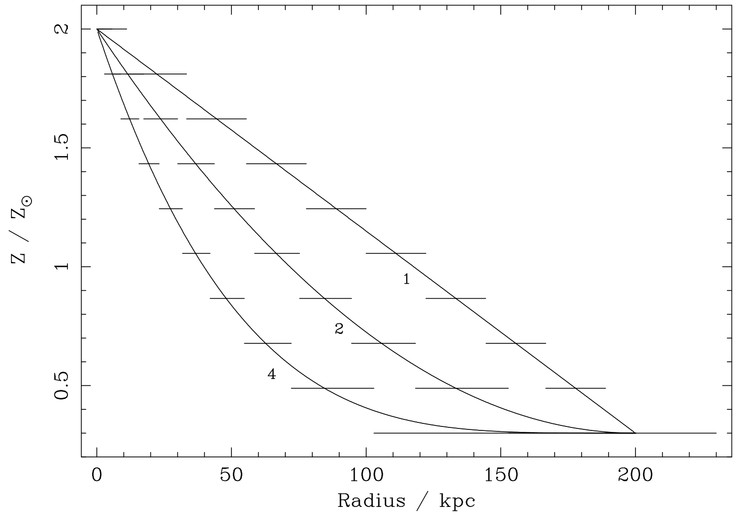


Figure 4.9: Examples of model abundance gradients. The smooth curves are parametrized according to equation (4.6), with $n_Z = 1, 2, 4$. Common parameters: $\bar{Z} = 0.3$; $Z_c = 2.0$; $r_Z = 200\,\mathrm{kpc}$. The stepped profiles show how each curve would be discretized into a finite number of abundance sets. In practice more steps would be used than are shown here. The innermost and outermost steps for each of the three cases are coincident in the plot. Beyond r_Z the abundance is constant.

That is, a power-law gradient of index $n_Z > 0$ inside some radius r_Z, coupled with a uniform metallicity $\bar{Z}$ outside this radius. Z_c is the central metallicity, $Z(r=0)$. Some examples are shown in Figure 4.9. Another possible form would be $Z \propto 1-(r/r_Z)^{n_Z}$ (e.g. Pellegrini and Ciotti, 2002). The precise details are not too important for my purposes.

4.4.2 Practical Implementation

In order to use an abundance gradient in the simulation, it is necessary to discretize it in some way. Since the cooling function and spectral emissivity are dependent on metallicity, the values of these quantities will vary spatially. Memory issues put a limitation on the number of different abundance sets for which these functions can be calculated and stored.

With an abundance gradient of the form represented by equation (4.6), Z varies from $\bar{Z}$ to Z_c. If the maximum number of abundance sets that can be stored is n_{abund}, then the best way to discretize the profile is into a stepped form, where each step is equal in Z (rather than in radius, say). That is, the discrete abundance sets used would be: $\bar{Z}$, $\bar{Z}+\Delta Z$, $\cdots$, Z_c; with $\Delta Z \equiv (Z_c - \bar{Z})/n_{\mathrm{abund}}$ (of course, to be precise this gives $n_{\mathrm{abund}}+1$ sets).

Once the abundance sets that will be used are established, it is then necessary to assign them to the appropriate radial ranges. This requires inversion of equation (4.6) to give $r(Z)$,

$$\frac{r(Z)}{r_Z} = 1 - \left(\frac{Z-\bar{Z}}{Z_c-\bar{Z}}\right)^{1/n_Z}. \tag{4.7}$$

I take the radial range over which some abundance Z is to apply as extending from $r(Z-{}^1\!/_2\Delta Z)$ to $r(Z+{}^1\!/_2\Delta Z)$ (with obvious exceptions at $r=0$, r_Z), since this gives the best centring (i.e. the discretized profile has the smallest deviation from the actual one). An alternative would be to have each zone extend halfway to the next crossing point in radius, but I feel this to be less desirable since it gives greater deviation between the discrete and continuous profiles. The adopted procedure is illustrated in Figure 4.9 on the facing page.

A minor complication occurs if the spatial resolution is relatively low (or, equivalently, if the abundance resolution is too high), because then one physical zone may correspond to more than one abundance zone. This means one has oversampled the abundance space, and a smaller ΔZ would be acceptable. Actually, this is not necessarily the case, since as can be seen from Figure 4.9 on the facing page the width of the abundance regions increases with radius as the gradient flattens off. So a ΔZ that oversamples at the inside may not do so at the outside. It is desirable to avoid this oversampling, so the code warns if it occurs. The situation must still be handled, so if a physical zone encompasses more than one abundance region, I use the abundance appropriate to the zone mid-point — the median abundance, if you will. There is no point taking the mean of the included Z sets when the available values of Z are quantized (and with the sets equally spaced in Z the mean will be very similar to the median anyway). If oversampling does occur, then there are unused abundance sets. For efficiency, I do not calculate Λ etc. for these unused sets.

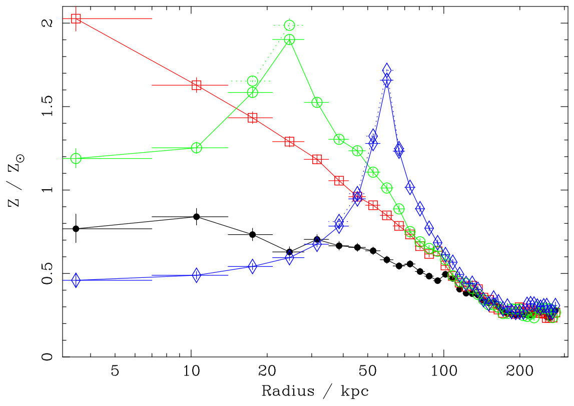
4.4.3 Two-component Gradients

The technique described in Section 4.1.2 is still applicable. In the gradient region, the metal-rich zones are assigned a metallicity of $Z(r)/f_{\rm hi}$, where $Z(r)$ is the abundance appropriate for the zone radius. Λ and other quantities are thus calculated for these values of Z, rather than those of the nominal abundance profile.

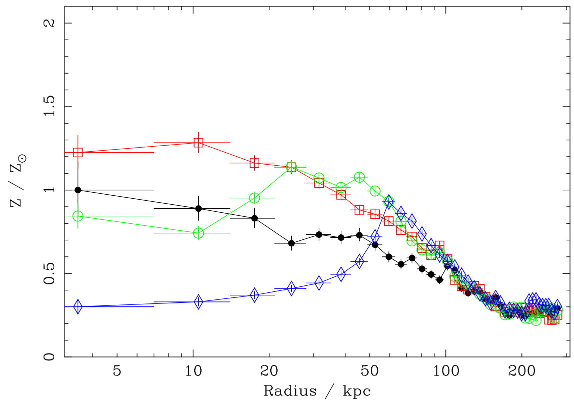
4.5 Results for the Evolution of the Intracluster Medium in the Gradient Case

In Figure 4.10 on the following page are shown the abundance profiles that were fitted to fake data produced from a simulated cluster with an abundance gradient. The basic properties of the system are the same as those of the cluster in Section 4.3. In terms of the parameters defined in equation (4.6), the true abundance profile has $\bar{Z} = 0.3\,{\rm Z}_\odot$, ${\rm Z_c} = 2.0\,{\rm Z}_\odot$, $r_{\rm Z} = 200\,{\rm kpc}$, and $n_{\rm Z} = 1.5$. In Figures 4.10(c) and 4.10(d), the actual metallicity profile smoothly follows that specified by equation (4.6); whereas in Figures 4.10(a) and 4.10(b) a fraction $f_{\rm hi} = 0.1$ of the gas is enriched to ten times the value of equation (4.6), and the remainder of the gas has zero metallicity.

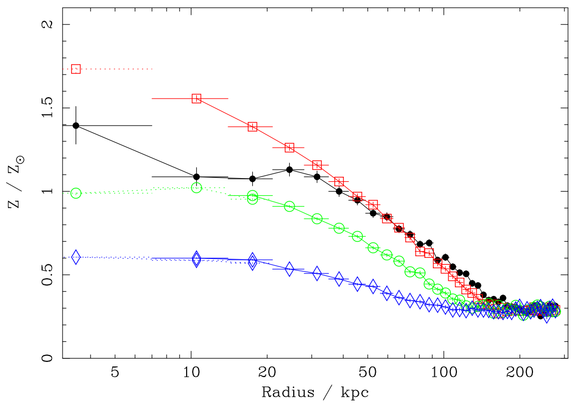
Once again, the two-component system develops an off-centre abundance peak, of a larger amplitude than before. This time, an off-centre peak (albeit not as intense) is reproduced when fitting solely to the high-energy part of the range, Figure 4.10(d). In the one-component system, on the other hand, no such peak develops; rather the abundance declines reasonably smoothly back down towards a flatter state. At no time is any form of off-centre peak produced. Note that the temporal evolution is in both cases faster, owing to the higher abundances that are present.



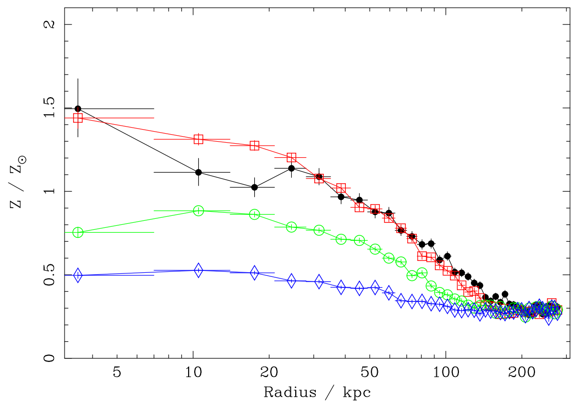
(a) Two-component, 0.3–7.0 keV



(b) Two-component, 3.0–7.0 keV



(c) One-component, 0.3–7.0 keV



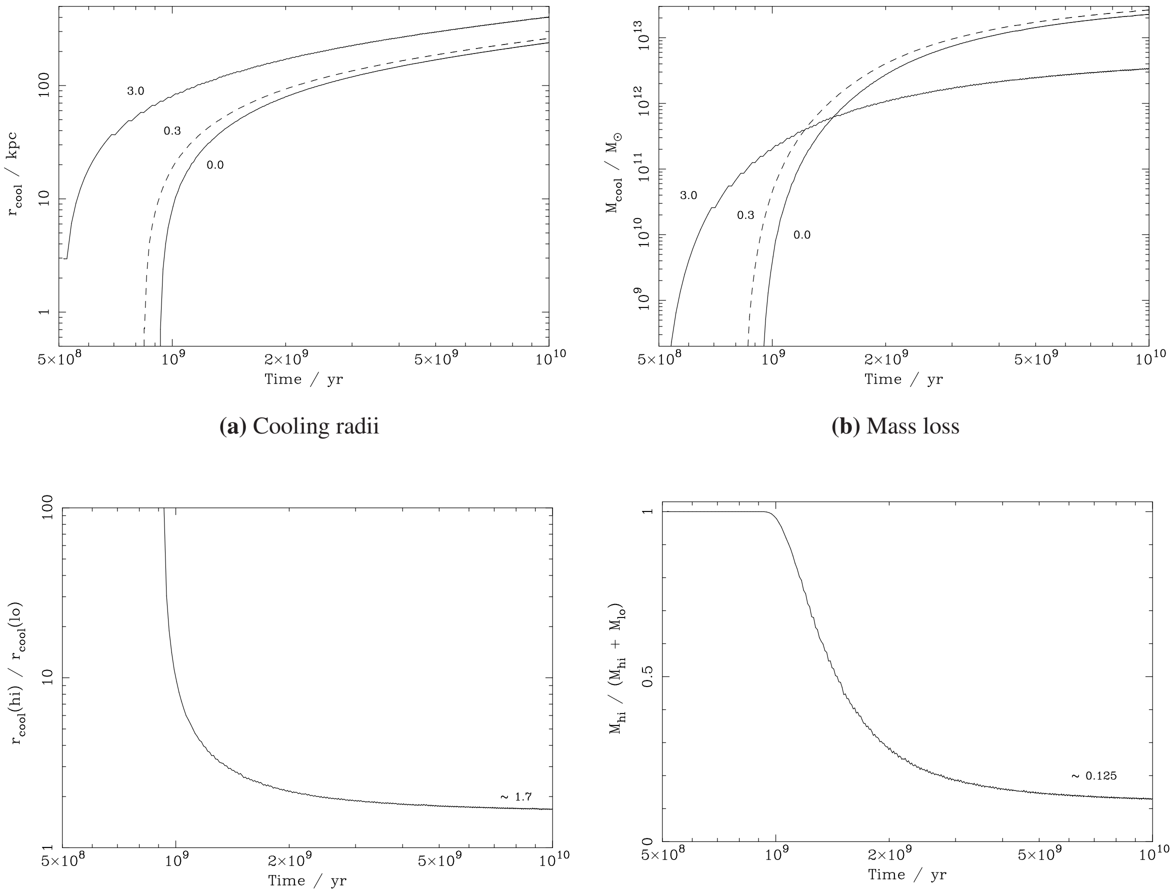
(d) One-component, 3.0–7.0 keV

Figure 4.10: Time evolution of the observed abundance profile for a cluster with a genuine abundance gradient. Vertical error bars are 1σ, horizontal bars show the extent of the spectral annuli. The different curves represent different times. (a, b) two metallicities: black filled circles 0.0 Gyr; red open squares 0.4 Gyr; green open circles 0.5 Gyr; blue open diamonds 0.8 Gyr. (c, d) one metallicity: black filled circles 0.0 Gyr; red open squares 1.3 Gyr; green open circles 2.5 Gyr; blue open diamonds 3.8 Gyr. (a, c) fits to the 0.3–7.0 keV region; dotted lines show the results from two-temperature fits where the reduced χ^2 for the single-temperature fit exceeded 2.0. (b, d) fits to the 3.0–7.0 keV region (i.e. excluding iron L). The simulated observation time was 25 ks with ACIS-S in each case. Compare with Figure 4.2 on page 90.

4.6 Cooling Flow Equivalent Widths

Figure 4.11 on the facing page shows how the cooling radii and mass deposited behave as a function of time for a one-metallicity ICM, and for each of the abundance sets in a two-component ICM. The simulated clusters have the same parameters as in Figure 4.2 on page 90. For the two-component ICM of Figure 4.11(b) on the facing page, summing the cold mass in the metal-rich and metal-poor components gives a profile that is (at late times) essentially indistinguishable from that of the homogeneous 0.3 Z$_\odot$ plasma shown in the same figure. This is another pleasing illustration of the close resemblance of the one- and two-component metallicity plasmas.

Figure 4.11(d) on the facing page shows that the fraction of cold mass originating from the metal-



(a) Cooling radii

(b) Mass loss

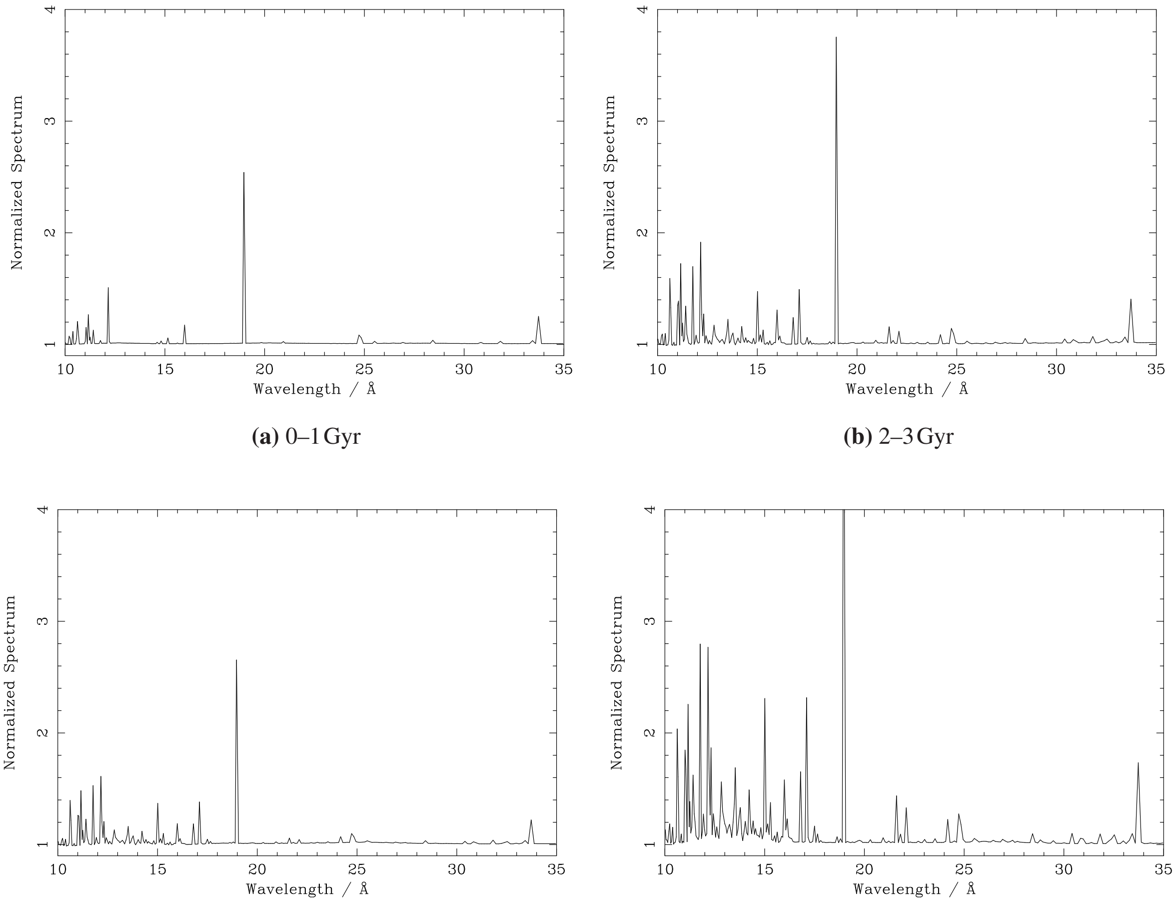
(c) Cooling radius ratio Z_{hi}:Z_{lo}

(d) High Z cold mass fraction

Figure 4.11: The evolution of the cooling radii and the deposited mass for an ICM with a one- or two-component metallicity. (a) solid lines show r_{cool} for each of the abundance sets in a two-component ICM with $f_{hi} = 0.1$, $Z_{hi} = 3\,Z_{\odot}$, $Z_{lo} = 0$; dashed line is for a one-component ICM with $Z = 0.3\,Z_{\odot}$. (b) evolution of the mass that has cooled out for each of the abundance sets of (a). (c) ratio of cooling radii of the high- and low-metallicity gas for the two-component ICM of (a). (d) fraction of the cold mass originating in the metal-rich phase. (a, b) labels are metallicities in solar units; (c, d) labels shows approximate asymptotic values. The parameters of the simulated clusters are described in Section 4.3.

rich phase asymptotes out to around one-eighth, rather than the value of one-tenth that might have been expected, given that the high-metallicity gas initially accounts for a fraction $f_{hi} = 0.1$ of the total mass. This is a consequence of the fact that the metal-rich gas has a larger cooling radius at any instant than the metal-poor gas, as shown in Figure 4.11(c). So at any instant, the volume in which appreciable cooling of the metal-rich gas has taken place over the age of the system is significantly greater than the corresponding volume for the metal-poor gas.

In practice, however, this effect will not have much influence on the observed spectra. The *XMM-Newton* RGS instrument functions essentially as a long-slit spectrometer, with a width of 5 arcmin in the cross-dispersion direction (Section 1.1). Although spectroscopy of extended sources is difficult, one can certainly extract spectra corresponding to the central arcmin, say (e.g. Tamura *et al.*, 2001). This usually encompasses a region significantly smaller than the cooling radius. There will of course be a slight projection effect from more distant regions, but I would not expect this to be greatly significant,



(a) 0–1 Gyr

(b) 2–3 Gyr

(c) 7–8 Gyr

(d) 7–8 Gyr, one-component *Z*

Figure 4.12: Spectra observed from the central Mpc of the two-metallicity cluster of Figure 4.11 on the previous page at different times. The wavelength range is that of the *XMM-Newton* RGS instrument, and has been deredshifted. Each plot is an average of the emission over the stated time period. The spectra have been normalized by dividing by a fitted continuum, so that the changes in line strength are more readily apparent. (d) shows the equivalent of (c) for a one-component metallicity ICM.

given the declining surface brightness profile.

The detailed picture is a little more complicated though. The actual spectrum observed from a two-component ICM such as that of Figure 4.11(b) on the preceding page evolves with time. For example, in the first Gyr of the simulation shown in that figure, there is very little deposition of metal-poor gas. The metal-rich gas (and its associated emission lines) consequently have a stronger influence on the spectrum in this time frame. At later times the contribution of the metal-poor gas increases. It therefore becomes necessary to calculate spectra by averaging the emission over some time period, and examining the differential mass loss for each abundance set in that time. Some examples of the changing spectra are shown in Figure 4.12. Note the changes in the O VIII 19 Å and Fe XVII 15 and 17 Å lines over time. Initially, there is little cool gas of any form; then the metal-rich gas begins to cool first; then at late times the metal-poor gas is cooling appreciably. Shown for comparison in Figure 4.12(d) is the spectrum from a homogeneous metallicity ICM, clearly exhibiting much stronger line emission.

4.7 Discussion of the Small-scale Metallicity Variation Hypothesis

To address the question of how small-scale ICM abundance variations might come about, supernovae can be considered as essentially point sources (on ICM scales) of extremely high metallicity gas. Of themselves, these have structures that are far from simple (e.g. Hughes *et al.*, 2000). One way in which escape of the enriched gas from the potential well of the host galaxy into the ICM may occur is via superwinds (e.g. Heckman, 2001) (although other processes such as ram-pressure stripping and pregalactic winds are doubtless significant). Superwinds have a complex, multi-phase structure in which it is difficult to probe directly the energetic, enriched gas driving the wind. There is, however, clear evidence for complicated structure in the wind on very small scales in the optical and soft X-ray (e.g. Strickland, 2002). In my opinion, the question ought to be posed the other way around: how might the ICM become uniformly enriched?

There are essentially two issues that determine whether or not the situation as I have chosen to model it is physically realistic: (i) how are the metals injected into the ICM, and (ii) once in the ICM, do the metals move to spread out over a wide area or remain confined? The answers to both these questions remain uncertain at present. To address the second question first, following Spitzer Jr (1962), one may express the deflection time (average time for a cumulative deflection of 90°) for particles of mass m and charge Ze diffusing through field particles of mass m_f, charge $Z_\mathrm{f}e$ and density n_f as

$$\tau = \frac{6\sqrt{3}\pi\varepsilon_0^2 k_\mathrm{B}^{3/2}}{e^4} \frac{m^{1/2}T^{3/2}}{Z^2 Z_\mathrm{f}^2 n_\mathrm{f} \mathcal{F}(x) \ln\Lambda}, \tag{4.8}$$

where

$$x = \sqrt{\frac{3m_\mathrm{f}}{2m}}, \tag{4.9}$$

$$\mathcal{F}(x) \equiv \operatorname{erf} x + \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\operatorname{erf} x}{2x}\right), \tag{4.10}$$

with erf the standard error function, $\ln\Lambda$ the Coulomb logarithm (Section 5.2), and assuming that all species are in thermal equilibrium at temperature T. The factor $\mathcal{F} \leq 1$ is the Chandrasekhar correction for the finite mass of the field particles. Numerically,

$$\tau \approx 0.26\,\mathrm{Myr}\left(\frac{m}{m_\mathrm{H}}\right)^{1/2}\left(\frac{T}{10^8\,\mathrm{K}}\right)^{3/2}\left(\frac{n_\mathrm{f}}{\mathrm{cm}^{-3}}\right)^{-1}\frac{1}{Z^2 Z_\mathrm{f}^2 \mathcal{F} \ln\Lambda}. \tag{4.11}$$

As was shown by Rephaeli (1978) for the case of sedimentation, in the ICM the contributions from field particles other than protons (particularly helium nuclei) are significant. Summing over the elements in a $0.3\,\mathrm{Z}_\odot$ plasma with iron nuclei as test particles, one finds that τ is reduced by around a factor of 2 from the value for a pure proton plasma. At 10^8 K, and with $n_\mathrm{H} = 10^{-3}\,\mathrm{cm}^{-3}$, I find the corrected deflection times for iron and helium nuclei are $\tau_\mathrm{Fe} \approx 0.3\,\mathrm{Myr}$ and $\tau_\mathrm{He} \approx 4\,\mathrm{Myr}$.

From standard kinetic theory, the rms three-dimensional distance a particle will diffuse in a time t is given by

$$r_{\rm rms} = \sqrt{2\lambda v_{\rm rms} t} = v_{\rm rms}\sqrt{2\tau t} = \sqrt{\frac{6k_{\rm B}T\tau t}{m}}, \tag{4.12}$$

where $\lambda = \tau v_{\rm rms}$ is the mean free path. Hence one can estimate that iron nuclei may diffuse about 20 kpc in 1 Gyr. To maintain small-scale abundance variations over significant time-scales in the ICM therefore requires a strong suppression of diffusion. This result is to be expected, as it is equivalent to the strong suppression of conduction that has traditionally been invoked for the multi-phase cooling flow picture (Chapter 6). Magnetic fields have normally been appealed to as the causative agent to dampen transport properties in an ionized gas. Early calculations (e.g. Tribble, 1989) appeared to show that a tangled magnetic field would indeed produce a strong reduction in transport. More recent calculations considering a chaotic field with turbulence extending over a wide range of length-scales indicate that the suppression may in fact be minimal (Narayan and Medvedev, 2001). I return to this issue in more detail in Section 5.5.3.

In practice, simple diffusion is probably unlikely to be the limiting factor for the spread of ICM metals. Convection, turbulence due to radio sources, galactic wakes, etc., will also play roles to varying degrees. If, however, the abundance drops observed with *Chandra* in the central regions of the ICM for several clusters (Section 4.3.3) are genuine, then this will imply limits on the amount of convection or mixing that can have taken place (else these features would have been smoothed out). It does not seem unreasonable to explore some of the consequences of small-scale metallicity variations, and to keep the possibility in mind during spectral analyses.

The question as to whether or not the metals in the intracluster and intergalactic media are homogeneously distributed remains an open one. There is a non-negligible scatter in Galactic stellar metallicities for stars of all ages (the scatter increases for low-metallicity stars), which has been taken to suggest that the Galactic disc has been chemically inhomogeneous throughout its development (e.g. McWilliam, 1997). Classical galactic chemical evolution models have tended to assume instantaneous dispersion of synthesized elements.

The *average* metallicity of the intracluster medium on large scales has for some time been reasonably well established at roughly one-third solar in most cases, for both nearby (Edge and Stewart, 1991) and distant (Mushotzky and Loewenstein, 1997) clusters. Until comparatively recently, however, there has been little information on how the ICM metals might be distributed on finer scales. There is, in my opinion, no particular reason why the whole ICM should be uniformly enriched to the same metallicity, although this might be one's natural assumption. As I discuss in Section 4.3.2, direct detection of small-scale abundance variations in the ICM will be very difficult. Consequently, I feel it is impossible to rule out such variations at the present time.

Recently, a deal of support has been given to the idea that thermal conduction might be operating in cluster cores at significant levels (this issue is discussed in Chapter 5). Whilst this runs contrary to the established multi-phase cooling flow picture (Chapter 6), it seems to have a degree of success in matching the observed temperature profiles. It is certainly the case, though, that at least in some regions of the ICM, conduction is highly suppressed. This is revealed by the "cold fronts" seen in several cluster cores, e.g. Abell 2142 (Markevitch *et al.*, 2000) — see Section 5.7. If thermal conduction is suppressed, then ion movement must be even more restricted, since net heat conduction may still take place without individual electrons travelling a significant distance. If thermal conduction is operating with a degree of efficiency, however, then the same need not necessarily apply to ion motion. If this were the case, though, there would be the possibility of sedimentation of the heavy elements (Fabian and Pringle, 1977; Rephaeli, 1978; Fabian *et al.*, 2002b). Cold fronts of course only provide direct evidence for inhibited thermal conduction across the fronts themselves. It is possible (though as yet far from proven) that conduction may reach the Spitzer Jr (1962) value elsewhere (e.g. Voigt *et al.*, 2002). If conduction is operating efficiently, it does not necessarily rule out small-scale metallicity variations, but it does weaken the case, both from the point of view of reducing/removing the motivation for them (suppressing the low-temperature cooling flow line emission by suppressing the low-temperature gas), and by making it less likely that such conditions can persist.

The modelling discussed here is simplistic, and it may be argued that the conditions as I have chosen to represent them are not physically relevant. For example, there is no possibility of segregation of elements based on weight (that is, it is not possible for the metal-rich gas to sink to the centre). This is a consequence of my adoption of the theoretical framework of Nulsen (1986), which requires that in a multi-phase flow the various phases co-move, i.e. there is a single velocity profile, as discussed in Section 6.1. Also, the time-scale over which the changes in the abundance profile occur is somewhat short. This is a result of the rather unphysical initial condition: starting out with regions enriched to $3\,Z_\odot$ and not allowing any replacement of cooling metals. A more realistic treatment would allow the metallicity to build up initially with time and then allow for some replenishment, and possibly also for different distributions of products of Types Ia and II supernovae (SNe). However, I am not concerned here with producing detailed models of the evolution, but rather in seeing what the general trends of behaviour might be. Note that because I have restricted attention to single-phase models, there is no real transport of material by the cooling flow, since all the mass loss (except for the metal-rich zones) takes place in the centre. Reisenegger *et al.* (1996) demonstrated the ability of multi-phase cooling flows to create (genuine) abundance gradients through transport of injected materials. Thus one may expect that a more sophisticated multi-phase treatment of this process might reveal more complex effects.

Given that the abundance gradients observed with *ASCA* and *Chandra* are seen clearly in the iron K line, the results of Section 4.3.4 imply that this cannot be solely due to a bimodal distribution of metallicities. The results are not, however, inconsistent with such a metal distribution, at least in the simple scenario outlined here. At the very least, I have presented another reason to be wary of the iron L complex when fitting to X-ray spectra (e.g. Finoguenov *et al.*, 2000). In this case, it is not due to the complex atomic physics of the L shell (the code used to generate the spectra, MEKAL, is the same as that used to fit them), but rather to the temperature dependence of the equivalent width. Recall that iron is really the only ICM element for which one has two strong spectral indicators (K and L); for other elements only the K lines are useful. Playing devil's advocate, one could therefore imagine a situation in which the observed gradients in the *iron* profile are genuine, but those in the other elements are due to a process such as the one outlined here. This would, for example, severely impact on estimates of the SNe Ia:II importance ratio. Of course, this simple model would not explain the matching radial scales of the variations for iron and the other elements in this case. Nor could it explain any correlation between the iron profile and the visible light in the central regions of the cluster.

Another process which may give rise to the same sort of radial abundance profile as those seen here (namely a peak in the abundance at an off-centre position) is resonant scattering (Gil'fanov *et al.*, 1987).

This acts to redistribute photons from the central regions of the cluster (where the optical depth is highest) to a surrounding ring. See the work of Mathews *et al.* (2001) for an application of these ideas to M87. Computation of detailed optical depths requires knowledge of the velocity structure of the gas along the line of sight. I will not comment on the resonant scattering issue here, except to say that to some extent the ideas of this work and those of resonant scattering are in conflict. As was pointed out by Wise and Sarazin (1992), clumping of the ICM reduces the amount of scattering that takes place. X-ray surface brightness profiles depend on the rms density $\sqrt{\langle n^2 \rangle}$, whereas electron scattering depends on the mean density $\langle n \rangle$. The rms of a set of numbers is always greater than their mean (unless the numbers are all equal). Thus increasing the degree of clumping in the plasma reduces the mean density relative to the rms density and so reduces the relative effect of scattering.

Another possibility is that the 'extra' metals in the central regions of many clusters are to be associated with the central cD galaxy. Makishima *et al.* (2001) have suggested an alternative explanation for the enhanced emission seen at the centre of many clusters, which has traditionally been attributed to the cooling flow phenomenon by many researchers. Instead, these authors suggest the excess may be associated with the inter-*stellar* medium of the central cD galaxy. One argument invoked against the cooling flow interpretation is the fairly frequent presence of metallicity increases near cluster centres. The results of this Chapter suggest a mechanism by which such effects may indeed be produced by cooling flows. The reality or otherwise of any central dips in abundance would be an important discriminant for these two interpretations.

Böhringer *et al.* (2002) also give some consideration to the possible effects of inhomogeneous metallicities on the observed abundances (section 2.2 of their paper). Unlike my consideration of individual emission lines, they look at the overall observed metallicity of the spectrum, as inferred from the global ratio of power in all emission lines to power in the continuum. These authors also discuss some comparisons between the shape of the spectrum around the 1 keV region for M87 and the predictions of the bimodal model. All hypotheses live or die through comparison with data, so such investigations are highly necessary, but there is a large parameter space to investigate. And as they point out, such checks must be made on a source-by-source basis. They find a poor fit between the actual spectral shape and the predictions of the bimodal hypothesis (e.g. their fig. 6). Note, however, that this is for the case where the normalizations of the metal-rich and metal-poor phases are "roughly equal". I would not expect such a division to be successful in reducing the EW of the low temperature lines by an appreciable amount.

The main result of this Chapter (in combination with Chapter 2) is that the observed equivalent-width suppression for low-temperature emission lines in a cooling flow spectrum due to a metal distribution which is inhomogeneous on small-scales is not as great as one might expect. It seems unlikely, therefore, that this method in isolation could produce a reduction in equivalent width equal to that seen in data, without pushing the bimodality of the metal distribution to extreme levels. There is an effect, but it is not large enough. I have also shown that small-scale metallicity variations can give rise to interesting effects in the observed abundance profiles (Section 4.3.3) as compared to the true profiles. Such effects would give rise to serious difficulties in terms of interpreting abundances in cluster central regions. The possibility of small-scale metallicity variations ought to be borne in mind when analysing high resolution X-ray spectra of cluster central regions.

It is rather our role to take what unrealistic factors that exist and to work them into a more sophisticated form that might be grounded in the grand scheme of reality. The doings of men run to unrealities.

A Wild Sheep Chase
HARUKI MURAKAMI

5
Some Effects of Thermal Conduction

Another possible solution to the lack of low-temperature line emission seen in cooling flow spectra is afforded by thermal conduction. Indeed, this idea has become extremely fashionable of late, though it is not by any means a new idea (e.g. Rephaeli, 1977a). In this Chapter, I make some discussion of this hypothesis. The numerical scheme of Chapter 3 is modified to include the effects of thermal conduction.

5.1 Basics of Thermal Conduction

The coefficient of thermal conductivity relates the heat flux $\boldsymbol{q}$ through a fluid to the temperature gradient in the fluid

$$\boldsymbol{q} \equiv -\kappa \nabla T, \tag{5.1}$$

with the dimensions of κ being those of energy time^{-1} length^{-1} temperature^{-1}. Thermal conduction is a diffusive process in which heat is transferred down temperature gradients.

The total heat power flowing out across a surface S bounding some volume V is given by

$$\iint \boldsymbol{q} \cdot \mathrm{d}\boldsymbol{S} = \iiint \nabla \cdot \boldsymbol{q} \; \mathrm{d}V, \tag{5.2}$$

making use of the divergence theorem, equation (A.6). Hence the conductive power loss per unit volume is

$$Q = \nabla \cdot \boldsymbol{q} = -\nabla \cdot (\kappa \nabla T) = -\kappa \nabla^2 T, \tag{5.3}$$

with the last equality only holding for a medium in which κ is independent of position.

5.2 The Coulomb Logarithm

The Coulomb logarithm, $\ln\Lambda$, appears in many of the expressions related to the transport properties of an ionized gas. At heart, all these expressions are based on simple collision theory, in which the encounter between two charged particles maps out a hyperbolic path (classical Rutherford scattering) An important parameter of such encounters is the *impact parameter*, p, defined as the distance of closest approach in the absence of any forces between the particles. Clearly, the smaller the impact parameter, the greater the deflection experienced in the encounter. It is readily shown (e.g. Spitzer Jr, 1962) that the impact parameter which gives rise to a 90° degree deflection is

$$p_0 = \frac{ZZ_{\mathrm{f}}e^2}{4\pi\varepsilon_0 m v^2}. \tag{5.4}$$

This is for the idealized case of a light test particle of mass m, charge Ze, speed at infinity v being deflected by a much heavier field particle of mass m_{f}, charge Z_{f} which is at rest. Corrections taking into account the motion of the centre of mass of the system were derived by Chandrasekhar (see e.g. Chandrasekhar, 1960). These corrections can safely be ignored in the case of thermal conduction, since electrons are many times less massive than protons, but they can be relevant in some cases, for example equation (1.9) and equation (4.8).

In a neutral gas, it is only necessary to consider collisions with $p \lesssim p_0$. This is an invalid assumption for an ionized gas, since the Coulomb force is effective over a large distance. If collisions with $p > p_0$ are neglected, the mean free path (for example) is seriously overestimated. In fact, integrals over all impact parameters for encounters between charged particles tend to diverge at infinity. This is because the basic picture of an isolated encounter between two charged particles becomes invalid for large separations. If p_{max} is that impact parameter at which the integration is truncated, then Λ of the Coulomb logarithm is defined as

$$\Lambda \equiv \frac{p_{\mathrm{max}}}{p_0}. \tag{5.5}$$

The rms velocity change on scattering turns out to be $\propto \ln\Lambda$, hence this quantity appears in various expressions for transport coefficients.

Spitzer Jr (1962) sets $p_{\mathrm{max}} = \lambda_{\mathrm{D}}$, where λ_{D} is the Debye shielding length mentioned in Section 1.3.1. Note that from equation (1.2) this is much greater than $n^{-1/3}$. Substituting $mv^2 = 3k_{\mathrm{B}}T$ in order to obtain a mean value for p_0, one arrives at

$$\Lambda = \frac{\lambda_{\mathrm{D}}}{p_0} = \frac{12\pi}{ZZ_{\mathrm{f}}}\left(\frac{\varepsilon_0 k_{\mathrm{B}}}{e^2}\right)^{3/2}\left(\frac{T^3}{n_{\mathrm{e}}}\right)^{1/2}, \tag{5.6}$$

for the case in which a light test particle (unsubscripted) encounters a heavy field particle (subscript f). This equation must in fact be modified at high temperatures because of quantum mechanical effects. A particle passing through a circular aperture of radius p is spread by diffraction through an angle $\lambda/2\pi p$, where $\lambda = h/mv$ is the particle de Broglie wavelength. If this exceeds the classical deflection ($\sim 2p_0/p$ at large p), then the expression for Λ must be modified. $\lambda \propto T^{-0.5}$ and $p_0 \propto T^{-1}$, hence as temperature increases both lengths decline, but p_0 does so more rapidly. The temperature above which quantum effects becomes important is therefore given by

$$\lambda > 4\pi p_0 \quad \Rightarrow \quad \frac{h}{mv} > \frac{ZZ_{\mathrm{f}}e^2}{\varepsilon_0 mv^2} \quad \Rightarrow \quad \left(\frac{3k_{\mathrm{B}}T}{m}\right)^{1/2} > \frac{ZZ_{\mathrm{f}}e^2}{\varepsilon_0 h}$$

$$\therefore \quad T > T_{\mathrm{c}} \equiv \left(\frac{ZZ_{\mathrm{f}}e^2}{\varepsilon_0 h}\right)^2 \frac{m}{3k_{\mathrm{B}}}. \tag{5.7}$$

When T exceeds this value, the value of Λ from equation (5.6) must be multiplied by a factor

$$f = \frac{2\alpha c}{v}, \tag{5.8}$$

where α is the fine-structure constant

$$\alpha \equiv \frac{e^2}{2c\varepsilon_0 h}. \tag{5.9}$$

$$\Rightarrow \quad \frac{2\alpha c}{v} = \frac{e^2}{\varepsilon_0 h}\left(\frac{m}{3k_{\mathrm{B}}T}\right)^{1/2} = \frac{1}{ZZ_{\mathrm{f}}}\left(\frac{T}{T_{\mathrm{c}}}\right)^{-1/2}.$$

To summarize,

$$\Lambda = \begin{cases} \dfrac{12\pi}{ZZ_{\mathrm{f}}}\left(\dfrac{\varepsilon_0 k_{\mathrm{B}}}{e^2}\right)^{3/2}\left(\dfrac{T^3}{n_{\mathrm{e}}}\right)^{1/2} & T < T_{\mathrm{c}} \\ \dfrac{1}{ZZ_{\mathrm{f}}}\left(\dfrac{T}{T_{\mathrm{c}}}\right)^{-1/2}\dfrac{12\pi}{ZZ_{\mathrm{f}}}\left(\dfrac{\varepsilon_0 k_{\mathrm{B}}}{e^2}\right)^{3/2}\left(\dfrac{T^3}{n_{\mathrm{e}}}\right)^{1/2} & T \geqslant T_{\mathrm{c}} \end{cases} \tag{5.10}$$

$$T_{\mathrm{c}} = \left(\frac{ZZ_{\mathrm{f}}e^2}{\varepsilon_0 h}\right)^2 \frac{m}{3k_{\mathrm{B}}} \tag{5.11}$$

Numerically,

$$\ln\Lambda = \begin{cases} 40.5 - \ln\left(ZZ_{\mathrm{f}}\right) - \dfrac{1}{2}\ln\left(\dfrac{n_{\mathrm{e}}}{10^{-3}\,\mathrm{cm}^{-3}}\right) + \dfrac{3}{2}\ln\left(\dfrac{T}{10^8\,\mathrm{K}}\right) & T < T_{\mathrm{c}} \\ 37.8 - \ln\left(ZZ_{\mathrm{f}}\right) - \dfrac{1}{2}\ln\left(\dfrac{n_{\mathrm{e}}}{10^{-3}\,\mathrm{cm}^{-3}}\right) + \ln\left(\dfrac{T}{10^8\,\mathrm{K}}\right) + \dfrac{1}{2}\ln\left(\dfrac{m}{m_{\mathrm{e}}}\right) & T \geqslant T_{\mathrm{c}} \end{cases} \tag{5.12}$$

$$T_{\mathrm{c}} = 4.2 \times 10^5\,\mathrm{K}\left(\frac{m}{m_{\mathrm{e}}}\right)\left(ZZ_{\mathrm{f}}\right)^2 \tag{5.13}$$

In the temperature range appropriate to the ICM, electrons are in the high-temperature regime where quantum effects are important, whereas ions are always in the low-temperature regime. The increased charge of most ions can compensate for the otherwise slightly higher $\ln\Lambda$ value appropriate to the low-temperature regime. Recalling the definition of Λ, equation (5.5), the fact that $\mathrm{e}^{38} \sim 10^{16}$ shows just how long-range the electrostatic force is. For extremely high densities and extremely low temperatures (on ICM scales), the theory breaks down and no value of Λ can be given.

Z	1	2	4	16	∞
δ_T	0.225	0.356	0.513	0.791	1.000
ε	0.419	0.410	0.401	0.396	0.400

Table 5.1: Numerical coefficients for the Spitzer conductivity. From Spitzer Jr and Härm (1953, table III).

5.3 The Spitzer Conductivity

Much early theoretical work on the transport properties of ionized gases was carried out in the 1950s[1] by a variety of authors (e.g. Cohen *et al.*, 1950; Spitzer Jr and Härm, 1953). The name most closely associated with this topic is that of Spitzer, due to his excellent summary of the subject (Spitzer Jr, 1962). The theoretical value for the thermal conductivity of an ionized gas in the absence of magnetic fields is frequently referred to as the Spitzer conductivity.

Classical thermal conduction is a diffusive process in which the electrons (due to their greater mobility) redistribute thermal energy through the gas. In an ionized gas, the presence of a temperature gradient sets up not only a flow of heat energy, but also an electric current. In a steady state, there is no net flow of current because the motion of charges sets up a secondary electric field which results in the cancellation of the temperature gradient current. The flow of heat is also reduced by this secondary electric field, so that the effective coefficient of conductivity is reduced by a factor $\varepsilon \sim 0.4$.

A useful concept when dealing with ionized gases is that of the "Lorentz gas". In such a gas, the electrons do not interact with each other, and the positive ions are at rest (this is therefore equivalent to the limit of infinite mass ions, $Z \to \infty$). It was shown by Spitzer Jr and Härm (1953) that the coefficient of thermal conductivity for such a gas is

$$\kappa_{\mathrm{L}} = 640 \left(\frac{2\pi}{m_{\mathrm{e}}}\right)^{1/2} \frac{k_{\mathrm{B}}\varepsilon_0^2}{e^4} \frac{(k_{\mathrm{B}}T)^{5/2}}{Z\ln\Lambda} \approx \frac{5.2\times10^9\,\mathrm{W\,K^{-1}\,m^{-1}}}{Z} \left(\frac{T}{10^8\,\mathrm{K}}\right)^{5/2} \left(\frac{\ln\Lambda}{37.8}\right)^{-1}, \tag{5.14}$$

where Ze is the ion charge. This expression does not include the factor of ε. Using equation (1.11), this can be written in terms of the electron mean free path λ,

$$\kappa_{\mathrm{L}} = \frac{640\times0.714}{9\sqrt{2\pi}} Z^3 k_{\mathrm{B}} n_{\mathrm{e}} \lambda \left(\frac{k_{\mathrm{B}}T}{m_{\mathrm{e}}}\right)^{1/2} \sim k_{\mathrm{B}} n_{\mathrm{e}} \lambda \bar{v}. \tag{5.15}$$

For a real gas, there is an additional numerical factor, so that the Spitzer conductivity (as it will be referred to hereinafter) is given by

$$\kappa_{\mathrm{S}} = \varepsilon\delta_{\mathrm{T}}\kappa_{\mathrm{L}}. \tag{5.16}$$

The coefficients ε and δ_{T} are functions of Z — see Table 5.1. The $Z = 1$ case is appropriate for the ICM where hydrogen dominates in terms of the electron number density, irrespective of metallicity (see Table B.2 on page 161).

It should be noted that this formula is only accurate to within about five or ten per cent under as-

[1] As with the studies of fluid dynamics carried out around the same time, work was often supported by the U. S. Atomic Energy Commission.

trophysical conditions, deriving as it does from an approximate solution of the full Boltzmann equation (Spitzer Jr and Härm, 1953[2]).

5.4 Saturated Conduction

Equation (5.1) implies that as the temperature gradient increases, so too does the rate at which energy is carried by the conductive heat flux. In practice, there must of course be some limit to this process. Since thermal conduction transfers energy through the diffusion of electrons, one would expect the thermal velocity of the electrons to act as a limit on the maximum rate at which heat can flow. The "classical" formula for the conductive heat flux, equation (5.1), and the associated derivation of equation (5.16), both assume that the electron mean free path is short compared to the temperature length scale $T/|\Delta T|$. If this assumption becomes invalid, then it is no longer acceptable to treat conductive energy transfer as a diffusive process. This process of "saturation" was described by Cowie and McKee (1977), who showed that

$$q_{\text{sat}} \approx 0.4 \left(\frac{2k_{\text{B}}T}{\pi m_{\text{e}}} \right)^{1/2} k_{\text{B}} n_{\text{e}} T \sim p v_{\text{thermal}} \tag{5.17}$$

is an appropriate value for the heat flux in this regime. It is important to stress that this is only an estimate, in that the numerical coefficient may not be exactly correct. This result is obtained taking into account two constraints. First, that the electron distribution be stable against plasma instabilities that would otherwise tend to restore isotropy; and second that there be no net current flow. By analogy with the classical case, the latter requirement is assumed to be fulfilled through the emergence of a self-consistent electric field. This is the origin of the leading factor of 0.4 (see Table 5.1 on the facing page for the $Z = 1$ case). The analysis also assumes a Maxwell–Boltzmann distribution (the actual choice of distribution function has only a small effect) of electron velocities, so that T in equation (5.17) is really the electron temperature, T_{e}. As discussed in Section 1.3.4, however, it is safe to assume that the ICM electrons and ions have the same temperature.

Since the mean free path in the ICM is typically short relative to other length scales (Section 1.3.6), conduction is not usually saturated. In practice, it can be preferable to smoothly interpolate between the classical and saturated limits. For example, Sarazin (1988) uses the form

$$\boldsymbol{q} = -\frac{\kappa T}{\dfrac{T}{|\nabla T|} + 4.2\lambda} \frac{\nabla T}{|\nabla T|}. \tag{5.18}$$

5.5 Effect of Magnetic Fields

5.5.1 Electron Motion

The force experienced by a particle of charge Ze in a magnetic field $\boldsymbol{B}$ is $Ze\boldsymbol{v} \times \boldsymbol{B}$. Motion parallel to the field lines is therefore unaffected, whereas the component of velocity perpendicular to the field

[2] The concluding remark of this paper is "In view of the lack of observational data in this field, development of a more refined theory does not seem worth the very considerable effort required."

experiences a centripetal force of magnitude $ZeBv_\perp$. The net result is a helical path that circles the field line with a gyroradius given by

$$r_g = \frac{mv_\perp}{ZeB} \tag{5.19}$$

Substituting for the thermal speed leads to

$$r_g = \frac{\sqrt{2mk_B T}}{ZeB} \approx \frac{3.1 \times 10^6\,\mathrm{m}}{Z} \left(\frac{m}{m_e}\right)^{1/2} \left(\frac{T}{10^8\,\mathrm{K}}\right)^{1/2} \left(\frac{B}{1\,\mu\mathrm{G}}\right)^{-1}. \tag{5.20}$$

5.5.2 Uniform Fields

The electron gyroradius given by equation (5.20) is many orders of magnitude smaller than the mean free path of equation (1.11). The proton gyroradius is some 43 times larger than that of the electron, but this length is still entirely negligible on cluster scales. For almost all purposes, it is therefore acceptable to regard the electrons and ions in the ICM as being tied to any magnetic field lines that are present, and constrained to move along them.

In the presence of a uniform magnetic field, then, thermal conduction perpendicular to the field lines is essentially non-existent, whereas conduction parallel to the field lines is unaffected[3]. In the more general case in which a uniform field lies at some angle θ to the temperature gradient, the classical heat flux parallel to the gradient is reduced by a factor of $\cos^2\theta$. One factor of $\cos\theta$ is due to the fact that only the component of the temperature gradient along the field direction is able to have any effect; the other is due to the projection of the resulting heat flux back into the direction of the gradient. Saturated conduction is independent of the temperature gradient, so that only the second factor of $\cos\theta$ applies in the saturated case (e.g. Cowie and McKee, 1977).

5.5.3 Tangled Fields

Observations of Faraday rotation measures show that the magnetic fields in clusters have coherence lengths that are $\sim 10\,\mathrm{kpc}$ (e.g. Carilli and Taylor, 2002). The uniform field model is therefore of little use in connection with (large-scale) studies of the ICM. A more appropriate picture is one of a tangled magnetic field (e.g. Jaffe, 1980); unfortunately the theoretical treatment of such fields is more difficult and probably still incomplete.

In early theoretical work, Tribble (1989) modelled a tangled magnetic field in terms of a random walk of field lines between two separated parallel plates. As the separation of the plates is increased, the fraction of magnetic field lines that manage to successfully cross the entire gap declines, so that conduction over distances larger than a few steps of the random walk is greatly reduced. Indeed, the conventional relationship between heat flux and temperature gradient is destroyed, in that for plates of fixed temperature, the resulting heat flux is no longer linearly proportional to their separation. Such a model also leads inevitably to a highly multi-phase ICM (Chapter 6) in which there is a wide range of temperatures at any given radial position, since separate magnetic field lines are effectively thermally isolated from one another.

[3] Narayan and Medvedev (2001) claim $\kappa_\parallel = {}^1\!/_3\kappa_S$, because "diffusion is in one dimension rather than three".

Qualitatively, the reduction in thermal conduction in the presence of a tangled magnetic field can be understood as a consequence of the increased distances an electron must travel in order to pass between any two given points, owing to its confinement to the randomly-wandering field line. The effective temperature gradient is therefore reduced.

Tribble (1989) also made some consideration of a stochastic magnetic field, in which the separation of neighbouring field lines diverges rapidly, but concluded that this effect was unlikely to be important in the ICM. Rechester and Rosenbluth (1978) developed the theory for a chaotic tangled magnetic field, in which the separation of neighbouring field lines increases with an exponential Kolmogorov–Lyapunov scaling with distance along the lines; that is,

$$r \sim r_0 \exp \frac{l}{l_\mathrm{s}}, \tag{5.21}$$

where r_0 is the initial separation of two field lines, and r is the separation after moving a distance l along one of the lines. l_s is a characteristic length scale of the problem (of the order of the magnetic field coherence length).

This exponentially increasing separation of field lines (i.e. electron paths) can mean that the suppression of conduction is not as severe as might otherwise be expected. Different authors, however, have come to different conclusions about whether or not this exponential divergence allows conduction to operate at non-negligible levels, as we shall see.

There is a second factor which acts to reduce the thermal conductivity — that of *magnetic mirroring* (e.g. Malyshkin and Kulsrud, 2001). An electron moving along a field line can be trapped by a region of high magnetic intensity (where field lines converge) which reflects it back along the field line. An electron can be bottled between two mirrors until collisions reduce its pitch angle to such an extent that it can escape. Malyshkin and Kulsrud (2001) found that magnetic mirrors with separations of the same scale as the electron effective mean free path are most important for impeding electron motion. In the limit in which the magnetic field decorrelates slowly (i.e. the length scale for separation is much greater than the electron mean free path), there is little mirroring effect.

Chandran and Cowley (1998) considered the effects of magnetic mirror trapping of electrons, and concluded that it (operating in conjunction with exponentially diverging field lines) would act to suppress thermal conduction in the ICM by factors of several hundred relative to the Spitzer value, to an extent that is "consistent with the neglect of thermal conduction in homogeneous cooling-flow models of intracluster plasma." Malyshkin and Kulsrud (2001), however, used a different approach to the problem and reached different conclusions: "whether or not electron thermal conductivity in clusters of galaxies is sufficiently inhibited that it can be ignored is still an open question." Malyshkin (2001) went on to examine thermal conduction more specifically in relation to galaxy clusters, concluding that "conduction should be included in hydrodynamical simulations of galaxy clusters, even when the reduction factor is large."

Narayan and Medvedev (2001) made the further extension of considering a *turbulent*, tangled magnetic field. The effect of turbulence is to produce a "multiscale" chaotic magnetic field. By this it is meant that the magnetic field no longer has a single coherence length, and as a consequence there is no longer a single scale length appropriate for equation (5.21). Narayan and Medvedev (2001) show that if the ICM turbulence extends over at least two orders of magnitude in scale, thermal conduction can operate at a high efficiency, within factors of a few of the Spitzer value. Although the ICM probably is a turbulent environment, due to the wakes of constituent galaxies, radio sources, and so on, I would feel it premature to say that the magnetic field must be turbulent over such a wide range of length scales as Narayan and Medvedev (2001) claim is necessary to give rise to effective conduction.

Indeed, the general issue of the transport properties of a chaotically magnetized ionized gas is a complex one, still deserving of study in its own right before the ideas can be applied with confidence to astrophysical situations. It has great relevance for laboratory plasma physics, for example in tokamaks (e.g. Rosenbluth and Kaufman, 1958; Rechester and Rosenbluth, 1978). Personally, I feel that our understanding of the theoretical underpinnings of this important issue is still incomplete.

5.6 The Field Stability Length

It was shown by Field (1965) that thermal conduction can suppress thermal instability in some cases. Conceptually, if the time taken for conductive heat transfer across a density fluctuation is less than the radiative cooling time of the fluctuation, then the growth of thermal instability is suppressed. Through a linearized perturbation analysis Field (1965) showed that exists a *critical wavenumber for stability*, conduction being able to suppress fluctuations with wavenumbers greater than this critical value. The stability analysis is somewhat more complicated in a bulk-flow environment (e.g. Balbus and Soker, 1989), but the basic principles are not invalidated.

The *Field critical wavelength for stability* is given by

$$\lambda_{\mathrm{c}} \equiv k_{\mathrm{c}}^{-1} \tag{5.22}$$

(with some ambiguity as to whether or not a factor of 2π is included in the wavelength–wavenumber conversion), so that fluctuations on wavelengths less than λ_{c} can be suppressed by thermal conduction. Böhringer and Fabian (1989) derive almost the same result for λ_{c} by a simpler method that considers the ratio of the timescales for thermal conduction and radiative cooling for a density fluctuation.

Algebraically,

$$k_{\mathrm{c}} \equiv \sqrt{k_\kappa (k_\rho - k_T)}, \quad \text{where} \tag{5.23}$$

$$k_\rho \equiv \frac{\mu m_{\mathrm{H}}(\gamma-1)\rho}{k_{\mathrm{B}} c_{\mathrm{s}} T}\left(\frac{\partial \mathcal{L}}{\partial \rho}\right)_T \qquad k_T \equiv \frac{\mu m_{\mathrm{H}}(\gamma-1)}{k_{\mathrm{B}} c_{\mathrm{s}}}\left(\frac{\partial \mathcal{L}}{\partial T}\right)_\rho \qquad k_\kappa \equiv \frac{k_{\mathrm{B}} c_{\mathrm{s}} \rho}{\mu m_{\mathrm{H}}(\gamma-1)\kappa}. \tag{5.24}$$

Using equation (3.21) to rewrite the specific luminosity $\mathcal{L}(\rho, T)$ in terms of the cooling function Λ, and using $c_{\mathrm{s}}^2 = \gamma p/\rho$ for the sound speed,

$$k_{\{\rho,T\}} = \frac{\gamma-1}{\sqrt{\gamma}}\left(\frac{\mu}{k_{\mathrm{B}}}\right)^{\frac{3}{2}} \frac{X_{\mathrm{H}}^2 r_{\mathrm{e}}}{\sqrt{m_{\mathrm{H}}}} \frac{\rho}{\sqrt{T}} \left\{\frac{\Lambda}{T}, \frac{\mathrm{d}\Lambda}{\mathrm{d}T}\right\} \qquad k_\kappa = \frac{\sqrt{\gamma}}{\gamma-1}\left(\frac{k_{\mathrm{B}}}{\mu m_{\mathrm{H}}}\right)^{\frac{3}{2}} \frac{\rho\sqrt{T}}{\kappa}. \tag{5.25}$$

Using these expressions,

$$k_{\mathrm{c}} = n_{\mathrm{H}} \sqrt{\frac{r_{\mathrm{e}}}{\kappa}} \sqrt{\frac{\Lambda}{T} - \frac{\mathrm{d}\Lambda}{\mathrm{d}T}}. \tag{5.26}$$

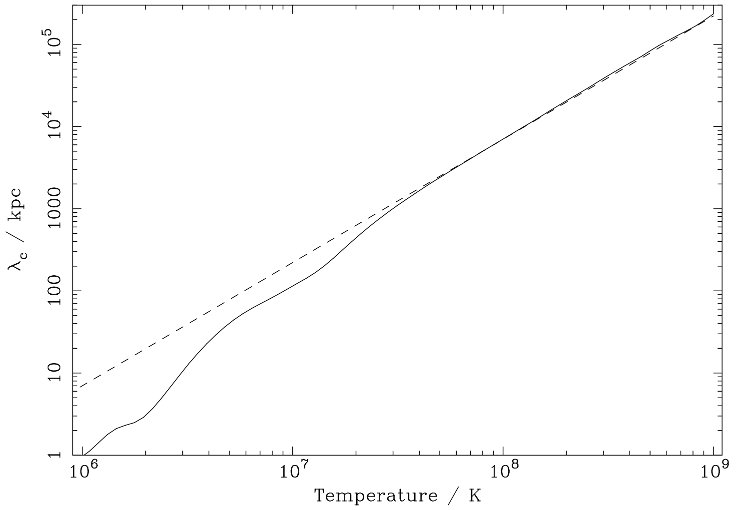


Figure 5.1: The temperature dependence (solid line) of the Field (1965) stability length, calculated as the inverse of k_c from equation (5.28). This is for a $0.3\,Z_\odot$ metallicity plasma with an electron density $n_e = 10^{-3}\,\mathrm{cm}^{-3}$, and thermal conductivity at the Spitzer value of equation (5.16). $\lambda \propto \sqrt{\kappa}/n_e$. The dashed line is a $T^{3/2}$ proportionality fit to the high-temperature range.

Substituting for κ from equation (5.16) leads to

$$k_\kappa = \frac{\sqrt{\gamma}}{\gamma - 1} \frac{e^4}{640\varepsilon_0^2\,(\mu m_H)^{3/2}} \sqrt{\frac{m_e}{2\pi}} \frac{Z \ln\Lambda}{\varepsilon\delta_T} \frac{\rho}{(k_B T)^2} \tag{5.27}$$

$$k_c = \frac{e^2}{8\varepsilon_0} \sqrt{\frac{r_e}{10 k_B}} \left(\frac{m_e}{2\pi}\right)^{\frac{1}{4}} \sqrt{\frac{Z\ln\Lambda}{\varepsilon\delta_T} n_H\,(k_B T)^{-\frac{5}{4}}} \sqrt{\frac{\Lambda}{T} - \frac{d\Lambda}{dT}}. \tag{5.28}$$

Numerically, for $\gamma = 5/3$, $\mu = 0.6$, $X_H = 0.7$, $r_e = 1.2$, $\ln\Lambda = 38$, and using data from Table 5.1 on page 110 for the $Z = 1$ case,

$$k_{\{\rho,T\}} \approx 4.5 \times 10^{-6}\,\mathrm{pc}^{-1} \left(\frac{n}{\mathrm{cm}^{-3}}\right) \left(\frac{T}{10^8\,\mathrm{K}}\right)^{-\frac{1}{2}} \left\{\frac{\Lambda}{T}, \frac{d\Lambda}{dT}\right\} \tag{5.29}$$

$$k_\kappa \approx 2.0 \times 10^{-3}\,\mathrm{pc}^{-1} \left(\frac{n}{\mathrm{cm}^{-3}}\right) \left(\frac{T}{10^8\,\mathrm{K}}\right)^{-2} \tag{5.30}$$

$$k_c \approx 9.3 \times 10^{-5}\,\mathrm{pc}^{-1} \left(\frac{n}{\mathrm{cm}^{-3}}\right) \left(\frac{T}{10^8\,\mathrm{K}}\right)^{-\frac{5}{4}} \sqrt{\frac{\Lambda}{T} - \frac{d\Lambda}{dT}}. \tag{5.31}$$

with $\frac{\Lambda}{T}$, $\frac{d\Lambda}{dT}$ in units of $2 \times 10^{-23}\,\mathrm{erg\,cm^3\,s^{-1}}$ per 10^8 K. Thus, for $n = 0.01\,\mathrm{cm}^{-3}$, $\lambda_c \approx 1\,\mathrm{Mpc}$.

If one makes the simplifying assumption that $\Lambda = \Lambda_0 T^\alpha$, then $\frac{\Lambda}{T} - \frac{d\Lambda}{dT} \to (1-\alpha)\frac{\Lambda}{T} \propto T^{\alpha-1}$, and consequently $\lambda_c \propto n^{-1} T^{7/4-\alpha/2}$. Thus at high temperatures with $\Lambda \propto T^{1/2}$, $\lambda_c \propto T^{3/2}$. The temperature dependence of λ_c is illustrated in Figure 5.1.

5.7 The Effect of Thermal Conduction on the Intracluster Medium

There are several issues that must be addressed in connection with the operation of thermal conduction in the ICM and its possible effect on cooling flows.

First, is it even possible for thermal conduction to have a significant effect on the intracluster gas? This is (from a theoretical point of view), the easiest question to answer. There are several approaches that can be taken. Figure 5.1 on the preceding page shows that in the density and temperature range appropriate for the ICM, conduction at the Spitzer level ought to be able to suppress density fluctuations on scales $\lesssim 100\,\mathrm{kpc}$, which is clearly significant.

Another way to assess the potential relevance of conduction is through calculation of an "effective conductivity". In this method, the total energy lost by a spherical volume of radius r, due to radiation, L, is converted to an equivalent energy flux that must flow in across the outer boundary, $L/4\pi r^2$, to maintain a steady state. Estimating the temperature gradient at the outer boundary, one may then obtain an effective conductivity, which is the level required if thermal conduction is to provide the requisite heat flux. That is,

$$\kappa_{\mathrm{eff}} = \frac{L(<r)}{4\pi r^2}\left(\frac{\mathrm{d}T}{\mathrm{d}r}\right)^{-1}. \tag{5.32}$$

If this effective conductivity is less than the Spitzer value, then it is energetically possible for conduction to play a role. This method was used by Fabian *et al.* (2002b), for example, who used a simple estimate $\mathrm{d}T/\mathrm{d}r \sim T/r$ for the temperature gradient, applied at the cooling radius, to obtain a bulk estimate of the importance of conduction. The result for the majority of clusters in their sample was a value between one and one-tenth of the Spitzer conductivity.

Bregman and David (1988) perform essentially the same analysis, but frame the result in terms of the maximum length scale of a temperature gradient that can be supported by conduction. If radiative losses are not to exceed conduction, then

$$\frac{L}{4\pi r^2} < \kappa \frac{\mathrm{d}T}{\mathrm{d}r}.$$

Using $L \sim \frac{4}{3}\pi r^3 n_{\mathrm{e}} n_{\mathrm{H}} \Lambda$, and $\frac{\mathrm{d}T}{\mathrm{d}r} \sim \frac{T}{r} \quad \Rightarrow r^2 < \frac{3\kappa T}{n^2 \Lambda}.$

At $10^8\,\mathrm{K}$, $\Lambda \approx 2\times 10^{-23}\,\mathrm{erg\,cm^3\,s^{-1}}$. Making use of equation (5.16) for κ shows that, for a density of $10^{-2}\,\mathrm{cm^{-3}}$, thermal conduction can suppress radiative losses for gradients with length scales less than around 0.5 Mpc. This considerably exceeds the cluster cooling radius.

Narayan and Medvedev (2001), on the other hand, assess the bulk effectiveness of conduction by comparing an estimate for the time required for heat to diffuse conductively across a sphere of given radius with the radiative cooling time. All these methods are at heart equivalent, and return the same global assessment of the energetics of thermal conduction. There is no dispute that the bulk of the intracluster gas lies outside the cooling radius at any instant, e.g. in Abell 1835 $\gtrsim 97$ per cent of the mass lies outside the central 100 kpc (Section 1.6). This gas constitutes a significant reservoir of thermal energy. If it is possible to efficiently channel this energy (or even a fraction of it) into the central regions where cooling is occurring, then analyses of the kind outlined above show that the cooling would be significantly reduced. In fact, there may be *too much* energy, in the sense that cooling may be completely prevented, which does not match with the observational X-ray picture at all.

At a slightly more detailed level, Fabian *et al.* (1994b) solve the cooling flow energy equation

$$n\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{5k_{\mathrm{B}}T}{2\mu m_{\mathrm{H}}}\right) = -n_{\mathrm{e}}n_{\mathrm{H}}\Lambda + \nabla\cdot(\kappa\nabla T) \tag{5.33}$$

for the special case of a steady-state with zero net enthalpy change. Assuming that $\Lambda \propto T^{\alpha}$, $\kappa \propto T^{5/2}$ (and therefore neglecting the slight temperature dependence of the Coulomb logarithm) leads to the result $T \propto r^{2/(11/2-\alpha)}$.

Fabian *et al.* (2001b) report on *Chandra* observations of Abell 1795, in which an 80 kpc long filament of X-ray emission is seen in the core. This feature coincides with a previously known Hα filament. There appears to be a direct connection of some kind between the X-ray gas and the optical emission. These authors point out that if thermal conduction was operating at the Spitzer level, the cold clouds in the filament would be destroyed by evaporation in the hot ambient gas in $\sim 10^8$ yr, and conclude that conduction must be suppressed by at least an order of magnitude.

The second question that must be examined relating to ICM conduction follows directly from the first. The quoted answers to that question have all assumed the Spitzer conductivity of equation (5.16), which is the value appropriate for an unmagnetized plasma. Magnetic fields clearly are present in the ICM (Section 1.3.2), so the second question is, how will these fields affect the thermal conductivity of the ICM plasma? As discussed in Section 5.5.3, in my opinion this is really a question of plasma physics that ought to have an answer outside of X-ray studies on the ICM. And yet the answers that have been given to this question in the context of ICM studies have (to some extent) tended to depend on what one believes about the intracluster medium. Proponents of the multi-phase cooling flow picture (Chapter 6) have generally argued for a strong suppression of conduction, as is required by that paradigm. Those unhappy with the large mass deposition rates inferred from some X-ray analyses have been more in favour of conduction operating efficiently to prevent mass drop out.

An important piece of evidence is provided by those *Chandra* observations of various clusters that have revealed the presence of so-called "cold fronts", which are abrupt jumps in the surface brightness profiles; e.g. in Abell 2142 (Markevitch *et al.*, 2000) and Abell 3667 (Vikhlinin *et al.*, 2001a,b). These features are extended edges where the surface brightness changes by a significant amount (factors of a few) on essentially unresolved scales ($\lesssim 10$ kpc). The density and temperature profiles exhibit jumps of similar magnitude, so that the fronts separate cold, dense regions from hotter, more rarefied ones. The pressure is essentially continuous, so that these edges represent contact discontinuities rather than shocks. Cold-fronts are generally interpreted as the signatures of cold, dense subclumps in the process of merging with the intracluster medium.

Ettori and Fabian (2000) used the presence of a sharp temperature jump in Abell 2142 to argue that thermal conduction across the edge must be highly suppressed. This result is obtained by comparing the time it would take for unimpeded conduction across the front to smooth out the temperature discontinuity with the core-crossing time of the dense clump. The analysis shows that for the extreme sharpness of the edge to be preserved over the lifetime of the clump, conduction must be suppressed by at least a factor of 250–2500. Mazzotta *et al.* (2002), however, appear to detect signs of a Kelvin–Helmholtz instability in the Abell 3667 cold front, which might imply that these objects are more transient structures.

The work of Ettori and Fabian (2000) is important, because it provides one of the few genuine observational constraints on thermal conduction in the ICM. It is important to stress, however, that cold-fronts are quite likely to be special cases, since they only correspond to one particular interface in the cluster gas. If these interfaces are in fact the boundaries between merging subclumps, then it is possible that each region retains a separate magnetic structure. The fact that thermal conduction *between* the regions of isolated magnetic field is highly suppressed would then be expected (since the electrons are well-confined to the magnetic field lines), and of itself does not reveal anything about the efficiency of conduction *within* each separate magnetic region (i.e. the bulk of the ICM).

The third question pertaining to thermal conduction concerns stability. Even if it is energetically feasible for thermal conduction to play a part, can it do so in a long-lived, stable fashion, and over the wide range of mass, and therefore temperature, scales that are required? Since *some* cool gas in seen in clusters, it is necessary to ensure that neither conduction nor radiation be allowed to gain the upper hand. Achieving a balance between these two competing processes, with their radically different density and temperature dependencies, is difficult indeed.

Ruszkowski and Begelman (2002) suggest that a combination of AGN heating and thermal conduction can stabilize cooling flows against excessive cooling, but this is an area that needs more investigation before firm conclusions can be reached about whether it can be effective over the requisite range of cluster properties. Zakamska and Narayan (2003) have carried out a stability analysis of a cluster with both conduction and some local heat source. They find that clusters without conduction are unconditionally unstable. For clusters with conduction, they recover the Field (1965) result that conduction can suppress instabilities shorter than some critical wavelength (Section 5.6). Their analysis is a local one, and considers only linear, tangential perturbations. The five actual clusters they analyse all appear to lie close to the unstable regime.

It is possible that a form of self-regulation may come into play. Initially, magnetic fields are highly disordered and conduction is ineffective in preventing the development of a cooling flow. The slow infall of cooling gas may then draw out the magnetic field lines, producing a field with a more significant radial component. This would allow enhanced thermal conduction that, whilst not stopping the cooling flow altogether, might be effective in reducing the mass-deposition rate below the level at which it would otherwise operate. Such a process might explain the correlation between Faraday rotation measure and X-ray mass deposition rate observed by Taylor *et al.* (2002, figure 12). It might be interesting to explore in simulations the effect of a two-component (one tangled, one radial) field, where the strengths of the components are time-dependent. A crude approximation to this would be to only turn on thermal conduction after a certain length of time had elapsed.

Actually, it is my belief that the second question (is it feasible for conduction to be energetically important in the ICM) is inadequate by itself. Fabian *et al.* (2002b) showed that it may simply be a consequence of the cluster gas fraction that the global effective conductivity for so many clusters seems to lie close to the Spitzer value. That is, the properties of clusters are such that they lie in an interesting regime where conduction *may* be important, if it is operating at levels close to Spitzer, but in of itself, this fact does not imply that efficient conduction *is* occurring. Simple global comparisons are inadequate as diagnostics — more detailed, small-scale analyses are required. The work of Voigt *et al.* (2002, and in preparation) is more interesting, because it shows that in certain clusters the effective conductivity derived at a range of radii (and therefore temperatures) appears to follow profiles close to some multiple ($\lesssim 1$) of the Spitzer value (but see Section 5.9 for more discussion of this issue).

Conduction may also have important effects on the *large-scale* structure and evolution of clusters (e.g. Dos Santos, 2001; Loeb, 2002).

For all these reasons, it is important to carry out *time-dependent* simulations of the ICM over a range of radii, to see how the effects of thermal conduction manifest themselves and evolve. Various authors have attempted to construct simulations of clusters incorporating conduction (e.g. Bregman and David, 1988; Meiksin, 1988; Gaetz, 1989; David *et al.*, 1992).

5.8 Addition of Thermal Conduction to the Numerical Scheme

Accepting that a theoretical solution to the question of the importance of thermal conduction in the ICM will be difficult, let us turn once more to numerical simulation to see what insights it can bring. I shall restrict attention to the single-phase case of Chapter 3. It is not clear to me how one might include the effects of thermal conduction in a multi-phase scenario (Chapter 6), or indeed if such a picture could be theoretically self-consistent, given the assumptions made in the standard multi-phase treatment (Section 6.1). Some discussion of this issue is made in Chapter 7.

From equation (5.3), and using equation (A.5) for the ∇^2 operator in spherical polar co-ordinates, the power transferred per unit mass by conduction is

$$\mathcal{Q} = \frac{1}{\rho r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2 q\right) \tag{5.34}$$

$$q = -\kappa\frac{\mathrm{d}T}{\mathrm{d}r}, \tag{5.35}$$

assuming spherical symmetry. q is the (radial) heat flux, here defined to be positive for radial outflow (since then $\mathrm{d}T/\mathrm{d}r$ is negative). With the addition of a conductive term, the one-dimensional single-phase energy equation, equation (3.19), thus becomes

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}t} = -p\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{\rho}\right) - \mathcal{L}(\rho,T) - \frac{1}{\rho r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2 q\right). \tag{5.36}$$

Note the choice of sign for the conductive term. $4\pi r^2 q$ is the power transmitted across a shell of radius r by a flux q. If this quantity has a positive radial derivative, more power is lost from a zone than enters, and hence the contribution of this term to $\mathrm{d}\varepsilon/\mathrm{d}t$ must be negative. $\mathcal{Q}$ as defined in equation (5.34) therefore represents the specific power *loss* due to conduction.

The heat flux is a function of temperature, $q = q(T)$ (there is also a weak density dependence that enters through the $\ln\Lambda$ term in the conductivity, or an explicit density dependence in the case of saturated conduction). Since $\varepsilon = \varepsilon(T)$ as well, one cannot obtain correct centring of the $\mathcal{Q}$ terms without making the scheme implicit. Indeed, the conductivity term behaves in exactly the same way as the luminosity term, $\mathcal{L}(\rho,T)$, which is discussed in Section 3.3.2.

As with the surface brightness (Section 3.8.2), we have the freedom to associate the heat flux q with either integer or half-integer mesh points in the numerical grid as is convenient. The functional dependence of the flux is $q = q(T, \mathrm{d}T/\mathrm{d}r)$. Given that T is defined on half-integer mesh-points, $\mathrm{d}T/\mathrm{d}r$ would naturally be given on half-integer mesh-points from the centred Euler method. T itself can be converted without much difficulty to integer mesh-points via interpolation. I therefore choose to associate

q with integer mesh-points. This also leads to the simple conceptual interpretation of q_j, q_{j+1} as the heat fluxes into and out of zone j, respectively, across its boundary shells j and $j+1$.

Inspection of equation (5.36) shows that the addition of the conduction term behaves exactly like a modification of the radiative power loss term, $\mathcal{L} \to \mathcal{L} + \mathcal{Q}$. The modified numerical scheme is therefore constructed by adding to $\mathcal{L}^n_{j+\frac{1}{2}}$ in equation (3.50) the term

$$\mathcal{Q}^n_{j+\frac{1}{2}} = \frac{1}{\rho^n_{j+\frac{1}{2}}} \frac{4}{(r^n_{j+1} + r^n_j)^2} \left[\frac{(r^n_{j+1})^2 q^n_{j+1} - (r^n_j)^2 q^n_j}{r^n_{j+1} - r^n_j} \right]. \tag{5.37}$$

The q terms in this expression are all heat fluxes, and should not be confused with the pseudoviscous pressure (denoted by q in Chapter 3). The magnitude of the heat flux is calculated using

$$|q^n_j| = \begin{cases} \kappa\left(T^n_j, n^n_{\mathrm{e,\,j}}\right) \left|\dfrac{\mathrm{d}T}{\mathrm{d}r}\right|^n_j & \text{classical} \\ 0.4 \left(\dfrac{2k_\mathrm{B}}{\pi m_\mathrm{e}}\right)^{1/2} k_\mathrm{B} n^n_{\mathrm{e,\,j}} (T^n_j)^{3/2} & \text{saturated} \end{cases} \tag{5.38}$$

The appropriate value is whichever is the smaller of the classical and saturated heat fluxes. The sign of q is then determined from the sign of $-\mathrm{d}T/\mathrm{d}r$. Since n_e and T are normally associated with half-integer mesh-points, values must be interpolated, e.g. $T^n_j = {}^1\!/_2(T^n_{j+\frac{1}{2}} + T^n_{j-\frac{1}{2}})$. κ is calculated according to equation (5.16); and the temperature gradient from

$$\left.\frac{\mathrm{d}T}{\mathrm{d}r}\right|^n_j = \frac{T^n_{j+\frac{1}{2}} - T^n_{j-\frac{1}{2}}}{{}^1\!/_2(r^n_{j+1} - r^n_{j-1})}. \tag{5.39}$$

5.8.1 Boundary Conditions

More boundary conditions must be added in the presence of thermal conduction. Quantities such as $T_{j-\frac{1}{2}}$ cannot be evaluated directly at $j = 0$, j_max. That is, the heat flowing across the inner and outer shells cannot be calculated without knowing what lies beyond them. Since the inner boundary is at $r_\mathrm{min} = 0$, no heat can flow across this surface. For the outer boundary, a dummy mesh-point of fixed properties is again used, as described in Section 3.3.7, except that now T and n_e are required. Given that the outer boundary is chosen such that no cooling reaches this point, fixing the heat flow across the outer boundary to be zero is essentially equivalent.

5.8.2 Time Step Considerations

The addition of thermal conduction to the difference scheme necessitates a further constraint on the time-step, beyond those discussed in Section 3.3.6. Based on experience with the radiative cooling time, equation (3.54), one might envisage adopting a time-step limit of the form

$$\Delta t = f_\mathrm{cond} t_\mathrm{cond} \qquad f_\mathrm{cond} < 1 \tag{5.40}$$

$$t_\mathrm{cond} \sim \frac{k_\mathrm{B} T}{\mu m_\mathrm{H} \mathcal{Q}}, \tag{5.41}$$

where $t_{\rm cond}$ gives a crude estimate of the time for a zone to lose all its energy through conduction. Given that κ decreases with T, one might expect $f_{\rm cond} \sim 1$ to give acceptable results. In practice, this does not form a very good control on the time step. Indeed, it forms no limit at all if conduction is causing a zone to gain energy rather than lose it. Conduction cannot really be considered in the same way as radiation, because it is a transport process that acts to *redistribute* existing energy between zones, whereas radiation always functions as a *sink* removing energy from the simulation. The inclusion of transport processes makes the world of the simulation more complex.

Another possibility is as follows. Consider three neighbouring zones, $j-1$, j, and $j+1$. Suppose that the central zone is initially the hottest, and is therefore losing energy due to conduction to its neighbours. In the absence of any other factors, the rate of conduction will slow as zone j cools and the temperature gradient decreases to zero, so that a steady, isothermal state is reached. If this is being modelled numerically and too large a time step is taken, the cooling of zone j will overshoot the equilibrium position, so that it becomes colder than its neighbours. At the next iteration, the direction of heat flow will be reversed, as the simulation tries to bring the temperature of zone j back up. If the time step is still too large, overshooting will occur again, leading to a series of oscillations about the mean point. It is desirable to prevent such oscillations. Not only are they artificial, but they actually slow the simulation down by introducing spurious temperature gradients and conductive heat fluxes.

In the above scenario, in a time Δt the temperature of zone j will change according to $k_{\rm B}T_j \to k_{\rm B}T_j - \Delta t \mu m_{\rm H} \mathcal{Q}_j$. The condition of equation (5.40) corresponds to the requirement that the final zone temperature shall be greater than zero. This is clearly inadequate. If overshooting is to be prevented, this requires the temperature of zone j after cooling to be greater than that of either of its neighbours after heating. Considering just the relationship between j and $j-1$, mathematically we require

$$k_{\rm B}T_j - \Delta t \mu m_{\rm H} \mathcal{Q}_j > k_{\rm B}T_{j-1} - \Delta t \mu m_{\rm H} \mathcal{Q}_{j-1}. \tag{5.42}$$

By construction, $T_j - T_{j-1} > 0$, $\mathcal{Q}_j > 0$ (zone j is cooling), and $\mathcal{Q}_{j-1} < 0$ (zone $j-1$ is heating), so that $\mathcal{Q}_j - \mathcal{Q}_{j-1} > 0$. Therefore, the inequality can be rearranged to give

$$\Delta t < \frac{k_{\rm B}}{\mu m_{\rm H}} \frac{T_j - T_{j-1}}{\mathcal{Q}_j - \mathcal{Q}_{j-1}}. \tag{5.43}$$

It is immediately apparent that this could be a much stricter constraint than equation (5.40). Considering the case in which zone j is initially cooler than zone $j-1$ leads to the same inequality. Similar relationships can be derived by considering the zone pair j, $j+1$. In practice, this is not necessary, as one can start in the centre, comparing zone 0 against zone 1, and then work outward.

In this method, one actually needs to take the factor $f_{\rm cond}$ to be slightly greater than unity. Otherwise, the simulation would become trapped in Zeno's paradox, continually taking ever-smaller time steps to approach fractionally closer to equilibrium. An alternative approach is to tolerate oscillations up to some level. Suppose in the framework of the above example that the temperature of zone j initially exceeds that of zone $j-1$ by some amount ΔT. If the temperature of zone j is T, let the oscillation be ignored if $\Delta T/T < f$ for some f (10^{-6}, say, to tolerate oscillations at the level of one part per million). Actually, fully ignoring even small oscillations is not acceptable. Taking too large a time step can potentially amplify any oscillation to unacceptable levels. A more robust approach is to scale the time-step limits

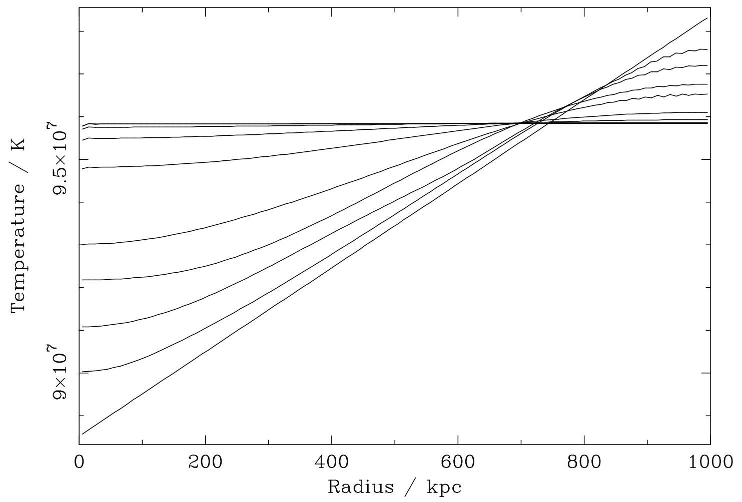


Figure 5.2: Illustrating the erasure of a temperature gradient by thermal conduction. The initial temperature profile has a linear increase with radius. Over time conduction produces isothermality by transporting heat from the outer regions to the centre. This example uses a zero-flux outer boundary condition, and has no radiation or gravitational potential.

derived from small oscillations, such that

$$\Delta t = \begin{cases} \Delta t & \left(\dfrac{T}{\Delta T} f\right) \leqslant 1 \\ \left(\dfrac{T}{\Delta T} f\right) \Delta t & \left(\dfrac{T}{\Delta T} f\right) > 1 \end{cases} \tag{5.44}$$

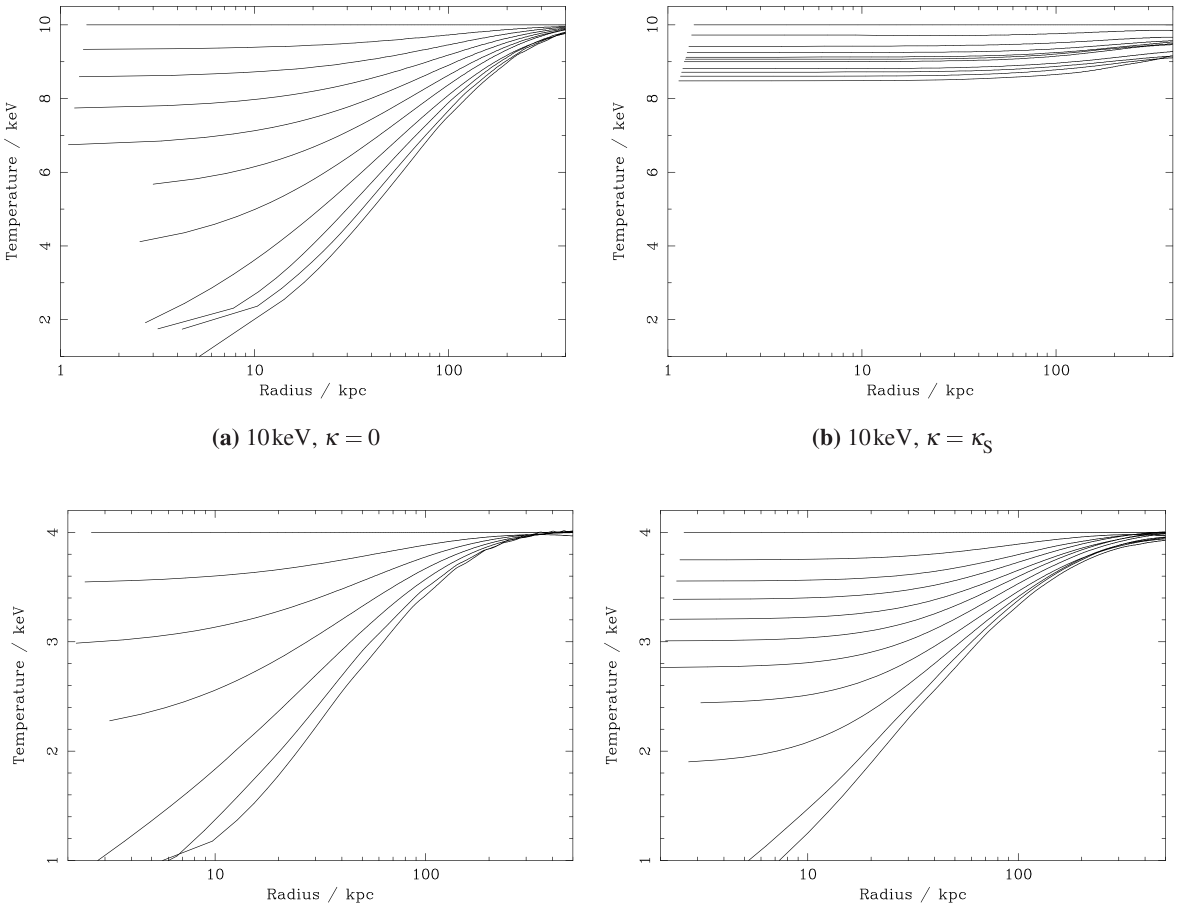
$$\text{i.e. } \Delta t = \max\left(\Delta t, \frac{T}{\Delta T} f \Delta t\right), \tag{5.45}$$

where Δt on the right-hand side of this relation is that from equation (5.43). Thus, for example, if $f = 10^{-6}$ is considered acceptable, then if some zone has a small oscillation $\Delta T / T = 10^{-7}$, the time limit this gives will be scaled up by a factor of ten.

Given that radiation and conduction effectively function as separate decay modes, it might also be worth considering some combined time-limit of the form $\Delta t_{\text{tot}}^{-1} = \Delta t_{\text{cool}}^{-1} + \Delta t_{\text{cond}}^{-1}$, in the case where conduction is causing a zone to lose energy (obviously this would not apply if conduction was causing a zone to gain energy).

5.9 Results of Simulations Including Thermal Conduction

I present here the results of some simulations involving thermal conduction. As a simple illustration of the general effect of conduction, Figure 5.2 shows the evolution of the temperature profile in a system with neither radiative energy losses, nor a gravitational potential. The initial linear temperature gradient is erased over time as a conductive heat flux transports energy from the outer, hotter regions to the inner, cooler regions. Eventually, isothermality is obtained. In general, of course, we can expect much more complicated behaviour when the effects of gravity and radiation are included, but the basic influence of conduction is always the same: a restoring force attempting to bring about an even distribution of thermal energy.



(a) 10 keV, $\kappa = 0$

(b) 10 keV, $\kappa = \kappa_S$

(c) 4 keV, $\kappa = 0$

(d) 4 keV, $\kappa = \kappa_S$

Figure 5.3: The effect of thermal conduction on the time evolution of the temperature profile. (a, b) 10 keV cluster; (c, d) 4 keV cluster. In (a, c) there is no thermal conduction; in (b, d) it operates at the Spitzer level. The various curves in each plot represent different elapsed times: (a) 0–5 Gyr in 10×0.5 Gyr steps; (b) 0-10 Gyr in 10×1 Gyr steps; (c) 0–3 Gyr in 6×0.5 Gyr steps; (d) 0–5 Gyr in 10×0.5 Gyr steps. All results are for a NFW profile cluster with a concentration parameter $c = 5$.

I have carried out simulations where the conductivity is fixed at some multiple of the Spitzer conductivity, equation (5.16). This is a technique that has been applied by many authors (e.g. Bregman and David, 1988) in order to parametrize our ignorance of the precise effect of the ICM magnetic field, as discussed in Section 5.5.3. Note that in line with the comments made at the end of Section 5.5.2, if the classical conduction is assumed to be reduced to some factor f of the Spitzer value, then the saturated conduction should be reduced by a factor $f^{1/2}$. In practice, this is not too important, since ICM conduction is rarely saturated. I am not aware of any considerations of the possibility that the effect of magnetic fields on thermal conduction might be more complex than a simple multiplicative factor on the Spitzer formula (for example, that the temperature dependence might be altered), although it might be interesting to speculate what kind of effect would be needed to produce stability in the ICM.

Figure 5.3 shows some illustrative examples of the effect that thermal conduction has on the evolution of the temperature profiles. Results are shown for clusters with virial temperatures of 4 and 10 keV, both without conduction, and with it operating at the Spitzer level. The scaling relations of Allen *et al.* (2001c) were used to assign the cluster mass profiles.

In the absence of thermal conduction, the initially isothermal temperature profiles develop a cooling core, and eventually reach a steady-state temperature profile with a strong gradient between the cool central regions and the outer regions where little cooling takes place (owing to the lower density and consequently much longer cooling time). The low-temperature cluster reaches this steady-state more rapidly.

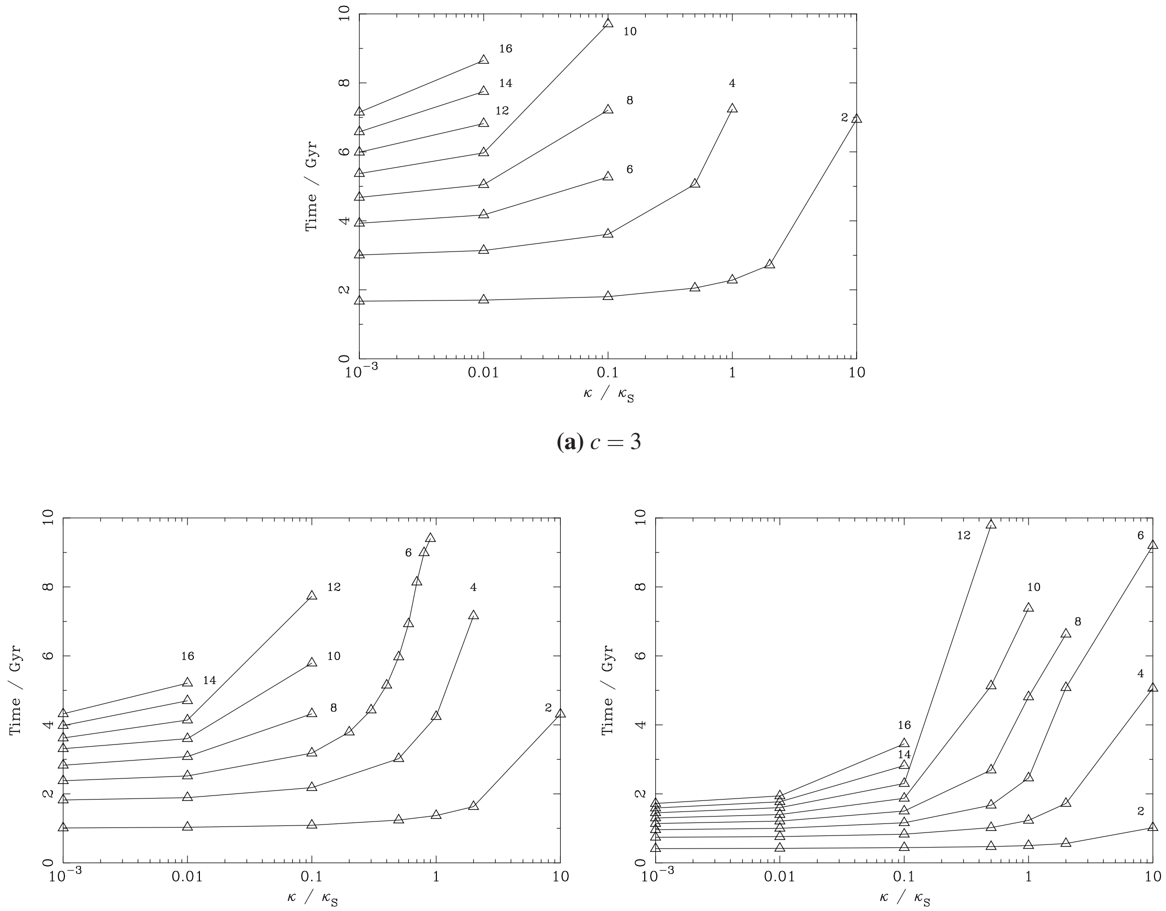
The addition of thermal conduction at the Spitzer level has a marked effect on the evolution of the high-temperature cluster. A cool core never really develops in this case. Instead, the effects of thermal conduction are almost strong enough to prevent the formation of any kind of temperature gradient and enforce a slow isothermal cooling. As it is, only a very slight gradient forms. The evolution of the system to lower temperatures is significantly slowed: compare the final states of Figure 5.3(a) and Figure 5.3(b) on the previous page, which show the states after 5 Gyr and 10 Gyr of evolution, respectively. Even allowing twice as much time, the cluster with conduction hardly cools at all. The addition of thermal conduction therefore has a very marked effect on high-temperature systems.

The situation for the low-temperature cluster is different. Whilst the operation of thermal conduction is able to slow the development of a strong temperature gradient to some extent, it is unable to prevent it. After 5 Gyr have elapsed, the states of the systems with and without conduction are very similar. This difference in behaviour between high- and low-temperature systems is obviously due to the different temperature dependencies of the effects of conduction and radiation. The former declines strongly with temperature, $\propto T^{5/2}$, from equation (5.14). Radiative losses, on the other hand, increase sharply at temperatures $\lesssim 10^7$ keV, as is clear from Figure 1.7(a) on page 17. The net result is that while conduction is able to dominate at high temperatures, at low temperatures radiative losses win out. Note once again that we seem to see clusters spanning an interesting range, where the dominance of one process over the other can be finely balanced.

It is possible to discern a slight difference in the temperature profiles of the low-temperature cluster with and without conduction at late times, at the largest radius shown in the plots. The temperature profile of the simulation without conduction seems to level off relatively sharply to isothermal, whereas that of the system with conduction folds over more gradually. This is presumably due to the need for some small gradient to exist to allow conduction to operate at all. It might be worth investigating the evolution of the cooling radius in simulations with and without conduction. Whilst this might prove interesting, it is difficult to see how this could form the basis of any kind of observational test to distinguish the operation of conduction in real clusters.

This is symptomatic of a general feature that hampers observational studies — the fact that any one object is of course only observed at one particular point in its evolution. *Statistical* arguments can be made through the observation of a large number of objects — for example, if a balance between conduction and radiative losses is unstable and persists only for a brief time in the life of any one object, as in the simulations presented here, then it is singularly unlikely that we will happen to observe many objects all at such a point in their evolution.

From my point of view, one of the chief difficulties in invoking thermal conduction as a solution to the reduced low-temperature line emission seen recently in cooling flow clusters is the difficulty in making the process stable over the wide range of cluster masses (or temperatures) required. It is necessary to produce simulations over a wide range of parameter space, but not feasible to examine the temperature



(a) $c = 3$

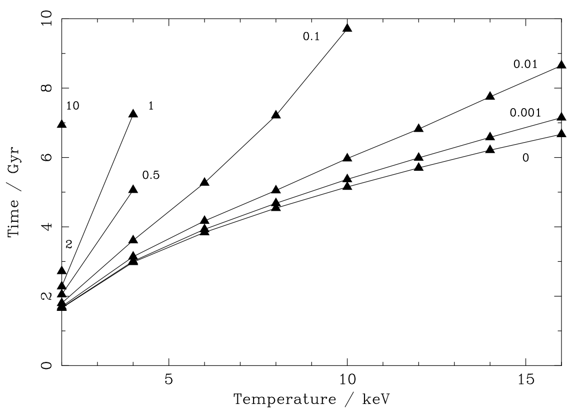
(b) $c = 5$

(c) $c = 8$

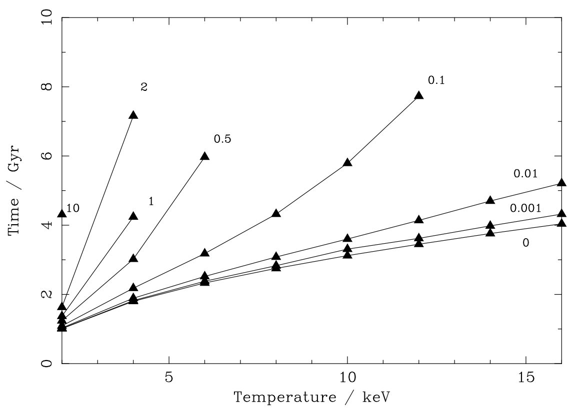
Figure 5.4: Time (if less than 10 Gyr) for central temperature to drop to one-third of its initial value in various clusters. The various plots (a, b, c) are for clusters of different concentration parameter, c. The abscissa in each case is the conductivity as a fraction of the Spitzer value. The various lines in any given plot correspond to different cluster virial temperatures: 2, 4, 6, 8, 10, 12, 14, 16 keV in order of increasing time. Times for each temperature are evaluated at conductivities which are 0.001, 0.01, 0.1, 0.5, 1, 2, 10 times the Spitzer level (with the exception of the $c = 5$, $T = 6$ keV case, for which there are more points). The absence of a point for any temperature indicates that the associated time was longer than 10 Gyr. See also Figure 5.5 on the following page.

profile in each case as a function of time, as was done for the examples of Figure 5.3 on page 123. In any case, the plots of that figure serve to illustrate the general concepts involved.

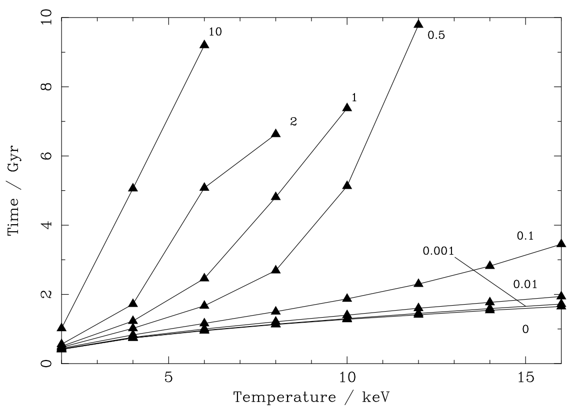
With this is mind, Figures 5.4 to 5.5 on pages 125–126 are intended to provide an overview of the effects of thermal conduction over a fairly wide region of parameter space. Clusters were simulated with a variety of different virial temperatures, concentration parameters and conductivities, and the evolution was followed for 10 Gyr. The figures show the time that elapsed in each case before the central temperature dropped to one-third of its starting value (the absence of a data point indicates that the associated time was longer than 10 Gyr). The results are not changed a great deal if one uses an absolute cooling criterion (e.g. time for central temperature to drop below 1 keV) rather than a relative one (time for central temperature to drop by some factor). Results are plotted for values of the conductivity that exceed the Spitzer value. This is of course unphysical (the Spitzer conductivity, being the value appropriate for



(a) $c = 3$



(b) $c = 5$



(c) $c = 8$

Figure 5.5: As for Figure 5.4 on the previous page, but with temperature rather than conductivity as abscissa. The curve labels refer to conductivity as a fraction of the Spitzer value.

an unmagnetized plasma, acts as an upper limit for the ICM with its magnetic fields), but is intended to allow investigation of just how special a value the Spitzer level is (e.g. if conduction were slightly more efficient, would clusters be able to cool at all?). In Figure 5.4 on the preceding page, results are plotted as a function of κ for various different temperatures; whereas in Figure 5.5 they are plotted against temperature for various levels of conduction.

Various results are apparent from these figures. First, it is more difficult to offset cooling in clusters with higher concentration parameters (due to their denser cores).

Visible in Figure 5.4 on the preceding page is the result derived by Fabian *et al.* (2002b) on general theoretical grounds; namely that the physical properties of clusters are such that they lie close to the region where conduction at the Spitzer level can be significant. This is revealed by the upward curve of the lines around the region where $\kappa = \kappa_S$. This shows that conduction at levels close to κ_S can offset cooling for a time that is significant compared with the likely age of these systems.

From Figure 5.5 we see that we require $\kappa \gtrsim 0.01\kappa_S$, and probably $\kappa \sim 0.1\kappa_S$, if there is to be a significant effect on the evolution of clusters (a pleasing confirmation of those results derived on general energetic grounds).

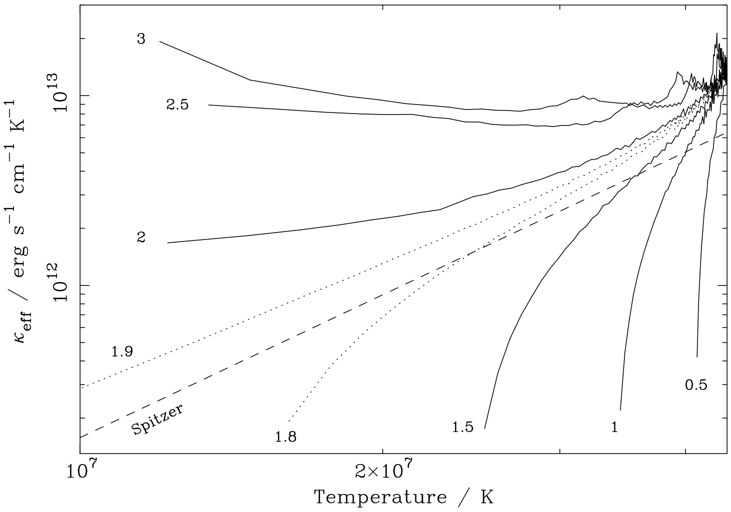


Figure 5.6: The time evolution of the "effective conductivity" for the cluster of Figure 5.3(c) on page 123. The dashed line shows the Spitzer conductivity, $\propto T^{5/2}$, from equation (5.16). Solid lines show the effective conductivity calculated from simulation output as per equation (5.32), at the same times (curve labels, Gyr) as plotted in Figure 5.3(c) on page 123. Dotted lines show the result for two additional times that straddle the Spitzer conductivity.

Returning to Figure 5.3 on page 123, the fact that the temperature profiles for the low-temperature case are in the end so similar means that it would not be possible to distinguish these cases using an effective conductivity plot of the type employed by Voigt *et al.* (2002); that is, a plot of $\kappa_{\rm eff}$ against temperature as derived from equation (5.32) at various radii in the cluster. Figure 5.6 shows a plot of this form as would have been produced from observations of the simulated cluster of Figure 5.3(c) on page 123, which does *not* include any actual thermal conduction. Also shown is the Spitzer conductivity of equation (5.16), with its $T^{5/2}$ temperature dependence. Initially, the $\kappa_{\rm eff}(T)$ profile is steep, which is an inevitable consequence of the isothermal initial conditions. At later time, the temperature decrease combined with the associated density and luminosity increase raises the effective conductivity needed to offset the radiative losses. The net result is that the $\kappa_{\rm eff}$ track swings round in a clockwise fashion. The inferred value swings past the Spitzer locus very quickly indeed (in around 0.1 Gyr). This is a clear illustration of the unstable nature of the balance between conduction and radiation (e.g. Fabian *et al.*, 1984). Indeed, the results of Voigt *et al.* (2002) are both puzzling and remarkable. In the $\kappa_{\rm eff}$ / T parameter space, we *ought* to find that the Spitzer locus is a desert. Either radiative losses should dominate over conduction leading to massive cooling flows (in the absence of other factors), or conduction should produce essentially isothermal clusters with no noticeable gradients. Considering Figure 5.6 once more, imagine *starting* at around the 1.9 Gyr mark, when it is possible for conduction and radiation to roughly balance. The results show that this symmetry must very rapidly be broken, one way or the other, carrying the effective conductivity track far from the Spitzer locus. The fact that this does *not* seem to happen in practice must indicate some missing physics in the simulation.

Note that the 10 keV cluster with $\kappa = \kappa_S$, Figure 5.3(b) on page 123, *is* still cooling, albeit at a greatly reduced rate. Given enough time, it would develop a temperature gradient of the same form as Figure 5.3(a) on page 123. Actually, the choice of outer boundary condition would have an effect here. Cooling to a steady state with a strong temperature gradient will inevitably occur if no heat is allowed to flow in across the outer boundary, since the total energy of the system is fixed and some is always being lost through radiation. If the outer boundary temperature is fixed, though, then this is equivalent to providing a truly infinite heat reservoir that can continue to supply energy to offset cooling indefinitely. In practice, this issue is not very important, since it only has an effect if no strong temperature gradient develops, and then only in the limit of unfeasibly long time scales.

The initial conditions also have relevance for the system evolution. With an initially isothermal profile, conduction is obviously ineffective at the start of the simulation, which allows radiative losses to obtain a foothold. Loeb (2002) has shown that efficient thermal conduction would in fact cause substantial *cooling* of the cluster gas, if the large-scale temperature profiles of clusters decline in the vicinity of the virial radius, as claimed by Markevitch *et al.* (1998) from *ASCA* studies, and De Grandi and Molendi (2002) from *BeppoSAX* data. As mentioned by Fabian *et al.* (2002b), this is perhaps evidence that efficient thermal conduction might be restricted solely to the cores of clusters, if it operates at all.

There is therefore a compelling argument to consider simulations where the efficiency of thermal conduction changes with time, as mentioned in Section 5.7. In such a model, conduction would be initially inefficient (presumably due to a highly tangled magnetic field) and would therefore allow a cooling flow to develop. The inflow of gas might draw out the associated magnetic field, producing a field with a more significant radial component, allowing for enhanced thermal conduction. In this way, a form of negative feedback might come into play that allows for the existence of a steady state with a significant temperature gradient. Such an approach was attempted by Bregman and David (1988), who found it unsuccessful on the whole, but the issue probably merits a reinvestigation in light of recent observational developments. I do not see how thermal conduction in isolation can ever satisfy the observational constraints, which show that there *is* significant cooling in the centres of many clusters, by roughly a factor of three, say, as compared to the outer regions. More work needs to be done in this area. Another obvious candidates for future investigation is adding a heat source of some form (e.g. a central AGN), as considered by Ruszkowski and Begelman (2002), which could probably be done in a relatively straightforward fashion within the numerical scheme used here.

What we see before us is just one tiny part of the world. We get into the habit of thinking, This is the world, but that's not true at all. The real world is a much darker and deeper place than this, and most of it is occupied by jellyfish and things. We just happen to forget all that.

The Wind-up Bird Chronicle
HARUKI MURAKAMI

6 Multi-phase Models of the Intracluster Medium

I present in this Chapter the adaptations necessary to extend the numerical model of Chapter 3 to multi-phase gas (following the general method of Thomas, 1987, 1988). This opens up a very wide parameter space for investigation. Recently, the strength of evidence in favour of a multi-phase ICM seems to be declining (e.g. Johnstone *et al.*, 2002). I have not made a great deal of use of multi-phase models, but include this Chapter more for reference and completeness. See the comments of Section 6.7.

6.1 The Multi-phase Concept

Within the paradigm of a spherically-symmetric cluster, a *single-phase* medium is one in which there exists a single, well-defined temperature and density at any given radius. In other words, the thermodynamic structure of the ICM can be completely described by one temperature profile and one density profile (the pressure profile obviously following from the equation of state). In contrast, in a *multi-phase* medium, gas at a variety of different densities and temperatures may coexist at any one radius. That is, there are a number of *phases* in the ICM, each with its own density and temperature profile. It is usually assumed that the phases coexist in pressure equilibrium, for reasons that will be outlined below, so that there is only one pressure profile, common to all phases.

Indeed, there is a reasonable argument for a single-phase ICM being an unstable state that is likely to fragment into a multi-phase medium. Since the process of radiative cooling is *thermally unstable* (see e.g. Field 1965; the radiation power increases strongly upon cooling due to the associated density increase), it might seem inevitable that any small density fluctuations which happen to exist initially in an almost single-phase medium (a perfect single-phase medium being a practical impossibility of course) would be amplified on the scale of the cooling time, leading to a multi-phase ICM.

There is, however, a powerful counter argument to this idea that a multi-phase ICM would "naturally" arise. It is possible that thermal conduction (as discussed in Chapter 5) could suppress the development of thermal instability in the ICM and so constrain the medium to remain single phase. Within the context of traditional multi-phase models, it is usually assumed that conduction is heavily suppressed. Indeed, it is not clear to me how these two concepts can be combined in one model. I make some discussion of this in Chapter 7.

If a single-phase gas is cooling in a gravitational potential well, then the coolest gas is (inevitably) always at the centre. Therefore, if any mass loss is to occur (i.e. if gas is to cool out), it will do so in the centre. The resulting surface brightness profile would be very highly peaked, more so (it was traditionally said) than would be compatible with deprojected surface brightness profiles (e.g. Fabian, 1994). See figure 11 of Johnstone *et al.* (2002) for a recent illustration of this issue. Equivalently, the fact that derived mass deposition rates were seen to increase with radius in a roughly linear fashion, i.e. $\dot{M} \propto r$, seems to indicate that gas must be cooling out not just at the centre, but over a range of radii, something which would naturally tend to imply a multi-phase medium. Recent observational evidence, however, leans towards the single-phase picture (e.g. Johnstone *et al.*, 2002). In my opinion, the situation is still unclear.

A classic theoretical treatment of a multi-phase, cooling ICM is that of Nulsen (1986). The basic picture, which treats thermal conduction as negligible, is as follows. The appropriate mental image for a multi-phase ICM is not, say, one of a mixture of ideal gases each occupying the full volume and having some partial pressure which is a fraction of the total pressure. Rather, one should picture a collection of plasma "blobs" or "clumps", each occupying some fraction of the total volume.

One of the main assumptions of the analysis is that, whilst the blobs coexisting at any radius may have different temperatures and densities, they remain in pressure equilibrium with each other. This requires that the sound crossing time of a blob be shorter than its cooling time, a usually safe assumption that is broken only for extreme density fluctuations (which in any case cool out extremely rapidly). See Section 1.3.4 for a discussion of some important ICM timescales.

A second important assumption is that the blobs are comoving — that is, there exists a single velocity profile, independent of phase. Dense blobs obviously have a tendency to fall ahead of the mean flow. It is shown by Nulsen (1986), however, that differential pressures forces rapidly build up on such an infalling blob. In the idealized case of a spherical blob, there is a pressure peak at the stagnation point on the leading surface, and pressure minima on either side where the flow past the blob is fastest, due to Bernoulli's principle. Behind the blob is a region of separated flow with intermediate pressure. The net result of these forces is a differential pressure gradient which acts to flatten out the infalling blob perpendicular to the direction of motion.

Perhaps one of the potential weaknesses of the treatment is that it is assumed there are no cohesive forces holding a blob together as a unit. If this is the case, then the pressure forces rapidly shred the blob into smaller fragments. These have a correspondingly lower terminal velocity, the net result being that blobs are broken up and decelerated back down to the mean flow velocity. Thus, assuming a single velocity profile is acceptable when considering global flow properties. It is not immediately obvious to me, however, that neglecting cohesive forces is acceptable. If it is not, then gravitational convection of over-dense blobs might have significant effects. The assumptions of pressure equilibrium and a single velocity profile are necessary, however, if the problem is to be analytically tractable.

A detailed plot of the various behaviours of density fluctuations in the two-dimensional phase space mapped out by density and size is presented in figure 1 of Nulsen (1986). Another illustrated effect is the stabilizing effect that magnetic fields can have, acting to pin blobs to the mean flow and so constituting another motivator for considering a single velocity profile. Once again, we see magnetic fields possibly playing an important role in the ICM. Fabian *et al.* (2002a) suggest that if cold, dense blobs were to fall ahead of the mean flow, this could have significant effects on the observed X-ray spectrum.

An important theoretical tool introduced by Nulsen (1986) is the "volume distribution function", hereinafter referred to as the *volume fraction of the density distribution* (vfdd). $\mathrm{vfdd}(\rho,\boldsymbol{r},t)\,\mathrm{d}\rho$ gives the fraction of the volume at position $\boldsymbol{r}$, time t that is occupied by gas in the density range ρ to $\rho+\mathrm{d}\rho$. Thus, the vfdd serves to describe the spectrum of density fluctuations, and parametrize the extent to which the ICM is multi-phase. The limit of a single-phase medium would be represented by a Dirac delta function.

6.2 Extending the Numerical Scheme to the Multi-phase Case

Since the phases are assumed to comove, and exist in pressure equilibrium, both u and p are independent of phase, and hence the motion equation (3.18) becomes

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -\left(g + \frac{1}{\bar{\rho}}\frac{\mathrm{d}p}{\mathrm{d}r}\right), \tag{6.1}$$

where $\bar{\rho}$ is the mean density of the phases coexisting at any point, i.e.

$$\bar{\rho} \equiv \frac{M_{\mathrm{total}}}{V_{\mathrm{total}}}. \tag{6.2}$$

The revised motion equation is of the same form as equation (3.18), with $\rho \to \bar{\rho}$, and consequently the corresponding numerical evolution equation can be obtained directly by analogy with equation (3.26),

$$u_j^{n+\frac{1}{2}} = u_j^{n-\frac{1}{2}} - g_j^n \Delta t - \frac{4\Delta t}{\left(\bar{\rho}^n_{j+\frac{1}{2}} + \bar{\rho}^n_{j-\frac{1}{2}}\right)} \frac{\left(p^n_{j+\frac{1}{2}} - p^n_{j-\frac{1}{2}}\right)}{\left(r^n_{j+1} - r^n_{j-1}\right)}. \tag{6.3}$$

The velocity equation is unchanged from the single-phase case, equation (3.17)

$$\frac{\mathrm{d}r}{\mathrm{d}t} = u, \tag{6.4}$$

and so the corresponding numerical solution is identical to the single-phase equation (3.27),

$$r_j^{n+1} = r_j^n + u_j^{n+\frac{1}{2}}\Delta t. \tag{6.5}$$

The continuity equation is modified in an analogous way to the motion equation, $\rho \to \bar{\rho}$, so that

$$\bar{\rho}^{n+1}_{j+\frac{1}{2}} = \left[\frac{(r^{n+1}_{j+1})^3 - (r^{n+1}_j)^3}{(r^n_{j+1})^3 - (r^n_j)^3}\right] \bar{\rho}^n_{j+\frac{1}{2}}. \tag{6.6}$$

This expresses the conservation of the total mass (summed over phases) between any two Lagrangian co-ordinates.

The energy equation is the only one of the fluid equations that requires more substantial modification on moving to a multi-phase treatment. The starting point is the most basic discretization of the one-dimensional energy equation, namely equation (3.24), applied to some phase i,

$$\frac{\varepsilon^{n+1}_{i,j+\frac{1}{2}} - \varepsilon^{n}_{i,j+\frac{1}{2}}}{\Delta t} = -p^{n}_{j+\frac{1}{2}} \frac{1}{\Delta t} \left(\frac{1}{\rho^{n+1}_{i,j+\frac{1}{2}}} - \frac{1}{\rho^{n}_{i,j+\frac{1}{2}}} \right) - \mathcal{L}^{n}_{i,j+\frac{1}{2}}, \tag{6.7}$$

where once again use has been made of the fact that the phases exist in pressure equilibrium.

The first difficulty is that the updated density for phase i, $\rho^{n+1}_{i,j+\frac{1}{2}}$ is unknown — only the updated *mean* density $\bar{\rho}^{n+1}_{i,j+\frac{1}{2}}$ is known. This problem obviously does not arise in the single-phase case, but in the multi-phase case the energy equation is implicit from the start.

Equation (6.7) is applicable to each of the $n_{\rm ph}$ phases separately, and thus provides $n_{\rm ph}$ equations involving $2n_{\rm ph}$ unknowns, $\varepsilon^{n+1}_{i,j+\frac{1}{2}}$ and $\rho^{n+1}_{i,j+\frac{1}{2}}$. Another $n_{\rm ph}$ equations involving these unknowns can be obtained quite simply, according to

$$p^{n+1}_{j+\frac{1}{2}} = (\gamma - 1)\varepsilon^{n+1}_{i,j+\frac{1}{2}}\rho^{n+1}_{i,j+\frac{1}{2}}. \tag{6.8}$$

Thus it is straightforward to eliminate either $\varepsilon^{n+1}_{i,j+\frac{1}{2}}$ or $\rho^{n+1}_{i,j+\frac{1}{2}}$ as desired from the equations represented by equation (6.7). This, however, comes at the expense of introducing one more unknown, the updated pressure $p^{n+1}_{j+\frac{1}{2}}$. A further equation can be obtained without the introduction of any more unknowns, namely

$$V_j = \sum_{i=1}^{n_{\rm ph}} V_{i,j} \qquad \text{or} \tag{6.9}$$

$$\frac{m_{j+\frac{1}{2}}}{\bar{\rho}^{n+1}_{j+\frac{1}{2}}} = \sum_{i=1}^{n_{\rm ph}} \frac{m_{i,j+\frac{1}{2}}}{\rho^{n+1}_{i,j+\frac{1}{2}}}, \tag{6.10}$$

which is essentially just the definition of the mean density, $\bar{\rho}^{n+1}_{j+\frac{1}{2}}$, of zone j. Note that the mass in each phase, $m_{i,j+\frac{1}{2}}$, is written as being independent of time. This is of course not necessarily true if mass loss is allowed to occur, but the extension is straightforward and does not change the mathematics of the solution, since the mass in each phase at any one time remains a known quantity.

Since the multi-phase difference equation for the energy evolution, equation (6.7), is implicit anyway, and since it is necessary to introduce the time-advanced pressure $p^{n+1}_{j+\frac{1}{2}}$ in its solution, there is no extra cost involved in correctly centring the pressure term, $p^{n}_{j+\frac{1}{2}} \to {}^{1}\!/_{2}\big(p^{n+1}_{j+\frac{1}{2}} + p^{n}_{j+\frac{1}{2}}\big)$. Making this change, and using equation (6.8) to eliminate the ε terms, equation (6.7) becomes

$$2\left(\frac{p^{n+1}_{j+\frac{1}{2}}}{\rho^{n+1}_{i,j+\frac{1}{2}}} - \frac{p^{n}_{j+\frac{1}{2}}}{\rho^{n}_{i,j+\frac{1}{2}}} \right) = -(\gamma-1)\left(p^{n+1}_{j+\frac{1}{2}} + p^{n}_{j+\frac{1}{2}}\right)\left(\frac{1}{\rho^{n+1}_{i,j+\frac{1}{2}}} - \frac{1}{\rho^{n}_{i,j+\frac{1}{2}}} \right) - 2(\gamma-1)\mathcal{L}^{n}_{i,j+\frac{1}{2}}\Delta t$$

$$\frac{1}{\rho^{n+1}_{i,j+\frac{1}{2}}}\left[(\gamma+1)p^{n+1}_{j+\frac{1}{2}} + (\gamma-1)p^{n}_{j+\frac{1}{2}}\right] = \frac{1}{\rho^{n}_{i,j+\frac{1}{2}}}\left[(\gamma-1)p^{n+1}_{j+\frac{1}{2}} + (\gamma+1)p^{n}_{j+\frac{1}{2}}\right] - 2(\gamma-1)\mathcal{L}^{n}_{i,j+\frac{1}{2}}\Delta t. \tag{6.11}$$

Multiplying by $m_{i,j+\frac{1}{2}}$ and summing over phases leads to

$$\left[(\gamma+1)p^{n+1}_{j+\frac{1}{2}}+(\gamma-1)p^{n}_{j+\frac{1}{2}}\right]\sum_{i=1}^{n_{\rm ph}}\frac{m_{i,j+\frac{1}{2}}}{\rho^{n+1}_{i,j+\frac{1}{2}}}=\left[(\gamma-1)p^{n+1}_{j+\frac{1}{2}}+(\gamma+1)p^{n}_{j+\frac{1}{2}}\right]\sum_{i=1}^{n_{\rm ph}}\frac{m_{i,j+\frac{1}{2}}}{\rho^{n}_{i,j+\frac{1}{2}}}$$
$$-2(\gamma-1)\Delta t\sum_{i=1}^{n_{\rm ph}}m_{i,j+\frac{1}{2}}\mathcal{L}^{n}_{i,j+\frac{1}{2}}.$$

Equation (6.10) can now be used to eliminate the phase densities $\rho_{i,j+\frac{1}{2}}$ in favour of the mean density $\bar{\rho}_{j+\frac{1}{2}}$,

$$p^{n+1}_{j+\frac{1}{2}}\left(\frac{\gamma+1}{\bar{\rho}^{n+1}_{j+\frac{1}{2}}}-\frac{\gamma-1}{\bar{\rho}^{n}_{j+\frac{1}{2}}}\right)=p^{n}_{j+\frac{1}{2}}\left(\frac{\gamma+1}{\bar{\rho}^{n}_{j+\frac{1}{2}}}-\frac{\gamma-1}{\bar{\rho}^{n+1}_{j+\frac{1}{2}}}\right)-\frac{2(\gamma-1)\Delta t}{m_{j+\frac{1}{2}}}\sum_{i=1}^{n_{\rm ph}}m_{i,j+\frac{1}{2}}\mathcal{L}^{n}_{i,j+\frac{1}{2}}. \tag{6.12}$$

This is the multi-phase version of the pressure evolution equation. It is identical to the single-phase version, equation (3.35), except with $\rho\to\bar{\rho}$, and the luminosity term replaced by a mass-weighted average

$$\bar{\mathcal{L}}^{n}_{j+\frac{1}{2}}\equiv\frac{\sum_{i=1}^{n_{\rm ph}}m_{i,j+\frac{1}{2}}\mathcal{L}^{n}_{i,j+\frac{1}{2}}}{\sum_{i=1}^{n_{\rm ph}}m_{i,j+\frac{1}{2}}}. \tag{6.13}$$

Once the updated pressure $p^{n+1}_{j+\frac{1}{2}}$ has been determined from equation (6.12), back substitution in equation (6.11) enables determination of the updated phase densities $\rho^{n+1}_{i,j+\frac{1}{2}}$.

6.3 Addition of Pseudoviscosity

The pseudoviscosity is dimensionally a pressure term, and is really a dynamical rather than a thermodynamical quantity, so like the true pressure, it is not a function of phase, but only of shell position and speed. In the multi-phase case the only change is $\rho\to\bar{\rho}$, so that equation (3.41) becomes

$$q^{n+1}_{j+\frac{1}{2}}=\begin{cases}\dfrac{2q_c^2}{\left(\dfrac{1}{\bar{\rho}^{n+1}_{j+\frac{1}{2}}}+\dfrac{1}{\bar{\rho}^{n}_{j+\frac{1}{2}}}\right)}\left(u^{n+1}_{j+1}-u^{n+1}_{j}\right)^2 & u^{n+1}_{j+1}-u^{n+1}_{j}<0\\ 0 & \text{otherwise.}\end{cases} \tag{6.14}$$

The version of equation (6.12) incorporating the pseudoviscosity is presented in equation (6.19).

6.4 Summary of the Multi-phase Numerical Scheme

The final multi-phase numerical scheme incorporating pseudoviscosity is presented here for reference. It can be compared with the single-phase scheme, equations (3.46) to (3.52). The differences are minor: $\rho\to\bar{\rho}$ in various places, and the luminosity term in the energy equation is replaced by a mass-weighted average.

$$u_j^{n+\frac{1}{2}} = u_j^{n-\frac{1}{2}} - g_j^n \Delta t - \frac{4\Delta t}{\left(\bar{\rho}^n_{j+\frac{1}{2}} + \bar{\rho}^n_{j-\frac{1}{2}}\right)} \frac{\left(p^n_{j+\frac{1}{2}} - p^n_{j-\frac{1}{2}} + q^{n-\frac{1}{2}}_{j+\frac{1}{2}} - q^{n-\frac{1}{2}}_{j-\frac{1}{2}}\right)}{\left(r^n_{j+1} - r^n_{j-1}\right)} \tag{6.15}$$

$$r_j^{n+1} = r_j^n + u_j^{n+\frac{1}{2}} \Delta t \tag{6.16}$$

$$\bar{\rho}^{n+1}_{j+\frac{1}{2}} = \left[\frac{(r^{n+1}_{j+1})^3 - (r^{n+1}_j)^3}{(r^n_{j+1})^3 - (r^n_j)^3}\right] \bar{\rho}^n_{j+\frac{1}{2}} \tag{6.17}$$

$$q^{n+1}_{j+\frac{1}{2}} = \begin{cases} \dfrac{2q_c^2}{\left(\dfrac{1}{\bar{\rho}^{n+1}_{j+\frac{1}{2}}} + \dfrac{1}{\bar{\rho}^n_{j+\frac{1}{2}}}\right)} \left(u^{n+1}_{j+1} - u^{n+1}_j\right)^2 & u^{n+1}_{j+1} - u^{n+1}_j < 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.18}$$

$$\begin{aligned} p^{n+1}_{j+\frac{1}{2}} \left(\frac{\gamma+1}{\bar{\rho}^{n+1}_{j+\frac{1}{2}}} - \frac{\gamma-1}{\bar{\rho}^n_{j+\frac{1}{2}}}\right) = p^n_{j+\frac{1}{2}} \left(\frac{\gamma+1}{\bar{\rho}^n_{j+\frac{1}{2}}} - \frac{\gamma-1}{\bar{\rho}^{n+1}_{j+\frac{1}{2}}}\right) - 2(\gamma-1) \times \\ \left[\frac{\Delta t}{m_{j+\frac{1}{2}}} \sum_{i=1}^{n_{\rm ph}} m_{i,j+\frac{1}{2}} \mathcal{L}^n_{i,j+\frac{1}{2}} + q^{n+1}_{j+\frac{1}{2}} \left(\frac{1}{\bar{\rho}^{n+1}_{j+\frac{1}{2}}} - \frac{1}{\bar{\rho}^n_{j+\frac{1}{2}}}\right)\right] \end{aligned} \tag{6.19}$$

$$\frac{1}{\rho^{n+1}_{i,j+\frac{1}{2}}} \left[(\gamma+1) p^{n+1}_{j+\frac{1}{2}} + (\gamma-1) p^n_{j+\frac{1}{2}}\right] = \frac{1}{\rho^n_{i,j+\frac{1}{2}}} \left[(\gamma-1) p^{n+1}_{j+\frac{1}{2}} + (\gamma+1) p^n_{j+\frac{1}{2}}\right] - 2(\gamma-1) \mathcal{L}^n_{i,j+\frac{1}{2}} \Delta t \tag{6.20}$$

$$T^{n+1}_{i,j+\frac{1}{2}} = \frac{\mu m_{\rm H}}{k_{\rm B}} \frac{p^{n+1}_{j+\frac{1}{2}}}{\rho^{n+1}_{i,j+\frac{1}{2}}} \tag{6.21}$$

$$\mathcal{L}^{n+1}_{i,j+\frac{1}{2}} = \rho^{n+1}_{i,j+\frac{1}{2}} \Lambda\left(T^{n+1}_{i,j+\frac{1}{2}}\right) \left(\frac{X_{\rm H}}{m_{\rm H}}\right)^2 r_{\rm e}\left(T^{n+1}_{i,j+\frac{1}{2}}\right) \tag{6.22}$$

6.5 Multi-phase Initial Conditions

The range and functional form of the vfdd is first specified in a dimensionless fashion, for example

$$\text{vfdd}(\rho) \propto \rho^{-4} \qquad\qquad (1 < \rho < 2),$$

which represents density fluctuations spanning a factor of two in density, and obeying a power-law. The vfdd is then discretized, by distributing the specified number of phases, $n_{\rm ph}$, evenly across the available (dimensionless) density range, $\rho_{\rm min}$–$\rho_{\rm max}$ (see Figure 6.1 on the facing page for an example). The vfdd is then calculated for each phase density, and the zone volumes are calculated from the radial shell positions. The fraction of each zone volume occupied by a given phase follows directly from the value of the vfdd for that phase density, suitably normalized so that the total volume of all the phases in any zone equals the zone volume of course.

Pressure equilibrium between phases is an essential requirement of the theoretical treatment. It is not possible for phases of different temperature to satisfy this requirement and simultaneously be in hydrostatic equilibrium. The approach I adopt is to place a "mean" phase density in hydrostatic equilibrium.

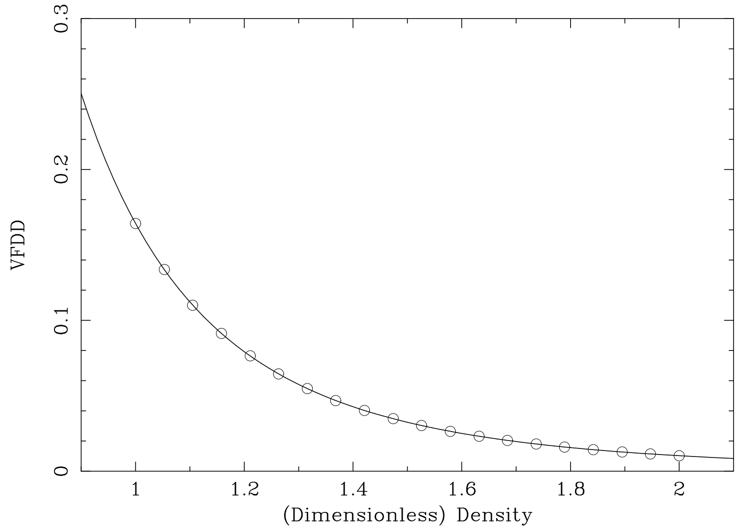


Figure 6.1: Example of the discretization of the volume fraction of the density distribution (vfdd) function. In this example, vfdd$(\rho) \propto \rho^{-4}$ and $n_{\mathrm{ph}} = 20$.

Specifically, a pressure profile is calculated as in the single-phase case, i.e. equation (3.77). The only difference is that $\rho(r_{200})$ is replaced by a mean density $\bar{\rho}(r_{200})$,

$$\bar{\rho}_{200} \equiv \frac{m_{\mathrm{H}}\mu}{k_{\mathrm{B}}} \frac{p_{200}}{\bar{T}} \qquad\qquad \bar{T} \equiv T_{\mathrm{virial}} \qquad \text{(in most cases)},$$

although the method of calculation is exactly the same. The dimensionless phase densities assigned from the vfdd are then scaled to have this mean at the outer boundary,

$$\rho_{200,i} = \rho'_{200,i} \frac{\bar{\rho}_{200}}{\sum\limits_i \frac{\rho'_{200,i}}{n_{\mathrm{ph}}}},$$

where primed quantities are dimensionless.

A density profile for each phase is then assigned according to equation (3.79). This ensures that the phases are in pressure equilibrium. The total mass of gas is calculated by summing the mass present in each phase in every zone. Normalization to a specific gas mass fraction is again achieved by adjusting the outer boundary pressure and repeating the density set-up process. The method adopted for specifying the initial density profiles has the convenient property that the total gas mass is linearly proportional to the outer boundary pressure ($\rho_{200,i} \propto \bar{\rho}_{200} \propto p_{200}$), making normalization easy.

Once the pressure and density profiles are established, other quantities are assigned in the usual way. For example, the phase temperatures are obtained from the equation of state,

$$T_{200,i} \equiv \frac{\mu m_{\mathrm{H}}}{k_{\mathrm{B}}} \frac{p_{200}}{\rho_{200,i}}.$$

These temperatures are independent of p_{200}, depending only on $\bar{T}$ and the number of phases:

$$\sum_i \frac{1}{T_i} = \frac{k_{\mathrm{B}}}{\mu m_{\mathrm{H}}} \frac{\sum_i \rho_{200,i}}{p_{200}} = \frac{k_{\mathrm{B}}}{\mu m_{\mathrm{H}}} \frac{n_{\mathrm{ph}} \bar{\rho}_{200}}{p_{200}} = \frac{n_{\mathrm{ph}}}{\bar{T}}. \tag{6.23}$$

The initial vfdd is assumed to be independent of radius. It might be interesting to relax this assumption, but the total gas mass would then be no longer proportional to the outer boundary pressure, so that normalization of the gas mass would not be straightforward.

6.6 Phase Loss

Moving to a multi-phase system obviously allows for the possibility of phase loss, in addition to the zone loss already discussed in Section 3.6. Phase loss is handled in a similar manner, as illustrated in Figure 6.2 on the facing page. When a phase is lost from zone x, the remaining phases in that zone expand into part of the freed space. Other zones expand into the remaining space. If V_{lost} is the volume of the lost phase, then this is divided into volumes V_{a} and V_{b} as in the zone loss case, so that from equation (3.88)

$$V_{\mathrm{a}} = V_{\mathrm{lost}} \left(1 + \frac{V_{\mathrm{o}}}{V_{\mathrm{i}}}\right)^{-1} \qquad V_{\mathrm{b}} = V_{\mathrm{lost}} \left(1 + \frac{V_{\mathrm{i}}}{V_{\mathrm{o}}}\right)^{-1}, \tag{6.24}$$

with V_{o} and V_{i} defined as per equation (3.87).

The zone where phase loss has taken place transforms as

$$V_{\mathrm{x}} \to V_{\mathrm{x}} - (1-a)V_{\mathrm{a}} - (1-b)V_{\mathrm{b}}, \tag{6.25}$$

which represents a volume *increase* for the surviving phases. The other zones transform as follows.

Interior zones:

$$\begin{aligned} V_{\mathrm{x}-1} &\to V_{\mathrm{x}-1} + ar\, V_{\mathrm{a}} \\ V_{\mathrm{x}-2} &\to V_{\mathrm{x}-2} + ar^2\, V_{\mathrm{a}} \\ &\vdots \\ V_{\mathrm{j}} &\to V_{\mathrm{j}} + ar^{x-j} V_{\mathrm{a}} \\ &\vdots \\ V_0 &\to V_0 + ar^x\, V_{\mathrm{a}} \end{aligned} \tag{6.26}$$

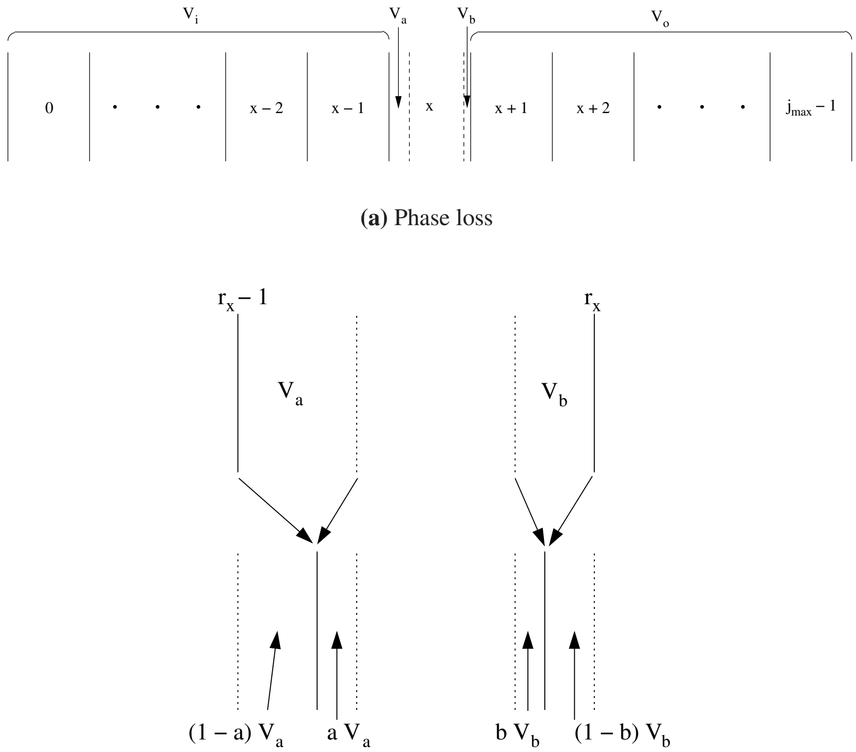
Exterior zones:

$$\begin{aligned} V_{\mathrm{x}+1} &\to V_{\mathrm{x}+1} + br\, V_{\mathrm{b}} \\ V_{\mathrm{x}+2} &\to V_{\mathrm{x}+2} + br^2\, V_{\mathrm{b}} \\ &\vdots \\ V_{\mathrm{j}} &\to V_{\mathrm{j}} + br^j\, V_{\mathrm{b}} \\ &\vdots \\ V_{\mathrm{jmax}-1} &\to V_{\mathrm{jmax}-1} + br^{j_{\mathrm{max}}-1-x} V_{\mathrm{b}} \end{aligned} \tag{6.27}$$

Compared with the zone-loss case, equations (3.91) and (3.92), there is one more term in each series, hence the normalization conditions are

$$a = \frac{1-r}{1-r^{x+1}}, \qquad b = \frac{1-r}{1-r^{j_{\mathrm{max}}-x}}. \tag{6.28}$$

This only specifies the updated *zone* volumes. The updated *phase* volumes in each of those zones can be calculated assuming the expansions are adiabatic, and using the condition of pressure equilibrium



(a) Phase loss

(b) Close-up of phase-loss region

Figure 6.2: Illustration of how phase loss is treated in the numerical implementation. A phase is lost from zone x, releasing a volume $V_{\rm a} + V_{\rm b}$, where $V_{\rm a} : V_{\rm b} = V_{\rm i} : V_{\rm o}$. The remaining phases in the zone where loss occurred expand into an extra volume $aV_{\rm a} + bV_{\rm b}$. Zones lying inward of the loss-site expand into an extra volume $(1-a)V_{\rm a}$, and those lying outward into an extra volume $(1-b)V_{\rm b}$.

between the phases in any one zone. If primed quantities denote values after expansion, then

$$p'_j V'^{\gamma}_{i,j} = p_j V^{\gamma}_{i,j} \tag{6.29}$$

$$\sum_i V'_{i,j} = V_{\rm new} = \sum_i V_{i,j} - V_{\rm lost} = V_{\rm old} - V_{\rm lost}. \tag{6.30}$$

These represent $n'_{{\rm ph},\,j} + 1$ equations for the corresponding number of unknowns in zone j, with $n'_{{\rm ph},\,j}$ the updated number of phases remaining in the zone after loss. Solving for the zone pressure and phase volumes,

$$p'_j = p_j \left(\frac{V_{\rm old} - V_{\rm lost}}{V_{\rm new}} \right)^{\gamma} \tag{6.31}$$

$$V'_{i,j} = V_{i,j} \left(\frac{V_{\rm new}}{V_{\rm old} - V_{\rm lost}} \right). \tag{6.32}$$

These equations apply to all zones. In the zone where phase loss has occurred, $V_{\rm lost}$ is the volume of the lost phase, whereas in the other zones it is zero. As with zone loss, if there are multiple instances of phase loss in any iteration, then the phase-loss algorithm is simply applied more than once.

6.7 Comments

In an ICM with efficient thermal conduction, it seems unlikely that multi-phase gas could exist. Indeed, given the unstable nature of radiative cooling that would tend to make a single-phase gas fragment into a multi-phase medium, perhaps the very lack of evidence for multi-phase gas in recent data is itself suggestive of efficient thermal coupling between the notional "phases".

The traditional multi-phase model does not seem to have much relevance in light of recent observations. For example, Johnstone *et al.* (2002) put strong limits (their figure 9) on the filling fraction of a second temperature component in the inner regions of Abell 2199. Sanders and Fabian (2002), however, do find some evidence of a second temperature component within the central few kpc of the Centaurus cluster. In a similar fashion, Buote *et al.* (2002, 2003) find multi-temperature fits are favoured within the central 30 kpc of the bright, nearby group NGC 5044.

In light of this, it would be wrong (at very least, premature) to cast aside entirely the multi-phase model; even more so given its detailed theoretical basis and long history. In its traditional form, in which the gas is strongly multi-phase over a wide range of radii, though, it clearly does not apply to current datasets. It must, therefore, be modified in some fashion to take account of changed conditions. I therefore have presented in this Chapter my implementation of the numerical machinery required to deal with the full multi-phase case. Whilst not directly applicable in its present form, it may form the basis of an updated treatment more applicable to current results. One attractive possibility is to combine thermal conduction with a (potentially) multi-phase ICM, in the hope that efficient heat transfer can suppress the divergence of phases until close to the centre of the cluster. It is not yet clear to me, however, how this might be achieved in principle, or even handled at a theoretical level. Given the complexity of the traditional multi-phase theory (Nulsen, 1986), I suspect this may well be a difficult problem that will require some time to deal with. The work presented in this Chapter is therefore necessarily incomplete, in that sense.

Someone may observe that no doubt the conclusion preceded the 'proofs'. For who gives himself up to looking for proofs of something he does not believe in or the prediction of which he does not care about?

Three Versions of Judas
JORGE LUIS BORGES

7 Conclusions and Future Work

7.1 A Critique of the Numerical Method

The numerical scheme adopted for this work is a simple one, which is obviously advantageous in terms of understanding the principles at work and interpreting the output, but has its limits in terms of speed and flexibility. I have not really been concerned with making detailed numerical predictions of the evolution of specific systems, but rather with investigating general trends and patterns of behaviour, and the possible effects of given physical assumptions. It has been a benefit to have a numerical scheme which is not so complex in its implementation that it is far divorced from reality.

Addressing the issue of flexibility, in the same way as an adaptive time step is used here, it is also possible to implement so-called Adaptive Mesh Refinement (AMR) codes (e.g. Berger and Colella, 1989; Yahagi and Yoshii, 2001; Teyssier, 2002), in which the system is regridded in the spatial co-ordinate as and when necessary. This technique is useful if part of the system requires a higher resolution than the rest. The use of AMR allows high spatial resolution to be obtained where necessary, without the computational (memory and time) costs of modelling the entire system in high spatial resolution. This is obviously at the expense of greater algorithmic complexity, though. Such an approach would be quite useful in this case, where high resolution is desired in the central regions, but not necessary beyond the cooling radius where the changes are slight.

With regards to speed, this is not so much of an issue. Nevertheless, particularly with the inclusion of thermal conduction, and with the desire to investigate an ever-growing parameter space, I do feel it might be useful to change to a faster, more sophisticated numerical method. An implicit method (Section 3.2) of some kind would probably be useful (indeed the system almost fits the text-book case for the application of such methods). There is a wide variety of ever-more exotic numerical schemes that have been applied to fluid flow (for an in-depth review, see Woodward and Colella, 1984).

There is of course also always the desire to extend the modelling from the current one-dimensional, spherically symmetric case, to two and three dimensions. A Lagrangian scheme such as the one used here, though, does not readily extend itself to a higher-dimensional formalism. With a grid of two or three dimensions it can be difficult to deal with the changing cell size and avoid holes appearing in the mesh. Lagrangian schemes, however, do have a natural advantage when dealing with multi-phase flows, since the extra information on the origin of each fluid particle makes separate species easier to handle. If one is willing to restrict attention to a single-phase treatment, one has more freedom of choice. Of all the limitations of the current method, I actually feel this one is the least important. Whilst it is clear that the detailed structures of cluster cores are far from spherically symmetric, important progress can still be made under this assumption. The possible range of interesting behaviour that can be explored is still far from exhausted. See Section 7.2, though, for some reasons why a two-dimensional approach would be desirable.

Making any of these changes would require a substantial investment of time, though. Almost certainly, it would be more cost-effective to try and adapt some existing hydrodynamic simulation code (of which there are many), rather than starting again from scratch.

7.1.1 Possible Extensions

Leaving these concerns aside, there are several smaller changes and improvements that one could consider making.

The inclusion of thermal conduction (Chapter 5) into the equations necessitates a consideration of the possible effects of magnetic fields. These will also have dynamic effects through the introduction of a non-thermal pressure component. Such effects may be significant in the cores of cooling flow clusters (e.g. Carilli and Taylor, 2002), and so one might consider introducing magnetic terms into the equations of motion in order to investigate this area.

The cooling gas is only followed down to some minimum temperature, T_{min}, at which point it is removed from the simulation (Section 3.6). One might consider associating some energy feedback with this process, e.g. to represent in a crude way the energy released by any star formation in the cold, collapsed gas. Only a very simplistic treatment would be possible within the current model, but it might produce significant effects.

The initial conditions in every case were taken to be isothermal. This is certainly a reasonable starting point for the processes which have been investigated here; but a more general treatment would consider a range of possible starting states. Then, however, one runs the risk of moving into cluster formation, which is a separate topic in its own right.

One area that would be easy to investigate is the effect of different gravitational potentials on the gas evolution. For example, one might make use of the mass profile of Moore *et al.* (1998), to see if the difference in slope at small scales between this and the NFW profile (Navarro *et al.*, 1997) has any important effect on the evolution of the cluster gas. Since the mass profile, potential, and acceleration are specified as external functions in the computational scheme, such an investigation would be straightforward.

A multi-phase medium with thermal conduction operating at non-negligible levels would be an interesting case to examine. For example, how would conduction modify the evolution of the density distribution function, what would be the effect on the mass deposition profile, and so on. I am not sure how to treat such a system theoretically though, or even if a self-consistent picture could be obtained.

Would one allow conduction to operate between the phases in any given spatial zone, or between one phase in a given zone and its counterparts in neighbouring zones, or both? The former would not really be possible numerically, since there is no spatial information on the location of phases in a given zone from which to calculate any kind of temperature gradient and thence a heat flux. Allowing conduction between the same phase in adjacent zones would be easier. This might be justifiable on theoretical grounds with separate phases representing separate magnetic structures with isolated transport properties. Perhaps such a picture might be best addressed in a two-dimensional model, though.

To some extent, we are hampered by the theoretical framework of Nulsen (1986). With its restriction to comoving phases, it does not allow for a treatment of sedimentation of heavy elements (e.g. Fabian and Pringle, 1977; Fabian *et al.*, 2002b), or a variety of other process one might wish to investigate, such as the possibility that cold gas might decouple from the mean flow and fall inwards. This was investigated by Fabian *et al.* (2002a) as another possible explanation for the missing soft X-ray luminosity in cooling flows. There is not, however, really any alternative comprehensive theoretical framework available. Indeed, the general problem is a very difficult one, which is why Nulsen (1986) makes a variety of simplifying assumptions. At the time, these were all readily defensible: in light of more recent evidence, some of them (particularly the suppression of conduction) seem less so. If these assumptions are abandoned, then specialized numerical schemes of the form outlined in Chapter 6, which rely on this theoretical framework, will become obsolete. Computational approaches would probably have to move to more general multi-purpose numerical grid codes.

Even the simplified treatment of Nulsen (1986), though, has not been used a great deal in its full form. Most observational analyses have been based on a comparison with `mkcflow`-type models. This is appropriate for either an isobaric, homogeneous flow, or locally for the high-density limit of an inhomogeneous flow (Nulsen, 1988).

The role of magnetic fields may well turn out to be key, either through their influence on transport properties, or the pinning (or otherwise) of cold dense gas to the mean flow, or both. They are the implicit background to much of the theoretical framework of Nulsen (1986). As discussed in Section 5.5.3, though, the influence of magnetic fields remains unclear. The correlation between Faraday rotation measure and X-ray mass deposition rate presented by Taylor *et al.* (2002) may be another important piece of the puzzle.

With regards to the work of Chapters 2 and 4 on the effects of inhomogeneities in the metal distribution, I have considered essentially just the simplest limiting case, namely that of a two-component metallicity. This undoubtedly captures the important aspects of behaviour, but a more refined treatment would allow a spread of metallicities in an arbitrary manner, controlled by some function analogous to the volume fraction of the density distribution function (Section 6.1) in the standard multi-phase model. The multi-phase picture almost invites us to consider a spread of chemical composition, relying as it does on suppressed transport properties between phases. Combining the two ideas (density phases and metallicity phases) into the same model would give rise to a very large parameter space for investigation. The simple case of two phases of differing abundance is very similar to the models investigated in Chapter 4. In a multi-phase picture, there is the possibility for metal transport (Reisenegger *et al.*, 1996), something that cannot really take place in the single-phase case. It would be very interesting to examine this in connection with the abundance gradients (Section 4.3.3) seen in observational data. As well as using a more sophisticated metallicity distribution, it might be better to allow metals to be replenished in some way during the course of a simulation.

7.2 The Lack of Low-temperature Line Emission in Cooling Flows

In some ways, I find it surprising that there is still a significant gap in our understanding of cooling flows. The physics ought to be comparatively simple, involving as it does merely the radiative cooling of hot gas in a gravitational potential. Yet it seems that nature retains the ability to surprise us. The problem is made all the more intriguing by the fact that it is so simple, and yet so fundamental as well. Similar processes (albeit on a smaller scale) must take place during galaxy formation. Indeed, as was remarked upon in Chapter 1, Fabian *et al.* (2002b) have used the recent work on thermal conduction in the intracluster medium to suggest a possible cause for the lack of super-massive galaxies.

The recent work of Peterson *et al.* (2003) (see also Kahn *et al.*, 2002) is extremely important, because it at last makes some steps towards putting the cooling flow problem on a quantitative footing. For a long time, it seemed as if the issue would often be stated as “there is no cool gas in the centres of clusters”. Later this changed to the slightly more descriptive “there is less cool gas than expected”. By themselves, these statements have only limited informational content. Peterson *et al.* (2003) have quantified the problem by plotting the ratio of the observed emission measure originating from gas in a given (fairly broad) temperature range to that predicted from the isobaric cooling flow model against temperature. There is emission originating from gas across the entire X-ray temperature range, but the deficit compared to the standard cooling flow model increases as the temperature decreases. In some ways, this is a more bizarre result than a complete lack of cold gas. There are various ways one could envisage stopping the ICM cooling altogether, but it is a lot more difficult to allow some cooling to continue, but at a greatly reduced level. The most intriguing result is that when the temperature abscissa in the emission measure plots is scaled according to the cluster virial temperature, the lines corresponding to various clusters lie fairly close to one another (i.e. there is a degree of similarity between clusters). It is not yet clear what this is telling us. If this result (based on 14 clusters) holds generally, then I feel that reproducing this trend, over the requisite wide range of absolute temperatures, will be a key test of any theoretical solution. I note that isothermal cooling flow models (Nulsen, 1998) are no better at reproducing the observed emission measure distribution.

Temperature drops of the order of a factor of three are seen between the centres and exteriors of clusters. Cooling is therefore not entirely prevented, but rather suppressed in some fashion. Given the vanishing probability of many clusters being observed in this state if it is not long-lived, we require some form of feedback to operate between the radiative cooling process and whatever is responsible for offsetting it. It is difficult to see how thermal conduction acting in isolation can fulfil this requirement. Given the opposite temperature dependencies of conductive heating and radiative cooling (the former becoming weaker as the latter grows in strength), one would expect either runaway cooling or establishment of an essentially isothermal state to rapidly take place.

I feel it important to point out that the small-scale metallicity variation hypothesis has not yet received a great deal of observational testing. Böhringer *et al.* (2002) have given some consideration to this issue in connection with M87, but only one particular case with a metal-rich and a metal-poor phase of “roughly equal” normalizations. I would not expect such an arrangement to be particularly effective at suppressing low-temperature line emission. Consequently, I do not feel this to be a serious test of the hypothesis.

There is a wide parameter space to explore, and, as discussed in Section 4.3.2, detection of such variations will be difficult, but this is no reason not to try. Indeed, Schmidt *et al.* (2002) have reported what may be the first signs of small-scale metallicity variations in the Perseus ICM, although the scale of the effect is at the limit of significance. I am certainly not aware of any strong observational constraint that rules out this hypothesis. Within the context of the traditional multiphase model, I would almost expect such a situation to come about. The motivation for such a model is admittedly weakened by the evidence in favour of thermal conduction. In some respects, I have discussed two contradictory hypotheses in this thesis. One requires suppression of transport properties, the other that they operate efficiently. As I point out in Section 4.7, however, the two ideas (metallicity inhomogeneities and conduction) are not necessarily mutually exclusive. Efficient transfer of heat does not require individual particles (electrons) to move too far, whereas the spread of metal ions through diffusion obviously does. Bearing in mind that ions are much more massive and consequently slower, it is not outside the bounds of possibility that reasonably efficient thermal conduction could coexist with an uneven metal distribution. I will admit to finding this idea somewhat aesthetically displeasing though.

There are of course other possible solutions that have been put forward that have not been discussed in any great detail in this thesis, but were mentioned in Section 1.6.1. For an X-ray astronomer, it is tempting to point the finger of blame at the atomic physics, but I do not regard this as a realistic possibility. One might be able to make some investigation of differential absorption within the context of the simple numerical scheme used in this work. For example, at any instant, the amount of absorption applied to the spectrum originating from gas at a given point could be modulated by the amount of cold gas that had previously dropped out of the cooling flow at that point. Another avenue that could be explored relatively easily with the method of this thesis is the addition of some form of heat source, either local or distributed.

I find the mixing hypothesis, in which cool ($\sim 10^7$ K) X-ray gas is cooled rapidly and non-radiatively to low temperatures by mixing with cold ($\sim 10^4$ K say) gas appealing, particularly in light of the results of Oegerle *et al.* (2001), as discussed in Section 1.6.1. I feel that more investigation of this idea in the context of the cooling flow problem is merited. A time-dependent approach will be necessary, because whilst one can envisage a steady state in which gas cools rapidly and non-radiatively, such a state requires a substantial reservoir of cold gas to exist already. This gas can of course itself only originate from initially hot intracluster gas. So an important question is, if a steady state is possible in this scenario, how is it arrived at? How long does the establishment of such a state take, and how long can it persist? This does highlight a general reason why a two-dimensional code would be desirable — processes such as mixing, sedimentation and convection cannot really be dealt with inside the confines of a one-dimensional model. The scheme presented in this thesis is therefore not applicable to such cases.

Judging simply by the number of papers published, thermal conduction is the mechanism which has the most support of the community at present. This would probably be my own hypothesis of choice at the moment, but with some caveats. I am largely pessimistic about the possibility of a solution to the cooling flow problem that involves *just* thermal conduction. Apart from the possibility of some form of self-stabilizing feedback developing between the cooling flow and the conduction process, I do not see how such a stable balance between conduction and radiation can be obtained over the required range of cluster masses and temperatures. It is possible that central heating by an AGN, combined with conduction, can produce a stable situation, as claimed by Ruszkowski and Begelman (2002), but I feel this is an area that needs more work.

The one thing that is clear is that there is no shortage of theories that seek to explain the recent X-ray observations of cooling flow clusters. We can spin as many theoretical castles in the air we wish, however, without advancing much closer to a solution. More observational testing of the various hypotheses is required (though the cry of "more data" is a common one, it is not always entirely without justification). It is not obvious to me, though, how one can prove that conduction (say) is taking place at a significant level. I find energy-balance arguments of the type put forward by Voigt *et al.* (2002) suggestive, but by no means compelling. As Fabian *et al.* (2002b) point out, and as I have tried to show in Chapter 5, such results may simply be a "coincidence". Just because the properties of clusters are such that thermal conduction *could* have a significant effect if it operated efficiently, is not evidence that it actually *is* happening.

ASTRO-E2, the replacement for the original *ASTRO-E* mission that was lost at launch on 10th February, 2000, is currently scheduled for launch in 2005. It will carry a new type of instrument, the X-ray Spectrometer (XRS), which will be the first X-ray microcalorimeter flown on an orbiting observatory. XRS will have an unprecedented energy resolution of $\sim 10\,\mathrm{eV}$ (FWHM) over the 0.4–10 keV energy band. Microcalorimeters are non-dispersive devices (unlike gratings, for example), and consequently have a very high efficiency. That of XRS will be nearly unity over the full bandpass, about an order of magnitude better than can be achieved with gratings. In light of the advances in knowledge that have been brought about by *XMM-Newton* and *Chandra*, we can look forward to the future of X-ray astronomy with optimism.

The process of research for this thesis has spanned a transitional period in our understanding of the intracluster medium. The *Chandra* satellite was launched a few months before the start of this thesis work, and *XMM-Newton* a few months after. Thanks to these superb instruments, our knowledge of X-ray astronomy has improved in many ways. Whenever understanding increases, so too does our appreciation of our ignorance. At least in connection with the intracluster medium, we are well on the way to, if not eliminating, at least quantifying our ignorance. There is much we do not understand, but in time we can hope to reclaim a little more land from the infinite unknown.

Appendices

The mathematics is not there till we put it there.

The Philosophy of Physical Science
ARTHUR STANLEY EDDINGTON

A

Useful Mathematical Results

This Appendix collates some useful mathematical formulae.

A.1 Power Series

$$\ln(1+x) = \int \frac{\mathrm{d}x}{1+x} = \int 1 - x + x^2 - \cdots \,\mathrm{d}x = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots \qquad (x \ll 1) \tag{A.1}$$

A.2 Geometric Series

$$\left.\begin{aligned} S_\mathrm{N} &\equiv a + ar + ar^2 + ar^3 + \cdots + ar^{N-1} \\ rS_\mathrm{N} &= \qquad ar + ar^2 + ar^3 + \cdots + ar^{N-1} + ar^N \end{aligned}\right\} \quad \Rightarrow \quad (1-r)S_\mathrm{N} = a(1-r^N)$$

$$S_\mathrm{N} = \frac{a(1-r^N)}{1-r} \qquad\qquad S_\infty = \frac{a}{1-r} \qquad (r < 1) \tag{A.2}$$

A.3 Vector Calculus

In spherical polar co-ordinates (r, θ, ϕ), with θ the polar angle and ϕ the azimuthal angle:

$$\nabla = \left(\frac{\partial}{\partial r}, \frac{1}{r}\frac{\partial}{\partial \theta}, \frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}\right) \qquad \text{gradient} \tag{A.3}$$

$$\nabla \cdot \boldsymbol{F} = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 F_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}(\sin\theta\, F_\theta) + \frac{1}{r\sin\theta}\frac{\partial F_\phi}{\partial \phi} \qquad \text{divergence} \tag{A.4}$$

$$\nabla^2 \equiv \nabla \cdot \nabla = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial \phi^2} \qquad \text{Laplacian} \tag{A.5}$$

The divergence theorem:

$$\iint \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} = \int F \, \mathrm{d}V, \tag{A.6}$$

where $\boldsymbol{S}$ is the surface bounding the volume V.

A.4 Special Functions

A.4.1 The Euler Gamma Function

$$\Gamma(z) \equiv \int_0^\infty \mathrm{e}^{-t} t^{z-1} \, \mathrm{d}t \qquad \mathrm{Re}(z) > 0 \quad \text{(for convergence at the lower limit)} \tag{A.7}$$

Recursion relation, and relation to the factorial function:

$$\Gamma(z+1) = -[\mathrm{e}^{-t} t^z]_0^\infty + z \int_0^\infty \mathrm{e}^{-t} t^{z-1} = z\Gamma(z) = z(z-1)\Gamma(z-1) = \cdots = z! \tag{A.8}$$

This equivalence for integer z is used to extend the definition of the factorial function to non-integers.

Alternative form:

$$\Gamma(z) = 2 \int_0^\infty \mathrm{e}^{-u^2} u^{2z-1} \, \mathrm{d}u \qquad t = u^z \tag{A.9}$$

Hence:

$$\Gamma(^1\!/_2) = 2 \int_0^\infty \mathrm{e}^{-x^2} \, \mathrm{d}x = \sqrt{\pi}. \tag{A.10}$$

A.4.2 The Beta Function

$$\beta(p,q) \equiv \int_0^1 t^{p-1} (1-t)^{q-1} \, \mathrm{d}t \tag{A.11}$$

$$= -\int_1^0 (1-u)^{p-1} u^{q-1} \, \mathrm{d}u = \int_0^1 u^{q-1} (1-u)^{p-1} \, \mathrm{d}u$$

$$\equiv \beta(q,p) \tag{A.12}$$

Making the substitution

$$t = \frac{u}{1+u} \qquad \Rightarrow \quad u = \frac{t}{1-t}, \quad \mathrm{d}t = \frac{\mathrm{d}u}{(1+u)^2}$$

leads to the alternative form

$$\beta(p,q) = \int_0^\infty \left(\frac{u}{1+u}\right)^{p-1} \left(\frac{1}{1+u}\right)^{q-1} \frac{\mathrm{d}u}{(1+u)^2} = \int_0^\infty \frac{u^{p-1} \, \mathrm{d}u}{(1+u)^{p+q}}. \tag{A.13}$$

The beta function is related to the product of gamma functions as follows.

$$\Gamma(p)\Gamma(q) = 4 \int_0^\infty \mathrm{e}^{-x^2} x^{2p-1} \, \mathrm{d}x \int_0^\infty \mathrm{e}^{-y^2} y^{2q-1} \, \mathrm{d}y = 4 \int_0^\infty \int_0^\infty \mathrm{e}^{-x^2-y^2} x^{2p-1} y^{2q-1} \, \mathrm{d}x \, \mathrm{d}y$$

$$
\begin{aligned}
&= 4\int_0^\infty \int_0^{\pi/2} (r\cos\theta)^{2p-1}\,(r\sin\theta)^{2q-1}\,\mathrm{e}^{-r^2}\,r\,\mathrm{d}r\,\mathrm{d}\theta \quad \text{(in plane polar co-ordinates)}\\
&= 4\int_0^\infty r^{2p+2q-1}\mathrm{e}^{-r^2}\,\mathrm{d}r \int_0^{\pi/2} \cos^{2(p-1)}\theta\,\sin^{2(q-1)}\theta\,\sin\theta\cos\theta\,\mathrm{d}\theta\\
&= \Gamma(p+q)\int_0^{\pi/2}(1-\sin^2\theta)^{p-1}(\sin^2\theta)^{q-1}2\sin\theta\cos\theta\,\mathrm{d}\theta\\
&= \Gamma(p+q)\int_0^1 (1-t)^{p-1}t^{q-1}\,\mathrm{d}t
\end{aligned}
$$

$$
\therefore \quad \Gamma(p)\Gamma(q) = \Gamma(p+q)\beta(p,q) \tag{A.14}
$$

A.5 The Equations of Fluid Dynamics

See for example Landau and Lifshitz (1987).

The first equation of fluid dynamics is the continuity equation, which expresses mass conservation. Consider a volume V containing a fluid of density ρ. Let $\boldsymbol{j} \equiv \rho\boldsymbol{u}$ be the mass flux (mass area^{-1} time^{-1}) of fluid flowing out of the surface S bounding the volume. Conservation of mass requires that

$$
\begin{aligned}
&\frac{\partial}{\partial t}\int \rho\,\mathrm{d}V + \iint \boldsymbol{j}\cdot\mathrm{d}\boldsymbol{S} = 0\\
&\int \frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\boldsymbol{u})\,\mathrm{d}V = 0 && \text{using the divergence theorem}\\
&\therefore \quad \frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\boldsymbol{u}) = 0\,, && \text{continuity equation} \qquad \text{(A.15)}
\end{aligned}
$$

or

$$
\left(\frac{\partial}{\partial t} + \boldsymbol{u}\cdot\nabla\right)\rho + \rho\nabla\cdot\boldsymbol{u} = 0\,. \tag{A.16}
$$

Mass sources and sinks may be added to the right-hand side as needed. For an incompressible fluid, where ρ is constant, the continuity equation reduces to $\nabla\cdot\boldsymbol{u} = 0$.

The second equation of fluid dynamics is the equation of motion. From Figure A.1 on the following page, the net force acting on a fluid element in the x direction due to any pressure gradient is

$$
\mathrm{d}F_x = -\,\mathrm{d}p\,\mathrm{d}A = -\frac{\partial p}{\partial x}\mathrm{d}x\;\mathrm{d}A = -\frac{\partial p}{\partial x}\mathrm{d}V\,.
$$

Hence the force per unit volume is $\boldsymbol{F} = -\nabla p$. The velocity field is a function of both position and time, $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x},t)$, and hence

$$
\frac{\mathrm{d}u_i}{\mathrm{d}t} = \frac{\partial u_i}{\partial t} + \frac{\partial u_i}{\partial x_j}\frac{\mathrm{d}x_j}{\mathrm{d}t} = \frac{\partial u_i}{\partial t} + (\boldsymbol{u}\cdot\nabla)u_i\,.
$$

Consequently, Newton's second law of motion per unit volume of fluid may be expressed as

$$
\rho\left(\frac{\partial}{\partial t} + \boldsymbol{u}\cdot\nabla\right)\boldsymbol{u} = -\nabla p\,. \qquad \text{motion equation} \tag{A.17}
$$

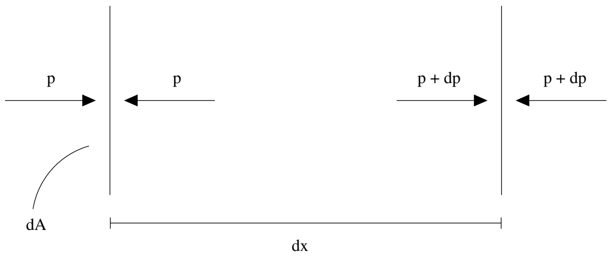


Figure A.1: Pressure differential acting across a fluid element.

Body forces (e.g. gravity) may be added to the right-hand side as needed.

The quantity

$$\frac{\mathrm{D}}{\mathrm{D}t} \equiv \frac{\partial}{\partial t} + \boldsymbol{u} \cdot \nabla \tag{A.18}$$

appearing in equations (A.16) and (A.17) is referred to as the *total derivative*, and expresses the rate of change of fluid properties due to both time and motion in space.

The third equation of fluid dynamics expresses conservation of energy.

$$\frac{\mathrm{D}\varepsilon}{\mathrm{D}t} + p\frac{\mathrm{D}V}{\mathrm{D}t} = 0, \qquad \text{energy equation} \tag{A.19}$$

where ε is the specific internal energy of the fluid, and V is the specific volume. Sources and sinks of heat may be added to the right-hand side as needed.

A.6 Properties of an Ideal Gas

The ideal gas equation of state relating the thermodynamic variables p, ρ, and T is

$$p = \mathrm{R}\rho T = k_{\mathrm{B}} n T, \tag{A.20}$$

with R the gas constant per unit mass, k_{B} Boltzmann's constant (gas constant per particle), and n the number density.

According to Boltzmann's law, a particle in thermodynamic equilibrium possesses $^1\!/_2 k_{\mathrm{B}} T$ of energy per degree of freedom f. Thus, the specific internal energy of an ideal gas particle is

$$\varepsilon = \frac{f}{2}\frac{k_{\mathrm{B}}T}{\mu m_{\mathrm{H}}} = \frac{f}{2}\frac{p}{\rho}. \tag{A.21}$$

The heat capacities at constant volume, pressure, and their ratio are defined as

$$C_V \equiv \left.\frac{\partial Q}{\partial T}\right|_V \qquad C_p \equiv \left.\frac{\partial Q}{\partial T}\right|_p \qquad \gamma \equiv \frac{C_p}{C_V}. \tag{A.22}$$

At constant volume,

$$\mathrm{d}Q = \mathrm{d}\varepsilon + p\,\mathrm{d}V = \mathrm{d}\varepsilon = \frac{f}{2}\frac{k_{\mathrm{B}}}{\mu m_{\mathrm{H}}}\mathrm{d}T \qquad \Rightarrow \qquad C_V = \frac{f}{2}\frac{k_{\mathrm{B}}}{\mu m_{\mathrm{H}}} \tag{A.23}$$

At constant pressure,

$$\mathrm{d}Q = \mathrm{d}\varepsilon + p\,\mathrm{d}V = \mathrm{d}\left(\varepsilon + \frac{p}{\rho}\right) = \left(\frac{f}{2}+1\right)\frac{k_\mathrm{B}}{\mu m_\mathrm{H}}\,\mathrm{d}T \qquad \Rightarrow \quad C_p = \left(\frac{f}{2}+1\right)\frac{k_\mathrm{B}}{\mu m_\mathrm{H}} \tag{A.24}$$

$$\therefore \quad \gamma = 1 + \frac{2}{f} \tag{A.25}$$

Hence from equation (A.21),

$$p = \frac{f}{2}\varepsilon\rho = (\gamma - 1)\varepsilon\rho. \tag{A.26}$$

For an adiabatic change,

$$\mathrm{d}Q = 0 = \mathrm{d}\varepsilon + p\,\mathrm{d}\left(\frac{1}{\rho}\right) = \frac{1}{\gamma-1}\,\mathrm{d}\left(\frac{p}{\rho}\right) + p\,\mathrm{d}\left(\frac{1}{\rho}\right)$$

$$\Rightarrow \frac{\mathrm{d}p}{p} = \gamma\frac{\mathrm{d}\rho}{\rho} \qquad \therefore \quad p \propto \rho^{\gamma}, \tag{A.27}$$

so that the ratio of the specific heats γ is also the adiabatic index.

A.7 Mass Fractions, Mean Molecular Weights

In a mixture of gases, the mass fraction X_i of some species i is defined by

$$\rho_i = X_i\rho, \tag{A.28}$$

where ρ is the total mass density, and ρ_i is that due to species i alone. In terms of the number densities,

$$X_i = \frac{\rho_i}{\sum_j \rho_j} = \frac{m_i n_i}{\sum_j m_j n_j}. \tag{A.29}$$

The mean molecular weight μ facilitates treating a mixture of ideal gases like a single ideal gas. For a single species, the molecular weight expresses the mass m of a gas particle in units of the hydrogen mass m_H,

$$m = \mu m_\mathrm{H}. \tag{A.30}$$

Thus the mass density ρ_i of some species i is related to the number density n_i as

$$\rho_i = m_i n_i = \mu_i m_\mathrm{H} n_i. \tag{A.31}$$

For a gas composed of a mixture of different constituent gases, the same relationship is used to define the mean molecular weight μ:

$$\rho \equiv \mu m_\mathrm{H} n, \tag{A.32}$$

where ρ and n are the total mass density and number density respectively. The mean molecular weight μ is related to the constituent molecular weights as follows.

Non-ionized Case

$$n = \sum_i n_i = \sum_i \frac{\rho_i}{\mu_i m_{\rm H}} = \frac{\rho}{\mu m_{\rm H}}$$
$$\Rightarrow \quad \frac{1}{\mu} = \sum_i \frac{\rho_i}{\rho}\frac{1}{\mu_i} = \sum_i \frac{X_i}{\mu_i} \tag{A.33}$$

Fully Ionized Case

$$n = n_{\rm e} + \sum_i n_i = \sum_i (Z_i+1) n_i$$
$$\Rightarrow \quad \frac{1}{\mu} = \sum_i \frac{X_i(Z_i+1)}{\mu_i} \tag{A.34}$$

This is sometimes approximated as

$$\frac{1}{\mu} \approx \frac{2X_{\rm H}}{1} + \frac{3X_{\rm He}}{4} + \frac{{}^1\!/_2 A X_{\rm metals}}{A} = 2X_{\rm H} + {}^3\!/_4 X_{\rm He} + {}^1\!/_2(1 - X_{\rm H} - X_{\rm He})$$
$$\therefore \quad \frac{1}{\mu} \approx {}^1\!/_2 + {}^3\!/_2 X_{\rm H} + {}^1\!/_4 X_{\rm He} \tag{A.35}$$

Since hydrogen and helium dominate in terms of both number and mass, a good approximation for μ is obtained from the fact that each hydrogen atom supplies two particles and one unit of mass (the electron mass is neglected throughout), and each helium atom three particles and four units of mass.

$$\mu \approx \frac{n_{\rm H} + 4n_{\rm He}}{2n_{\rm H} + 3n_{\rm He}} \approx \frac{1+0.4}{2+0.3} \approx 0.61. \tag{A.36}$$

The number fraction is related to the mass fraction as

$$n_i = \frac{\rho_i}{\mu_i m_{\rm H}} = \frac{X_i \rho}{\mu_i m_{\rm H}} = n\mu \frac{X_i}{\mu_i} \tag{A.37}$$

For the electrons (assuming full ionization as usual),

$$n_e = \sum_i Z_i n_i = n\mu \sum_i \frac{Z_i X_i}{\mu_i} \tag{A.38}$$

This can be approximated as

$$n_e \approx n\mu \left(\frac{X_{\rm H}}{1} + \frac{2X_{\rm He}}{4} + \frac{{}^1\!/_2 A X_{\rm metals}}{A} \right) = n\mu \left[X_{\rm H} + {}^1\!/_2 X_{\rm He} + {}^1\!/_2 (1 - X_{\rm H} - X_{\rm He}) \right]$$
$$\therefore \quad n_e \approx {}^1\!/_2\, n\mu (1 + X_{\rm H}) \tag{A.39}$$

Since the free electrons are due almost entirely to hydrogen and helium, and since the cosmic abundance of helium is around 10 per cent, then

$$n_{\rm e} \approx n_{\rm H} + 2n_{\rm He} \approx 1.2 n_{\rm H}. \tag{A.40}$$

A.8 Density Profiles

Poisson's equation relating gravitational potential Φ and mass density ρ:

$$\nabla^2 \Phi = 4\pi G \rho. \tag{A.41}$$

A.8.1 The Navarro, Frenk, and White Profile[1]

$$\rho(x) = \frac{\rho_{\rm N}}{x(1+x)^2}, \tag{A.42}$$

where $x = r/r_{\rm N}$ is the dimensionless radius. Note that $\rho \to \infty$ as $x \to 0$, i.e. this is a *cuspy* profile.

The enclosed mass is given by

$$M(r) \equiv \int_0^r 4\pi r^2 \rho(r)\,{\rm d}r \tag{A.43}$$

$$= 4\pi r_{\rm N}^3 \rho_{\rm N} \left[\int_0^x \frac{x\,{\rm d}x}{(1+x)^2} = \int_0^x \frac{1}{1+x} - \frac{1}{(1+x)^2}\,{\rm d}x\right]$$

$$= 4\pi r_{\rm N}^3 \rho_{\rm N} \left[\ln(1+x) - \frac{x}{1+x}\right] \tag{A.44}$$

Note that $M(0) = 0$ (as expected since $\rho \sim x^{-1}$ as $x \to 0$), and that the enclosed mass diverges logarithmically at infinity.

Using equation (A.41) and the spherically symmetric form of equation (A.5)

$$\frac{1}{x^2}\frac{\rm d}{{\rm d}x}\left(x^2 \frac{{\rm d}\Phi}{{\rm d}x}\right) = \frac{4\pi G r_{\rm N}^2 \rho_{\rm N}}{x(1+x)^2}, \qquad x^2 \frac{{\rm d}\Phi}{{\rm d}x} = 4\pi G r_{\rm N}^2 \rho_{\rm N} \int \frac{x\,{\rm d}x}{(1+x)^2}$$

Hence the magnitude of the gravitational acceleration

$$g = 4\pi G r_{\rm N} \rho_{\rm N} \frac{1}{x^2}\left[\ln(1+x) + \frac{1}{1+x} + c_1\right] \tag{A.45}$$

$$[\cdots] \sim x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + \left(1 - x + x^2 - x^3 + \cdots\right) + c_1$$

$$\therefore \quad g \sim 4\pi G r_{\rm N} \rho_{\rm N} \frac{1}{x^2}\left[(1+c_1) + \frac{x^2}{2} - \frac{2x^3}{3}\right] \qquad x \ll 1 \tag{A.46}$$

$$\frac{{\rm d}\Phi}{{\rm d}x} = 4\pi G r_{\rm N}^2 \rho_{\rm N} \frac{1}{x^2}\left[\ln(1+x) - \frac{x}{1+x} + (1+c_1)\right]$$

$$\therefore \quad \Phi = 4\pi G r_{\rm N}^2 \rho_{\rm N}\left[-\frac{\ln(1+x)}{x} + \int \frac{(1+c_1)}{x^2}\,{\rm d}x\right]$$

[1] Navarro *et al.* 1997

$$\Phi = -4\pi G r_{\mathrm{N}}^2 \rho_{\mathrm{N}} \left[\frac{\ln(1+x)}{x} + \frac{(1+c_1)}{x} + c_2 \right] \tag{A.47}$$

$$\Phi \sim -4\pi G r_{\mathrm{N}}^2 \rho_{\mathrm{N}} \left[(1+c_2) + \frac{(1+c_1)}{x} - \frac{x}{2} \right] \qquad x \ll 1 \tag{A.48}$$

Since $M(0) = 0$, $\left.\frac{\mathrm{d}\Phi}{\mathrm{d}x}\right|_{x=0}$ must be finite, and hence, from equation (A.46), $c_1 = -1$. Alternatively, this result and equation (A.45) follow directly from $g = GM/r^2$ applied to equation (A.44). c_2 remains arbitrary, setting the zero level of the potential. For $\Phi \to 0$ as $x \to \infty$, $c_2 = 0$. Note that then $\Phi(0) = -4\pi G r_{\mathrm{N}}^2 \rho_{\mathrm{N}}$; and $g(0) = 2\pi G r_{\mathrm{N}} \rho_{\mathrm{N}}$.

Velocity Dispersion

It can be shown (e.g. Binney and Tremaine, 1987) that the associated (singular isothermal sphere) velocity dispersion of such a density profile is

$$\sigma_{\mathrm{SIS}} = {}^2\!/_3 r_{\mathrm{N}} \sqrt{\pi G \rho_{\mathrm{N}}}. \tag{A.49}$$

A.8.2 The King Profile

See for example Binney and Tremaine (1987).

$$\rho(x) = \frac{\rho_{\mathrm{K}}}{(1+x^2)^{3/2}}, \tag{A.50}$$

where $x = r/r_{\mathrm{K}}$ is the dimensionless radius. Note that $\rho \to \rho_{\mathrm{K}}$ as $x \to 0$, i.e. this is a *cored* profile.

The enclosed mass is given by

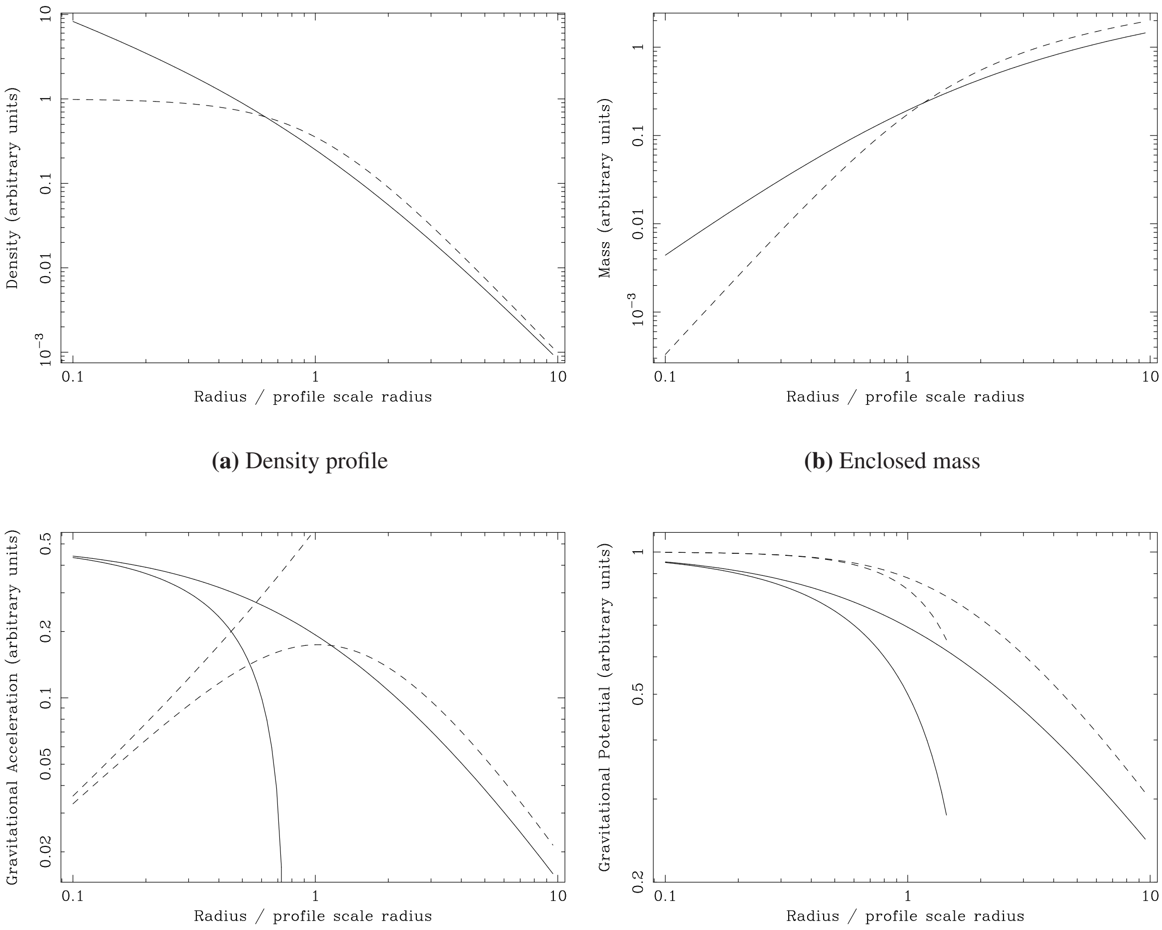
$$M(r) = 4\pi r_{\mathrm{K}}^3 \rho_{\mathrm{K}} \left[\int_0^x \frac{x^2\,\mathrm{d}x}{(1+x^2)^{3/2}} \equiv I_1 \right].$$

$$\begin{aligned} x = \tan\theta \quad \Rightarrow \quad I_1 &= \int \frac{\tan^2\theta \sec^2\theta\,\mathrm{d}\theta}{\sec^3\theta} = \int \frac{\sin^2\theta}{\cos\theta}\,\mathrm{d}\theta = \int \sec\theta - \cos\theta\,\mathrm{d}\theta \\ &= \int \frac{\sec^2\theta + \sec\theta\tan\theta}{\sec\theta + \tan\theta} - \cos\theta\,\mathrm{d}\theta = \ln(\sec\theta + \tan\theta) - \sin\theta \\ &= \ln\left[\sqrt{\tan^2\theta + 1} + \tan\theta\right] - \frac{\tan\theta}{\sqrt{\tan^2\theta + 1}} \end{aligned}$$

$$\therefore \quad M(r) = 4\pi r_{\mathrm{K}}^3 \rho_{\mathrm{K}} \left\{ \ln\left[\sqrt{1+x^2} + x\right] - \frac{x}{\sqrt{1+x^2}} \right\} \tag{A.51}$$

Note that $M(0) = 0$, and that the mass is logarithmically divergent (and asymptotically equal to that for the NFW profile, since $\ln(2x) = \ln 2 + \ln x \sim \ln x$, $x \to \infty$) at infinity; and that $\ln\left[\sqrt{1+x^2} + x\right] = \sinh^{-1} x$.

$$x^2 \frac{\mathrm{d}\Phi}{\mathrm{d}x} = 4\pi G r_{\mathrm{K}}^2 \rho_{\mathrm{K}} \int \frac{x^2\,\mathrm{d}x}{(1+x^2)^{3/2}},$$



(a) Density profile

(b) Enclosed mass

(c) Gravitational acceleration

(d) Gravitational potential

Figure A.2: Functional form of some cluster density profiles. Solid line NFW; dashed King. Profiles are plotted in arbitrary units against radius in units of the profile scale radius. (c, d) also show the small radius approximations to the profiles.

$$g = 4\pi G r_{\mathrm{K}} \rho_{\mathrm{K}} \frac{1}{x^2} \left\{ \ln\left[\sqrt{1+x^2}+x\right] - \frac{x}{\sqrt{1+x^2}} + c_3 \right\} \tag{A.52}$$

$$\{\cdots\} \sim \left[\ln\left(1+\frac{x^2}{2}-\frac{x^4}{8}+\cdots+x\right) - x\left(1-\frac{x^2}{2}+\frac{x^4}{8}-\cdots\right) + c_3 \right]$$

$$\sim \left(x+\frac{x^2}{2}-\frac{x^4}{8}\right) - \frac{1}{2}\left(x^2+x^3+\frac{x^4}{4}\right) + \frac{1}{3}\left(x^3+\frac{3x^4}{2}\right) - x\left(1-\frac{x^2}{2}+\frac{x^4}{8}\right) + c_3$$

$$\therefore \quad g \sim 4\pi G r_{\mathrm{K}} \rho_{\mathrm{K}} \frac{1}{x^2}\left[c_3+\frac{x^3}{3}+\frac{x^4}{4}\right] \qquad x \ll 1 \tag{A.53}$$

$$\frac{\mathrm{d}\Phi}{\mathrm{d}x} = 4\pi G r_{\mathrm{K}}^2 \rho_{\mathrm{K}} \frac{1}{x^2} \left\{ \ln\left[\sqrt{1+x^2}+x\right] - \frac{x}{\sqrt{1+x^2}} + c_3 \right]$$

$$\therefore \quad \Phi = 4\pi G r_{\mathrm{K}}^2 \rho_{\mathrm{K}} \left\{ -\frac{1}{x}\ln\left[\sqrt{1+x^2}+x\right] + \int \frac{1}{x}\frac{\left[x(1+x^2)^{-1/2}+1\right]}{\sqrt{1+x^2}+x} - \int \frac{\mathrm{d}x}{x\sqrt{1+x^2}} - \frac{c_3}{x} - c_4 \right\}$$

$$\Phi = -4\pi G r_{\rm K}^2 \rho_{\rm K} \left\{ \frac{1}{x} \ln\left[\sqrt{1+x^2}+x\right] + \frac{c_3}{x} + c_4 \right\} \tag{A.54}$$

$$\{\cdots\} \sim \frac{1}{x}\left[x - \frac{x^3}{6} + {\rm O}(x^4)\right] + \frac{c_3}{x} + c_4$$

$$\therefore \quad \Phi \sim -4\pi G r_{\rm K}^2 \rho_{\rm K} \left[(1+c_4) + \frac{c_3}{x} - \frac{x^2}{6}\right] \qquad x \ll 1 \tag{A.55}$$

For the correct relationship between gravitational acceleration and mass, $c_3 = 0$. For $\Phi \to 0$ as $x \to \infty$, $c_4 = 0$. Note that then $\Phi(0) = -4\pi G r_{\rm K}^2 \rho_{\rm K}$; and $g(0) = 0$.

A.9 Inversion of the Surface Brightness Relationship

From the definition of surface brightness, equation (3.97),

$$S(b) \equiv \int_b^\infty \frac{\varepsilon(r) 2r\,{\rm d}r}{\sqrt{r^2-b^2}} \quad \Rightarrow \quad \int_{a^2}^\infty \frac{S(b)\,{\rm d}b^2}{\sqrt{b^2-a^2}} = \int_{a^2}^\infty \frac{{\rm d}b^2}{\sqrt{b^2-a^2}} \int_{b^2}^\infty \frac{\varepsilon(r)\,{\rm d}r^2}{\sqrt{r^2-b^2}}.$$

The order of the integrations can be exchanged according to

$$\int_{a^2}^\infty {\rm d}b^2 \int_{b^2}^\infty {\rm d}r^2 = \int_{a^2}^\infty {\rm d}r^2 \int_{a^2}^{r^2} {\rm d}b^2,$$

(which is a special case of the interchange in equation (3.102) with $b_2 \to \infty$). Hence

$$\int_{a^2}^\infty \frac{S(b)\,{\rm d}b^2}{\sqrt{b^2-a^2}} = \int_{a^2}^\infty {\rm d}r^2 \varepsilon(r) \left[\int_{a^2}^{r^2} \frac{{\rm d}b^2}{\sqrt{b^2-a^2}\sqrt{r^2-b^2}} \equiv I\right].$$

Making the substitution

$$c^2 \equiv \frac{b^2-a^2}{r^2-a^2} \quad \Rightarrow \quad I = \int_0^1 \frac{(r^2-a^2)\,{\rm d}c^2}{\sqrt{r^2-a^2}\,c\sqrt{r^2-a^2}\sqrt{1-c^2}}$$

$$= \int_0^1 \frac{{\rm d}c^2}{c\sqrt{1-c^2}} = 2\int_0^1 \frac{{\rm d}c}{\sqrt{1-c^2}} = 2\left[\sin^{-1} c\right]_0^1 = \pi.$$

$$\therefore \quad \int_{a^2}^\infty \frac{S(b)\,{\rm d}b^2}{\sqrt{b^2-a^2}} = \pi \int_{a^2}^\infty {\rm d}r^2 \varepsilon(r)$$

$$\frac{\rm d}{{\rm d}a} \int_{a^2}^\infty \frac{S(b)\,{\rm d}b^2}{\sqrt{b^2-a^2}} = -2\pi a \varepsilon(a)$$

$$\therefore \quad S(b) = \int_{b^2}^\infty \frac{\varepsilon(r)\,{\rm d}r^2}{\sqrt{r^2-b^2}} \qquad \Rightarrow \qquad \varepsilon(r) = -\frac{1}{2\pi r}\frac{\rm d}{{\rm d}r} \int_{r^2}^\infty \frac{S(b)\,{\rm d}b^2}{\sqrt{b^2-r^2}} \tag{A.56}$$

This relationship is a special case of the *generalized Abel equation* that relates a pair of *Abel transforms*,

$$f(x) = \int_0^x \frac{g(t)\,{\rm d}t}{(x-t)^\alpha} \qquad 0 < \alpha < 1 \qquad \Leftrightarrow \qquad g(t) = \frac{\sin(\pi\alpha)}{\pi}\frac{\rm d}{{\rm d}t} \int_0^t \frac{f(x)\,{\rm d}x}{(t-x)^{1-\alpha}}\text{ , or,} \tag{A.57}$$

$$f(x) = \int_x^\infty \frac{g(t)\,\mathrm{d}t}{(t-x)^\alpha} \qquad 0 < \alpha < 1 \qquad \Leftrightarrow \qquad g(t) = -\frac{\sin(\pi\alpha)}{\pi}\frac{\mathrm{d}}{\mathrm{d}t}\int_t^\infty \frac{f(x)\,\mathrm{d}x}{(x-t)^{1-\alpha}}. \tag{A.58}$$

For $\alpha = 1/2$ this is the surface-brightness–emissivity relationship.

A.10 Equations of Cosmography

See for example Hogg (1999), Peebles (1993).

The *redshift* z relates emitted and observed frequencies and wavelengths, being defined as

$$z \equiv \frac{\nu_\mathrm{e}}{\nu_\mathrm{o}} - 1 = \frac{\lambda_\mathrm{o}}{\lambda_\mathrm{e}} - 1 = \sqrt{\frac{1+\beta}{1-\beta}} - 1 \qquad \beta \equiv \frac{v}{c}, \tag{A.59}$$

with v the radial velocity.

The *Hubble constant* H_0 expresses the proportionality between distance and recession velocity,

$$v = H_0 d \tag{A.60}$$

$$H_0 = 50\, h_{50}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}. \tag{A.61}$$

The *Hubble time* is the reciprocal of the Hubble constant,

$$t_\mathrm{H} \equiv \frac{1}{H_0} = 6.17 \times 10^{17}\, h_{50}^{-1}\,\mathrm{s} = 19.6\, h_{50}^{-1}\,\mathrm{Gyr}. \tag{A.62}$$

The dimensionless parameters Ω_M, Ω_Λ, and Ω_k are defined by

$$\Omega_\mathrm{M} \equiv \frac{8\pi G\rho_0}{3H_0^2} \tag{A.63}$$

$$\Omega_\Lambda \equiv \frac{\Lambda c^2}{3H_0^2} \tag{A.64}$$

$$\Omega_\mathrm{M} + \Omega_\Lambda + \Omega_\mathrm{k} = 1, \tag{A.65}$$

where ρ is the mass density of the Universe, Λ is the cosmological constant, and subscript 0 refers to quantities evaluated at the present epoch.

The redshift dependence of the *Hubble parameter* is

$$H(z) = E(z)H_0 \tag{A.66}$$

$$E(z) \equiv (1+z)\sqrt{1 + z\Omega_\mathrm{M} + \Omega_\Lambda\left[\frac{1}{(1+z)^2} - 1\right]}. \tag{A.67}$$

The *line-of-sight comoving distance* D_C to an object at redshift z is given by

$$D_\mathrm{C} = D_\mathrm{H}\int_0^z \frac{\mathrm{d}z'}{E(z')}. \tag{A.68}$$

The *transverse comoving distance* D_M between two objects at redshift z is related to D_C as

$$D_\mathrm{M} = \frac{D_\mathrm{H}}{\sqrt{\Omega_\mathrm{k}}} \sinh\left(\sqrt{\Omega_\mathrm{k}}\frac{D_\mathrm{C}}{D_\mathrm{H}}\right) \qquad (\text{all } \Omega_\mathrm{k}) \tag{A.69}$$

This can be rewritten in various cases as follows:

$$\sinh x \equiv {}^1\!/\!_2(\mathrm{e}^x - \mathrm{e}^{-x}) = {}^1\!/\!_2[(1+x+\cdots)-(1-x+\cdots)] \sim x \qquad \therefore \quad \lim_{x\to 0}\frac{\sinh x}{x} = 1$$

$$\therefore \quad D_\mathrm{M} = D_\mathrm{C}\frac{D_\mathrm{H}}{D_\mathrm{C}\sqrt{\Omega_\mathrm{k}}}\sinh\left(\frac{D_\mathrm{C}\sqrt{\Omega_\mathrm{k}}}{D_\mathrm{H}}\right) = D_\mathrm{C} \qquad (\Omega_\mathrm{k} = 0) \tag{A.70}$$

$$\sinh(\mathrm{i}x) \equiv {}^1\!/\!_2\left(\mathrm{e}^{\mathrm{i}x} - \mathrm{e}^{-\mathrm{i}x}\right) = \mathrm{i}\sin x$$

$$\therefore \quad D_\mathrm{M} = \frac{D_\mathrm{H}}{\mathrm{i}\sqrt{|\Omega_\mathrm{k}|}}\sinh\left(\mathrm{i}\sqrt{|\Omega_\mathrm{k}|}\frac{D_\mathrm{C}}{D_\mathrm{H}}\right) = \frac{D_\mathrm{H}}{\sqrt{|\Omega_\mathrm{k}|}}\sin\left(\sqrt{|\Omega_\mathrm{k}|}\frac{D_\mathrm{C}}{D_\mathrm{H}}\right) \qquad (\Omega_\mathrm{k} < 0) \tag{A.71}$$

It can be shown that

$$D_\mathrm{M} = D_\mathrm{H}\frac{2\left[\Omega_\mathrm{M}z + (2-\Omega_\mathrm{M})\left(1-\sqrt{1+\Omega_\mathrm{M}z}\right)\right]}{\Omega_\mathrm{M}^2(1+z)} \qquad (\Omega_\Lambda = 0) \tag{A.72}$$

$$\therefore \quad D_\mathrm{M} = 2D_\mathrm{H}\left(1 - \frac{1}{\sqrt{1+z}}\right) \qquad (\Omega_\Lambda = 0,\ \Omega_\mathrm{M} = 1) \tag{A.73}$$

The *angular diameter* distance D_A is defined as the Euclidean distance at which at an object of extent l subtends an angle θ,

$$D_\mathrm{A} \equiv \frac{l}{\theta} \tag{A.74}$$

$$D_\mathrm{A} = \frac{D_\mathrm{M}}{1+z}. \tag{A.75}$$

The *luminosity distance* D_L is defined as the Euclidean distance from which an object of bolometric luminosity L is seen to produce a flux F,

$$D_\mathrm{L} \equiv \sqrt{\frac{L}{4\pi F}} \tag{A.76}$$

$$D_\mathrm{L} = (1+z)D_\mathrm{M} = (1+z)^2 D_\mathrm{A}. \tag{A.77}$$

For those who want some proof that physicists are human, the proof is in the idiocy of all the different units which they use for measuring energy.

The Character of Physical Law
RICHARD FEYNMAN

B
Physical Constants and Symbols

This Appendix collates some useful physical data and lists some common symbols used in this thesis.

B.1 Unit Conversions

Density

$$1\,\mathrm{kg\,m^{-3}} = \left(10^{3}\,\mathrm{g}\right)\left(10^{-2}\,\mathrm{cm}\right)^{-3} = 10^{-3}\,\mathrm{g\,cm^{-3}}$$

Energy

$$1\,\mathrm{J} = 1\,\mathrm{N\,m} = 1\,\mathrm{kg\,m^{2}\,s^{-2}} = \left(10^{3}\,\mathrm{g}\right)\left(10^{-2}\,\mathrm{cm}\right)^{-2}\mathrm{s^{-2}} = 10^{7}\,\mathrm{erg} = \left(\frac{e}{\mathrm{J\,eV^{-1}}}\right)^{-1}\mathrm{eV} = 6.242\times10^{18}\,\mathrm{eV}$$

Specific energy

$$1\,\mathrm{J\,kg^{-1}} = 10^{7}\,\mathrm{erg}\left(10^{3}\,\mathrm{g}\right)^{-1} = 10^{4}\,\mathrm{erg\,g^{-1}}$$

Luminosity

$$1\,\mathrm{W} = 10^{7}\,\mathrm{erg\,s^{-1}} \tag{B.1}$$

Surface Brightness

$$1\,\mathrm{W\,m^{-2}} = 10^{7}\,\mathrm{erg\,s^{-1}}\left(10^{2}\,\mathrm{cm}\right)^{-2} = 10^{3}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$$

Cooling Function

$$1\,\mathrm{W\,m^3} = 10^7\,\mathrm{erg\,s^{-1}}\left(10^2\,\mathrm{cm}\right)^3 = 10^{13}\,\mathrm{erg\,s^{-1}\,cm^3}$$

Temperature

$$1\,\mathrm{K} = \left(\frac{k_\mathrm{B}}{\mathrm{J\,K^{-1}}}\right)\mathrm{J} = \left(\frac{k_\mathrm{B}}{\mathrm{J\,K^{-1}}}\right)\left(\frac{e}{\mathrm{J\,eV^{-1}}}\right)^{-1}\mathrm{eV} = 10^{-3}\left(\frac{k_\mathrm{B}}{\mathrm{J\,K^{-1}}}\right)\left(\frac{e}{\mathrm{J\,eV^{-1}}}\right)^{-1}\mathrm{keV} = 8.620\times10^{-8}\,\mathrm{keV}$$

$$\frac{T}{\mathrm{keV}} = 0.862\left(\frac{T}{10^7\,\mathrm{K}}\right) \qquad \frac{T}{\mathrm{K}} = 1.16\times10^7\left(\frac{T}{\mathrm{keV}}\right) \tag{B.2}$$

Force

$$1\,\mathrm{N} = 1\,\mathrm{kg\,m\,s^{-2}} = 10^3\,\mathrm{g}(10^2\,\mathrm{cm})\,\mathrm{s^{-2}} = 10^5\,\mathrm{dyn}$$

Pressure

$$1\,\mathrm{N\,m^{-2}} = 1\,\mathrm{J\,m^{-3}} = \left(\frac{k_\mathrm{B}}{\mathrm{J\,K^{-1}}}\right)^{-1}\mathrm{K}\left(10^2\,\mathrm{cm}\right)^{-3} = 7.241\times10^{16}\,\mathrm{K\,cm^{-3}}$$

$$\frac{p}{\mathrm{K\,cm^{-3}}} = 7.241\times10^{16}\,\mathrm{N\,m^{-2}} \qquad \frac{p}{\mathrm{N\,m^{-2}}} = 1.381\times10^{-17}\,\mathrm{K\,cm^{-3}} \tag{B.3}$$

Magnetic Flux Density

$$1\,\mathrm{T} = 10^4\,\mathrm{G}$$

Energy – Wavelength

$$\frac{E}{\mathrm{J}} = \left(\frac{h}{\mathrm{Js}}\right)\left(\frac{c}{\mathrm{m\,s^{-1}}}\right)\left(\frac{\lambda}{\mathrm{m}}\right)^{-1}$$

$$\left(10^3\frac{e}{\mathrm{J\,eV^{-1}}}\right)\frac{E}{\mathrm{keV}} = \left(\frac{h}{\mathrm{Js}}\right)\left(\frac{c}{\mathrm{m\,s^{-1}}}\right)\left(10^{-10}\frac{\lambda}{\mathrm{\AA}}\right)^{-1}$$

$$\frac{E}{\mathrm{keV}} = 12.4\left(\frac{\lambda}{\mathrm{\AA}}\right)^{-1} \qquad \frac{\lambda}{\mathrm{\AA}} = 12.4\left(\frac{E}{\mathrm{keV}}\right)^{-1} \tag{B.4}$$

Solid Angle

$$1\,\mathrm{steradian} = 1\,\mathrm{sr} = 1\,\mathrm{rad}^2 = \left(\frac{180}{\pi}\right)^2\mathrm{deg}^2 = \left(\frac{60\times180}{\pi}\right)^2\mathrm{arcmin}^2 \tag{B.5}$$

$$\therefore\quad 4\pi\,\mathrm{sr} = 4\pi\left(\frac{180}{\pi}\right)^2\mathrm{deg}^2 \approx 41253\,\mathrm{deg}^2$$

Quantity	Symbol	Value	Unit
Speed of light	c	2.998×10^{8}	$\mathrm{m\,s^{-1}}$
Parsec	pc	3.086×10^{16}	m
Boltzmann's constant	k_{B}	1.381×10^{-23}	$\mathrm{J\,K^{-1}}$
Planck's constant	h	6.626×10^{-34}	$\mathrm{J\,s}$
Gravitational constant	G	6.673×10^{-11}	$\mathrm{m^{3}\,kg^{-1}\,s^{-2}}$
Permittivity of free space	ε_{0}	8.854×10^{-12}	$\mathrm{F\,m^{-1}}$
Electron charge	e	1.602×10^{-19}	C
Electron mass	m_{e}	9.110×10^{-31}	kg
Proton mass	m_{H}	1.673×10^{-27}	kg
Solar mass	$\mathrm{M}_{\odot}$	1.989×10^{30}	kg

Table B.1: Values of and symbols for some important physical constants.

Element	Atomic Number	Mass Number	Number Abundance	Mass Fraction	Electron Fraction[a]
H	1	1	1.00×10^{0}	7.05×10^{-1}	8.27×10^{-1}
He	2	4	9.77×10^{-2}	2.76×10^{-1}	1.62×10^{-1}
C	6	12	3.63×10^{-4}	3.07×10^{-3}	1.80×10^{-3}
N	7	14	1.12×10^{-4}	1.11×10^{-3}	6.48×10^{-4}
O	8	16	8.51×10^{-4}	9.60×10^{-3}	5.63×10^{-3}
Ne	10	20	1.23×10^{-4}	1.73×10^{-3}	1.02×10^{-3}
Na	11	23	2.14×10^{-6}	3.47×10^{-5}	1.95×10^{-5}
Mg	12	24	3.80×10^{-5}	6.43×10^{-4}	3.77×10^{-4}
Al	13	27	2.95×10^{-6}	5.62×10^{-5}	3.17×10^{-5}
Si	14	28	3.55×10^{-5}	7.01×10^{-4}	4.11×10^{-4}
S	16	32	1.62×10^{-5}	3.66×10^{-4}	2.14×10^{-4}
Ar	18	40	3.63×10^{-6}	1.02×10^{-4}	5.40×10^{-5}
Ca	20	40	2.29×10^{-6}	6.46×10^{-5}	3.79×10^{-5}
Fe	26	56	4.68×10^{-5}	1.85×10^{-3}	1.01×10^{-3}
Ni	28	59	1.78×10^{-6}	7.40×10^{-5}	4.12×10^{-5}

[a] Full ionization

Table B.2: Solar abundances (Anders and Grevesse, 1989) and atomic properties of some astrophysically important elements.

B.2 Physical Constants

See Table B.1.

B.3 Plasma Abundance Data

See Tables B.2 to B.4 on pages 161–162.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Number	H	He	O	C	Ne	N	Fe	Mg	Si	S	Ar	Al	Ca	Na	Ni
Mass	H	He	O	C	Fe	Ne	N	Si	Mg	S	Ar	Ni	Ca	Al	Na
Electrons	H	He	O	C	Ne	Fe	N	Si	Mg	S	Ar	Ni	Ca	Al	Na

Table B.3: The most important (decreasing in importance left – right) species in the ICM, according to different criteria: number of ions; mass of ions; number of electrons (full ionization).

Metal abundance (solar units)	μ_{neutral}	μ_{ion}	n_{e}	X_{H}	X_{He}	X_{met}
3.0	1.34	0.63	1.24	0.68	0.27	0.056
1.0	1.29	0.61	1.21	0.71	0.28	0.019
0.3	1.27	0.61	1.20	0.71	0.28	0.0059
0.0	1.27	0.61	1.20	0.72	0.28	0.00

Table B.4: Effect of metallicity on some bulk plasma properties. H and He are fixed at the solar abundance in all cases.

B.4 Glossary of Common Symbols

See Table B.5 on the facing page for a glossary of the common symbols used in this thesis.

r	radius
b	projected radius
t	time
t_{cool}	isobaric, radiative cooling time
t_{stab}	Courant stability time = $\Delta x / c_{\mathrm{s}}$
$\mathcal{C}$	Courant number = $\Delta t / t_{\mathrm{stab}}$
u	speed
p	pressure
q	pseudoviscous pressure, heat flux, specific energy
T	temperature
ε	specific internal energy, or emissivity
c_{s}	sound speed
μ	molecular weight
m	mass
$\mathcal{V}$	specific volume, $1/\rho$
ρ	mass density
n	number density
n_{e}	electron number density
n_{H}	hydrogen number density
r_{e}	relative electron density, n_{e} / n_{H}
Δ	overdensity, ρ / ρ_{c}
γ	adiabatic index
f_{hi}	mass fraction of metal-rich phase
X_{i}	mass fraction of species i
Z_{i}	atomic number of species i
Z	metallicity relative to solar
r_{i}	number density of species i relative to hydrogen
$S(b)$	surface brightness at projected radius b
$x^n_{j+\frac{1}{2}}$	x at point $j+\frac{1}{2}$, time n
$x^n_{i,j+\frac{1}{2}}$	x for phase i at point $j+\frac{1}{2}$, time n
$\bar{x}^n_{j+\frac{1}{2}}$	$x^n_{i,j+\frac{1}{2}}$ averaged over phases
g	magnitude of gravitational acceleration
Φ	gravitational potential
κ	coefficient of thermal conductivity
Λ	cooling function
$\ln\Lambda$	Coulomb logarithm
$\mathcal{Q}$	specific conductive heat power loss
$\mathcal{H}$	enthalpy
$\mathcal{L}$	specific radiative power loss

Table B.5: Glossary of common symbols as used in this thesis.

BIBLIOGRAPHY

Abell G. O., 1965, ARA&A, 3, 1

Allen S. W., Ettori S., Fabian A. C., 2001a, MNRAS, 324, 877

Allen S. W., Fabian A. C., 1994, MNRAS, 269, 409

Allen S. W., Fabian A. C., 1998, MNRAS, 297, L63

Allen S. W., Fabian A. C., Johnstone R. M., Arnaud K. A., Nulsen P. E. J., 2001b, MNRAS, 322, 589

Allen S. W., Schmidt R. W., Fabian A. C., 2001c, MNRAS, 328, L37

Allen S. W., Schmidt R. W., Fabian A. C., 2002, MNRAS, 335, 256

Ames W. F., 1992, *Numerical Methods for Partial Differential Equations*, 3rd edn., Academic Press, Inc.

Anders E., Grevesse N., 1989, Geochim. Cosmochim. Acta, 53, 197

Arnaud K. A., 1996, in Jacoby G. H., Barnes J. (eds.), *Astronomical Data Analysis Software and Systems V*, vol. 101, p. 17, ASP Conf. Ser., San Francisco

Arnaud K. A., 1999, OGIP Memo, 92-009, `http://heasarc.gsfc.nasa.gov/docs/heasarc/ofwg/ofwg_recomm.html`

Arnaud K. A., George I. M., Tennant A. F., 1998, OGIP Memo, 92-007, `http://heasarc.gsfc.nasa.gov/docs/heasarc/ofwg/ofwg_recomm.html`

Arnaud M., Raymond J., 1992, ApJ, 398, 394

Arnaud M., Rothenflug R., 1985, A&AS, 60, 425

Aschenbach B., 1985, Rep. Prog. Phys., 48, 579

Bahcall N. A., 1977, ARA&A, 15, 505

Balbus S. A., Soker N., 1989, ApJ, 341, 611

Bayer-Kim C. M., Crawford C. S., Allen S. W., Edge A. C., Fabian A. C., 2002, MNRAS, 337, 938

Becker R., 1922, Z. Phys., 8, 321

Begelman M. C., Fabian A. C., 1990, MNRAS, 244, 26P

Berger M., Colella P., 1989, J. Comp. Phys., 82, 64

Binney J., Tremaine S., 1987, *Galactic Dynamics*, Princeton University Press, Princeton, New Jeresy

Biviano A., 2000, IAP, (astro-ph/0010409)

Böhringer H., Fabian A. C., 1989, MNRAS, 237, 1147

Böhringer H., Matsushita K., Churazov E., Ikebe Y., Chen Y., 2002, A&A, 382, 804

Boldt E., McDonald F. B., Riegler G., Serlemitsos P., 1966, Phys. Rev. Lett., 17, 447

Bradt H. V. D., Ohashi T., Pounds K. A., 1992, ARA&A, 30, 391

Bregman J. N., David L. P., 1988, ApJ, 326, 639

Bregman J. N., David L. P., 1989, ApJ, 341, 49

Brickhouse N., *et al.*, 1995, Legacy, 6, 4

Brüggen M., Kaiser C. R., Churazov E., Enßlin T. A., 2002, MNRAS, 331, 545

Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559

Buote D. A., Canizares C. R., Fabian A. C., 1999, MNRAS, 310, 483

Buote D. A., Lewis A. D., Brighenti F., Mathews W. G., 2002, ApJ, submitted (astro-ph/0205362)

Buote D. A., Lewis A. D., Brighenti F., Mathews W. G., 2003, ApJ, submitted (astro-ph/0303054)

Byram E. T., Chubb T. A., Friedman H., 1966, Science, 152, 66

Carilli C. L., Taylor G. B., 2002, ARA&A, 40, 319

Cavaliere A., Fusco-Femiano R., 1978, A&A, 70, 677

Chandran B. D. G., Cowley S. C., 1998, Phys. Rev. Lett., 80, 3077

Chandrasekhar S., 1960, *Principles of Stellar Dynamics*, Dover, New York

Churazov E., Brüggen M., Kaiser C. R., Böhringer H., Forman W., 2001, ApJ, 554, 261

Churazov E., Sunyaev R., Forman W., Böhringer H., 2002, MNRAS, 332, 729

Cohen R. S., Spitzer Jr L., Routly P. M., 1950, Phys. Rev., 80 (2), 230

Cole S., Lacey C., 1996, MNRAS, 281, 716

Cowie L. L., McKee C. F., 1977, ApJ, 211, 135

Crawford C. S., Allen S. W., Ebeling H., Edge A. C., Fabian A. C., 1999, MNRAS, 306, 857

Crawford C. S., Fabian A. C., 1992, MNRAS, 259, 265

David L. P., Hughes J. P., Tucker W. H., 1992, ApJ, 394, 452

David L. P., Nulsen P. E. J., McNamara B. R., Forman W., Jones C., Ponman T., Robertson B., Wise M., 2001, ApJ, 557, 546

De Grandi S., Molendi S., 2001, ApJ, 551, 153

De Grandi S., Molendi S., 2002, ApJ, 567, 163

Dere K. P., Landi E., Mason H. E., Monsignori Fossi B. C., Young P. R., 1997, A&AS, 125, 149

Donahue M., Voit G. M., 1993, ApJL, 414, L17

Doron R., Behar E., 2002, ApJ, 574, 518

Dos Santos S., 2001, MNRAS, 323, 930

Drake S. A., 1992, Legacy, 1

Dressler A., 1984, ARA&A, 22, 185

Dupke R. A., White III R. E., 2000, ApJ, 537, 123

Edge A. C., 2001, MNRAS, 328, 762

Edge A. C., Ivison R. J., Smail I., Blain A. W., Kneib J.-P., 1999, MNRAS, 306, 599

Edge A. C., Stewart G. C., 1991, MNRAS, 252, 414

Edge A. C., Stewart G. C., Fabian A. C., 1992, MNRAS, 258, 177

Ettori S., Fabian A. C., 1999, MNRAS, 305, 834

Ettori S., Fabian A. C., 2000, MNRAS, 317, L57

Ettori S., Fabian A. C., Allen S. W., Johnstone R. M., 2002, MNRAS, 331, 635

Ettori S., Fabian A. C., White D. A., 1998, MNRAS, 300, 837

Evrard A. E., Metzler C. A., Navarro J. F., 1996, ApJ, 469, 494

Ezawa H., Fukazawa Y., Makishima K., Ohashi T., Takahara F., Xu H., Yamasaki N., 1997, ApJ, 490, L33

Fabian A. C. (ed.), 1988, *Cooling Flows in Clusters and Galaxies*, Kluwer Academic Publishers, Dordrecht, The Netherlands

Fabian A. C., 1994, ARA&A, 32, 277

Fabian A. C., 2002a, in Avila-Reese V., Firmani C., Frenk C., Allen C. (eds.), *Galaxy Evolution: Theory and Observations*, RevMexAA (SC), in press (astro-ph/0210150)

Fabian A. C., 2002b, in Gilfanov M., Sunyaev R., Churazov E. (eds.), *Lighthouses of the Universe: The Most Luminous Celestial Objects and their Use for Cosmology*, p. 24, Springer-Verlag, Berlin

Fabian A. C., Allen S. W., Crawford C. S., Johnstone R. M., Morris R. G., Sanders J. S., Schmidt R. W., 2002a, MNRAS, 332, L50

Fabian A. C., Arnaud K. A., Bautz M. W., Tawara Y., 1994a, ApJL, 436, L63

Fabian A. C., Canizares C. R., Böhringer H., 1994b, ApJ, 425, 40

Fabian A. C., Mushotzky R. F., Nulsen P. E. J., Peterson J. R., 2001a, MNRAS, 321, L20

Fabian A. C., Nulsen P. E. J., Canizares C. R., 1982, MNRAS, 201, 933

Fabian A. C., Nulsen P. E. J., Canizares C. R., 1984, Nat, 310, 733

Fabian A. C., Nulsen P. E. J., Canizares C. R., 1991, A&AR, 2, 191

Fabian A. C., Pringle J. E., 1977, MNRAS, 181, 5P

Fabian A. C., Sanders J. S., Ettori S., Taylor G. B., Allen S. W., Crawford C. S., Iwasawa K., Johnstone R. M., 2001b, MNRAS, 321, L33

Fabian A. C., Voigt L. M., Morris R. G., 2002b, MNRAS, 335, L71

Fabian A. C., *et al.*, 2000, MNRAS, 318, L65

Felten J. E., Gould R. J., Stein W. A., Woolf N. J., 1966, ApJ, 146, 955

Field G. B., 1965, ApJ, 142, 531

Finoguenov A., David L. P., Ponman T. J., 2000, ApJ, 544, 188

Friedman H., Lichtman S. W., Byram E. T., 1951, Phys. Rev., 83, 1025

Fukazawa Y., Ohashi T., Fabian A. C., Canizares C. R., Ikebe Y., Makishima K., Mushotzky R. F., Yamashita K., 1994, PASJ, 46, L55

Gaetz T. J., 1989, ApJ, 345, 666

George I. M., Arnaud K. A., 1998a, OGIP Calibration Memo, 92-002a, `http://heasarc.gsfc.nasa.gov/docs/heasarc/ofwg/ofwg_recomm.html`

George I. M., Arnaud K. A., 1998b, OGIP Memo, 92-007a, `http://heasarc.gsfc.nasa.gov/docs/heasarc/ofwg/ofwg_recomm.html`

George I. M., Arnaud K. A., Pence B., Ruamsuwan L., 1998, OGIP Calibration Memo, 92-002, `http://heasarc.gsfc.nasa.gov/docs/heasarc/ofwg/ofwg_recomm.html`

Giacconi R., Gursky H., Paolini F. R., Rossi B. B., 1962, Phys. Rev. Lett., 9, 439

Gil'fanov M. R., Syunyaev R. A., Churazov E. M., 1987, Pis'ma Astron. Zh., 13, 7

Giovannini G., Feretti L., Venturi T., Kim K.-T., Kronberg P. P., 1993, ApJ, 406, 399

Govoni F., Feretti L., Giovannini G., Böhringer H., Reiprich T. H., Murgia M., 2001, A&A, 376, 803

Grevesse N., Sauval A. J., 1998, Space Sci. Rev., 85, 161

Grevesse N., Sauval A. J., 1999, A&A, 347, 348

Harris D. E., Stern C. P., Willis A. G., Dewdney P. E., 1993, AJ, 105, 769

Heckman T. M., 2001, in Mulchaey J., Stocke J. (eds.), *Extragalactic Gas at Low Redshift*, vol. 254, p. 292, ASP Conf. Ser., San Francisco

Henriksen M. J., Tittley E. R., 2002, ApJ, 577, 701

Hogg D. W., 1999, astro-ph/9905116

Horner D. J., Mushotzky R. F., Scharf C. A., 1999, ApJ, 520, 78

Hughes J. P., Rakowski C. E., Burrows D. N., Slane P. O., 2000, ApJL, 528, L109

Iwasawa K., Fabian A. C., Allen S. W., Ettori S., 2001, MNRAS, 328, L5

Jaffe W., 1980, ApJ, 241, 925

Johnstone R. M., Allen S. W., Fabian A. C., Sanders J. S., 2002, MNRAS, 336, 299

Johnstone R. M., Fabian A. C., Edge A. C., Thomas P. A., 1992, MNRAS, 255, 431

Johnstone R. M., Fabian A. C., Nulsen P. E. J., 1987, MNRAS, 224, 75

Kaastra J. S., Ferrigno C., Tamura T., Paerels F. B. S., Peterson J. R., Mittaz J. P. D., 2001, A&A, 365, L99

Kaastra J. S., Mewe R., 1993, Legacy, 3, 16

Kahn S. M., *et al.*, 2002, in Bowyer S., Hwang C.-Y. (eds.), *Matter and Energy in Clusters of Galaxies*, ASP Conf. Ser., San Francisco, (astro-ph/0210665)

Karzas W. J., Latter R., 1961, ApJS, 6, 167

Katz J. I., 1976, ApJ, 207, 25

Kauffmann G., Colberg J. M., Diafero A., White S. D. M., 1999, MNRAS, 303, 188

Landau L. M., Lifshitz E. M., 1987, *Fluid Mechanics*, 2nd edn., Butterworth–Heinemann

Large M. I., Mathewson D. S., Haslam C. G. T., 1959, Nat, 183, 1663

Liedahl D. A., Osterheld A. L., Goldstein W. H., 1995, ApJL, 438, L115

Loeb A., 2002, NewA, 7, 279

Longair M. S., 1992, *High Energy Astrophysics 1: Particles, Photons and their Detection*, 2nd edn., Cambridge University Press, Cambridge

Makishima K., *et al.*, 2001, PASJ, 53, 401

Malyshkin L., 2001, ApJ, 554, 561

Malyshkin L., Kulsrud R., 2001, ApJ, 549, 402

Markevitch M., Forman W. R., Sarazin C. L., Vikhlinin A., 1998, ApJ, 503, 77

Markevitch M., *et al.*, 2000, ApJ, 541, 542

Mathews W. G., Buote D. A., Brighenti F., 2001, ApJ, 550, L31

Mathiesen B. F., Evrard A. E., 2001, ApJ, 546, 100

Matsumoto H., Koyama K., Awaki H., Tomida H., Tsuru T., Mushotzky R., Hatsukade I., 1996, PASJ, 48, 201

Matteucci F., Fusco-Femiano R. (eds.), 2002, *Chemical Enrichment of Intracluster and Intergalactic Medium*, vol. 253, ASP Conf. Ser. 253, San Francisco

Mazzotta P., Fusco-Femiano R., Vikhlinin A., 2002, ApJ, 569, L31

McWilliam A., 1997, ARA&A, 35, 503

Meiksin A., 1988, ApJ, 334, 59

Mewe R., Gronenschild E. H. B. M., 1981, A&AS, 45, 11

Mewe R., Kaastra J. S., Liedahl D. A., 1995, Legacy, 6, 16

Mitchell R. J., Culhane J. L., Davison P. J. N., Ives J. C., 1976, MNRAS, 176, 29P

Mitton S., 1970, MNRAS, 153, 133

Moore B., Governato F., Quinn T., Stadel J., Lake G., 1998, ApJL, 499, L5

Morris R. G., Fabian A. C., 2002, in Matteucci and Fusco-Femiano (2002), pp. 85–90

Morris R. G., Fabian A. C., 2003, MNRAS, 338, 824

Mushotzky R., Loewenstein M., Arnaud K. A., Tamura T., Fukazawa Y., Matsushita K., Kikuchi K., Hatsukade I., 1996, ApJ, 466, 686

Mushotzky R. F., Loewenstein M., 1997, ApJ, 481, L63

Mushotzky R. F., Szymkowiak A. E., 1988, in Fabian (1988), pp. 53–62

Narayan R., Medvedev M. V., 2001, ApJ, 562, L129

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493 von Neumann J., Richtmyer R. D., 1950, J. Appl. Phys., 21, 232

Nevalainen J., Markevitch M., Forman W., 2000, ApJ, 532, 694

Nulsen P. E. J., 1986, MNRAS, 221, 377

Nulsen P. E. J., 1988, in Fabian (1988), pp. 175–187

Nulsen P. E. J., 1998, MNRAS, 297, 1109

Oegerle W. R., Cowie L., Davidsen A., Hu E., Hutchings J., Murphy E., Sembach K., Woodgate B., 2001, ApJ, 560, 187

Pacholczyk A. G., 1970, *Radio Astrophysics: Nonthermal Processes in Galactic and Extragalactic Sources*, A Series of Books in Astronomy and Astrophysics, W. H. Freeman, San Francisco

Peebles P. J. E., 1993, *Principles of Physical Cosmology*, Princeton Series in Physics, Princeton University Press

Pellegrini S., Ciotti L., 2002, in Matteucci and Fusco-Femiano (2002), pp. 65–70

Peterson J. R., Ferrigno C., Kaastra J. S., Paerels F. B. S., Kahn S. M., Jernigan J. G., Bleeker J. A. M., Tamura T., 2002, in Jansen F. (ed.), *New Visions of the X-Ray Universe in the XMM-Newton and Chandra era*, ESA SP-488, ESA Publications Division, Noordwijk, (astro-ph/0202108)

Peterson J. R., Kahn S. M., Paerels F. B. S., Kaastra J. S., Tamura T., Bleeker J. A. M., Ferrigno C., Jernigan J. G., 2003, ApJ, accepted (astro-ph/0210662)

Peterson J. R., *et al.*, 2001, A&A, 365, L104

Phillips K. J. H., Mewe R., Harra-Murnion L. K., Kaastra J. S., Beiersdorfer P., Brown G. V., Liedahl D. A., 1999, A&AS, 138, 381

Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, MNRAS, 338, 14

Press W. H., Flannery B. P., Teukolsky S. A., Vetterling W. T., 1990, *Numerical Recipes*, Cambridge University Press, Cambridge

Raymond J. C., Smith B. W., 1977, ApJS, 35, 419

Rechester A. B., Rosenbluth M. N., 1978, Phys. Rev. Lett., 40, 38

Reisenegger A., Miralda-Escude J., Waxman E., 1996, ApJ, 457, L11

Rephaeli Y., 1977a, ApJ, 218, 323

Rephaeli Y., 1977b, ApJ, 212, 608

Rephaeli Y., 1978, ApJ, 225, 335

Richtmyer R. D., Morton K. W., 1967, *Difference Methods for Initial Value Problems*, 2nd edn., Interscience Publishers, New York

Rosenbluth M. N., Kaufman A. N., 1958, Phys. Rev., 109, 1

Ruszkowski M., Begelman M. C., 2002, ApJ, 581, 223

Rybicki G. B., Lightman A. P., 1979, *Radiative Processes in Astrophysics*, Wiley

Sanders J. S., Fabian A. C., 2002, MNRAS, 331, 273

Sarazin C. L., 1988, *X-Ray Emission from Clusters of Galaxies*, Cambridge Astrophysics Series, Cambridge University Press

Schlegel E. M., 2002, *The Restless Universe: Understanding X-Ray Astronomy in the Age of Chandra and Newton*, Oxford University Press, Oxford

Schmidt R. W., Allen S. W., Fabian A. C., 2001, MNRAS, 327, 1057

Schmidt R. W., Fabian A. C., Sanders J. S., 2002, MNRAS, 337, 71

Smith R. K., Brickhouse N. S., Liedahl D. A., Raymond J. C., 2001, ApJ, 556, L91

Spitzer Jr L., 1962, *Physics of Fully Ionized Gases*, Interscience Tracts on Physics and Astronomy, 2nd edn., Wiley, New York

Spitzer Jr L., Härm R., 1953, Phys. Rev., 89 (5), 977

Stewart G. C., Fabian A. C., Nulsen P. E. J., Canizares C. R., 1984, ApJ, 278, 536

Strickland D., 2002, in Matteucci and Fusco-Femiano (2002), pp. 387–396

Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253

Tamura T., *et al.*, 2001, A&A, 365, L87

Taylor G. B., Fabian A. C., Allen S. W., 2002, MNRAS, 334, 769

Teyssier R., 2002, A&A, 385, 337

Thomas P. A., 1987, *Cooling Flows and Galaxy Formation*, Ph.D. thesis, Univ. Cambridge

Thomas P. A., 1988, MNRAS, 235, 315

Tribble P. C., 1989, MNRAS, 238, 1247

Tucker W., Giacconi R., 1985, *The X-Ray Universe*, Harvard University Press, Cambridge, Massachusetts

Vikhlinin A., Markevitch M., Murray S. S., 2001a, ApJ, 549, L47

Vikhlinin A., Markevitch M., Murray S. S., 2001b, ApJ, 551, 160

Voigt L. M., Schmidt R. W., Fabian A. C., Allen S. W., Johnstone R. M., 2002, MNRAS, 335, L7

Wechsler R. H., Bullock J. S., Primack Joel R., Kravtsov A. V., Dekel A., 2002, ApJ, 568, 52

Willson M. A. G., 1970, MNRAS, 151, 1

Wise M. W., Sarazin C. L., 1992, ApJ, 395, 387

Woodward P., Colella P., 1984, J. Comp. Phys., 54, 115

Yahagi H., Yoshii Y., 2001, ApJ, 558, 463

Young P. R., Del Zanna G., Landi E., Dere K. P., Mason H. E., Landini M., 2003, ApJS, 144, 135

Zakamska N. L., Narayan R., 2003, ApJ, 582, 162

Science is not everything, but science is very beautiful.

J. ROBERT OPPENHEIMER

Well, somebody's got to write these things. And the same can be said for collecting garbage and shovelling snow. It doesn't matter whether you like it or not — a job's a job.
For three and a half years, I've been making this kind of contribution to society. Shovelling snow. You know, cultural snow.

Dance Dance Dance
HARUKI MURAKAMI