Future X-ray Spectroscopy Missions

Frits Paerels

SRON Laboratory for Space Research, Sorbonnelaan 2,
3584CA Utrecht, The Netherlands,

and

Columbia Astrophysics Laboratory, Columbia University,
538 West 120th Street, New York, NY 10027, USA

Abstract. This article provides a description of the high-resolution X-ray spectrometers for astrophysics that will become operational in the near future. The emphasis is on the physical principles of operation.

1 Introduction

The title of this Chapter refers to future missions for astrophysical X-ray spectroscopy. At the time of this writing (Spring 1998), the future is especially close, and we expect it to be bright. Within the next two years, high-resolution X-ray spectrometers will be placed in orbit on *AXAF*, *XMM*, *Spectrum X/γ*, and *Astro-E*. For the first time, we will have the sensitivity to spectroscopically detect a wide variety of diagnostic physical effects, from essentially all types of cosmic X-ray source. This article therefore concentrates on this immediate future, although at the end I will briefly discuss what is under study for the 21st century.

In selecting topics for this chapter, I have tried to emphasize the physics of the instruments, especially of the grating spectrometers on *AXAF* and *XMM*, rather than enumerating their properties and giving examples of simulated spectroscopy for various types of X-ray source (which might suggest that this field is intellectually already pretty much covered). A brief description of some X-ray spectroscopic diagnostics provides an estimate for astrophysically interesting ranges of resolving power, and I largely leave the astrophysical applications to your imagination.

Careful quantitative spectroscopy requires an understanding of the operation of the spectrometers at various levels of detail. A description of the physical principles may help to appreciate the properties of the instruments, the reasons for various design choices, and the physical and practical limits to their resolving power and efficiency (which will be pushed in future generations of instruments). As such, this chapter aims to provide an introduction to what is sometimes referred to as 'the theory of the experiment' in experimental physics.

J. van Paradijs, J.A.M. Bleeker (Eds.): LNP 520, pp. 347 - 434, 1999
© Springer-Verlag Berlin Heidelberg 1999

2 Resolving Powers of Interest in Astrophysical X-ray Spectroscopy

In this section, I will collect spectroscopic diagnostics of interest in the X-ray band, and the spectral resolving power required to use the diagnostic. You have seen a lot of this material before. I have added some more, and I will put it in the definite context of the spectroscopic capabilities of the X-ray spectrometers to be put in orbit in the near future, in the next section.

2.1 Ionization stage spectroscopy

At the lowest level of spectrosopic analysis is the ability to distinguish the emission spectra of the various elements and their ionization stages. The most basic parameters to be measured with this 'diagnostic' are of course the elemental abundances themselves, but an accurate abundance measurement must rely on an accurate measurement of the ionization balance as well. The ionization balance itself is an important diagnostic for the physical conditions in the emitting gas. It contains information on the excitation mechanism (collisional or recombination), the temperature and density, and the thermal history of the gas. You can find a nice set of scaling relations in George Fraser's contribution, one of which gives the relative energy splittings between the strongest transitions in the hydrogenic and helium-like ions as a function of nuclear charge Z, but I will repeat the numbers here for a few astrophysically important elements. Table 1 lists the energies of the principal $n = 2 - 1$ transitions in the hydrogenic and helium-like ions of O, Si, and Fe, which span the $0.8 - 7$ keV band. I also list the resolving power $\mathcal{R} \equiv E/\Delta E$ required to distinguish the H- and He-like transitions (which scales like Z).

Table 1. Energies of the principal $n = 2 - 1$ lines in H- and He-like ions

element	H-like $1s - 2p$	He-like $1s^2 - 1s2p$	$\mathcal{R}$
O	654 eV	574 eV	8
Si	2005	1865	14
Fe	6960	6701	27

2.2 Excitation mechanism

Another very basic diagnostic is the dominant line excitation mechanism. As discussed extensively in Duane Liedahl's contribution, there are very clear

spectroscopic differences between the emission spectra arising from collisionally or photoionized media. In the first case, line emission results primarily from radiative decays following collisional excitation, in the second case, radiative decays following recombination dominate. For a gas of cosmic composition in photoionization equilibrium, this leads to enhanced emission from the low- and mid-Z hydrogenic and He-like ions with respect to the Fe L ions, as compared to emission from a collisional plasma of comparable mean ionization.

In addition, for low enough electron temperatures the radiative recombination continua appear as narrow, line-like features at the series limits. To detect these features clearly separated from (at least) Lyβ in hydrogenic ions (to take the simplest case) requires resolving power $\mathcal{R} = 9$.

Another, subtler diagnostic is based solely on the spectroscopy of the Fe L ions, and it therefore also works in cases where the electron temperature is too high to make the narrow recombination continua detectable. As first described by Liedahl et al. (1990), the recombination spectra of the Fe L ions are markedly different from the collisionally excited spectra. This is shown in Fig.1, which shows the spectra of Fe XVI-XIX under conditions of coronal and photoionization equilibrium. The differences between these two different spectra should be evident even at resolving powers of order $\mathcal{R} \gtrsim 10 - 20$.

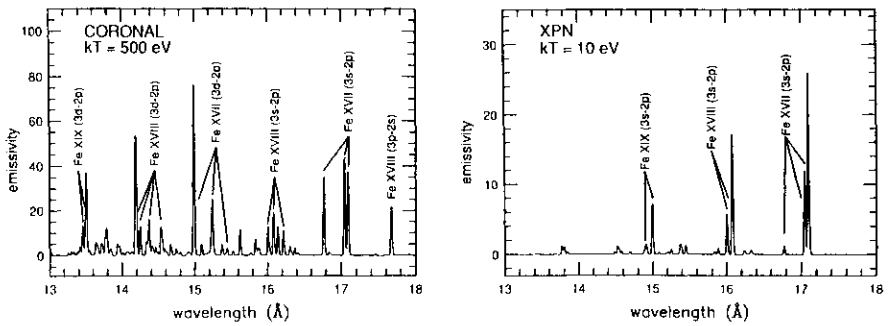

Fig. 1. Line spectra emitted by Fe XVI-XIX under conditions of coronal and photoionization equilibrium. From Liedahl et al. (1990).

2.3 Density diagnostics

He-like ions The intensities in the $n = 2 - 1$ transitions in the helium-like ions are sensitive to the density of the plasma. The physical principle underlying this diagnostic is treated in the Chapters by Rolf Mewe and Duane Liedahl, but I include a brief description for completeness. The standard reference is Gabriel & Jordan (1969). The diagnostic rests on the competition

between spontaneous and collisional transition rates out of the various $n = 2$ levels. The first are determined purely by atomic structure, the second scale with the electron density, and a measurement of the excitation balance in the $n = 2$ sublevels must therefore be sensitive to density. Four transitions are used in this diagnostic: $1s^2\ ^1S_0$ to $1s2p\ ^1P_1, 1s2p\ ^3P_{1,2}$ and $1s2s\ ^3S_1$. The first is an allowed transition, the second two involve a change in the total spin-angular momentum of the two electrons ('spin-flip'), the fourth is dipole-forbidden. The transitions are therefore commonly referred to as the resonance line ('R'), the 'intercombination' (between the triplet and singlet systems) lines ('I' – the two lines are too closely spaced to be resolved in practical cases), and the forbidden line ('F').

At low densities, the intensities in the lines are purely determined by the collisional excitation rates into the various upper levels: every collisional excitation is followed by a spontaneous radiative decay (in the case of a photoionized plasma, the intensities would be determined by the rate at which the upper levels are populated by recombination and cascades, instead). But at high densities, the collisional rate from the upper level of the F line to the upper level of the I lines becomes comparable to or exceeds the slow spontaneous radiative decay rate out of the upper level of the F line, and the ratio between the I and the F lines changes. The dividing line occurs when the two rates out of $1s2s\ ^3S_1$ are equal, and the density for which this happens is referred to as the critical density for the ion. Critical densities span the range 10^9 to 10^{17} cm^{-3} for the helium-like ions from O to Fe.

In order to apply this diagnostic, the I and F lines need to be resolved. Line energies and resolving powers for three important helium-like ions are listed in Table 2.

Table 2. Energies of the $n = 2 - 1$ R, I, and F lines in He-like ions

element	R	I	F	$\mathcal{R}$
O	574 eV	569 eV	561 eV	72
Si	1865	1853	1840	155
Fe	6701	6673	6634	240

Fe L Ions Again, the Fe L ions provide an alternative diagnostic in a photoionized plasma (Liedahl et al. 1992). At sufficiently high density, low-lying metastable excited states of the ions develop a significant population through collisional excitation. These excited ions have recombination spectra that are different from those in the ground state, as illustrated in Fig.2 for Fe XX. To

separate the strongest of the many lines that appear with increasing density from the lines present at low density requires $\mathcal{R} \gtrsim 150 - 200$ (for the specific but representative case of Fe XX).

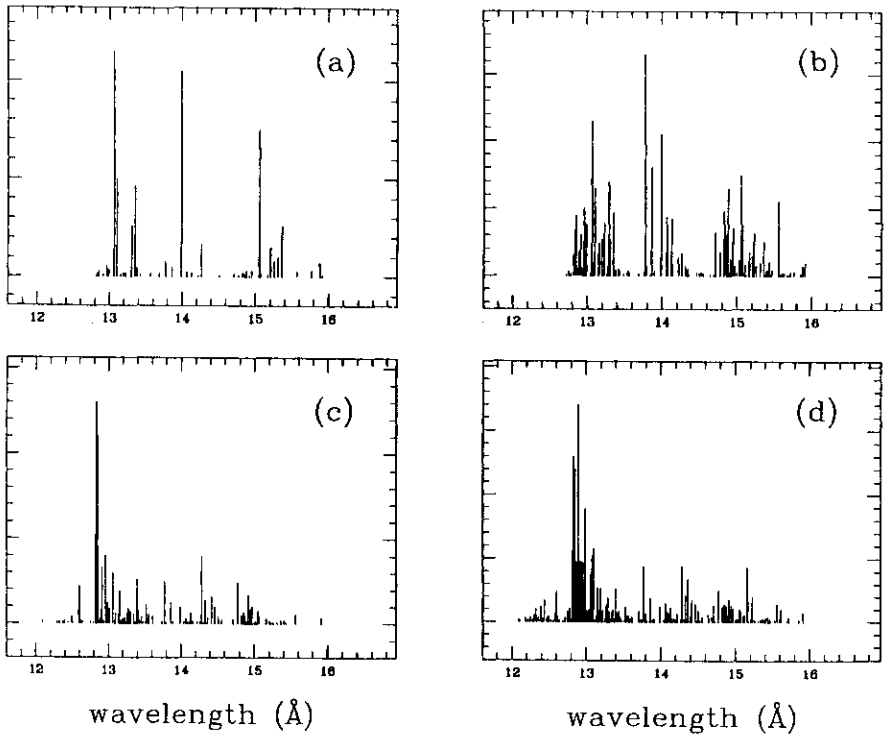

Fig. 2. Line spectra of Fe XX, under conditions of photoionization [(a),(b)] and coronal equilibrium [(c),(d)]. Panel (a) shows the recombination spectrum at low density, panel (b) the spectrum above the critical density of $n \sim 10^{13.5}$ cm^{-3}. From Liedahl et al. (1992).

2.4 Satellite line spectroscopy

A powerful electron temperature diagnostic, proven in terrestrial plasma experiments and solar spectroscopy, is spectroscopy of dielectronic satellite lines (Rolf Mewe, Duane Liedahl, this volume). Dielectronic recombination onto a target ion of (net) charge $z + 1$ leaves an ion in charge state z in a doubly excited state, with one electron, the 'spectator', in a high Rydberg level. This spectator provides some shielding of the nuclear field, and so the stabilizing

downward radiative transition involving the other electron has an energy that is slightly lower than the corresponding transition in an ion of charge $z+1$ (with no spectator); there will be a satellite for each possible value of the principal quantum number of the spectator. The ratio of the intensities in the satellite line to that of the 'parent' transition (in charge state $z+1$) is sensitive to the electron temperature, because the satellite intensity is proportional to the dielectronic recombination rate onto charge state $z+1$, while the parent intensity is equal to its collisional excitation rate, and both rates have different temperature dependence. The strength of the method is that it does not depend on knowing the ionization balance: the target charge state for both the parent transition and the dielectronic satellites is charge state $z+1$.

For definiteness, we examine the difference in the energies of the dielectronic satellites in the lithium-like ions of Fe and O, and the resonance transitions in the corresponding helium-like ions. A direct measurement of the satellites to the helium-like Fe resonance line at 6701 eV (1.850 Å) by Beiersdorfer et al. (1992) finds the $n=3$ satellite at a distance of $\Delta\lambda = 1.82\times10^{-3}$ Å towards longer wavelengths, so to unambiguously resolve at least this resonance requires high resolving power, $\mathcal{R}\sim1000$. For the $n=3$ satellite to the He-like O resonance line at 574 eV (21.60 Å), Mewe, Gronenschild & van den Oord (1985) list a wavelength difference of 0.034 Å, so a resolving power of $\mathcal{R}=640$ is required to apply the diagnostic.

2.5 Radiative recombination continuum spectroscopy

As explained in detail in Duane Liedahl's contribution, the characteristic that most distinguishes a plasma in photoionization equilibrium from one in collisional ionization equilibrium is the comparatively low electron temperature. As a consequence, the radiative recombination continuum photons will all be 'piled up' just above the ionization limit, because the free electrons all have kinetic energies much smaller than the ionization energy. The resulting narrow, quasi-discrete recombination continuum is a strong spectroscopic signature of photoionization. It can be used to straightforwardly measure the electron temperature in the zone where the radiative recombination rate onto a given ion species peaks: the emissivity depends on photon energy E as $\exp(-\Delta E/kT_e)$, with $\Delta E = E-\chi$, χ the ionization potential, T_e the electron temperature. The recombination continuum therefore has a characteristic width of order kT_e, and to resolve it requires roughly $\mathcal{R}\gtrsim\chi/kT_e$. As a practical example, you need $\mathcal{R}\sim140$ to resolve the continuum in hydrogenic Neon, at a (representative) electron temperature of ~10 eV. Higher temperatures are easier to resolve, but the contrast at the ionization edge goes down proportionally as well (which is why this diagnostic does not work for the hot plasmas in collisional equilibrium).

2.6 Thermal Doppler broadening

A straightforward ion temperature diagnostic is the thermal Doppler width of emission lines. The radial velocity distribution in a thermal plasma is Gaussian, with a variance[1] $\sigma^2 = kT_\mathrm{i}/m_\mathrm{i}$, with T_i and m_i the ion temperature and mass, respectively. For the resolving power corresponding to this width, we find $\mathcal{R} = \lambda/\Delta\lambda = c/\sigma = (m_\mathrm{i}c^2/kT_\mathrm{i})^{1/2} = 1000A^{1/2}T_7^{-1/2}$, with A the atomic weight of the ions, T_7 the ion temperature in units 10^7 K. This indicates that this is not a trivial diagnostic—for all ions of interest, $A > 12$, so a high spectral resolving power is required.

2.7 Compton scattering effects

Monochromatic photons of wavelength λ_0 scattering off of stationary electrons suffer an energy loss due to the Compton recoil effect. The maximum corresponding wavelength shift $\Delta\lambda_\mathrm{max}$ in a single scattering interaction is equal to twice the Compton wavelength of the electron, $\lambda_\mathrm{C} = h/m_\mathrm{e}c$, so $\Delta\lambda_\mathrm{max}/\lambda_0 = -2h/m_\mathrm{e}c\lambda_0 = -2E_0/m_\mathrm{e}c^2$ (for complete backscattering). To resolve the Compton-downscattered photons, you therefore need a resolving power of order $\mathcal{R} \sim 511/E_\mathrm{keV}$, with E_keV the photon energy in keV.

A measurement of the intensity of downscattered photons yields an estimate of the electron column density through which the photons passed. In order for the scattering probability to be non-negligible, you need to have the Thomson depth $\tau_\mathrm{T} \sim 1$, or $N_\mathrm{H} \sim \sigma_\mathrm{T}^{-1} = 1.5 \times 10^{24}$ cm^{-2} (with N_H the total ionized Hydrogen column density, and σ_T the Thomson cross section). At such large column densities, the photoelectric opacity at lower energies is large, so that the usefulness of this spectroscopic diagnostic is probably limited to the Fe K lines.

Photons passing through a large column density τ_T will suffer on average $\overline{n} = \tau_\mathrm{T}^2/2$ collisions, so a line at energy E_0 will appear downshifted by $\Delta E/E_0 \sim \overline{n}E_0/m_\mathrm{e}c^2$. The line will also appear broadened, due to statistical fluctuations both on the number of scattering interactions, as well as on the energy shift per scattering. After passing through a column equivalent to an average number of $\overline{n}$ scatterings, the r.m.s. width of the line is $\Delta E/E_0 = \overline{n}^{1/2}(7/5)^{1/2}E_0/m_\mathrm{e}c^2$ [the fundamental papers on Comptonization in cold gas are Ross et al. (1978) and Illarionov et al. (1979); one factor $\overline{n}^{1/2}$ comes from the variance in the number of scattering interactions, the distribution of energy shifts per scattering supplies the other factor $(\frac{2}{5}\overline{n})^{1/2}$].

If the scattering electrons have a finite temperature, the Doppler shifts due to the velocities of the electrons have to be taken into account. For a Maxwellian velocity distribution of temperature T_e, the characteristic width of the scattered line radiation, expressed as a variance or r.m.s. width, is

[1] not to be confused with the usual, 'official' definition of the frequency 'Doppler width', $\Delta\nu_\mathrm{D} \equiv \sqrt{2}\nu_0\sigma/c$ (Rybicki & Lightman 1979, p.288)

$\Delta E/E_0 \approx (2kT_e/m_ec^2)^{1/2}$ (Illarionov et al. 1979; compare this with the r.m.s. width of ionic emission lines given above—the additional factor two arises from two successive Lorentz transformations, into the electron rest frame and back to the observer's frame). The net line *shift* due to scattering from a finite-temperature medium can be both positive and negative, depending on whether the average electron energy is larger or smaller than one-fourth of the line photon energy.

The characteristics of the Compton scattering effects can conveniently be summarized as follows (cf. McCray 1984, and correcting a typo): passing through a column density of free electrons of temperature T_e, equivalent to an average number of scatterings $\overline{n} = \tau_T^2/2$, line photons of energy E_0 will experience a net shift

$$\Delta E_{\rm shift}/E_0 \approx \overline{n}(4kT_e - E_0)/m_ec^2, \tag{1}$$

and an r.m.s. broadening of the scattered radiation

$$\Delta E_{\rm broadening}/E_0 \approx \overline{n}^{1/2}(7E_0^2/5 + 2kT_em_ec^2)^{1/2}/m_ec^2, \tag{2}$$

and so the relevant resolving power is still set by $\mathcal{R} \gtrsim 511/E_{\rm keV}$, except for warm to hot plasmas ($kT_e \gtrsim 10$ eV).

2.8 Raman scattering

If X rays are absorbed in a neutral medium, you may observe, instead of Compton scattering, the effects of scattering by bound electrons, that is, Rayleigh scattering and Raman scattering. Since hydrogen is the most abundant element, you would in practice only expect to see Rayleigh and Raman scattering off hydrogen atoms, although very sensitive experiments might also detect the effects of scattering off helium atoms.

Rayleigh scattering involves excitation of the bound electron into an intermediate state (not a stationary state of the atom), from which it decays back to the initial state. The scattered photon therefore emerges with unchanged energy (coherent scattering), only its direction has been changed. If the photon has an energy larger than the excitation energy of a discrete transition between stationary states in the atom, however, the excited atom may deexcite to the excited level, and the outgoing photon has less energy than the incoming photon, the difference being the excitation energy of the atomic transition (Raman scattering). Since most Raman scattering events will involve a transition $n = 1-2$ in hydrogen (Lyα), the signature of Raman scattering of X-ray lines would be the presence of scattered photons at an energy 10.2 eV (the energy of the Lyα transition) below an emission line.

Again, since the scattering cross section is small (a fraction of the Thomson cross section), only absorbers with very large neutral column densities

will have finite Raman scattering optical depth, and these will be photoelectrically opaque at low energies. Raman scattering will therefore probably only be detected in Fe K photons, and it requires a resolving power $\mathcal{R} \sim 6400\text{eV}/10\text{eV} = 640$ to do so.

A careful study of the Raman spectrum (including the angular dependence, and the effects of the distribution of the velocity of the bound scattering electron—on average of order $\alpha c \sim 2000$ km s^{-1} in the ground state of Hydrogen! [α is the fine structure constant]) will reveal detailed information on the geometrical distribution and amount of scattering hydrogen atoms with respect to the emission line source and the observer. Sunyaev & Churazov (1996) give a (characteristically) complete calculation of the effect, and describe an application to the study of the Galactic Center Region. For more general applications of Raman spectroscopy in astrophysics, see Nussbaumer, Schmid & Vogel (1989).

2.9 Fluorescence spectroscopy

From ions with a filled K-shell, and at least a partially filled L-shell, you may get fluorescent X rays. The radiative transitions follow an inner-shell ionization, either collisional or by photoionization, with a probability y, the fluorescence yield, which is a strong function of nuclear charge. For instance, $y_{\rm K} \approx 0.3$ for Fe Kα, but only a few percent for O Kα. L-shell fluorescence is very inefficient. Despite the low yield, fluorescent emission from low-Z elements may still be detectable, from photoionized plasmas. The fluorescence emissivity $j_{\rm fl}$ is the product of the yield with the photoionization rate: $j_{\rm fl} = y_{\rm K} n_i \int_\chi^\infty dE\, \sigma(E) S(E)$ photons cm^{-3} s^{-1}, with n_i the density of element i, $\sigma(E)$ the photoionization cross section, $S(E)$ the ionizing radiation field (photons cm^{-2} s^{-1} keV^{-1}), and χ the ionization potential. The low yield $y_{\rm K}$ for the light elements is offset by higher abundances (as compared to Fe), by larger photoabsorption cross sections, and by the fact that for a typical AGN ionizing radiation field (for instance) the integral actually favors low-Z elements [$S(E)$ a steeply falling powerlaw in E, and lower χ's for low-Z elements].

Careful spectroscopy of the fluorescent spectrum may provide information on the ionization state of the fluorescing gas, its ionization history, and its physical and chemical state. To date, this has only been attempted at Fe K, although fluorescent emission from lighter elements has now also been detected (e.g., the buried active nucleus in NGC 6552, cf. Fukazawa et al. 1994).

Ionization state Precise determination of the (rest frame) energy of fluorescent emission lines or absorption edges in principle allows the determination of the ionization stage in which the transitions arise. This is no different in principle from the the Ionization Stage Spectroscopy mentioned above, except that in ions more neutral than helium-like, the transition energies vary

only slowly with charge state, requiring medium to high spectral resolution, and a very reliable calibration of the spectrometer energy scale to uniquely separate them. Also note that in any astrophysical source, charge states that differ little in ionization potential (like the Fe M shell ions) are likely to coexist, so that the emission spectrum shows the superposition of the individual spectra of the separate charge states. Rather than subtle energy shifts, the major effect may be to broaden the transitions by superposition.

To get some idea of which distinctions become important or measurable at what resolving power, let us look at the Fe K spectrum in particular. The dependence of the energies of the various Kα transitions on ionization stage is fairly steep for the Fe L ions. For instance, it takes $\mathcal{R} \approx 200$ to distinguish the $1s - 2p_{3/2}$ transitions arising in the various L-shell ions (Decaux et al. 1997). Higher resolving power is required to do the same for the more neutral species. The average separation between the $1s - 2p_{3/2}$ transitions arising in the ions Fe X-XVII is approximately $\Delta\lambda = 1.5 \times 10^{-3}$ Å per ionization stage (Decaux et al. 1995), so a resolving power $\mathcal{R} \gtrsim 1300$ is required to distinguish these stages uniquely (the wavelength of all Fe Kα transitions is approximately $\lambda \approx 1.93$ Å). At this resolution, a lot of the intrinsic complexity of the Fe K spectra of individual charge states will also be resolved (e.g. the fine structure split between Kα_1 and Kα_2).

Finally, at high resolving power, other peculiar effects become important. For the ionization stages below Fe XI, accidents of atomic structure make the energies of the Kα transitions actually go slightly below those for neutral Fe, with decreasing ionization stage, before going up again (Decaux et al. 1995). This underlines the importance of having an accurate and stable wavelength scale calibration for the spectrometer, as well as accurate and reliable rest frame energies for the transitions under study, or else the small effect I just mentioned might easily masquerade as a spurious velocity field in the source with an amplitude of several hundred km s^{-1}!

Physical state of the fluorescing material This issue is actually almost completely unexplored in astrophysics: astrophysical X-ray spectrometers never had the required resolving power. The following remarks are therefore somewhat sketchy—time will tell how important these issues really are. Because this field is only now becoming of interest to general astrophysics, accurate calculations and laboratory data that would be of peculiar interest to astrophysics are not always available (this applies to the requirements of pure astrophysical spectroscopy, as well as to laboratory calibration of X-ray astrophysical spectrometers!). The situation is changing with the advent of new experiments expressly designed to meet this need, though.

Precise measurement of the wavelengths of the fluorescent emission lines provides information on the physical state of the fluorescing atoms. In principle, there is a difference in the energy of a given transition depending on whether it occurs in a free atom, or in an atom bound to other atoms, as

in a molecule or a solid. There are small shifts and broadenings caused by the interaction of atomic electrons with the other charges in the molecule or solid. Moreover, the shifts and broadenings depend on the chemical constitution as well. These issues are already important for the ground calibration of the grating spectrometers on *AXAF* and *XMM*. These have resolving powers ranging from several hundred to several thousand, at which point the finite (but poorly characterized) *width* of the popular (for calibration purposes) Al and Mg Kα characteristic X-ray lines excited in a solid (approximately 1 eV, at 1.49 and 1.25 keV, respectively) is already resolved. Information on transition *shifts* has been tabulated by Sevier (1979), from which I take the data listed in Table 3.

Table 3. Energy Levels and Line Energies in Oxygen and Iron: Physical and Chemical Shifts

	Oxygen			Iron		
energy	gas	oxide	$\mathcal{R}$	gas	oxide	$\mathcal{R}$
$1s$	545.4 eV	532.0 eV	40	7124 eV	7113 eV	650
$2p_{1/2}$	16.40	7.1		733	721.2	
$2p_{3/2}$	—	—	—	720	708.1	
Kα_1	529.0	524.9	130	6391	6392	6400
Kα_2	—	—	—	6404	6405	6400

First note that in situations where near-neutral gas is viewed against a sufficiently strong background X-ray continuum source, a precise measurement of the energy of absorption edges already provides information on the physical state of the absorbing atoms. At least for the low-Z elements, only modest resolving power is required (e.g. $\mathcal{R} = 40$ to distinguish absorption by monatomic gaseous oxygen, from oxygen bound in oxide). For this technique to work, however, you require a bright 'backlighter', and the column density of the absorbing element should be such that the optical depth at the absorption edge is of order unity, to have good contrast. By contrast, a measurement of the energy of fluorescent line emission may be more widely applicable, but requires higher resolving power, and will only be practicable for low Z elements.

Fluorescence spectroscopy is relevant for all elements for which there is evidence that they are easily bound in dust grains in free space [see for instance Snow & Witt (1996), and references therein]. Just to be able to show that material is indeed locked in dust, and how much of it is locked in dust as opposed to being in the gas phase, is already important information when studying certain environments (e.g. the cold gas near the nuclei of AGN).

2.10 EXAFS spectroscopy

Continuum photoelectric absorption by atoms bound in a solid or in molecules may show the curious EXAFS effect ('Extended X-ray Absorption Fine Structure'), which appears as a slow, wavelike modulation of the atomic photoelectric absorption cross section just above an ionization threshold. This modulation is the result of quantum interference, and the physical mechanism works as follows. A photon of energy E ionizes an atom, and produces a photoelectron of momentum $p = [2m_e(E-\chi)]^{1/2}$, with χ the ionization potential. If E is only slightly larger than χ, the de Broglie wavelength of the photoelectron is large, and may be as large as the lattice period of the absorbing solid (or the interatomic distance, in molecules). The outgoing photoelectron probability wave is scattered by neighboring atoms, and if the de Broglie wavelength matches the lattice period, strong interference may result between the outgoing wave and the scattered waves. If the resulting interference is such that the photoelectron wavefunction amplitude is suppressed near the origin, the photoabsorption transition probability is reduced with respect to that for a free atom, and the opposite for constructive interference.

To estimate the characteristic energy scale involved in the effect, consider the following simplified argument [for a complete quantitative description, see Lee et al. (1981); astrophysical implications are discussed by Martin (1970), Evans (1986), and Woo (1995)]. The photon energies corresponding to successive interference resonances are given roughly by the condition $2ka = 2n\pi$, with $k = p/\hbar$ the photoelectron wavevector, a the lattice period or interatomic distance of the absorbing material, and n a positive integer. This neglects any phase shifts in the photoelectron wave function associated with the Coulomb interaction of the photoelectron with the ion, as well as with the scattering atoms. Using $\hbar^2k^2/2m_e = E - \chi$, we have for the resonance energies

$$E_n \sim \chi + \frac{h^2}{8m_e a^2}n^2 \tag{3}$$

so the spacing between the first two peaks in the ionization cross section is, very roughly,

$$\Delta E \sim \frac{h^2}{8m_e a^2} = 38(a/1\ \text{Å})^{-2}\ \text{eV}. \tag{4}$$

To see the effect at the oxygen K edge, assuming a characteristic lattice spacing $a = 1$ Å, therefore requires resolving power $\mathcal{R} = \chi_{\text{O K}}/\Delta E = 14$. Similarly, at the silicon edge, you need $\mathcal{R} = 50$.

The precise shape and amplitude of the modulation is sensitive to the exact structure and composition of the absorbing crystals or molecules, through the phase shifts incurred by Coulomb scattering of the photoelectron waves by the atoms neighboring the absorbing atom. That is, the effect can tell, for instance, whether photoelectrically absorbing Si atoms are all surrounded by other Si atoms, or by O atoms. The exact crystalline properties of interstellar dust is an unsolved problem, and EXAFS spectroscopy is a possible

technique to address the issue. Again, just to be able to detect the presence of dust or molecular material, and its abundance relative to material in the gas phase, may in itself provide important information.

Given the fact that photoelectric absorption is the physical effect employed in many types of X-ray detector, it is not surprising that EXAFS is a major concern for instrument calibration. In fact, many detectors contain two elements likely to show the astrophysical effect: silicon and oxygen—and the instrumental effect may mask the astrophysical effect (for examples of the instrumental effects, see Owens et al. 1997).

2.11 Radial-velocity spectroscopy

Radial-velocity spectroscopy is one of the oldest applications of spectroscopy in astronomy, and is obviously also of great interest to X-ray astronomy. With sufficiently high sensitivity, one could for instance detect binary motion (both in binaries containing coronal X-ray sources, as well as in classical X-ray binaries), motion in accretion flows, shock velocities in supernova remnants, bulk relative motion in merging clusters, etc.

Despite the conceptual simplicity of the measurement, accurate radial-velocity measurements are actually not trivial in practice. As a rough criterion, you need resolving power $\mathcal{R} \sim c/v$ to detect a velocity field of amplitude v, and this criterion says that velocity spectroscopy is hard: to detect $v \sim$ few hundred km s^{-1}, characteristic of many astrophysical dynamical situations, would require $\mathcal{R} \gtrsim 1000$.

To beat this limit, the following argument is often invoked: if an emission line has been detected with N photons, one can determine its centroid wavelength with an accuracy of order $\Delta\lambda = \Delta\lambda_{\rm sp}/\sqrt{N}$, with $\Delta\lambda_{\rm sp}$ the wavelength resolution of the spectrometer. Translated into velocity, this gives $v = c\Delta\lambda/\lambda = c/(\mathcal{R}_{\rm sp}\sqrt{N})$, so in principle you can do better in v/c than the spectrometer resolution by a factor $\sqrt{N}$. From the previous discussions, it will be clear that you have to be careful with this argument. For it to work, you need to know the laboratory wavelengths to high precision, and you need an accurate and stable wavelength calibration of the spectrometer.

In Figure 3, I have plotted representative values of the resolving powers we just discussed, as a function of X-ray photon energy. As you can see, the diagnostics are literally all over the diagram. Unfortunately, as we will see in the next section, there is no physical effect that will allow you to construct a spectrometer with a nearly energy-independent (and high) resolving power across the entire wavelength band. No single spectrometer will probably ever cover all of these diagnostics simultaneously, and so either compromises are made, or observatories are designed to carry multiple, complementary instruments.

This was only a very rough overview, which will mainly serve as an aid to appreciate the capabilities of the spectrometers to be put into orbit over the next few years. Several important classes of diagnostic have not been

discussed, mainly in order for the scope of the article not to blow up out of all proportion. For instance, nothing has been said about spectroscopy of hot, high-density plasmas (atmospheres of neutron stars and white dwarfs). It is also reasonable to expect that there will be surprises: unsuspected physical effects with clear spectroscopic signatures, waiting to be discovered.

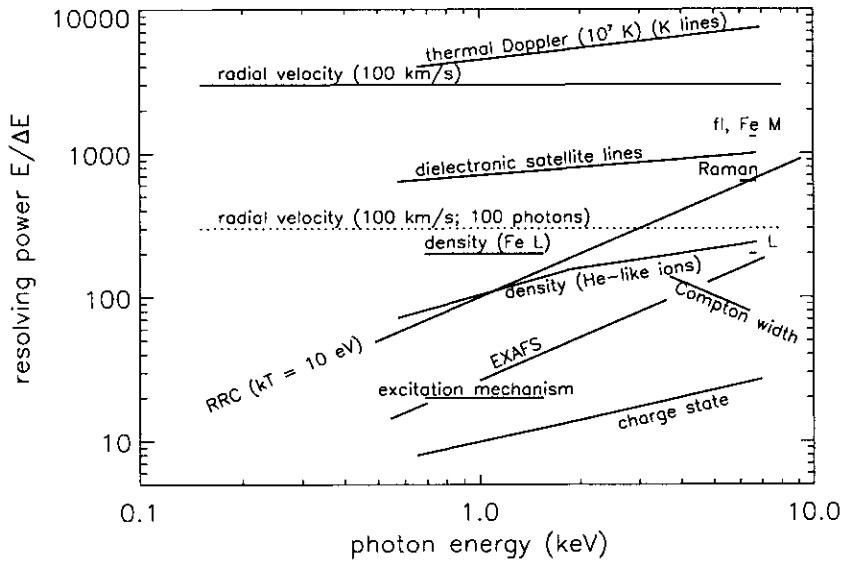

Fig. 3. Resolving power required for some spectroscopic diagnostics, as a function of the photon energy of the diagnostic feature. 'fl, Fe M' indicates the location of the resolving power required to distinguish between the Kα lines in the different Fe M-shell ions, and 'L' indicates the same for the Fe L-shell ions. 'RRC' stands for 'radiative recombination continuum'.

3 X-ray Astrophysical Spectrometers

To set the stage for the discussion of the real instruments on future observatories, we very briefly review the principles of the main types of spectrometer in use in X-ray astronomy.

X-ray spectrometers are usually divided into diffractive and non-diffractive instruments; the terms 'dispersive' and 'non-dispersive' are also used, as well as 'wavelength-dispersive' and 'energy-dispersive', or, less appropriately, 'constant-$\Delta\lambda$' and 'constant-ΔE' spectrometers. The first rely on diffraction of X rays, and comprise grating and crystal spectrometers. The second

comprise ionization detectors, calorimeters, and superconducting tunneling junctions, and as yet to be invented other devices that rely on conversion of photon energy into some other measurable quantity.

3.1 Diffractive spectrometers

Grating spectrometers Grating spectrometers are in many ways the simplest of all possible X-ray spectrometers. In a standard application, a grating or set of gratings, either transmission or reflection, is placed behind a focusing telescope. This is usually referred to as an 'objective grating spectrometer', although that term is actually, strictly speaking, reserved for configurations with the dispersing element located *in front* of the telescope. Figure 4 shows a schematic of the arrangement. First, let us look at transmission gratings. From the condition for constructive interference for a grating of period d, illuminated by light of wavelength λ incident at an angle χ, you derive the dispersion equation

$$m\lambda = d(\sin\theta - \sin\chi) \tag{5}$$

where θ is the dispersion angle, and m the spectral order (Fig. 4). If the incident beam has an intrinsic angular spread $\Delta\chi$, the diffracted beam will have a corresponding angular spread, $\Delta\theta = \Delta\chi$, at fixed wavelength and order. That $\Delta\theta$ corresponds to a wavelength width, according to the dispersion equation:

$$\Delta\lambda = \frac{d}{m}\cos\theta\Delta\theta, \tag{6}$$

and, for small θ, this is independent of λ for a fixed telescope angular resolution $\Delta\theta$—hence the name 'constant-$\Delta\lambda$ spectrometer'. The resolving power is $\mathcal{R} = \lambda/\Delta\lambda = \tan\theta/\Delta\theta \approx \theta/\Delta\theta$. You can increase the resolving power by either increasing θ (increasing the line density of the grating, or going to higher order m), or by decreasing $\Delta\theta$ (better telescope). We will return to this issue when we discuss the relative merits of transmission and reflection gratings.

Transmission gratings were flown on the *Einstein* and *EXOSAT* observatories (Brinkman et al. 1980; Seward et al. 1982). Spectra of a variety of cosmic X-ray sources were obtained, of which the results on stellar coronal emission deserve special mention (see Mewe 1991, for a review). Transmission gratings are also at the heart of the high-resolution spectrometers on *AXAF*, to be discussed later.

The corresponding geometry for a reflection grating is shown in Figure 5. Defining α as the angle of incidence with respect to the grating plane, and β as the dispersion angle, you derive the dispersion equation

$$m\lambda = d(\cos\beta - \cos\alpha). \tag{7}$$

For $m = 0$ (zero order) you obviously have $\alpha = \beta$, i.e. the grating acts as a mirror. For $m < 0$, the dispersed ray is inside the triangle defined by

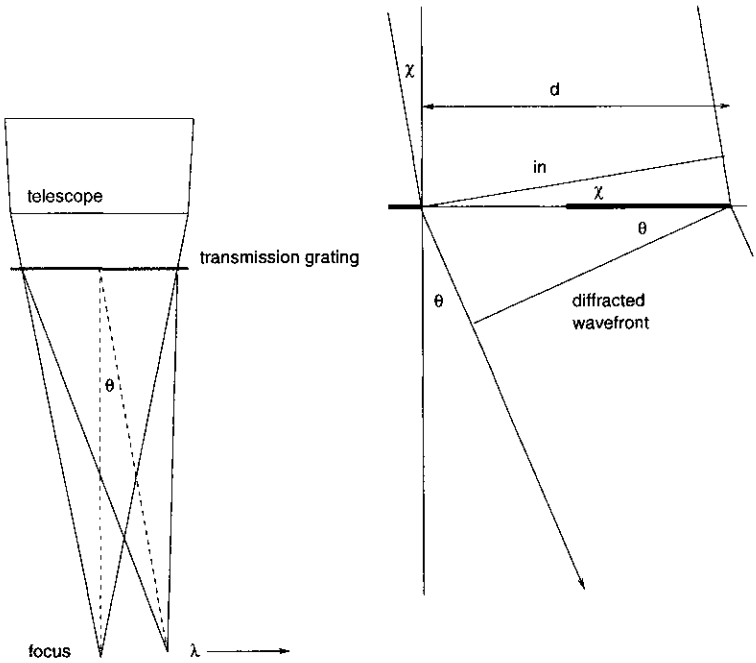

Fig. 4. Schematic arrangement for an X-ray transmission grating spectrometer. A grazing-incidence telescope focuses radiation, which is diffracted by a transmission grating placed in the focused beam (the grating bars are oriented perpendicular to the plane of the page). Focused X-rays are dispersed in the plane of the page, and detected by a position-sensitive detector in the focal plane.

the incoming ray and the zero order ray; these orders are referred to as the 'inside orders'. Outside orders ($m > 0$) don't always exist; for certain ranges of λ and given d, m and α, the dispersion equation may not have a solution ($\cos\beta > 1$).

When illuminated by a beam of finite angular resolution $\Delta\alpha$, the diffracted beams have an angular spread as well, which according to the dispersion relation corresponds to a wavelength resolution

$$\Delta\lambda = \frac{d}{m}\sin\alpha\Delta\alpha. \tag{8}$$

This shows that a reflection grating instrument is also a 'constant-$\Delta\lambda$' spectrometer. The resolving power is

$$\mathcal{R} = \lambda/\Delta\lambda = \frac{\cos\alpha - \cos\beta}{\sin\alpha\Delta\alpha} \tag{9}$$

and this shows that you can increase the resolving power by using a good telescope (small $\Delta\alpha$), or by lowering the angle of incidence. What counts is

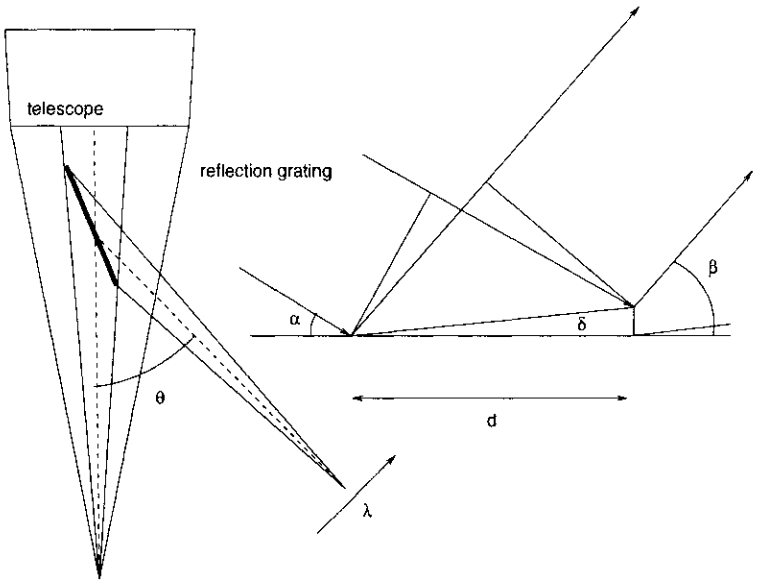

Fig. 5. Diffraction geometry for a reflection grating spectrometer.

the 'projected grating period', $d/\sin\alpha$. At grazing incidence, you can reach fantastically high effective ruling densities this way; for example, at $\alpha =$ 1 degree, and a moderate line density of 500 lines/mm, the effective line density is $\sim 29,000$ lines/mm! Reflection gratings therefore have the potential advantage over transmission gratings of delivering high resolving power at moderate, easy to fabricate ruling density.

But note that there is a price to pay: the grating must be extremely flat. A piece of the grating with the wrong slope, off by $\Delta\alpha$ from the nominal incidence angle α, will produce diffracted light in the wrong direction for a fixed λ, thereby broadening the outgoing diffracted beam, which implies a decrease in resolving power. A change in α, to $\alpha + \Delta\alpha$, produces a different dispersion angle, β', with $\cos\beta' = \cos(\alpha + \Delta\alpha) + m\lambda/d$, but β' refers to the *rotated* plane, not the nominal plane of the grating (Figure 6). To get the change in β with respect to the nominal dispersion direction, you have to add $\Delta\alpha$: $\Delta\beta = \beta' + \Delta\alpha - \beta$. Using the dispersion relation, you find $\Delta\beta = (1 + \sin\alpha/\sin\beta)\Delta\alpha$, and the apparent wavelength shift is $\Delta\lambda = d(\sin\alpha + \sin\beta)\Delta\alpha/m$. To get an idea, use $\lambda = 15$Å, $d = 15,000$ Å (666 lines/mm), $\alpha = 1$ degree. To keep $\Delta\lambda < 0.03$ Å(or resolving power 500), you need $\Delta\alpha \lesssim 6$ arcsec. Similarly, when using an array of gratings to cover the telescope beam, you need to align the gratings with respect to each other to similar precision.

The other obvious disadvantage of having to use very small angles of incidence on the grating is that its area projected to the incoming beam is

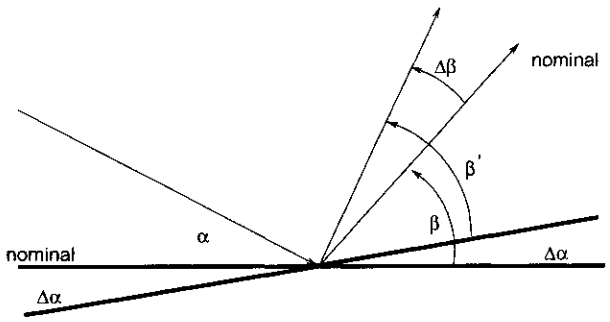

Fig. 6. A small piece of a reflection grating has the wrong orientation, and the incoming beam makes an angle $\alpha + \Delta\alpha$ with the local grating plane. The ray is dispersed to β' according to the dispersion equation, but with respect to the tilted plane. The change in angle with respect to the nominal dispersion angle β is $\Delta\beta = \beta' + \Delta\alpha - \beta$.

very small: you either have low throughput, or else you have to make many gratings to cover the telescope beam.

X-ray reflection gratings will be flown for the first time on *XMM*, and we will describe the Reflection Grating Spectrometer in more detail later. The Extreme Ultraviolet Explorer (*EUVE*) carried three reflection gratings, which provided spectra longward of $\lambda 70$Å (Bowyer & Malina 1996). A reflection grating spectrometer is also currently being considered for NASA's fleet of spectroscopic X-ray observatories *Constellation-X* (most recent information: `http://constellation.gsfc.nasa.gov`), to provide high resolving power in the soft X-ray band.

Crystal spectrometers The most venerable of all X-ray spectrometers is the crystal spectrometer, which was used by Friedrich, Knipping, and von Laue in 1912 to demonstrate the diffraction of X rays by periodic structures, proving conclusively that X rays are electromagnetic waves [for a historical account, read Compton & Allison (1935), pp. 20-38]. The reason this works is the coincidence that typical crystal lattice spacings are of order 1 Å, of the same order of magnitude as X-ray wavelengths.

Referring to Figure 7, it is straightforward to show that constructive interference between rays reflected off two planes spaced by the crystal lattice spacing d will occur if

$$2d \sin\theta = m\lambda, \tag{10}$$

the *Bragg condition*. Note that this is not true dispersion: for a chosen Bragg angle θ, only certain wavelengths are reflected: λ, $\lambda/2$ (in second order), etc. The crystal acts more like a narrow-bandpass interference filter, whereas a grating will simultaneously diffract all wavelengths, so you can record a

complete spectrum all at once. To obtain a spectrum in a finite-width band, a crystal spectrometer has to scan through a range of Bragg angles.

The resolution of a crystal spectrometer is

$$m\Delta\lambda = 2d\cos\theta\Delta\theta, \tag{11}$$

with $\Delta\theta$ the combined angular spread in the Bragg angle due to the finite angular resolution of the incoming beam, convolved with misorientations of different pieces of the crystal, and the effects on the diffraction pattern of absorption and dispersion within the crystal. The combined contributions from the crystal are referred to as the 'rocking curve'. If you illuminate the crystal with monochromatic light, and slowly rotate ('rock') the crystal through the Bragg condition, you will see the intensity sharply peak around the Bragg angle. The finite width of the intensity vs. angle curve is partly due to the fact that the crystal is not perfect, and the narrower this 'rocking curve', the higher the resolution of the crystal. With good crystals, very high resolving powers can be achieved ($\mathcal{R} \gtrsim$ several thousand). Crystal spectrometers are also 'constant $\Delta\lambda$' spectrometers, for a small range in θ.

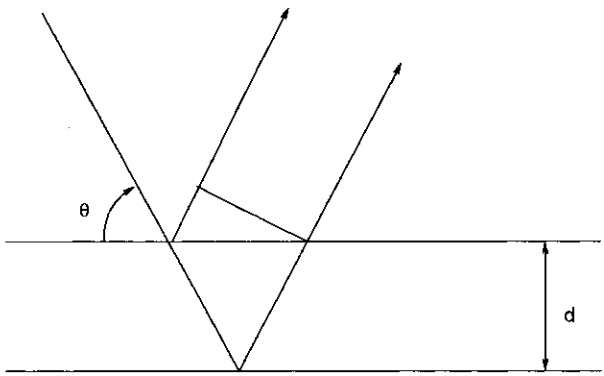

Fig. 7. Geometry for constructive interference between waves diffracted by successive crystal planes, with distance d.

There have been a small number of crystal spectrometers on rockets, and the Focal Plane Crystal Spectrometer (FPCS) on *Einstein* provided the highest resolution X-ray spectra of cosmic sources to date. The first astrophysical detection of an emission line (O VIII Lyα from hot gas in the supernova remnant Puppis A) was in a crystal experiment performed by Zarnecki & Culhane (1978). For a review of the beautiful FPCS results, see Canizares (1990). A crystal spectrometer has been flown on the Shuttle to investigate the emission from the hot ISM in the 44-85 Å range (Diffuse X-ray Spectrometer; Sanders, Edgar & Liedahl 1996). A crystal specrometer of novel design (OXS) will be flown on the *Spectrum X/γ* observatory (see below).

3.2 Non-diffractive spectrometers

Ionization detectors Ionization detectors rely on conversion of the energy of a photon into free electrons, through photoelectric absorption. They include various kinds of proportional counters and solid-state devices. We will briefly recall the properties of CCD detectors only, because the other detectors have too low resolving power to be of interest in the present context. A review of semiconductor detectors can be found in George Fraser's Chapter, and in Fraser (1989).

A photon produces a primary photoelectron, which produces more electrons by collisional processes. There is no amplification in a CCD, so that, very roughly, the number of electrons is

$$N = E/w \tag{12}$$

with E the photon energy, and w the average energy needed to produce one secondary electron (more appropriately, N is the number of electron-hole pairs in the semiconductor). For Silicon, $w = 3.62$ eV. The charge due to a single absorbed photon is collected and measured, and the (apparent) photon energy can be calculated. The energy resolution is set by the statistical fluctuation on N, for a given photon energy. The ionization event also produces other excitations, with their own expectation values and fluctuations. The fluctuation on the number of electron-hole pairs and on the other excitations are correlated, because energy must be conserved in the conversion process, and this implies that the fluctuation on N is smaller than that given by Poisson statistics, $\sigma = N^{1/2}$. This reduction can be described by a 'Fano factor' F, such that the variance in the number of electrons is

$$\sigma^2 = FN \quad (F < 1). \tag{13}$$

For Si, $F \approx 0.1$, and the maximum ('Fano-limited') energy resolution is $\Delta E(\mathrm{FWHM}) = 2.35w(FE/w)^{1/2} \approx 45(E/1\ \mathrm{keV})^{1/2}$ eV (for a Gaussian distribution, FWHM $= 2.35\sigma$). This goes like $E^{1/2}$, so a CCD is an 'almost-constant-ΔE' detector—$E^{1/2}$ varies by a factor ~ 3 between 1 and 10 keV, and other, small sources of noise tend to weaken the dependence of resolution on energy even further.

The first CCD's to be used extensively for X-ray astrophysics are the detectors in the Solid-state Imaging Spectrometers on *ASCA* (Tanaka, Inoue & Holt 1994). Jelle Kaastra's chapter provides a comprehensive overview of results obtained with these instruments.

Superconducting tunneling junctions The fundamental idea behind STJ's is to use a physical process with a very small w, in order to create a large number N of 'countable objects', with correspondingly small statistical fluctuations. Instead of photoionization, the STJ relies on the breaking of Cooper pairs in a superconductor, which have a binding energy of order 10^{-3}

eV, instead of the $\sim$ 1 eV atomic binding energies that are characteristic of photoionization.

The general mood among experimenters, however, seems to be one of pessimism that the early promises of these detectors will be realized in practical high-resolution devices, at least in the near future, for a variety of technical reasons. STJ's still have great potential as superb photon-counting detectors for IR/optical/UV applications (Rando et al. 1996).

Microcalorimeters *Very* loosely speaking, a microcalorimeter counts not photoelectrons or broken Cooper pairs, but phonons—heat. X rays are absorbed by a tiny cooled sensor with very low heat capacity, and the resulting rise in temperature is sensed with a thermometer. The limiting energy resolution corresponds to the fundamental statistical fluctuations in the total energy of the sensor volume, and can be made almost arbitrarily small by reducing the heat capacity. The expectation is that with high-sensitivity thermometers and clever electronic readout schemes, microcalorimeters can be made to deliver $\Delta E \sim 2$ eV (approximately constant), within the next few years.

The microcalorimeter to be flown on *Astro-E* will have $\Delta E \approx 12$ eV. We will come back to this instrument in a later section.

One astrophysical spectrum has already been obtained with a microcalorimeter. A detector flown on a rocket by the Wisconsin/NASA Goddard collaboration has recently detected (expected) emission lines in the spectrum of the hot ($\sim 10^6$ K) phase of the interstellar medium (Deiker et al. 1997). Microcalorimeters are also planned for NASA's future *Constellation-X* observatories, and for the European observatory *XEUS* (Turner et al. 1996).

3.3 Comparison with astrophysically significant resolving powers

Let us choose some representative parameters for each type of spectrometer, and overlay the resulting resolving power on the resolving powers required for the diagnostics discussed in the previous section. This is a crude procedure, in the sense that sensitivity to a particular effect is not just a function of resolving power, but also of signal-to-noise, crowdedness of the spectral region of interest (confusion), etc., but it will at least allow us to establish a broad overview.

For a Fano-limited CCD we had $\Delta E \approx 45(E/1\ \mathrm{keV})^{1/2}$ eV, so $\mathcal{R} \approx 22(E/1\ \mathrm{keV})^{1/2}$ (see Figure 8). The microcalorimeter on *Astro-E* has $\Delta E \approx 12$ eV, so $\mathcal{R} \approx 83(E/1\ \mathrm{keV})$. The 'microcalorimeter of the future' (*Constellation X* and *XEUS*) may have $\Delta E \approx 2$ eV, so $\mathcal{R} \approx 500(E/1\ \mathrm{keV})$. With a telescope response of $\Delta\theta \approx 1$ arcsec, and a line density of 5000 lines/mm, a transmission grating spectrometer has $\Delta\lambda = d\Delta\theta \approx 0.01$ Å, so $\mathcal{R} = 1240(E/1\ \mathrm{keV})^{-1}$. With a line density of 1000 lines/mm, the resolving power is $\mathcal{R} = 250(E/1\ \mathrm{keV})^{-1}$. These numbers are roughly representative of

the two grating spectrometers on *AXAF*. We will discuss the performance of these instruments, and of the Reflection Grating Spectrometer on *XMM* in more detail later. Finally, a crystal spectrometer could reach $\mathcal{R} \sim$ several thousand, but given the fact that each crystal can only scan a fairly narrow band (the arrangements on a satellite don't allow large, variable ranges of Bragg angles), an extensive array of crystals with different d-spacings would be needed to cover a wide energy band. In practice, crystals are therefore targeted for a certain narrow band of interest.

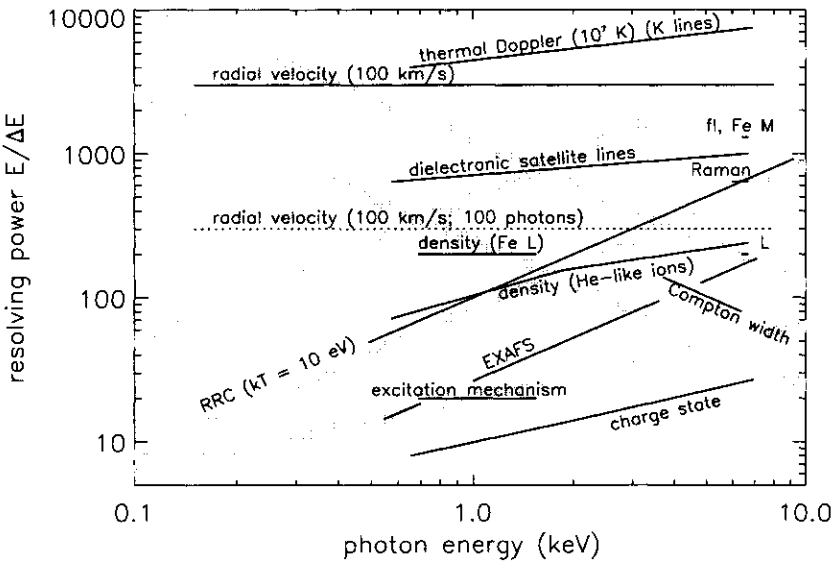

Fig. 8. Same as Figure 3, but with roughly representative instrumental resolving powers overlaid (gray bands). The lowest band corresponds to the performance of a Fano-limited CCD, the two curves rising with energy correspond to a $\Delta E = 12$ eV and a 2 eV microcalorimeter, and the two curves falling with energy correspond to a low- and high-dispersion transmission grating. All spectroscopic diagnostics below a given gray band are accessible in principle to the corresponding instrument.

As you can see, there is not a single type of instrument that covers the entire 0.1 – 10 keV band with uniformly high resolving power. Diffractive and non-diffractive spectrometers 'cross over' in $\mathcal{R}$ in the range around 1 keV, and the crossover is likely to remain around those energies to within a factor ~ 2, for a variety of practical reasons. On a fundamental level, the competition is between microcalorimeters and diffractive spectrometers, because the concept underlying either type of instrument does not contain a

fundamental energy scale (like the ionization energy of the absorbing material). Decisions on what type of instrument to use will therefore be based on practical considerations (need for large numbers of diffracting elements vs. need for cryogenic equipment, etc.). But note that microcalorimeters, by their nature, in principle also allow for imaging spectrometers, a concept that does not come natural to diffractive spectrometers.

As a final remark, note that for any finite redshift z of the sources of interest, the loci of the spectroscopic diagnostics in Fig. 8 shift to the left by a factor $1 + z$, and at significant redshifts $z \gtrsim 1$, this would emphasize the low-energy performance of any given spectrometer as its crucial characteristic. The result would be a complicated set of tradeoffs. At lower energies, it is easier to manufacture high resolving power diffractive spectrometers. On the other hand, some of the most interesting objects to study at such redshifts are clusters of galaxies, whose finite angular extent would complicate the problem (requires either large dispersion angles with diffractive spectrometers to preserve spectral resolution, or imaging *arrays* of non-diffractive spectrometers, coupled with good low-energy response).

3.4 The Rowland circle

In this section, I want to briefly discuss the concept of the Rowland circle. It is integral to the design of modern diffractive spectrometers (though not fundamental to the concept underlying their operation), so we will save time when we discuss the grating spectrometers on *AXAF* and *XMM*. The reason I am putting it in here at all is that you will not usually find the idea explained in modern texts [but see Michette (1986) for an algebraic derivation], and it is tricky enough to cause confusion.

Nineteenth-century spectroscopists were faced with the problem that you can not get the full resolving power from a flat grating (usually a reflection grating in this case) if it is not illuminated with collimated light (plane parallel waves). In order to collimate the light from a source at a finite distance away, you either need lenses or mirrors, or a system of narrow slits. The first cause large losses of light, especially at short wavelengths, while the latter are also an obvious waste of precious light in the instrument. After being diffracted by the grating, the beam has to be refocused onto a detector, again at the expense of light. The problem is how to design a spectrometer that will work with a diverging beam of light. The idea is to use a curved grating that automatically corrects for the aberrations associated with the diverging beam, and also 'refocuses' the dispersed light onto the detector. Of the various possible solutions proposed, the one devised by Rowland is the best known.

The idea is the following (see Figure 9). A source at S illuminates a grating, placed at a distance L. A ray of wavelength λ strikes the grating at A and is dispersed into an angle θ. Place a detector at F, on a circle of radius $R = L/2$ (the Rowland circle). You can show that a different ray of the same

wavelength and spectral order, which strikes the grating at a different place (A') will intersect with the first ray *almost* at F, provided (1) the grating has a radius of curvature $L = 2R$, (2) the grating period is constant along a plane tangent to the grating at the apex A. The residual aberrations are of order l^2/R^2 (l the length of the grating), *and are to first order independent of the wavelength*. The proof of these statements is messy; you can find it, for instance, in the massive classical review of grating spectrometers by Stroke (1967). You can appreciate that the aberrations are small as long as l/R is small, if you mentally tilt triangle SAF around S. Point A describes a circle of radius $2R$ centered on S, which should be the grating surface, and triangle $SA'F$ is almost congruent with triangle SAF as long as the tilt angle remains small.

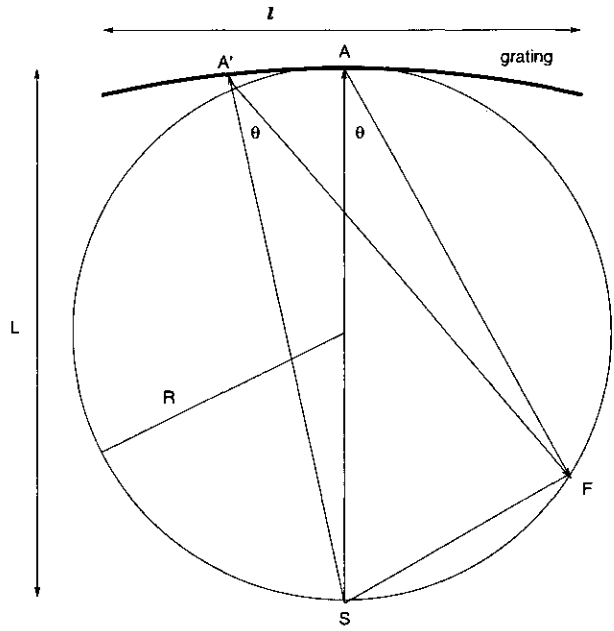

Fig. 9. Schematic of the Rowland circle geometry. A source is placed at S, rays of arbitrary wavelength strike the grating at A and A', and come to a common focus at F, if the grating has radius of curvature $L = 2R$, and the source, grating, and focus are placed on a circle of radius R.

You could reduce or eliminate the residual aberrations further, by choosing a clever line space variation along the grating face, or a more complicated geometry than the circles in Figure 9. But such improvements will turn out to only work for one particular wavelength and spectral order. The importance of the compromise Rowland circle geometry is that it is independent of

wavelength, so that a focused spectrum at *all* wavelengths is obtained simultanously. The focus is not perfect, but since the aberrations scale like l^2/R^2, you can limit the resulting blurring by scaling up R for given l (at the obvious expense of loss of light—the larger R, the smaller the fraction of light from S that will strike the grating).

It is now clear how to incorporate these ideas into the design of X-ray spectrometers. Figure 10 gives the equivalent Rowland geometries for both transmission and reflection grating spectrometers, with the gratings placed behind a grazing incidence focusing X-ray telescope —by reversing the direction of the rays SA and SA' in Fig. 9.

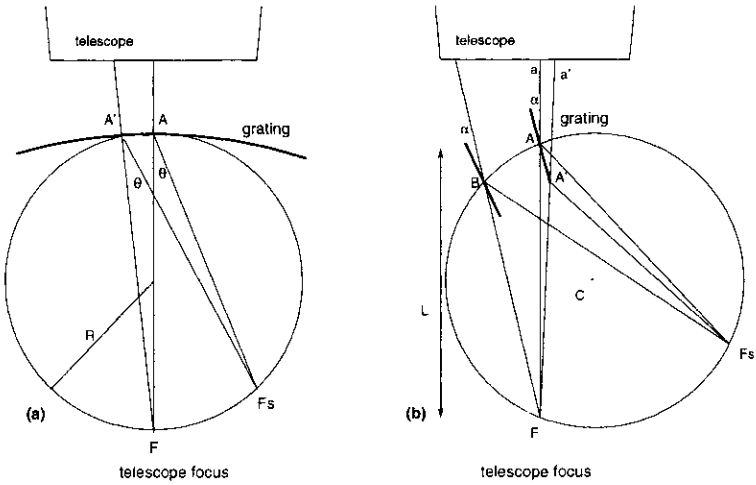

Fig. 10. Relation of the Rowland circle geometries for X-ray transmission grating spectrometers *(a)* and reflection grating spectrometers *(b)*, to the classical Rowland geometry of Figure 9.

Figure 10(a) displays the transmission grating case. Rays from the telescope converge on the telescope focus F (the equivalent of the source in the classical Rowland geometry), the grating bars are perpendicular to the plane of the page, and dispersion is in the plane of the page. Again, the grating and the focus are placed on a circle of radius R, and the grating has radius of curvature $2R$; the optimum spectroscopic focus F_s appears on the circle, for arbitrary wavelength and spectral order.

Figure 10(b) shows the equivalent geometry for the reflection grating case. The grating grooves are perpendicular to the plane of the page. Ray a strikes the grating a A, and is dispersed towards the spectroscopic focus F_s. Ray a' strikes at A', and, just like in the classical geometry, the grating has to be curved in order to make both rays converge on a common spectroscopic

focus. Instead of using a physically curved grating, it is easier to introduce a variation in the grating period along its face, which has the same effect (Hettrick & Bowyer 1983). The telescope and spectroscopic foci, F and F_s, and the center of the grating, A, are again placed on a Rowland circle, but the radius of the circle now depends on the chosen line density gradient—the smaller the line density gradient, the further away F_s will be from A, and the larger the radius of the Rowland circle. The optimum-focus spectrum again appears along the circle. In order to cover the focused telescope beam, other gratings have to be placed with their centers on the same circle, like the grating at B, at the same angle of incidence α for the focused ray passing through their centers. In principle, each grating should have its own unique line density gradient (because each grating has its own unique distance to F_s), but in practice the aberrations resulting from having identical gratings are small (as long as typical distances AB are small compared to AF_s), and are typically less important than the effect of the finite angular resolution of the telescope, and the finite accuracy of alignment of the gratings with respect to each other.

Finally, note that all Rowland circle grating spectrometers are intrinsically astigmatic (i.e. rays from a monochromatic point source do not pass through a single point in the focal plane, but instead have different foci in the dispersion, and in the cross-dispersion directions). The telescope has a finite extent out of the plane of the paper, and the telescope beam is 'filled' with grating elements by 'rocking' the Rowland circle back and forth around an axis passing through the telescope focus, lying in the plane of the paper, perpendicular to the optical axis. According to Fermat's principle, the telescope focuses on a cirle centered at A, of radius equal to the length of AF. Focused rays traveling in different planes rotated around F out of the plane of the paper will therefore converge in the *cross-dispersion* direction on that circle. As you can see from Fig. 10(a), that circle is always outside the Rowland circle for a transmission grating spectrometer (the two circles intersect at the telescope focus), which implies that light will focus in the dispersion direction on the Rowland circle, but is defocused in the cross-dispersion direction. The monochromatic image of a point source appears at F_s as a 'stripe' perpendicular to the dispersion direction. Similarly, a circle of radius $|AF|$ centered on A will not coincide with the Rowland circle for the reflection grating case either [Fig 10(b)], so the reflection grating spectrometer is also astigmatic.

This astigmatism does not affect the resolving power of the spectrometer, but the actual two-dimensional size of the image does determine the detector and diffuse sky background level in a resolution element, so you want to keep the astigmatism to a minimum, if possible.

4 The High Resolution X-ray Spectrometers on *AXAF*

4.1 Introduction

NASA's *Advanced X-ray Astrophysics Facility* (*AXAF*) is currently scheduled for launch on December 3, 1998. It is built around a high-resolution grazing-incidence telescope. The overal angular resolution of the telescope, across the X-ray band up to 10 keV ($\lesssim$ 1 arcsec FWHM, driven by the desire to explicitly resolve the point source contribution to the 2-10 keV diffuse X-ray background), implies superb high-resolution imaging and, with transmission gratings placed in the focused X-ray beam, high-resolution spectroscopy. In the following, we will briefly look at the various instruments on *AXAF*, and then discuss the grating spectrometers in detail.

Scientific operation of the observatory is the responsibility of the *AXAF* Science Center in Cambridge, MA (`http://asc.harvard.edu`). A good source of information on the observatory and instruments is the *AXAF Observatory Guide* (`http://asc.harvard.edu/USG/docs/docs.html`). An older standard reference is Weisskopf et al. (1987).

The heart of the observatory is the High Resolution Mirror Assembly (HRMA; Telescope Scientist: Leon van Speybroeck, Smithsonian Astrophysical Observatory). It consists of 4 pairs of iridium-coated paraboloid- hyperboloid shells in a Wolter I configuration. The focal length of the telescope is 10.066 m. The figure and the surface smoothness of the mirror shells (glass) were very accurately controlled to ensure the very high angular resolution and high-quality focus.

There are four separate detectors in the focal plane, two designed for imaging observations, and two designed for reading out the spectra obtained with the gratings. One imaging and one spectroscopic detector consist of CCD's [the *AXAF* CCD Imaging Spectrometer detectors, ACIS-I (imaging), and ACIS-S (Spectroscopy), Instrument PI Gordon Garmire, Pennsylvania State University], and the other imaging and spectroscopic detector are microchannel plate detectors [the High Resolution Camera detectors HRC-I (imaging) and HRC-S (spectroscopy), Instrument PI Stephen Murray (SAO)]. All four instruments can be moved into the telescope focus by means of a mechanism for lateral motion. In addition, the detectors can be moved along the optical axis of the telescope to optimize the focusing (the focal depth of the telescope is only 200 micron! and a small change in the length of the 10 m telescope tube, if uncorrected, would easily defocus the system). A schematic of the focal plane is shown in Figure 11. The full field of view of ACIS-I is 16.9 $\times$ 16.9 arcmin, of HRC-I is 31 $\times$ 31 arcmin.

The ACIS-I and HRC-I cameras were designed with different goals in mind. ACIS obviously offers spatially resolved CCD spectroscopy, whereas the HRC-I offers higher spatial resolution (ACIS slightly undersamples the telescope response), large field of view, high time resolution, and an extended

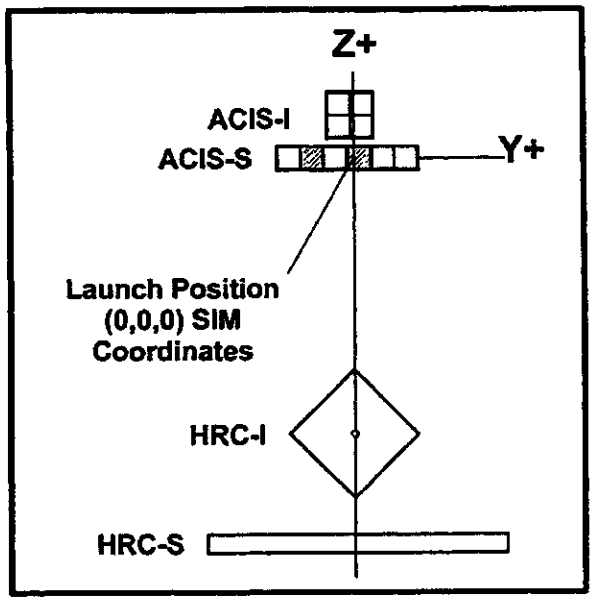

Fig. 11. Schematic of the *AXAF* focal plane, looking down on the focal plane from the position of the mirrors. The SIM (Scientific Instrument Module) can be moved along the vertical line; the focus can be moved in the perpendicular direction by offset-pointing the entire telescope. The dispersion direction for the two gratings is the $\pm Y$ direction (source: *AXAF Observatory Guide*, Ch. 1).

sensitivity to soft photons (down to $\sim$ 100 eV), but no intrinsic energy resolution.

The ACIS-S camera was specifically designed to read out the spectra obtained with the High Energy Transmission Grating (HETG) in the beam, while the HRC-S was designed specifically for use with the Low Energy Transmission Grating (LETG), although either detector can be used with either grating for specific non-standard applications. For extensive information on the detectors and their predicted performance in combination with the HRMA, I refer you to the *AXAF Observatory Guide.*

Either one of two sets of transmission gratings can be rotated into the focused X-ray beam, behind the HRMA, and together with the HRMA and the focal-plane detectors, these make up the High Energy Transmission Grating Spectrometer [HETGS, Instrument PI Claude Canizares (MIT)], and the Low Energy Transmission Grating Spectrometer [LETGS, Instrument PI Albert Brinkman (SRON/Utrecht)]. These will be the first astrophysical *truly* high-resolution X-ray spectrometers, and from even just a cursory glance at Figure 8, you can see what a tremendous increase in sensitivity to diagnostic

physical effects they represent. Both grating spectrometers take advantage of the very high angular resolution of the mirrors to attain high spectral resolution, even at moderate dispersion angles.

In the following we will discuss both spectrometers, work out the efficiency of a transmission grating and how it can be optimized for a chosen range of photon energies, and calculate the effect of random fluctuations in the grating properties on the performance.

4.2 The High Energy Transmission Grating Spectrometer

The HETGS (Canizares et al. 1987; Markert et al. 1994; and `http://space.mit.edu/HETG`) consists of two different sets of gratings, with different periods. These form the High and the Medium Energy Transmission Gratings (HETG and METG). Due to the finite size of the ACIS-S detector, the HETG's first-order bandpass is limited to photon energies $E > 800$ eV. The METG has lower dispersion and complements the HETG down to 400 eV. The gratings actually consist of small rectangular flat elements arranged on four separate annuli, covering the four hollow-cone shaped focused beams emerging from the HRMA's four mirror shells. The HETG gratings are arranged behind the inner two shells, the METG gratings behind the outer two shells, because the inner shells have higher throughput at the highest photon energies (smaller diameter, hence smaller graze angle [for fixed focal length], which implies high reflectivity up to higher photon energies). This arrangement is shown in Figure 12.

The gratings are mounted in a Rowland configuration slightly different from the one shown in Figure 10(a): all elements are placed on the Rowland circle itself (diameter 8633.69 mm), perpendicular to the focused rays, instead of on a surface of radius equal to twice the Rowland circle radius. As long as the diameter of the Rowland circle is large compared to the diameter of the grating (approximately 1000 mm), and the grating elements are kept small, the aberrations arising from the approximation of the toroidal surface by finite, flat elements of constant period are small (Beuermann, Bräuniger & Trümper 1978).

The gratings themselves are made of gold, have rectangular bars, and are supported on thin polyimide films. The period is 2000.81 Å (HETG) and 4001.41 Å (METG). The bar widths are 1200 Å (HETG) and 2080 Å (METG).

At high photon energies, the photoelectric absorption coefficient of gold drops approximately as $E^{-3.5}$, and eventually any grating becomes transparent to the radiation, and doesn't diffract at high energies. The optical depth at wavelength $\lambda \equiv 2\pi/k$ through a gold bar of thickness z_0 is $\tau = k\gamma z_0 \approx (z_0/5000\ \text{Å})(E/3\ \text{keV})^{-2.5}$ for $E \gtrsim 2$ keV. Here, γ is the imaginary part of the complex index of refraction, $n \equiv 1 - \delta + i\gamma$ [tables of optical constants as a function of photon energy for all elements can be found at `http://xray.uu.se/hypertext/henke.html`, which contains the tables

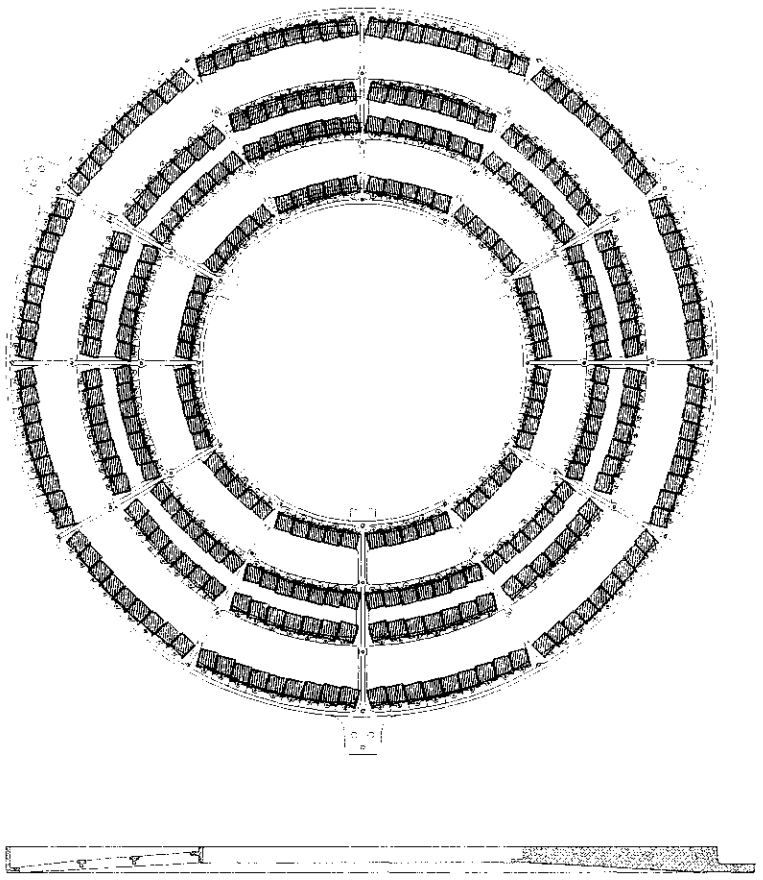

Fig. 12. The High Energy Transmission Grating assembly. The four annuli cover the beams from the four mirror shells; the high dispersion gratings are on the inner two annuli, the low dispersion gratings on the outer two shells (source: *AXAF Observatory Guide*, Ch. 7).

published by Henke, Gullikson & Davis (1993) in digital form]. The HETG has a bar thickness $z_0 = 5100$ Å, so it starts to become optically thin above 3 keV. Fortunately, the real part of the index is still non-zero in this regime, so in addition to small attenuation, the radiation suffers a significant phase shift on passing through the bars. Interference between this phase-shifted part of the wavefronts with those parts that passed through the slits is considerable, leading to significant diffraction efficiency. In this regime, the grating operates effectively as a phased array. The phase shift through the bars depends on wavelength as $\Delta\phi = kz_0 \cdot Re(n-1) = kz_0\delta$, so by optimizing the bar thickness you can optimize the diffraction efficiency for a chosen energy band, as we will explicitly calculate in the next section.

The requirements of high dispersion (small period), and significant phase shift at energies up to 10 keV, leads to the extreme aspect ratio of $\sim 4/1$ of the HETG grating bars. The manufacture of these novel gratings by a photolithographic process is described by Schattenburg et al. (1994). Laboratory efficiency measurements (which cleverly use the interference properties of the large aspect ratio of the grating bars, as a function of angle of incidence) are described by Nelson et al. (1994). Preliminary results on the ground calibration of the HETGS are given by Dewey et al. (1997) and Marshall et al. (1997).

As discussed above, the HETGS was designed to take advantage of the very high angular resolution of the HRMA, so its resolving power in first order is very nearly given by the expression we used to produce Figure 8. The actual resolving power, as predicted from the ground calibrations, is shown in Figure 13. At large dispersion angles, the effect of small grating-to-grating variations Δd in grating period slightly widen the resolution. From the dispersion relation, we have $\Delta\lambda/\lambda = \Delta d/d$, so this effect grows linearly with wavelength. At short wavelengths (small dispersion angles), on the other hand, slight errors in the telescope attitude reconstruction and focusing contribute to the width of the spectral image.

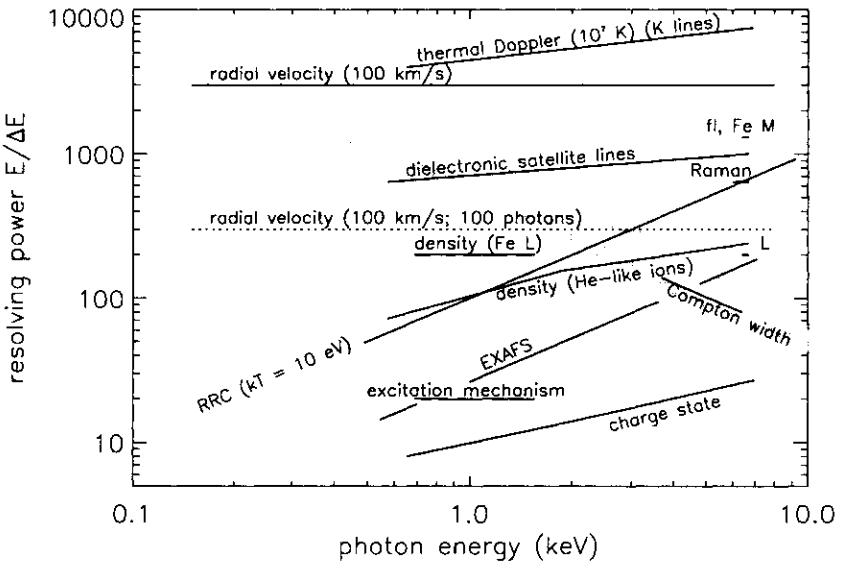

Fig. 13. Same as Figure 8, but with the predicted resolving power for the *AXAF* HETGS overlaid (gray bands). The upper gray band corresponds to the HETG, the lower band to the METG (source: *AXAF Observatory Guide*, Ch. 7).

The spectrum can be read out by either the ACIS-S or HRC-S detector, but the HETGS was designed for use with ACIS-S. The six individual chips that make up ACIS-S are 'folded around' the Rowland circle. Four chips are conventional 'front-illuminated' (FI) devices, two are back-illuminated (BI), which gives them higher quantum efficiency at low energies (but poorer energy resolution). One BI chip was placed at the on-axis position of the zero order, one in the middle of the spectral range on one side. The dispersion directions of the METG and HETG are offset from each other by a small angle, so that the two spectra (high dispersion and low dispersion) form a shallow 'X' on the detector.

From the dispersion equation you conclude that different spectral orders will overlap in the spectral image: the first-order position for wavelength λ coincides with the mth order for wavelengths λ/m. These higher orders 'contaminate' the first order, but they also contain interesting information, because they have resolving powers m times higher than first order (the higher orders are faint though). Separation, or at least unique identification, of the higher orders is therefore desirable, and for this, the intrinsic energy resolution of ACIS-S can be used. The CCD resolution is good enough to uniquely assign events in the focal plane to the correct spectral order to which they belong. This is illustrated in Figure 14, which shows a ground calibration spectrum obtained at the *AXAF* X-ray Calibration Facility (XRCF) at Marshall Space Flight Center. The calibration spectrum has been summed over the cross-dispersion direction, and each event has been plotted in a diagram of CCD event pulse height vs. position along the dispersion direction. For any given position along the dispersion direction, photons belonging to mth order dispersed light of wavelength λ/m are offset vertically with respect to each other. Conversely, the different spectral orders for a given wavelength are offset horizontally with respect to each other. The resulting curved spectra are the different spectral orders of the same incident spectrum. The CCD therefore acts in the same way as the 'cross-disperser' of an echelle grating.

In addition, the pulse height-dispersion plane nicely visualizes the effects of 'pileup' of events in the CCD. If two or more photons arrive in neighboring pixels during a single CCD frame exposure time, the so-called event reconstruction algorithm combines them into a single event in the output, with the sum of the individual charges assigned to this single, 'reconstructed' event. The reason the event reconstruction has to be applied to the CCD frames is that charge from individual photons may physically leak to neighboring pixels, and if this is not corrected for (each pixel treated as a separate event), the energy resolution of the CCD is degraded, and a fraction of fake events with low apparent charges is generated. The unavoidable side effect, though, is the unwanted 'pileup' of true single pixel photons referred to above. But since pileup is likely to occur only in intense parts of the spectrum (emission lines), the piled-up events will appear distinctively patterned and easily recognizeable in the pulse height-dispersion plane, as a series of dots offset

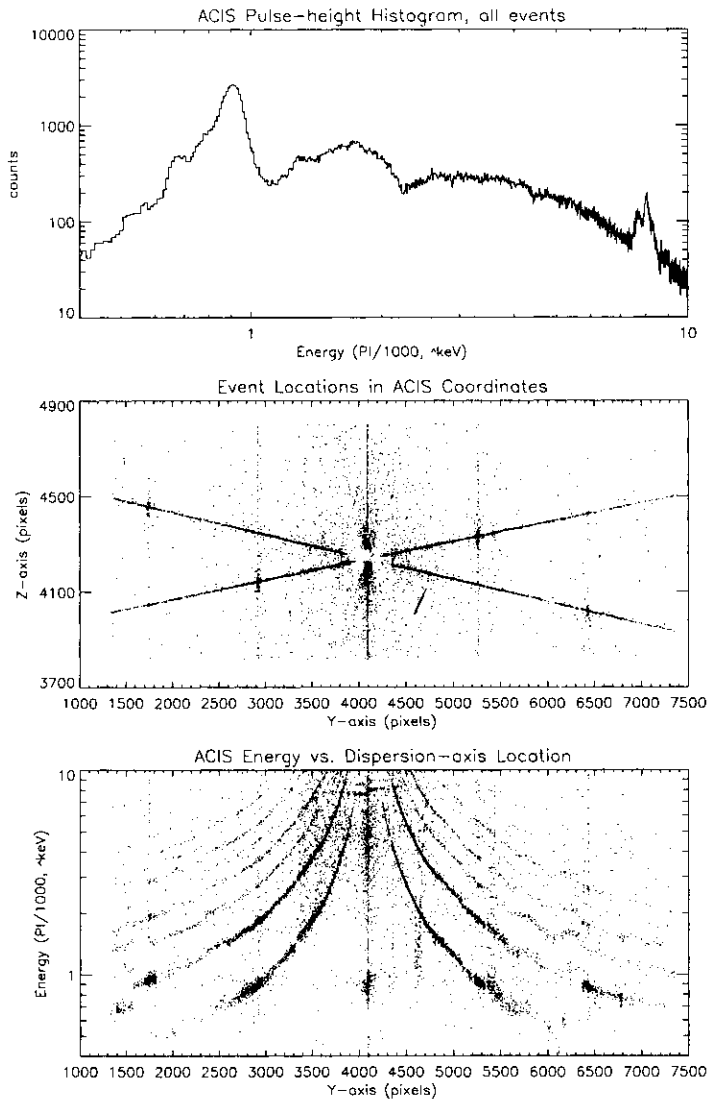

Fig. 14. HETGS spectrum of a Cu target bombarded by electrons, consisting of a bremsstrahlung continuum and fluorescent Cu L and K lines (0.93 and 8.0 keV, respectively). The top panel shows the CCD spectrum for all events combined, the middle panel the HETGS spectral image (you can see the 'X' shape formed by the HETG and METG spectra). The dark spots are the emission line photons (the vertical lines are artefacts from photons arriving at the detector during a 'frame transfer' of the CCD). The bottom panel shows the location of all events on the CCD pulse height vs. dispersion angle plane. The different spectral orders are clearly separated in the pulse height direction. The bright zero-order image has been suppressed (source: *AXAF Observatory Guide*).

in the vertical direction, at one, two, three, etc. times the true single photon charge. In actual flight operation of the HETGS, this effect is expected to be small, because of the low expected photon rates.

The pulse height-dispersion plot also clearly separates dispersed and non-dispersed (background) photons. Non-dispersed photons are sky and intrinsic detector background, and light scattered by the telescope (and possibly by the gratings). In other words, by using the CCD pulse height information, you can reduce the background in the spectrum by requiring that the pulse height and the dispersion angle match, for a given wavelength. Any events that don't match are background, and are smoothly distributed in between the orders in Figure 14.

For accurate quantitative analysis of actual spectra, it will probably be a good idea to take advantage of these powerful features of the HETGS, and perform both data analysis and spectral simulation in this 2D pulse height-dispersion space.

All statements so far regarding the resolving power assume that the source is pointlike. If the source has a finite extent, like a cluster or supernova remnant, the angular width $\Delta\theta$ in the expression for the resolving power becomes the angular size of the source, and the resolving power in principle degrades more or less linearly by a factor (source extent/telescope response). This is a severe effect with *AXAF*, because the telescope resolution is so high. For moderately extended sources, however, that radiate primarily in a small number of distinct, well-separated strong emission lines, the spectral image will look like a discreet set of partially overlapping monochromatic images of the source. Under these circumstances, it may be possible to perform spectrally resolved imaging, i.e. imaging in isolated emission lines. Figure 15 shows an example of this idea, the HETGS spectral image of the supernova remnant N132D in the Large Magellanic Cloud.

Let us discuss one more specific example that nicely illustrates the novel spectroscopy that will become available with the HETGS, the predicted 1-10 keV spectrum of the galactic X-ray binary Cygnus X-3. Cyg X-3 is a highly unusual X-ray binary, with properties unique among the galactic population. It is currently suspected to be the only binary we happen to catch in a rapid, short-lived evolutionary phase of massive X-ray binaries, where the compact object rapidly 'spirals into' the extended atmosphere of its massive companion (van den Heuvel 1994). The ultimate outcome of the binary evolution may be the formation of a white dwarf/neutron star binary, or a neutron star/neutron star binary.

Near-infrared spectroscopy of the IR counterpart shows distinctive emission lines from He I and II, characteristic of Wolf-Rayet stars (van Kerkwijk et al. 1992). The compact object is therefore likely accreting from a massive stellar wind, if the primary really is a WR star. Direct evidence for this conjecture was provided by the 1-10 keV spectrum of Cyg X-3 obtained with *ASCA* (Kitamoto et al. 1994). The spectrum shows a rich discrete emission

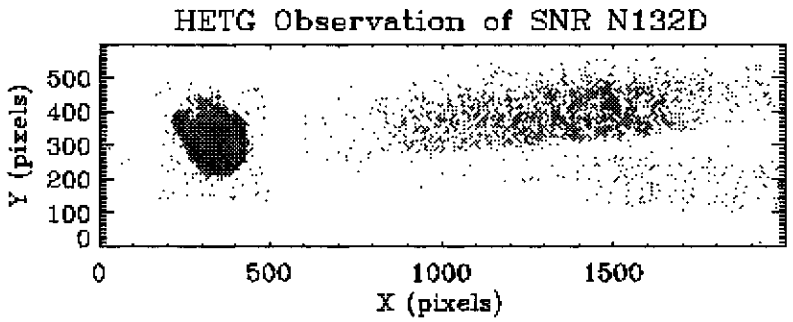

Fig. 15. Partial HETGS spectral image of the SNR 132D in the LMC (about 2 arcmin diameter). The dark image at left is the zero-order image of the remnant (a straight image). The dark band is the METG spectrum (the HETG spectrum below it is very faint because the remnant spectrum is very soft). You can see monochromatic emission line images of the SNR superimposed on the continuum in the MEG spectrum (courtesy of John Houck and the HETGS group, MIT).

spectrum superimposed on a strong, hard continuum which originates near the compact object. The discrete spectrum shows strong line emission from H- and He-like Mg, Si, S, Ar, Ca, and Fe. The spectrum also clearly shows narrow radiative recombination continua (RRC), of which the H-like Si, and especially the H-like S RRC can be seen directly in the data (Liedahl & Paerels 1996). The presence of these narrow features in the spectrum provides unambiguous evidence for the presence of X-ray photoionized gas in the system, probably relatively tenuous material in the WR wind, ionized by the strong central continuum.

With *ASCA*, the RRC's were distinguishable, and possibly just barely resolved, which indicates electron temperatures in the photoionized wind of order $kT_e \lesssim 80$ eV (Liedahl & Paerels 1996; Kawashima & Kitamoto 1996). With the HETGS, these features will be clearly resolved, and a detailed analysis of the spectrum will provide constraints on the conditions in various parts of the ionized stellar wind (density, temperature, abundances, possibly flow velocity). To illustrate this, I show a simulation in Figure 16, obtained by folding the recombination spectra for the He- and H-like ions of Ne, Mg, Si, S, Ar, Ca, and Fe, with values for the free parameters that fit the *ASCA* spectrum, through the HETGS response. The RRC's are clearly resolved, which will give us accurate measurements of the electron temperature in the various ionization zones. The 'triplets' (resonance, intercombination, and forbidden lines) of the He-like ions should also be resolved in principle, up to He-like Ca, which will provide limits on the electron density in the various ionization regions in the wind.

Unfortunately, heavy interstellar and circumsource absorption preclude the detection of low-energy photons from Cyg X-3. A measurement of the abundances of the low-Z elements in the wind would have given us an independent constraint on the low-Z photospheric abundances, and a direct window on the nucleosynthesis and mass loss history of at least this WR binary.

4.3 The diffraction efficiency of an X-ray transmission grating

We will calculate the diffraction efficiency of a transmission grating, show how it is optimized to yield maximum efficiency in a chosen wavelength band, and indicate how the efficiency calibration of the spectrometer is established in terms of a physical model for the instrument.

As long as the wavelength of the radiation is much smaller than the physical size of the diffracting elements (the grating period, for a grating), and at large distance away from the grating, Fraunhofer diffraction applies (Born & Wolf 1959), and the diffraction pattern is simply calculated by Huygens' principle: you just add up the complex phases of waves originating from the various parts of the grating, for a given dispersion angle. It is straightforward to show that the intensity of the angular pattern is

$$I(p,q) = \frac{1}{D^2}\frac{\sin^2((p-q)D/2)}{\sin^2((p-q)d/2)}|f(p,q)|^2 \tag{14}$$

with

$$f(p,q) = a\frac{\sin((p-q)a/2)}{(p-q)a/2}\exp(-i(p-q)a/2) + \int_0^b \mathrm{d}s \exp i\Delta(s). \tag{15}$$

Here, d is the grating period, D is the total width of the grating, b is the width of the grating bars, and a is the width of the slits between the bars. The variables p and q are defined as $p = k\sin\chi$, $q = k\sin\theta$, where χ and θ are the angle of incidence and the dispersion angle, respectively, and k the radiation wave number. The first factor in the expression for $I(p,q)$ represents the rapid modulation due to the large number of periods in the grating. The two terms in the expression for $f(p,q)$ are the contributions of the slits and the bars to the amplitude, respectively. The phase shift $\Delta(s)$ in the contribution of the bars is

$$\Delta(s) = (p-q)s + k(n-1)z(s) \tag{16}$$

with s a coordinate in the plane of the grating, perpendicular to the bars, and $n \equiv 1 - \delta + i\gamma$ is the complex index again. The function $z(s)$ describes the cross-sectional shape of the bars: it is the bar thickness, as a function of s, measured perpendicular to the grating plane.

The efficiency of the grating is defined as the intensity for a given wavelength and spectral order, integrated over an angular region containing the sharp interference peak plus the weak, rapidly modulated 'wings' due to the

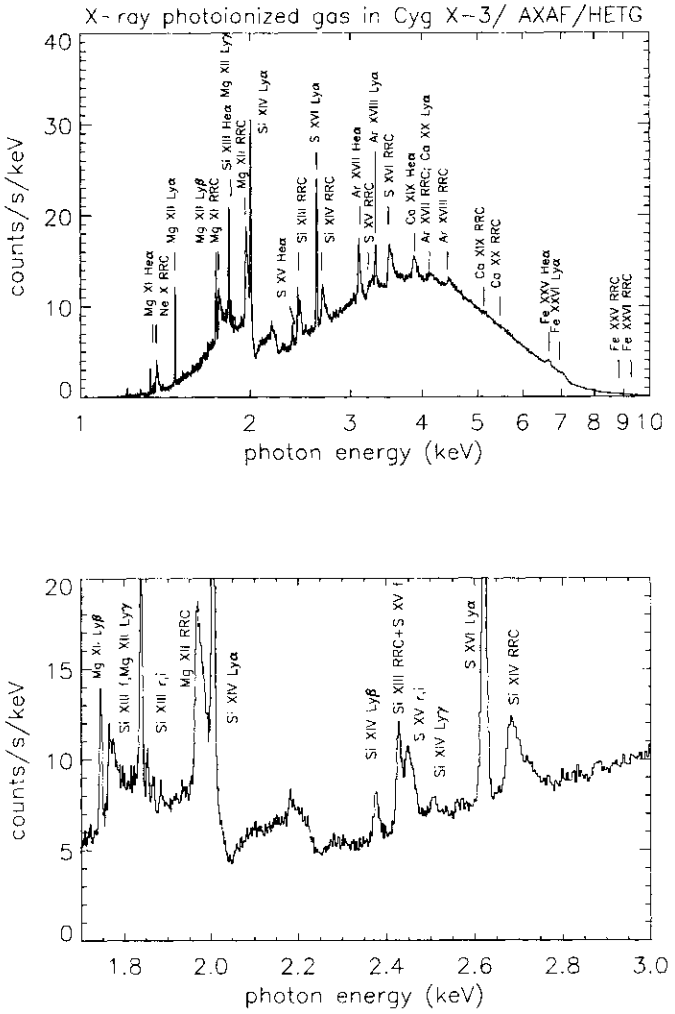

Fig. 16. Simulated spectrum of Cyg X-3, in a 40,000 sec exposure with the HETGS; the HETG spectrum is shown (positive and negative spectral orders summed). Data have been binned in approximately 0.005 Å bins (corresponding to one ACIS pixel). The lower panel shows a blow-up of the 1.8-3.0 keV region.

finite size of the grating, normalized to the intensity of the incoming radiation.

Let us examine the efficiency of the simplest possible grating: square bar cross sectional shape, bar thickness z_0. Inserting the condition for constructive interference, $p - q = -2\pi m/d$, it is easy to show that with the proper normalization, Eq.(14) reduces to

$$\text{for } m = 0\text{: } \eta = \frac{a^2}{d^2}\left\{1 + \frac{b^2}{a^2}e^{-2\gamma k z_0} + 2\frac{b}{a}e^{-\gamma k z_0}\cos\delta k z_0\right\}, \tag{17}$$

$$\text{for } m \neq 0\text{: } \eta = \frac{a^2}{d^2}\frac{\sin^2(m\pi a/d)}{(m\pi a/d)^2}\left\{1 + e^{-2\gamma k z_0} - 2e^{-\gamma k z_0}\cos\delta k z_0\right\} \tag{18}$$

These expressions were first given by Schnopper et al. (1977); note that for $m \neq 0$ they refer to the diffraction efficiency into *one* (i.e., positive or negative) spectral order only.

The efficiency of a grating with rectangular bars has several interesting properties. First, note that for a particular choice of bar width ($a/d = b/d = 1/2$, or bar width equal to slit width), diffraction into the even orders disappears. This is usually a desirable feature—more light is diffracted into the odd (first !) orders, and there is less overlap of orders. Second, the efficiency is a sensitive function of wavelength (through the wavelength dependence of the complex index n) at wavelengths short enough that the bars are partially transparent. This behavior is contained in the factor $\cos\delta k z_0$. By tuning the thickness z_0 of the bars, the efficiency can be enhanced over its value for an opaque grating, for a chosen range of wavelengths. The grating is said to be 'blazed' for that particular wavelength range (the origin of the term 'blazing' will become clearer when we discuss reflection gratings, below). Moreover, the dependence of the efficiency on wavelength, apart from an overall order-dependent constant, is identical for all orders. At long wavelengths, the bars are opaque, and the diffraction efficiency depends only on the bar-to-slit ratio. All these properties are illustrated in Fig.17, where I show the diffraction efficiency for a grating with gold bars, of rectangular cross-sectional shape, slit-to-period ratio 1/2, and two different bar thicknesses (0.35 and 0.60 μm). Notice how the thinner grating has much higher efficiency in the 1-2 keV band, while the thicker grating is (of course) superior above 2 keV. The resonance feature at 100 Å is due to a local minimum in the absorption coefficient of gold at those wavelengths, causing interference between the bars and the slits.

The *AXAF* LETG was optimized to provide enhanced efficiency in the 1-2 keV band, while the thicknesses of the bars of the HETG and METG were chosen to provide optimum efficiency at energies $E \gtrsim 2$ keV.

Things change when the bar cross sectional shape is no longer strictly rectangular (either by design or because of small imperfections in the manufacturing process). Once you realize that the edges of such bars don't have the same optical depth as the centers of the bars (usually smaller—the bar cross

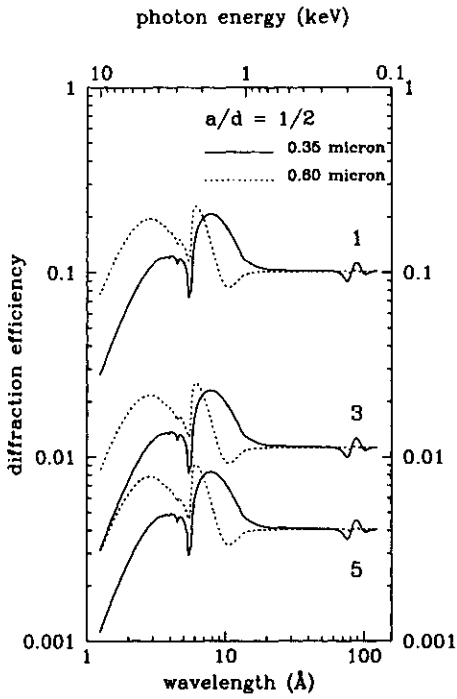

Fig. 17. Diffraction efficiency of a gold transmission grating, of slit-to-period ratio 1/2, rectangular bar cross sectional shape, for two different bar thicknesses.

section may look like a trapezoid, or even an ellipsoid), it is obvious that, for instance, the exact cancellation of the even-order diffraction no longer works for wavelengths at which the bars are partially transparent, even for bar-to-slit ratio equal to unity. In fact, the even order diffraction efficiency becomes extremely sensitive to bar shape and thickness at short wavelengths. Also, the ratio of the efficiencies in the various orders becomes a function of wavelength, even in the odd orders. These properties are illustrated in Fig.18, where I show efficiencies for two gratings with identical parameters (slit-to-period ratio 1/2, bar thickness 0.35 μm), but different bar cross-sectional shape (either rectangular or ellipsoidal). The 'ellipsoidal' grating shows the expected diffraction into the even orders (in fact, the even orders become more efficient than the odd orders at high energies). The higher the order is, the more sensitive is the efficiency to the details of the bar properties.

Finally, if there is a distribution in the properties of the bars across the grating (as will always be the case to some extent), the simple properties of a strictly uniform grating obviously no longer apply. The simplest example is a

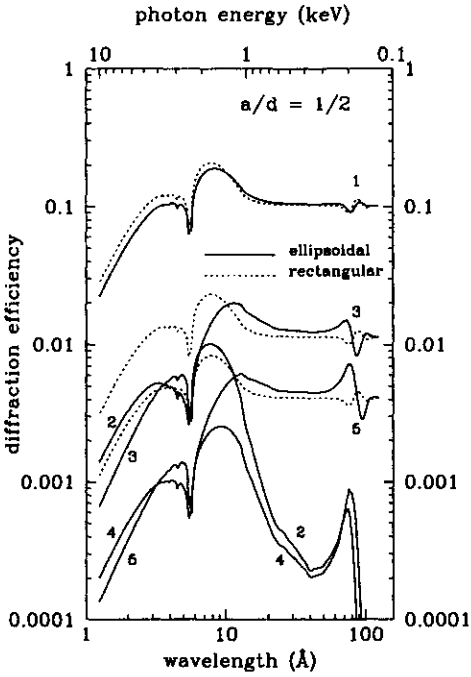

Fig. 18. Diffraction efficiency of a gold transmission grating, of slit-to-period ratio 1/2, bar thickness 0.35 μm, and two different bar cross sectional shapes (rectangular or ellipsoidal). The grating period is 1.0 μm.

grating designed with $a/d = 1/2$, which exhibits finite even-order diffraction due to the fact that the bar-to-slit ratio varies slightly around the mean over the grating.

With these features in mind, you will appreciate the complications associated with trying to uniquely establish the efficiency calibration of an X-ray grating. Calibrating all different parts of the grating, at all wavelengths (and possibly different angles of incidence) is obviously out of the question for practical reasons, so one typically relies on a physical model for the grating efficiency, with a (small) number of free parameters, whose values are constrained by comparing calibration data with model calculations. This is a difficult and not necessarily unique procedure. Past experiments (the gratings on *Einstein* and *EXOSAT*) for practical reasons had to rely almost exclusively on a finite set of efficiency measurements, averaged over the entire grating, at a discrete set of characteristic X-ray wavelengths (Seward et al. 1982; Paerels, Kahn & Wolkovitch 1998), which did not always admit of a

fully consistent, simple and unique model for the grating and it efficiency. The *AXAF* grating spectrometer calibration relies on a much wider range of data, including optical and IR period measurements and synchrotron X-ray efficiency measurements for many or all individual grating elements, and is ultimately expected to produce agreement between data and the best physical model for the gratings on the order of or better than a few percent.

4.4 The Low Energy Transmission Grating Spectrometer

The Low Energy Transmission Grating Spectrometer (LETGS; Brinkman et al. 1987) consists of four annular sets of free-standing gold transmission grating elements, mounted behind the HRMA mirror shells in a Rowland configuration identical to that of the HETGS (cf. Section 4.2). It is designed for use with the HRC-S imaging camera, which consists of three separate microchannel plate detectors, placed tangent to the Rowland circle. The grating period is 9912.5 Å, the bar thickness 0.5μ. The low dispersion, as compared to the HETGS, implies that, given the size of the HRC-S, the LETGS provides spectroscopy out into the EUV band, out to 170 Å. At the lowest photon energies, the LETGS reaches resolving powers in excess of $\mathcal{R} \sim 2000$, the highest resolving power on *AXAF*. The resolving power is shown overlaid on the spectroscopic diagnostics we discussed at the beginning, in Fig.19. The somewhat wiggly appearance of the resolving power curve at the lowest energies is due to the fact that the HRC-S camera elements are flat, and do not precisely follow the Rowland circle, which leads to a small amount of defocusing at some wavelengths.

The LETGS has unique capabilities for performing high-resolution spectroscopy of (very) soft sources (either intrinsically soft or sources with low interstellar absorption). Soft coronal emitters can be studied at high spectral resolution, which may actually allow the detection of bulk velocity fields, or binary motion in short-period active binaries (RS CVn stars). Other prime targets are cataclysmic binaries, white-dwarf stars, and thermal emission from hot neutron stars. The LETGS may also provide unique access to the spectra of the low-Z elements in sources with a significant redshift.

With these goals in mind, the gratings were again (like their predecessors on *Einstein* and *EXOSAT*) designed to be free-standing (no support film), and the design of the HRC-S was optimized with the lowest photon energies in mind. Instead, the grating is supported by two gold support grids, one (the fine support grid) at right angles to the X-ray grating bars, the other (the coarse support) actually consists of three grids at 120 degree angles. These support grids have very large periods, and very large slit-to-period ratios a/d, so while they do disperse light at an angle to the X-ray dispersion direction out of the X-ray spectral image, they do so at very low efficiency.

The actual gratings were produced at the Max Planck Institut für Extraterrestrische Physik in Garching, Germany, under the direction of Peter Predehl. Details on the design, manufacturing, and laboratory calibration,

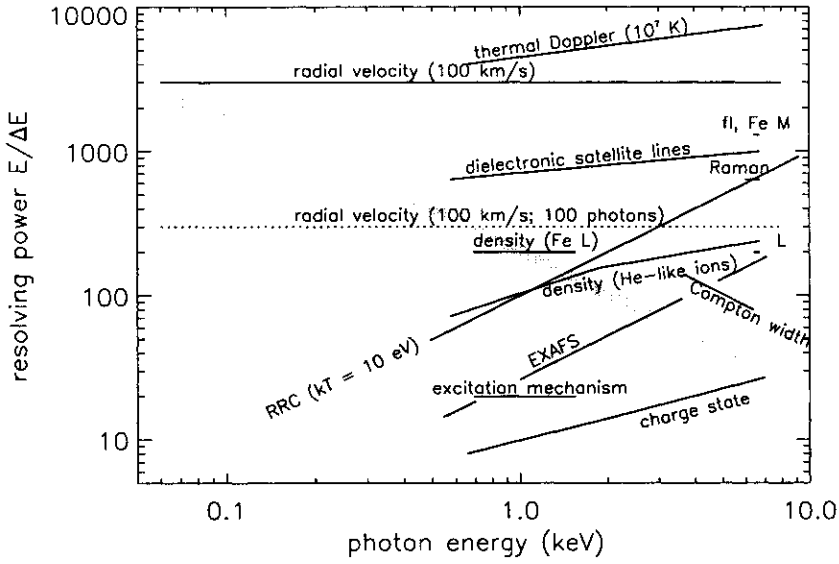

Fig. 19. Same as Figures 8 and 13, but with the predicted resolving power of the LETGS overlaid (gray band) (source: *AXAF Observatory Guide*, Ch. 8)

as well as a stunning color image of the complete, assembled LETG with all the facets shining like the rainbow in diffracted optical light, can be found at `http://www.rosat.mpe-garching.mpg.de/axaf/LETG_description.html`. A preliminary analysis of the ground calibration can be found in Brinkman et al. (1997) and Predehl et al. (1997).

Many of the considerations mentioned earlier that apply to the HETGS, apply equally to the LETGS, with one important exception. Since the HRC-S has no intrinsic energy resolution, the overlapping spectral orders cannot be separated by applying the detector energy resolution to the individual photons, as is the case with the HETGS/ACIS-S combination [imaging detectors with sufficient energy resolution and high quantum efficiency at low energies ($E \lesssim 500$ eV) did not yet exist at the time the instrument was designed]. This situation has received considerable attention (somewhat surprising, since the *Einstein* and *EXOSAT* gratings exhibited the same feature, so experience with this type of data in astrophysics exists), and a separate optical element was designed, the High Energy Suppression Filter (HESF), that can be inserted into the optical path to provide additional data in case spectral-order ambiguity precludes quantitative interpretation of a given dataset. It must be emphasized that the relative efficiencies of the various orders of the LETGS have accurately and extensively been characterized, so that the relative in-

tensities in the various orders, for an assumed incident spectrum, can always be predicted with precision, even if the individual photons cannot be sorted by order using detector energy information. No quantitative information is ever lost by the overlapping of orders.

It is generally agreed that order confusion may arise in sources with bright lines at relatively high energies ($E \gtrsim 1$ keV). The third orders of these lines may contribute significantly to the spectrum at energies below 1 keV, especially if the first-order low-energy spectrum is faint (for instance, due to heavy interstellar absorption). The HESF was designed with this situation in mind. It cleverly uses the notion of the critical angle for total external reflection (cf. Dick Willingale's chapter), by having the focused and dispersed light bounce off a special mirror made of low-Z material. This mirror is oriented along the dispersion direction, and is slightly offset from the HRC-S in the cross dispersion direction. By tilting the entire telescope slightly in the cross dispersion direction, the dispersed light can be made to bounce off the HESF before reaching the detector in the focal plane. This additional reflection suppresses photons with energies higher than the critical energy that corresponds to the angle of incidence on the HESF, and their contaminating effect in higher order is reduced. The HESF was incorporated in the *AXAF* instrument package at a late date, and there has not been much time to carry out measurements to characterize its performance, so experience with astrophysical data will probably have to tell how useful it will turn out to be in practice.

Ironically, the cross dispersion properties of the fine support grid could have supplied the additional information required to assign spectral orders to bright emission line images—the extent of the faint cross dispersion image scales linearly with wavelength, and in principle this allows another measurement of the wavelength of the photons in the image (a true cross disperser). However, this cross dispersion image will usually be very faint in astrophysical spectra.

As an example of the power of spectroscopy with the LETGS, Figure 20 displays a simulated LETGS spectrum of the famous Seyfert 1 galaxy NGC 5548, obtained in 40,000 sec exposure. This object has a complex soft X-ray/EUV spectrum. Continuum emission originating in the innermost regions of the AGN is seen through a partially ionized absorber (Fabian et al. 1994; Mathur, Elvis & Wilkes 1995), which itself also emits detectable amounts of X-ray line radiation (Kaastra, Roos & Mewe 1995). These spectral components are present in the simulation as well, at parameter values consistent with recent observational constraints. The line emission, specifically, has been represented by a source in collisional equilibrium at a temperature of $\sim 6 \times 10^5$ K, consistent with the discrete emission seen in the EUV spectrum.

All features in the $40 - 100$ Å band are due to the L-shell ions of the low- and mid-Z elements, which are formed in a range of ionization conditions that has not yet extensively been probed spectroscopically. Not much is known

about the properties of the medium that gives rise to the emission, and it has been speculated that this is the same medium that gives rise to the highly-ionized Ne emission lines seen in the UV spectra of high-redshift quasars (Hamann et al. 1998). Spectroscopy with the LETGS will also allow further quantitative investigation of the ~ 3800 km s^{-1} line broadening seen in the *EUVE* spectrum, as well as the ~ 1200 km s^{-1} bulk velocity field seen in the UV absorption spectrum of the absorbing medium.

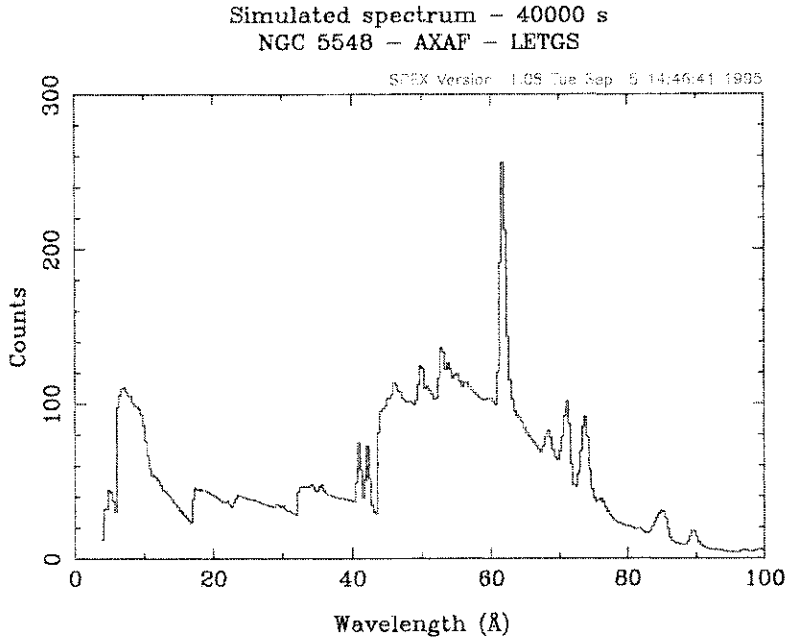

Fig. 20. Simulated 40,000 sec exposure *AXAF* LETGS spectrum of NGC 5548, illustrating the novel spectroscopy that can be expected longward of the interstellar Carbon edge (44 Å) in very soft-spectrum objects, such as AGN (courtesy Jelle Kaastra, SRON).

4.5 In Von Laue and Debye's footsteps: scattering by random fluctuations in the properties of a transmission grating

As another example of how basic physical principles can be applied directly to a calculation of the properties of a diffraction grating, I discuss the presence of small amounts of scattered light in the diffraction patterns of the HETG and LETG. The calculation also provides yet another illuminating example of the

physics of scattering mechanisms, which gives rise to very similar concepts in widely differing physical systems.

During calibration measurements of the *AXAF* grating spectrometers, it was found that the HETG and LETG exhibit small amounts of scattering, visible as a very faint, nearly continuous distribution of light in between the various diffracted orders of a given emission line. You can see this light in Figure 21, which shows a LETG calibration spectrum of the Al Kα line, as recorded with ACIS. From top to bottom, I plot the two-dimensional spectral image (each photon is represented by a single dot; the zero-order position is at the origin), followed by the ACIS pulse height spectrum of all events. The third panel shows the position of all photons in the dispersion-pulse height plane, and here you can clearly see the separation between true continuum radiation (the curved shapes running away from the lowest spectral orders) and monochromatic, Al Kα radiation spread out horizontally between the orders (all photons the same pulse height). Finally, the bottom panel shows a monochromatic grating spectrum, obtained by integrating the spectral image over the cross dispersion coordinate, and including only events in the $1.4-1.8$ keV CCD pulse height range.

The Al Kα light detected between the orders is the puzzle: light scattered by the telescope should be distributed circularly symmetrically around each spectral order in the two-dimensional spectral image, which is not what is seen. The dispersion-pulse height diagram shows that the inter-order light is not continuum radiation, so scattering by the grating is logically the only possible explanation. What causes the scattering ?

Small random fluctuations in the grating properties (the period, bar thickness, bar cross sectional shape) cause scattering. The simplest way to understand this is the following; we will specialize to scattering by period fluctuations, because it is the easiest case.

Arguing from Huygens' principle, you find that the diffraction pattern of a grating is formally equal to the squared modulus of the Fourier transform of the grating pattern (the pattern of slits and bars), as you can see as follows. Adding the complex amplitudes from all parts of the grating (assuming an opaque grating for simplicity) gives

$$I = \left| \int_{\text{grating slits}} \mathrm{d}s \exp i(p-q)s \right|^2 = \\ = \left| \int_{-\infty}^{\infty} \mathrm{d}s \; G(s) \; \exp i(p-q)s \right|^2 \qquad (19)$$

where $G(s)$ is the grating pattern [$G(s) \equiv 1$ for the slits, $G(s) = 0$ otherwise], which shows that I is the square of the transform of G.

Now suppose that the grating period were itself slowly modulated, by a slow sinusoidal deviation in the period. That means that $G(s)$ contains another period, other than the average grating period, which causes additional interference in the diffraction pattern. Right next to each bright diffraction

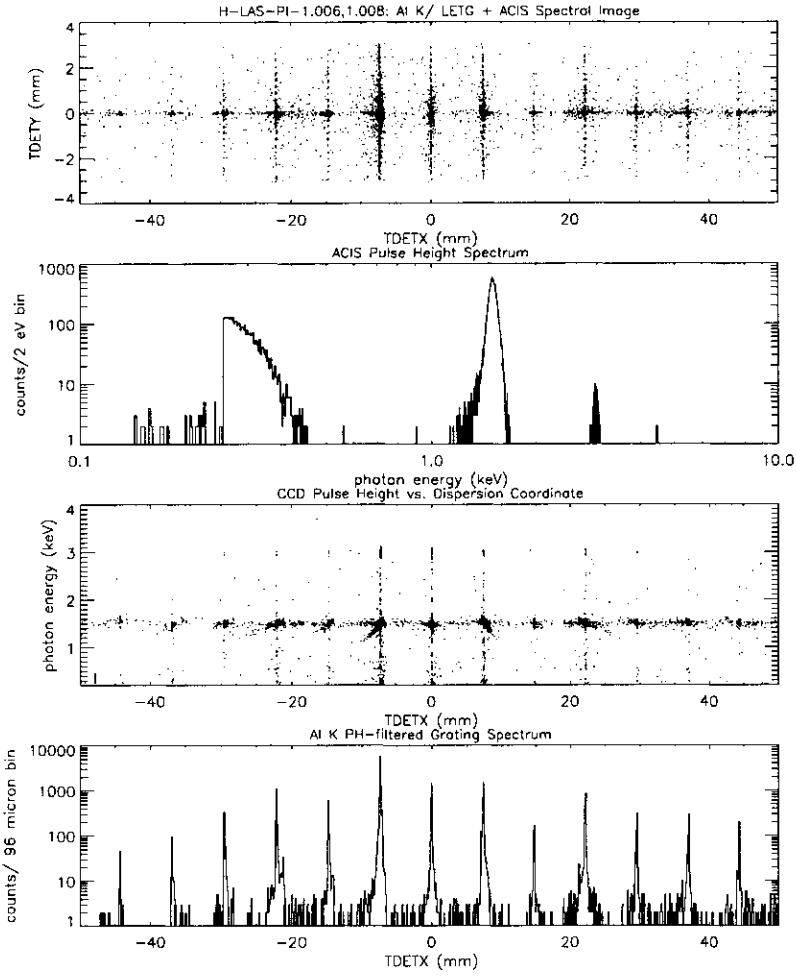

Fig. 21. LETG Al Kα calibration spectrum, recorded with ACIS-S. Top to bottom: two-dimensional spectral image, CCD pulse height spectrum, position of all photons in the dispersion-pulse height plane, and grating spectrum with the continuum suppressed by PHA filtering. The dispersion-pulse height plane shows faint scattered light in between the spectral orders as a horizontal band.

peak you will see smaller peaks, corresponding to 'beats' with the main grating period. Such subsidiary images are called 'ghost images', and the ones due to slow sinusoidal modulation of the grating period are called 'Rowland ghosts', after Henry Rowland, who first gave the correct explanation for their existence, and managed to produce ghost-free gratings.

The general case of small random fluctuations in the grating period can be represented by a Fourier decomposition, with random phases, for the fluctuations. Each of the Fourier components produces a set of ghosts outside the main diffraction peak, at an angular separation determined by the spatial wavelength of the Fourier component. The entire superposition of Fourier components therefore produces a continuous distribution of ghosts, or a scattering 'wing' on the main diffraction peaks. And as long as the amplitude of the period fluctuations is small, you intuitively guess that the shape of the scattering wing will reflect the distribution of power over the various Fourier components in the grating pattern. This is the basic mechanism behind scattering by diffraction gratings.

It is easy to estimate the total amount of light thus scattered. For simplicity, assume an opaque grating with bar/slit ratio equal to unity (it is easy to relax these conditions). Slit number n contributes a complex amplitude

$$\begin{aligned} A_n &= \int_{\text{slit } n} \mathrm{d}s \exp(-iks\sin\theta) \\ &= \qquad\qquad (q \equiv k\sin\theta) \\ &= \int_{(2n+1)d/2}^{(2n+2)d/2} \mathrm{d}s \exp(-iqs) \\ &= \frac{1}{-iq}\Big(\exp\big(-iq(2n+2)d/2\big) - \exp\big(-iq(2n+1)d/2\big)\Big), \end{aligned} \tag{20}$$

using the same notation as earlier (d is the grating period). The intensity is (we are not worrying about the correct normalization because we will be interested only in the fraction of light scattered)

$$\begin{aligned} I = AA^* &= \sum_{n=-N}^{N} \sum_{m=-N}^{N} A_n A_m^* \\ &= \frac{4\sin^2(qd/4)}{q^2} \sum_{n=-N}^{N} \sum_{m=-N}^{N} \exp\big(-iq(n-m)d\big) \end{aligned} \tag{21}$$

The intensity has its main peaks at the zeros of $1 - e^{iqd}$ (or $d\sin\theta = m\lambda$, the dispersion relation). This factor appears in the denominator in the final expression for I, and arises from the sums over $\exp(inqd)$.

Now apply a small random perturbation Δd_n to the location of slit n; this amounts to applying a small phase shift to A_n:

$$A_n = \frac{1}{-iq}\Big(\exp\big(-iq[(2n+2)d/2 + \Delta d_n]\big) - \exp\big(-iq[(2n+1)d/2 + \Delta d_n]\big)\Big) \tag{22}$$

(Here you see that the result will not depend on period-to-slit ratio; the phase shift is the same regardless of the slit-to-period ratio). The diffraction pattern is

$$I = AA^* = \frac{4\sin^2(qd/4)}{q^2} \sum_{n=-N}^{N} \sum_{m=-N}^{N} \exp\bigl(-iq(\Delta d_n - \Delta d_m)\bigr) \exp\bigl(-iq(n-m)d\bigr) \tag{23}$$

The observable quantity is the statistical average of I, averaged over all possible realizations of the Δd_n:

$$\langle I \rangle = \frac{4\sin^2(qd/4)}{q^2} \sum_{n=-N}^{N} \sum_{m=-N}^{N} \Bigl\langle \exp\bigl(-iq(\Delta d_n - \Delta d_m)\bigr) \Bigr\rangle \exp\bigl(-iq(n-m)d\bigr) \tag{24}$$

Now assume that the perturbations are small: $q\Delta d_n \ll 1$, and expand the exponential in $\langle\ \rangle$:

$$\begin{aligned} &\Bigl\langle \exp\bigl(-iq(\Delta d_n - \Delta d_m)\bigr) \Bigr\rangle = \\ &= \Bigl\langle 1 - iq(\Delta d_n - \Delta d_m) - \frac{1}{2}q^2(\Delta d_n - \Delta d_m)^2 + \ldots \Bigr\rangle \\ &\cong 1 - q^2\sigma^2 \end{aligned} \tag{25}$$

since $\langle \Delta d_n \rangle = \langle \Delta d_m \rangle = 0$, $\langle \Delta d_n \Delta d_m \rangle = 0$, and we define the variance of the perturbations $\langle \Delta d_n^2 \rangle \equiv \sigma^2$. Therefore, to first order, the diffraction pattern is reduced in intensity by a factor $1 - q^2\sigma^2$, which implies that a fraction $q^2\sigma^2 = k^2\sigma^2 \sin^2\theta$ of the light has been scattered out of the main diffraction peak. Recalling that $\sin\theta = m\lambda/d$ (m the order number), you find that the total amount of scattered light, as a fraction of the light in a given diffraction peak, scales like $f_m = 4\pi^2\sigma^2 m^2/d^2$; there is no scattering in zero order.

This calculation closely parallels Debye's famous calculation for scattering of X-rays by thermal vibrations of the ions in a crystal lattice in crystal diffraction experiments (Debye 1914). You can show that if the Δd_n are normally distributed, then the statistical average of $\exp\bigl(-iq(\Delta d_n - \Delta d_m)\bigr)$ (the reduction in the intensity of the diffraction pattern, or the fraction of unscattered light) is exactly equal to $\exp(-q^2\sigma^2)$, and this is the so-called Debye-Waller factor, by which the intensity of the sharp diffraction peaks decreases, and which occurs in all kinds of scattering calculations. Our factor $1 - q^2\sigma^2$ is of course the first-order approximation to this expression. Anecdote has it that when von Laue first discussed with Debye his idea of looking for diffraction of X-rays by crystals, the latter raised the objection that thermal vibrations of the ions around the lattice sites would destroy the interference pattern, and that von Laue wouldn't see anything. Von Laue proposed he would do the experiment anyway, if Debye would calculate the

effect of the vibrations. That led to the Debye-Waller factor and the conclusion that scattering would not significantly affect the sharp diffraction peaks, while von Laue got the Nobel prize for the experiment ...

John Davis (MIT) has given an elegant and mathematically powerful alternative derivation of the scattering effect, where he also calculates the angular distribution of the scattered radiation explicitly, to arbitrary order of approximation, by assuming a normal distribution of completely uncorrelated (bar-to-bar) period fluctuations (Davis 1997).

How does this compare to the observed scattering properties of the LETG? In order to measure the intensity of the scattered light in the lowest few orders, I took the Al K spectrum shown in Fig. 21(d), and divided the spectral range into equally sized subranges centered on the spectral orders. I approximately fitted a copy of the telescope point response function, integrated over the cross-dispersion direction, to the diffraction peaks, plus a Gaussian to represent the scattered light. The Gaussian was integrated over the relevant detector subrange, to obtain the scattered light associated with each order. These estimated scattered intensities, normalized to the intensities in the sharp diffraction peaks, are shown in Fig. 22, as a function of spectral order. The dotted line in this figure shows the prediction based on the expression derived above, $f_m = 4\pi^2(\sigma/d)^2 m^2$, with the r.m.s. fluctuation in the period σ normalized at $m = 4$.

As you can see, the qualitative behavior of the scattered light in low order is roughy consistent with our prediction, but two features stand out. First, there is some scattering in $m = 0$, which should not be there, and second, there seems to be an even/odd asymmetry in m of the measured f_m, with respect to the predicted f_m. Both these features have a natural explanation. Scattering near $m = 0$ is partially an artefact of my analysis, beacuse some light scattered through large angles from $m = \pm 1$ has leaked into the analysis region centered on $m = 0$. I can correct for this effect if I knew the theoretical angular distribution of the scattered light. To obtain some idea of how large this correction might work out, I used the explicit shape of the full diffraction pattern (including scattering) given by Davis (1997), which was derived under the assumption that the period fluctuations can be described by completely uncorrelated, Gaussian-distributed shifts of the bar locations. I integrated the intensity of this pattern over the spatial regions centered on the spectral orders, and normalized to the theoretical efficiencies of the orders. The shape of the scattering distribution turns out to depend on a/d, just like the diffraction efficiencies themselves, and so I adjusted a/d and σ/d to approximately fit the measured scattering fractions. The result is displayed in Fig. 22 as the solid histogram. The agreement is improved, and the physical interpretation of the scattering as being due to random fluctuations in the period appears to be validated. The parameters a/d and σ/d are not very well constrained, though, and the fact that the best-fitting a/d does not entirely agree with the value determined from IR

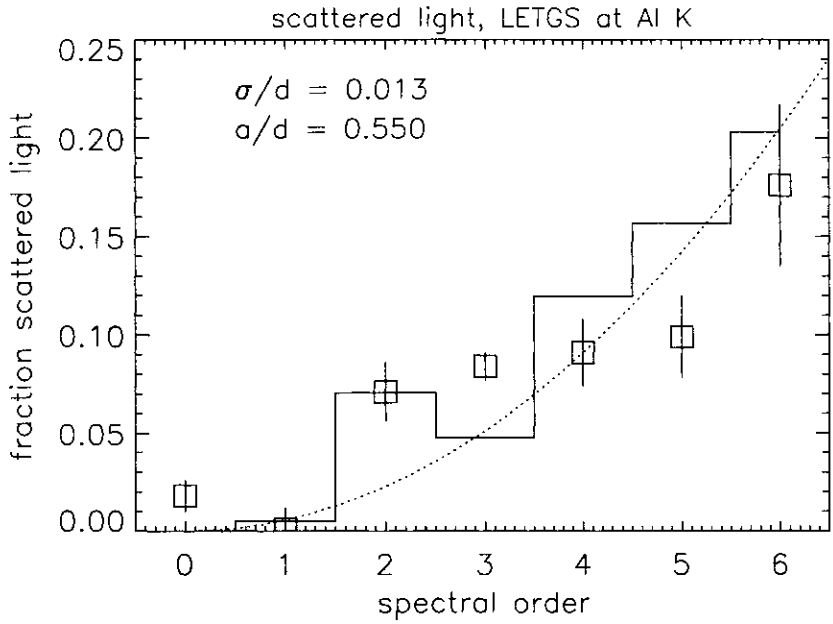

Fig. 22. Measured fraction scattered light, as a function of spectral order, for low-order Al K radiation. The dotted line is the simple first-order estimate of the scattering fraction ($f_m = 4\pi^2(\sigma/d)^2 m^2$), normalized at $m = 4$, while the solid histogram is based on an explicit shape for the distribution of scattered light.

reflectivity measurements of the individual grating elements is therefore not a serious concern. Also, the exact shape of the scattering distribution I have used is reasonable, but has not yet been independently verified.

The 'leakage' correction at $m = 0$ turns out to be small, and we have to conclude that there appears to be true scattered light associated with $m = 0$. This could be due to fluctuations in the bar thickness, which also cause scattering, and, in contrast to the period fluctuations, also cause scattering in zero order. The calculation of the bar thickness fluctuation effect is much nastier than the period fluctuation effect, because a perturbation to the bar thickness is not equivalent to a simple phase shift to the complex amplitude contributed by a given bar+slit. In addition, there is true absorption in the bars, so you cannot repeat the simple calculation leading up to Eq. (25): the apparent decrease in the diffraction efficiency can be due both to scattering as well as true absorption. In order to derive the fraction scattered light, you therefore have to calculate the diffraction pattern explicitly, to extract the scattered light. If you are curious, the details of the lowest-order approximate calculation are given in Paerels (1997).

It turns out that the bar-thickness-fluctuation scattered light depends strongly on wavelength (as it should—the effect should disappear at long wavelengths, where the bars are opaque), occurs in zero order as well as in orders $m \neq 0$, and does not depend on order for the orders $m \neq 0$. Taken at face value, the measured fraction scattered light in $m = 0$ at Al K indicates a relative r.m.s. fluctuation in bar thickness of ~ 0.08, and the theory predicts that the scattering by this effect should be very small ($f_m \lesssim 0.6\%$) in the orders $m \neq 0$ at Al K.

You may wonder why this scattering is a concern at all, given that it is measured to be a sub-1% effect in $m = 1$. The reason is, of course, that if the scattering would turn out to be strongly wavelength- or order-dependent, then the scattered light could significantly contaminate the true continuum light (and these components cannot be separated with the HRC-S detector, which has no intrinsic energy resolution), especially in line-dominated spectra, such as the cooler coronal plasmas ($T \lesssim 10^7$ K). Since absolute abundance measurements depend on line intensities relative to the continuum, the scattering could systematically bias abundance measurements.

And this turns out to be the case. From our calculation we find that $f_m \propto m^2$, whereas Eq.(18) shows that the diffraction efficiencies are approximately proportional to m^{-2}. So in the lowest orders, each order has the *same* absolute amount of scattered light! The higher orders of intense short-wavelength lines can therefore contaminate the continuum at longer wavelengths. At some high m, f_m must of course saturate to $f_m = 1$ through the Debye-Waller factor, and all the diffracted light is scattered, at which point the scattered light must again decrease like η_m itself. The actual degree of contamination and its effect on abundance measurements is still under investigation.

5 The Reflection Grating Spectrometers on *XMM*

5.1 Introduction

ESA's X-ray Multimirror Mission (*XMM*; `http://astro.estec.esa.nl/XMM/xmm_top.html`) was defined to be the high-throughput X-ray spectroscopy cornerstone of the Horizon-2000 strategic scientific program. Launch is currently scheduled for August 1999. The core of the observatory are three identical high-throughput, medium angular resolution ($\lesssim 15$ arcsec half-power spot diameter [HPD, or half energy width HEW], and approximately 6 arcsec FWHM) X-ray telescopes, of peak effective area ~ 1500 cm^2 at 1 keV, ~ 500 cm^2 at 5 keV (per telescope), and sensitivity out to 10 keV (Telescope Scientist: Bernd Aschenbach, Max Planck Institut für Extraterrestrische Physik). The focal length is 7500 mm. All three telescopes have CCD focal-plane cameras, collectively referred to as the European Photon Imaging Camera (EPIC; PI: M. Turner, Leicester University). Two CCD cameras are MOS-type devices, the third is a pn-junction device. The mirror assemblies represent a breakthrough in the fabrication

of lightweight X-ray optics: each assembly consists of 58 separate, densely nested gold-coated thin Ni shells of Wolter-I geometry. Details and pictures can be found at the site mentioned above; a set of technical references is given on `http://astro.estec.esa.nl/XMM/user/xmmpub_top1.html`.

Two of the telescopes will be equiped with an array of reflection gratings each, which, together with two dedicated CCD cameras at the spectroscopic focus, constitute the Reflection Grating Spectrometer (RGS) experiment (PI: Albert Brinkman, SRON). The Reflection Grating Arrays (RGA) will be mounted permanently behind their mirror assemblies, where they intercept approximately half the light for medium- to high-resolution spectroscopy in the 0.3-2.5 keV band; the other half of the light goes to the EPIC detectors at the prime focus. Therefore, an RGS spectrum will be obtained for each target observed with *XMM*.

Finally, the Optical Monitor (OM) experiment (PI: Keith Mason, Mullard Space Science Laboratory) will simultaneously cover the X-ray field of view in the UV/optical band (1700−6000 Å) with a 30 cm aperture Ritchey-Chrétien telescope.

In order to ensure a uniform, optimum, and timely analysis of the 'Serendipitous Survey' formed by the data content of all of the imaged fields, an *XMM* Survey Science Center has been formed (PI: Mike Watson, Leicester University), a collaboration of institutes in the UK, France, and Germany. The SSC will analyze all fields, and survey the *XMM* X-ray sky.

5.2 Properties of reflection gratings, and design of a grazing-incidence reflection grating spectrometer

As stated above, *XMM* was designed around high-throughput, medium angular resolution optics. The desired instrument package was to include a high-resolution spectrometer for the soft X-ray band, with high throughput and a resolving power of at least $\mathcal{R} = 250$ (also in the HEW sense, just like the telescope angular resolution specification) at 1 keV. This resolving power was chosen with the density diagnostics in the He-like ions in mind, among other things. At the time these recommendations were formulated (mid-eighties), the only way to achieve these goals was with diffractive spectrometers.

But with a moderate telescope angular resolution, the dispersion angles need to be large in order to achieve $\mathcal{R} \sim 250$; for instance, with a $20''$ telescope blur, a transmission grating needs to disperse to $250 \times 20'' = 1.4°$, which, for 1 keV photons, implies a line density of $\sim 20,000$ lines mm^{-1} ! Instead, as we have seen before, it is much easier to achieve these high dispersions with grazing-incidence reflection gratings, a solution first advocated for *XMM* by Kahn & Hettrick (1985).

In addition, the reflectivity properties of reflection gratings have certain attractive features. First, the efficiency in the soft X-ray band can be high (> 30% in first order), and second, the wavelength at which this maximum occurs can again be tuned by optimizing the grating groove shape. To understand

these properties, let us look at how to calculate the reflectivity of a reflection grating, an exercise in the application of electrodynamics to the solution of a real problem (as opposed to a textbook EM problem).

There is a fundamental distinction between the operation of a transmission grating and a reflection grating, in the regime $\lambda/d \ll 1$ that applies to the X-ray band for moderate line density structures. All a transmission grating does in that case is block and possibly phase shift parts of the incoming wave fronts; Huygens' principle is sufficient, and you don't need Maxwell's equations to work out the diffraction efficiency. But with a reflection grating, you have to apply the correct boundary conditions to the fields at the reflective surface in order to obtain the intensity of the reflected waves in terms of the intensity of the incident waves. For a flat mirror surface of given material properties, this is trivial, and the calculation is in every textbook. But there is no mathematically simple way to impose general boundary conditions at a periodically modulated (as opposed to a flat) surface. Hence, a general solution in closed analytical form to the problem of determining the reflectivity of a reflection grating does not exist.

Only approximate solutions to the problem exist. Let us look at a well-known, but as it turns out for our application physically unrealistic choice of boundary condition. If you ad-hoc impose the condition of perfect reflectivity at the grating surface, that is, the amplitudes of the fields inside the grating and at the surface are zero everywhere, each part of the grating surface radiates in phase with the incident wavefronts, and the efficiency calculation reduces again to applying Huygens' principle to the periodic surface. This is also referred to as 'scalar diffraction theory', because the actual dynamics of the electromagnetic field (its tensor character and the interaction with the atoms of the reflecting material) is entirely ignored.

As a benchmark example, a straightforward application of Huygens' principle to the case of a grating of period d, illuminated at incidence angle α, with triangular-shaped grooves of tilt angle δ (see Figure 5 for the geometry) yields the following expression for the scalar diffraction efficiency as a function of wavelength and order m:

$$\epsilon_m = (4\sin^2\theta\sin\alpha\sin\beta_m)^{-1} g^2 P_m^2 (\sin Q_m/Q_m)^2 \tag{26}$$

with

$$g = \sin\alpha/\sin(\alpha+\delta) \tag{27}$$

$$P_m = \sin\theta\{\sin(\alpha+\delta) + \sin(\beta_m - \delta)\} \tag{28}$$

$$Q_m = (\pi g d/\lambda)\sin\theta\{\cos(\alpha+\delta) - \cos(\beta_m - \delta)\} \tag{29}$$

with β_m the dispersion angle in order m, and θ the angle of the incident ray with the groove direction.

By inspecting this expression, you see that it peaks at wavelengths satisfying

$$\alpha + \delta = \beta - \delta \tag{30}$$

because $Q_m \to 0$. When this condition is fulfilled, the incident and dispersed ray make the same angle with the groove surface, and the grating facets act kind of like tiny mirrors—the grating literally 'blazes' in reflected light, and is said to 'be blazed' at this wavelength and spectral order. The blaze can be tuned by varying the facet angle δ.

But we know that mirror reflectivities in the soft X-ray band are definitely finite, and less than unity. As a stopgap measure, one therefore multiplies ϵ_m by the appropriate mirror reflectivity R_λ to 'correct' for this effect. This is physically not consistent, because it violates the choice of boundary conditions underlying Eq.(26), and the conceptual crisis is evident once again when it comes to choosing the incidence angle at which to evaluate the reflectivity correction factor: except at the blaze, the angle of incidence on the facets $\alpha + \delta$ and the 'reflection' angle $\beta - \delta$ are not equal. Often, the mean of R_λ evaluated at both angles is used: $R_\lambda \equiv [R_\lambda(\alpha + \delta)R_\lambda(\beta_m - \delta)]^{1/2}$, and the full grating efficiency is defined as $\eta_m = R_\lambda \cdot \epsilon_m$.

Theories that incorporate solutions to the full Maxwell equations are designated 'vector theories', to distinguish them from the scalar theory outlined above. All vector theories of grating reflectivity (Stroke 1967, Sect. V, and the references therein) employ the periodicity in the direction along the plane of the grating, and start with a Fourier expansion, with constant coefficients for the far fields; the components of the reflected field are the various diffraction orders. Following Rayleigh, one makes the assumption that this expansion is valid everywhere (that is, that the coefficients do not depend on the coordinate direction perpendicular to the grating plane), including near and at the grating surface. Imposing the boundary conditions on the grating surface leads to an infinite set of coupled linear equations, which, when suitably cut off at some maximum harmonic, can in principle be solved numerically. A further traditional simplification arises when one assumes that the grating material has infinite conductivity (i.e., imaginary part of the complex index of refraction equal to zero).

But if you have to resort to using a computer to evaluate the reflectivities in these approximate vector theories, you might as well solve the Maxwell equations directly and exactly, and apply the correct boundary conditions with no restrictions on the properties of the grating material. Various schemes have been worked out to accomplish this (Petit 1980). In the conceptually simplest of these, a Fourier expansion is again substituted in Maxwell's equations, but this time one allows for the general dependence of the expansion coefficients on the vertical distance to the grating. The resulting set of coupled ordinary differential equations for the Fourier coefficients can again be solved numerically with a suitable cutoff (see for instance Jark & Nevière 1987). These numerical procedures also naturally allow for arbitrary grating groove shapes.

Figure 23 displays an example comparison between the scalar diffraction and the EM theory efficiencies for a Gold grating of parameters very similar to those for the gratings in the RGS. The general features of both calculations are roughly the same (such as the blazing behavior - the blaze was chosen to be at 15 Å in $m = -1$), but for accurate predictions the scalar theory is clearly inadequate.

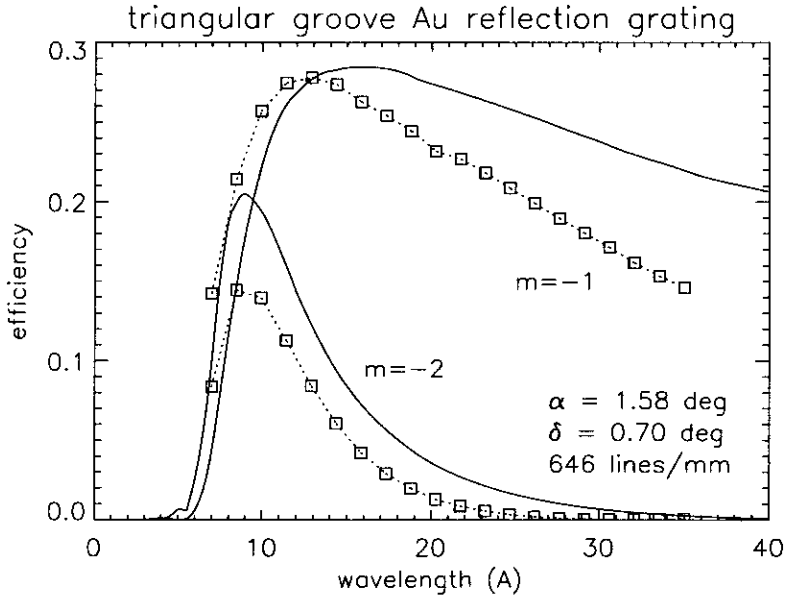

Fig. 23. Absolute efficiency of a gold reflection grating of triangular profile, illuminated at an angle $\alpha = 1.58$ deg; the line density is 646 lines/mm, the groove facet tilt angle $\delta = 0.70$ deg. The solid line gives the scalar diffraction result, the datapoints were calculated from a numerical solution to Maxwell's equations developed for the RGS project.

Reflection grating spectrometers have a few more free parameters than transmission grating spectrometers (angle of incidence α, groove shape, spacing between adjacent gratings in the array), so optimizing the design is a little more complicated. We will discuss the optimization of the design for the RGS as an example.

The primary design drivers are of course the chosen wavelength band of operation, and the resolving power at some chosen fiducial wavelength. You require that the design deliver the optimum throughput. For the RGS, the (minus) first-order wavelength band was chosen to be $5 - 35$ Å. The short wavelength cutoff is dictated by the properties of suitable grating materials.

All elements near $Z = 80$ (Au, Pt, Ir) have deep M-shell absorption edges near 2.5 keV, which provides a natural boundary. The upper limit was loosely chosen with atomic physics, and the absorbing properties of the interstellar medium in mind; many sources are very faint well beyond the O K edge at 23 Å. Also, a much wider band sampled at resolving powers of at least $\mathcal{R} = 250$ would have required a physically much larger focal-plane detector. The $5 - 35$ Å band contains the K-shell spectra of the abundant elements C through Si (for C, only the Hydrogenic species is in the band), as well as the diagnostic-rich Fe L spectra. The blaze wavelength was chosen in the middle of the band, $\lambda_B = 15$ Å, roughly where the Fe L $n = 3 - 2$ emission lines occur.

The resolving power at blaze, for given telescope angular resolution $\Delta\alpha$, depends of course on the angle of incidence α and the grating period d. It will be convenient to rewrite it as (Eq. 9):

$$\begin{aligned}\mathcal{R}_B = \lambda_B/\Delta\lambda &= \frac{\cos\alpha - \cos\beta_B}{\sin\alpha\Delta\alpha} = \\ &= \frac{(1-\sin^2\alpha)^{1/2} - (1-\sin^2\beta_B)^{1/2}}{\sin\alpha\Delta\alpha} \approx \\ &\approx \frac{(\sin\alpha + \sin\beta_B)(\sin\beta_B - \sin\alpha)}{2\sin\alpha\Delta\alpha} \approx \\ &\approx \frac{\gamma}{\Delta\alpha}\left(\frac{1}{\eta} - 1\right) \end{aligned} \tag{31}$$

where we have defined the graze angle γ on the groove facets, $\gamma \equiv \alpha + \delta = \beta_B - \delta$, used small angle approximations, and defined the ratio

$$\eta \equiv \sin\alpha/\sin\beta_B \tag{32}$$

(don't confuse this η with the efficiency of a grating; the coincidence in nomenclature is unfortunate, but the use of η for both is common in papers on the RGS). For a given telescope angular resolution $\Delta\alpha$, the resolving power at blaze therefore depends on η and γ.

Next, we calculate the throughput at blaze. From Eq.(26), you see that the diffraction efficiency ϵ_B at blaze is equal to $\epsilon_B = \sin\alpha/\sin\beta_B \equiv \eta$, and to obtain the grating reflectivity we multiply by the mirror reflectivity at the graze angle γ, at the chosen blaze wavelength, $R_{\lambda_B}(\gamma)$ (we use scalar theory for analytical convenience). The gratings cannot be placed infinitely close together—rays dispersed at large enough angles will be intercepted at the back of the next grating in the array, so there is a limit to how densely you should cover the focused beam in order to intercept the optimum amount of light. Packing the gratings more densely intercepts a larger fraction of the beam, but also increases the vignetting. For the RGS, the gratings were spaced such that all rays up to the first-order blaze wavelength escape shadowing (ray dispersed from the top of one grating barely grazes the bottom of the back side of the next grating in the array). That implies that the fraction

of the beam that is intercepted is equal to η, and the balance $1 - \eta$ passes unhindered through the grating array and can be used by the detectors at the prime focus. So the total throughput at blaze, t_B, is the product of the fraction light intercepted times the reflectivity of the gratings:

$$t_B = \eta^2 R_{\lambda_B}(\gamma). \tag{33}$$

You can see that there is a tradeoff between resolving power and throughput: lowering η increases the resolving power, but decreases the throughput, and *vice versa*.

The optimization might proceed as follows. Choose the resolving power at blaze; the telescope resolution $\Delta\alpha$ is given. Now examine the throughput at blaze as a function of η: for every value of η, Eq.(31) gives the corresponding value of γ, and η and γ together determine the throughput at blaze. For $\Delta\alpha = 30''$ (the original optimization for the RGS was based on a figure of $\Delta\alpha = 30''$), a resolving power at blaze of 300, and a 15 Å blaze, I plot this throughput as a function of η in Figure 24; the grating material is gold. With increasing η, the throughput first rises roughly as η^2. As η approaches unity, γ rapidly increases, and therefore the reflectivity $R_{\lambda_B}(\gamma)$ sharply decreases. Therefore, there is a maximum in the curve, in this case at $\eta = 0.50$. Given this optimum η, γ follows from Eq.(31); for $\eta = 0.50$, we find $\gamma = 2.50°$. From $\alpha + \delta = \beta_B - \delta = \gamma$, we find

$$\delta = \gamma \frac{1-\eta}{1+\eta} \tag{34}$$

so in our case, the facet tilt angle $\delta = 0.833$ deg. The grating period follows from the dispersion equation at blaze, $\lambda_B = 2d \sin\gamma \sin\delta$: $d = 11823$ Å, or a density of 846 lines mm^{-1}. Finally, the angle of incidence is $\alpha = \gamma - \delta = 1.67°$.

The actual optimization for the RGS was done slightly differently; there, the physical length of the detector (nine CCD chips long) was held constant instead of the resolving power at blaze. This yields an alternative relation between η and γ. The optimum value for η was chosen as a compromise between throughput in $m = -1$ and $m = -2$, and lies at $\eta = 0.53$, consistent with the overall requirement that approximately half the light should go to the EPIC detectors. The resulting actual design parameters for the RGS are listed in Table 4. Note that the zero order does not fall on the detector, so that the exact wavelength scale will have to be derived from accurate knowledge of the geometry, and the pointing direction of the satellite.

The design is completed by choosing the arrangement of the gratings in the focused beam. In the RGS, the gratings are placed on a Rowland circle as shown in Figure 10(b). As discussed earlier, the actual radius of the Rowland circle is arbitrary for the grazing-incidence reflection grating instrument, as long as the line density gradient on the gratings matches the Rowland circle radius. For the RGS, the center of the grating array was placed at 6700 mm from the telescope focus. The distance to the spectroscopic focus at the

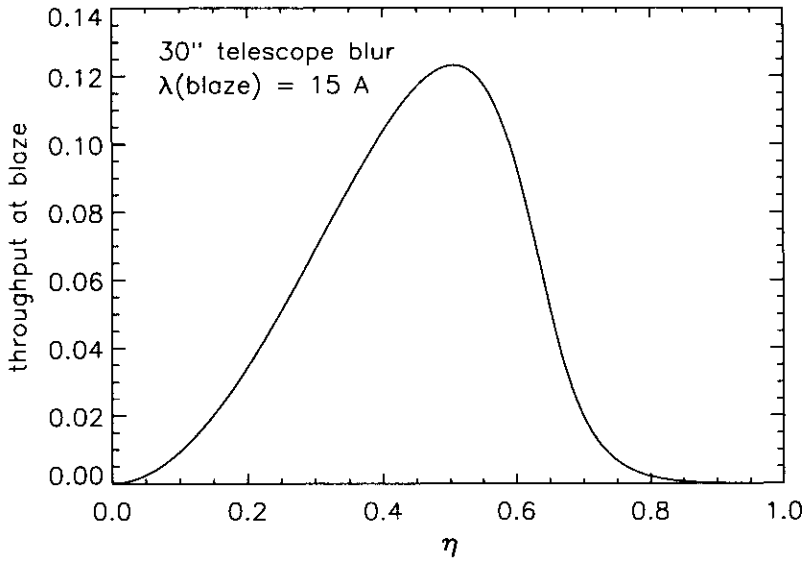

Fig. 24. Throughput at blaze, for an array of gold reflection gratings, with blaze wavelength $\lambda_B = 15$ Å, a resolving power at blaze of 300 with a 30″ telescope blur, as a function of the parameter $\eta = \sin\alpha / \sin\beta_B$.

blaze wavelength was chosen to be equal to this same distance, which implies that the Rowland circle geometry is nicely symmetric between the focused telescope and first-order blaze beams. The entire set of gratings lies on the toroidal surface generated by rotating the geometry around an axis passing through the telescope focus and first-order blaze focus; this arrangement minimizes the astigmatism of the spectrometer. In all, with a grating size of 100 × 200 mm, of order 200 gratings per array are needed to cover the beam.

5.3 Implementation of the design, and actual performance of the RGS

Implementation The gratings were produced by replication in gold on an epoxy layer from a mechanically ruled master grating. The master grating was ruled directly into gold with an interferometrically controlled ruling machine, which also allowed direct implementation of the required line density variation (using the same density gradient on all gratings introduces only a negligible aberration, as it turns out). The groove shape is approximately triangular, with a tilt angle close to the design value of $\delta = 0.70°$. X-ray performance tests of master gratings are described in Bixler et al. (1991). Performance of replicated gratings, and a preliminary physical model for the

Table 4. Design Parameters of the Reflection Grating Spectrometer on *XMM*

Mean Line Density	646 lines mm^{-1}
Blaze Angle δ	0.70 deg
Mean Graze Angle γ	2.28 deg
Angle of Incidence α	1.58 deg
Dispersion Angle at Blaze β_B	2.97 deg
Blaze Wavelength λ_B	15 Å
Fraction of Beam Intercepted η	0.53
Resolving Power at Blaze	290

grating efficiency, is decribed in Kahn et al. (1996). The primary variable in this model is the actual average groove shape, loosely characterized by the mean slope δ of the grating facets. From the measured angular, wavelength, and order dependence of the efficiency, a more detailed shape has been derived, with turns out to have small but noticeable departures from the triangular shape.

As discussed in Sect. 3.1, the gratings need to be very flat, and very accurately aligned in order to preserve the full resolution implied by the angular width of the telescope beam. In practice, this amounted to a combined flatness and alignment requirement of $\Delta\phi \lesssim 5$ arcsec. Flatness was achieved by replicating gratings onto lightweight, very flat SiC substrates, which were stiffened in the dispersion direction with thin support 'ribs' on the back of the substrates. In practice, the flatness of the substrates is such that this contribution to the spectrometer resolution is almost negligible. The grating-to-grating alignment requirement (both with respect to position and orientation) was one of the drivers on the design of the overall 'integrating structure' in which the gratings are held, directly behind the mirror module in the focused X-ray beam. The required angular precision is achieved by a 'multi-scale' design, where the actual 'grating box' provides low-precision mounts for high-precision elements (which in turn hold the gratings), which are interferometrically aligned to micron accuracy (Kahn et al. 1996). Figure 25 shows a picture of a complete RGA. The RGA's were designed and built by Columbia University (PI: Steven M. Kahn).

Resolving power and shape of the response In practice, the telescope blur and finite accuracy of grating-to-grating angular aligment contribute approximately equally to the resolution of the RGS. The actual performance of the telescopes has steadily and significantly improved over its original specification, whereas the grating alignment is fundamentally limited by the accuracy of the interferometric aligment procedure. The contribution of the

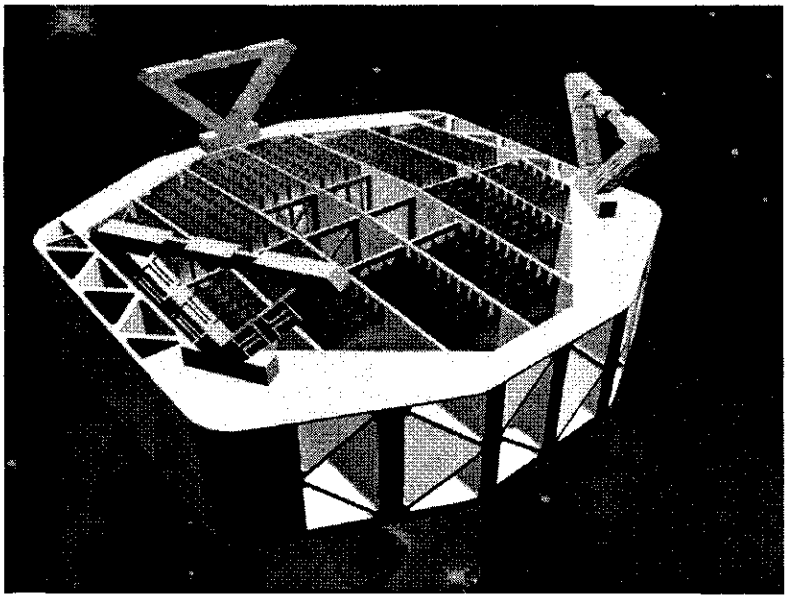

Fig. 25. Schematic of a single Reflection Grating Array for the RGS. The focused X-ray beam is incident from the top. The 182 gratings are placed on a Rowland toroidal surface. The lightweight structure holding the gratings is manufactured from beryllium, and is mounted with the triangular mounts directly behind the mirror module. The reflective faces of the gratings face forward; the five stiffening ribs on the back of each grating are also visible.

telescope blur according to Eq.(8) scales like

$$\Delta\lambda = (d/m)\sin\alpha\Delta\alpha, \tag{35}$$

which is strictly independent of wavelength, for given α and telescope blur $\Delta\alpha$. Relative misalignments, characterized by a angular spread $\Delta\phi$ produce a weak dependence on wavelength (cf. Section 3.1):

$$\Delta\lambda = (d/m)\sin\alpha(1+\sin\beta/\sin\alpha)\Delta\phi. \tag{36}$$

and the telescope contribution becomes progressively less important with increasing dispersion angle. As it is, the telescope blur and misalignment contributions are about equal at the short wavelength end of the band. Figure 26 shows the predicted resolving power of the RGS (based on the FWHM of the profile, not the HEW, to facilitate comparison with the other instruments), calculated with a dedicated raytrace code. Both first- and second-order resolving power are plotted this time, because the second-order efficiency of the instrument is actually appreciable.

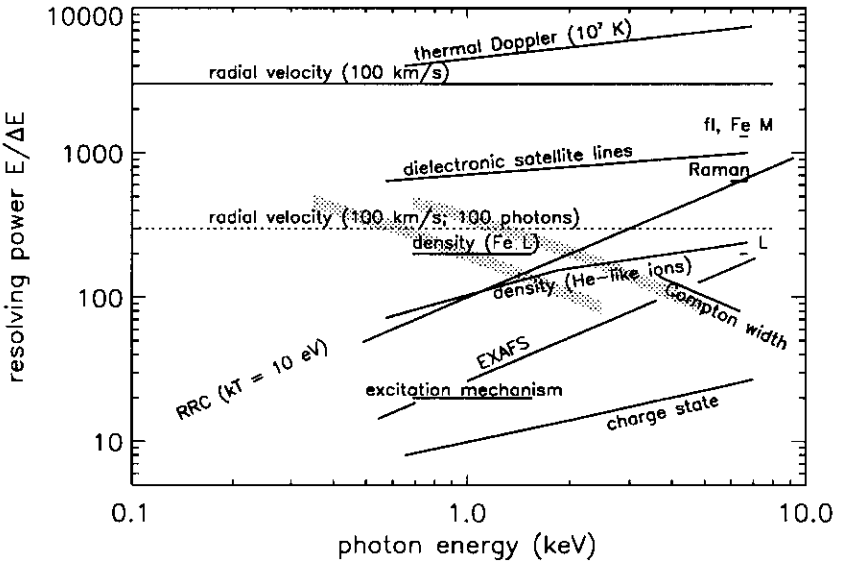

Fig. 26. Predicted resolving power of the RGS, in orders $m = -1$ (*lower gray band*) and $m = -2$ (*upper gray band*), overlaid on the set of spectroscopic diagnostics. The second-order bandpass formally extends up to 5 keV, but the effective area is small above 2.5 keV. (courtesy Jean Cottam, Columbia University).

The gratings exhibit scattering, in this case (just like for a mirror surface) by low-amplitude microroughness on the surface, which gives rise to a separate, wider component to the spectrometer profile (period fluctuation scattering is negligible, because the spacing of the grooves was actively controlled during the ruling of the master grating). The calculation of the effect of the roughness on a grating surface is entirely analogous to the calculation for grazing-incidence X-ray scattering by a mirror (e.g., Church, Jenkinson & Zavada 1977, and the references given by Klos 1985). In that case, one imagines the perturbations to the flat surface to be decomposed into their equivalent Fourier series. Each sinusoidal component in this expansion basically represents a sinusoidal grating on the mirror surface, which diffracts light out of the specularly reflected beam, into a series of ghost images on either side of the specular beam. Again, a distribution of sine waves with different spatial frequencies will produce a distribution of ghosts, i.e., a continuous scattering distribution, with the higher spatial frequency sine waves scattering light further from the core (through the dispersion relation for the light diffracted by the sine waves). The shape of the scattering distribution will again reflect the power spectral density of the surface perturbations, an intrinsic property

of the mirror surface. One often characterizes this power spectral density by a spatial-frequency bandwidth, which is then inversely proportional to the corresponding correlation length of perturbations on the mirror surface. The total amount of scattered light is again completely determined by the r.m.s. roughness of the surface (r.m.s. amplitude of the perturbations), just as with the transmission grating scattering.

Just as with calculating the reflectivity of a reflection grating, one has to face the fundamental question of how to impose the boundary conditions to the electromagnetic field on the scattering surface, and there are again 'scalar' and 'vector' scattering theories. The simplest calculation is of course the scalar calculation, assuming very-low-amplitude long spatial-wavelength perturbations, such that only the first-order diffraction by the Fourier components of the roughness has to be taken into account. This calculation is straightforward (although the standard papers on the subject are somewhat confusing). The extension to a periodically modulated surface (a grating) is also straightforward, and is decribed in Spodek et al. (1998), Kahn et al. (1996), and Paerels et al. (1994).

The expression for the fraction of light scattered in the lowest-order scalar diffraction scattering theory is closely analogous to the expression we found for the fraction of light scattered by period fluctuations in a transmission grating (σ is the r.m.s. amplitude of surface perturbations):

$$f_m = k^2\sigma^2(\sin\alpha + \sin\beta_m)^2 \tag{37}$$

which exhibits a rather strong wavelength- and spectral-order dependence. Just as with the transmission grating scattering, we expect that you can prove (under the usual assumptions) that this factor is just the lowest-order approximation to a Debye-Waller factor of the form

$$f_m = 1 - \exp\left(-k^2\sigma^2(\sin\alpha + \sin\beta_m)^2\right) \tag{38}$$

The characteristic angular width Γ of the scattering distribution scales inversely proportional to the correlation length l of the perturbations:

$$\Gamma \equiv (kl\sin\beta_m)^{-1} \tag{39}$$

As a general rule of thumb, with decreasing wavelength or increasing spectral order, the fraction scattered light increases, while the angular distribution narrows.

The RGS gratings, we found, exhibit scattering on two distinct angular scales, corresponding to correlation lengths of order $l \sim 10 - 12$ μm and $l \lesssim$ few grooves, respectively. The first gives rise to a component of scattered light close to the core of the spectrometer profile (the angular scale corresponds to ~ 8 resolution elements at blaze), the second basically produces a faint, almost constant background of inter-order light. The long-correlation length roughness has an amplitude of $\sim 11 - 13$ Å, which implies a fraction scattered light of $\sim 13\%$ at blaze. The effect of this scattered light has not been included

in the resolving power plot (Fig. 26). Because the profile consists of a sharp core, and fainter, more extended scattering wings, the width of this core is the determining factor when deciding whether a particular set of spectral features can be resolved.

The small-correlation length scattering still poses some conceptual problems. The small-angle approximation is not valid when you try to calculate its exact angular distribution—in a sense, the scattered photons do not 'belong' to a particular order anymore, which means that you really have to calculate the statistical average of the entire diffraction pattern, assuming a statistical distribution for the high-spatial-frequency surface perturbations. However, the fraction of light removed from the sharp diffraction peaks is well calibrated, and corresponds to a surface roughness of $\lesssim 11$ Å. Since the RGS is read out with CCD's, this scattered light is neatly separated from true continuum radiation in a pulse height vs. dispersion coordinate plot, and so the lack of a precise calculation for its angular distribution may not matter very much in practice. The large-angle scattering wil show up mostly as a 'lost light' correction factor to the grating efficiencies (and that is how it was found in the first place).

Effective area Fig. 27 shows the effective area of the RGS (mirror + RGA + detector), for the two modules combined. The Figure also shows the effective areas for three other high-resolution spectrometers, the *AXAF* HETGS and LETGS and the microcalorimeter on *Astro-E* ('X-ray Spectrometer', or XRS, to be discussed below). For the RGS we show both first and second order, because the second order actually has appreciable area. All curves are based on data and models which were available before the final instrument calibrations were performed, so they may (have) become slightly inaccurate.

All of the above instrument properties have been verified and characterized at the long beam X-ray testing facility 'Panter' of the Max Planck Institut für Extraterrestrische Physik near Munich (Rasmussen et al. 1998). The same physical models we have described here were also used to predict the performance of the instruments in this facility; the free parameters in the models were set to the values derived from sub-assembly calibrations, and fine tuned until agreement of the entire instrument performance with the data was reached. This roundabout 'calibration' is necessary, because the laboratory configuration is not fully representative of the flight configuration; with a 7.5m telescope focal length, you can still tell the difference between a source at a finite distance in the laboratory (125 m), and a source at infinity (the same was true for the *AXAF* calibration, with the additional complication of the effect of terrestrial gravity on the shape of the mirrors).

5.4 Examples

As a first example of the RGS performance, we show a laboratory calibration spectrum, obtained with a gold-anode electron impact X-ray source. Figure

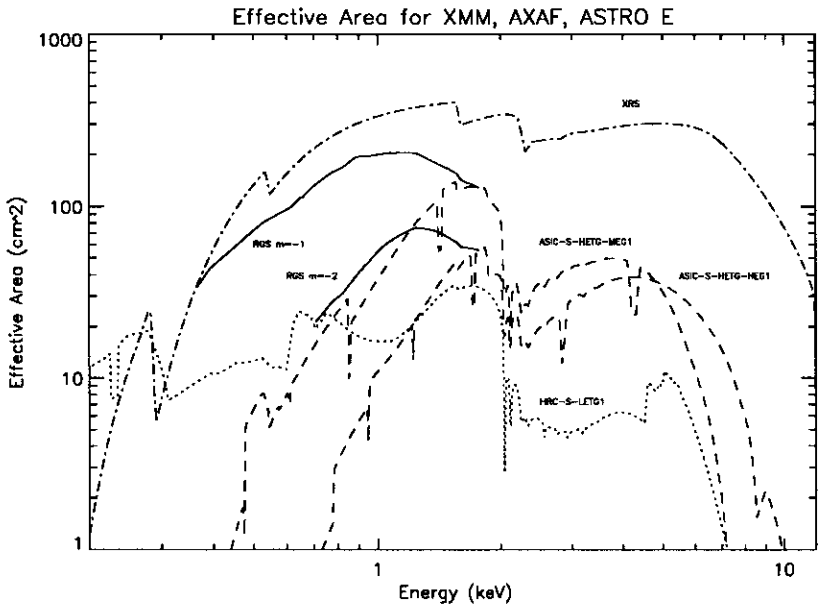

Fig. 27. Effective area of the RGS, compared with three other high-resolution spectrometers (courtesy Jean Cottam, Columbia University).

28 shows the CCD pulse height vs. dispersion coordinate diagram. This spectrum is interesting, because it actually looks a little bit like an astrophysical spectrum, and because the pulse height/dispersion plot once again demonstrates the power of performing analysis in that plane. CCD pulse height increases vertically, dispersion runs horizontally, with wavelength increasing to the left. The left edge of the image corresponds approximately to 35 Å in first order, the right edge to approximately 5 Å and the horizontal scale is approximately linear in dispersion angle β. You can see the small gaps between the nine individual CCD chips if you look closely. The curved dark bands are the continuum again (compare Fig. 14 for the *AXAF* HETGS); spectral orders $m = -1$ through -4 are visible in the continuum. Superimposed are Au fluorescence lines, the strongest being an $n = 4-3$ transition at approximately 6 Å (at the right edge of the plot). In the horizontal direction you can distinguish five spectral orders of this line, as well as the large-angle scattering referred to above. Notice how the fourth order nearly coincides in position along the dispersion direction with the first order O K emission line. In the vertical direction, centered on each emission line, you see the CCD response to monochromatic radiation, in addition to a small number of piled-up events.

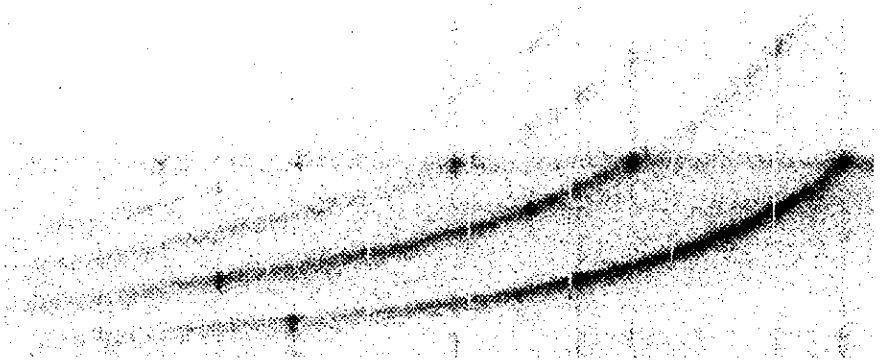

Fig. 28. Pulse height/dispersion coordinate diagram, for a laboratory calibration Au continuum spectrum. CCD pulse height increases vertically, dispersion runs horizontally, with wavelength increasing to the left. The left edge of the image corresponds approximately to 35 Å in first order, the right edge to approximately 5 Å (courtesy Masao Sako, Columbia University).

Finally, an astrophysical example. With the advent of CCD spectroscopy on *ASCA*, it was found, as had been suspected, that the soft X-ray spectra of Seyfert 2 galaxies are dominated by emission lines (e.g., Kunieda et al. 1994; Matt 1996; Ueno et al. 1994; Iwasawa et al. 1994; Iwasawa et al. 1997). The current view on the different properties of Seyfert 1 and 2 objects is that these are merely due to different viewing angles into a highly intrinsically anisotropic environment. The nuclear region may be surrounded by an absorbing torus of cool material, and in Seyfert 2's, the strong X-ray continuum arising in the region near the central compact object is blocked from direct view, leaving a large equivalent width discrete emission spectrum evidently associated with a region still in view. Such regions should also exist in Seyfert 1's, but their line emission is harder to detect against the bright nuclear continuum.

There is great interest in determining the location and origin of the line emission: it could be associated with hot, shocked gas in a strong circumnuclear starburst (detected at IR wavelengths), or in a cooler X-ray photoionized region associated with the very nucleus of the object, or a combination of both. A complete understanding of the relative contributions of both emission mechanisms would shed light on the nature of the currently suspected connection between Seyfert-type nuclear activity in galaxies, and starbursts. As we have seen in Section 2, the soft X-ray line spectrum is highly sensitive to the ionization and excitation mechanism that drives the emission.

Figure 30 shows a simulated EPIC spectrum of the famous Seyfert 2 NGC 4945 (Done, Madejski & Smith 1996), which basically looks like a much brighter version of the *ASCA* spectrum. You can see the continuum, scat-

tered into our line of sight (the intrinsic nuclear continuum does not start to become detectable through the obscuring material until you reach photon energies $E \gtrsim 10$ keV), a strong fluorescent Fe K complex, and unresolved soft X-ray line emission in the 500-1500 eV band. At this resolution, it is impossible to uniquely distinguish between the various excitation mechanisms for these lines, let alone measure average abundances, velocity fields, or diagnose radiative-transfer effects expected to be associated with an origin in X-ray photoionized gas.

Figure 29 shows the corresponding RGS spectrum, where the discrete emission detected with *ASCA* has been represented with emission from a collisional plasma at $kT_e = 0.35$ keV and solar abundances, normalized to produce the correct amount of line power. With spectra of this quality, we can finally address the questions mentioned above directly, and in detail. For faint objects, like the Seyfert 2's, the large throughput of the RGS will be required to obtain well-exposed spectra.

Some of the most interesting objects for soft X-ray spectroscopy are, or may be, extended: supernova remnants, clusters, the circumnuclear regions in nearby AGN, ordinary galaxies. As mentioned in Section 4.2, the resolution of a diffraction grating spectrometer degrades more or less linearly with source extent, if the source is resolved by the telescope. But since the RGS relies on a medium resolution telescope (15″ HEW), it is relatively insensitive to this effect. Sources up to a sizeable fraction of an arcminute still yield an RGS spectrum with much higher resolving power than a CCD; examples are the compact supernova remnants in the LMC, and cooling flows in clusters.

6 The Objective Crystal Spectrometer on *Spectrum X/γ*

The Objective Crystal Spectrometer (OXS) will be flown on the *Spectrum X/γ* observatory, which is a collaboration between Russia and a large number of institutions in Europe and the US. The experiments on it cover the entire UV through hard X-ray bands. Of these, the OXS is interesting in the present context, and in fact will reach the highest resolving power of all experiments discussed. The launch date for *Spectrum X/γ*, unfortunately, is highly uncertain. You can find information on the project at `http://hea.iki.rssi.ru/SXG/ SXG-home.html` (Institute for Space Research (IKI), Moscow, Russia).

The OXS (PI: H. Schnopper, Danish Space Research Institute) consists of a 4000 cm^2 panel of flat crystals, which can be placed in front of the aperture of the SODART (Soviet-Danish X-ray Telescope), such that the diffracted radiation is sent into the focusing telescope for imaging at the focal plane. A number of different energy bands are covered with different crystals, centered on different important emission line complexes. LiF crystals cover the Fe K band ($5.0 - 7.4$ keV at resolving power $\mathcal{R} = 1250$), Si crystals cover the S and Ar K complexes ($2.3 - 4.6$ keV at resolving power $\mathcal{R} = 3200$), and RAP crystals cover the O K band ($0.55 - 0.81$ keV at resolving power $\mathcal{R} = 770$).

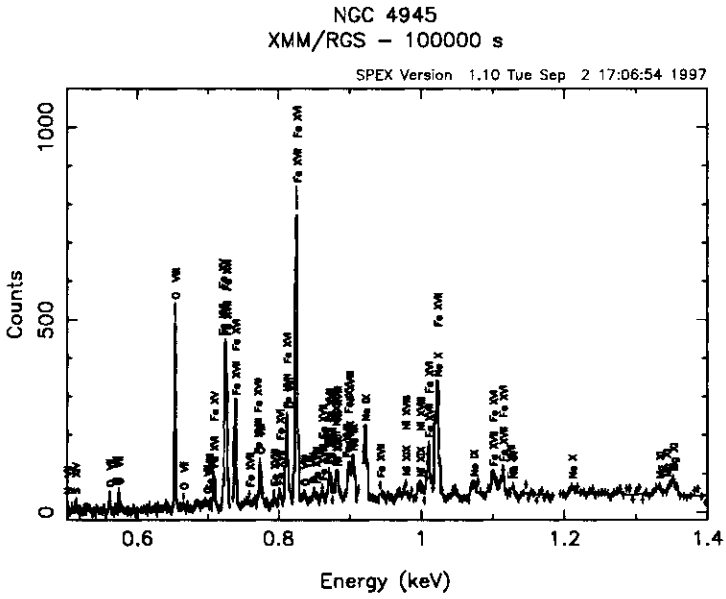

Fig. 29. Simulated RGS spectrum for the Seyfert 2 galaxy NGC 4945 (courtesy Jelle Kaastra, SRON).

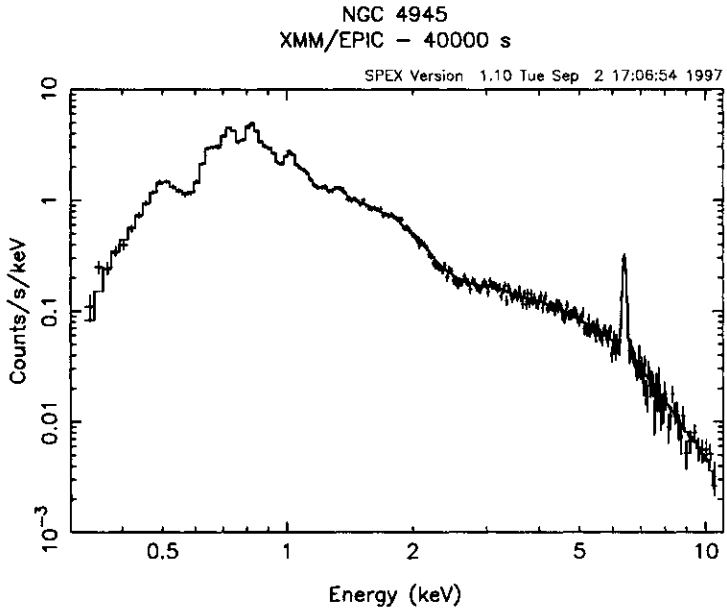

Fig. 30. Simulated EPIC spectrum for the Seyfert 2 galaxy NGC 4945 (courtesy Jelle Kaastra, SRON).

Finally, there is a Co/C multilayer which covers the band below the C K edge (0.175 – 0.280 keV, at resolving power $\mathcal{R} = 50$).

These resolving powers apply to observations of point sources, where one has to scan through a range of Bragg angles by moving the crystal panel with respect to the source direction; the telescope plus focal plane detector merely act as a very sensitive photometer. The resolving powers for the three highest-energy bands are plotted in Figure 31. As you can see, they are sufficient to detect almost all spectroscopic diagnostics we have discussed. The observations will be very slow, though: to scan the He-like S 'triplet' requires scanning a $\gtrsim$ 30 eV band at 2500 eV. At a resolving power of 3200, that requires $\gtrsim 30 \times (3200/2500) = 38$ separate exposures (Bragg angle steps)!

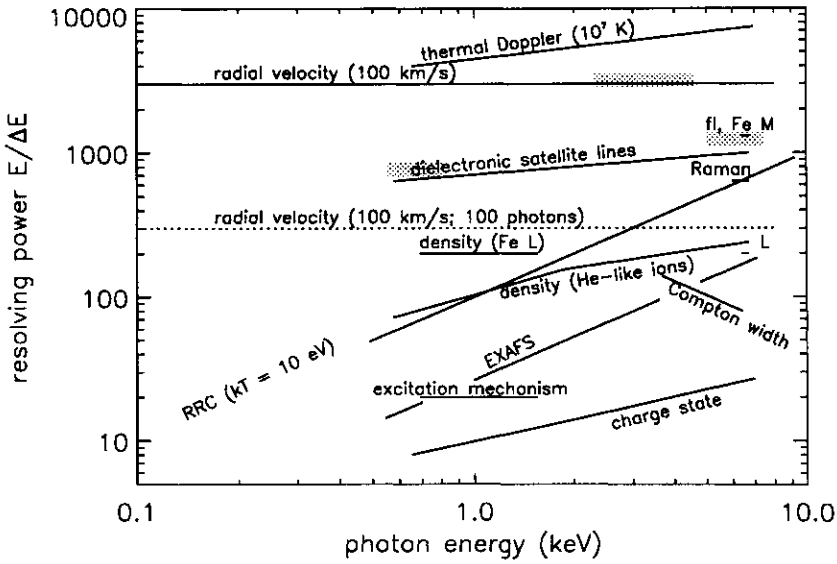

Fig. 31. Resolving powers for the three highest-energy bands of the OXS.

For an extended source, the OXS has an interesting multiplexing property. Light from different parts of the source strikes the crystals at different angles, which means that different photon energies are reflected and imaged from different parts of the source simultaneously (Figure 32). For a given orientation of the crystal/telescope, a weird image appears in the focal plane: it is monochromatic, but in different wavelengths in different spots! By changing the orientation, you can scan through a range of Bragg angles, and obtain a spectrum for every spot in the source. You cannot do this with a grating—since a grating diffracts all wavelengths at the same time, the image of an

extended source in a slitless spectrometer like the grating spectrometers on *AXAF* and *XMM* has the wavelength and spatial information completely scrambled. But note that now that the source is no longer a point source, your ability to separate two different wavelengths is effectively limited by the angular resolution of the system (crystal rocking curve plus telescope resolution), whereas the spectral resolving power for a point source is limited by the rocking curve only.

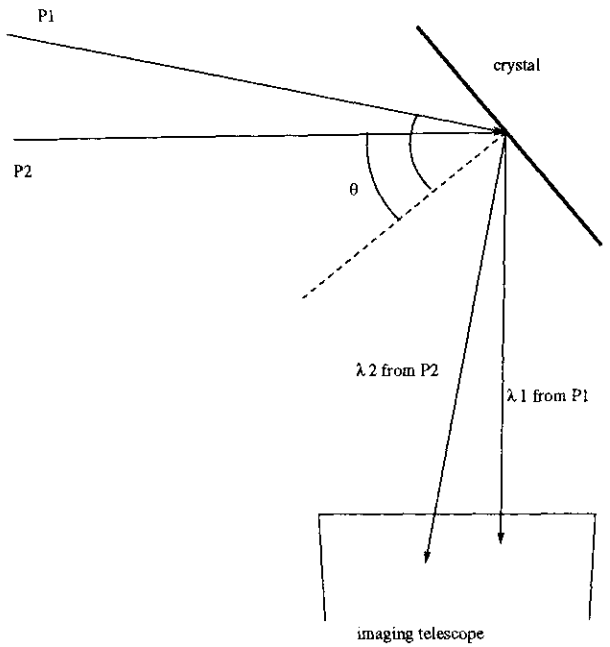

Fig. 32. Monochromatic imaging of an extended source with the OXS: light of wavelength λ_1 from direction P_1 is imaged in a separate spot from light of wavelength λ_2 from direction P_2. To record a spectrum for both P_1 and P_2, you again scan through a range of Bragg angles θ.

7 The Microcalorimeter Experiment on *Astro-E*

7.1 Introduction

The principle behind the operation of a microcalorimeter is simple: a photon is absorbed in a piece of material. Its energy is converted into heat, and the temperature rise is a measure of the photon energy (Moseley, Mather & McCammon 1984). An idealized detector consists of an absorber with a thermometer, connected to large heat bath at constant temperature.

A simplified theory of the device proceeds as follows. Assume the photon energy is instantly and completely converted into heat, and thermalized. The thermometer reacts instantly to this change in temperature. In the absence of incident X-ray photons the total energy of the system formed by the heat reservoir and absorber plus thermometer is constant, and the statistical properties of the entire system can be described with the microcanonical ensemble (Kittel 1958, Ch. 4). From this, the energy of any small subsystem, such as the X-ray absorber, is found to fluctuate spontaneously, and evidently there is a continuous exchange of energy between the subsystem and the heat reservoir. The subsystem is described by the canonical ensemble, and the calculation of the amplitude of the energy fluctuations is straightforward. For completeness, we will do the calculations in the next subsection. In this idealized instrument, the energy resolution is determined by the spontaneous thermodynamic fluctuations in the energy of the absorber.

But first, a subtle point. The photon energy is determined from the temperature rise in the absorber, so, strictly speaking, it is spontaneous fluctuations in the *temperature* of the absorber we want to know about. In a real microcalorimeter, energy flows back and forth across the link to the reservoir, and the total thermal energy in the absorber fluctuates. If the time constant associated with the heat conduction is long compared to the time scale for thermalization of the energy in the absorber, the temperature in the absorber will fluctuate. This situation is not described by the canonical ensemble, which assumes complete thermodynamic equilibrium. However, Mather (1982) has shown that as long as the temperature gradients are small, the simple thermodynamic result still applies, multiplied by a factor of order unity, which depends on the temperature dependence of the thermal conductivity of the link. For simplicity, we will ignore this factor here.

It will turn out that at very low temperatures the fluctuations in the energy of the absorber become very small (in fact, they vanish at zero temperature) due to the quantized nature of the vibrational motions (heat) in the absorber crystal.

7.2 Thermodynamic fluctuations

A system in thermodynamic equilibrium in contact with a large heat reservoir at constant temperature can be described by the canonical ensemble. The probability of finding a member of the ensemble, or a single system at a given instant in time, in a state n of total energy E_n is given by

$$p_n = \frac{e^{-\beta E_n}}{\sum_n e^{-\beta E_n}} \tag{40}$$

where $\beta \equiv 1/kT$. The quantity $\mathcal{Z}$,

$$\mathcal{Z} \equiv \sum_n e^{-\beta E_n} \tag{41}$$

is the partition function, which carries all the information about the thermodynamic properties of the system through the enumeration of the E_n.

We can derive an expression for the variance of the energy fluctuations as follows. Taking the derivative with respect to β, we find

$$\frac{\partial \mathcal{Z}}{\partial \beta} = -\sum_n E_n e^{-\beta E_n} \tag{42}$$

so that the average energy $\langle E \rangle$ of the system is

$$\begin{aligned} \langle E \rangle &= \frac{\sum_n E_n e^{-\beta E_n}}{\sum_n e^{-\beta E_n}} = \\ &= -\frac{1}{\mathcal{Z}} \frac{\partial \mathcal{Z}}{\partial \beta} \end{aligned} \tag{43}$$

The variance of the fluctuations in the energy is defined as

$$\begin{aligned} \langle \delta E^2 \rangle &\equiv \langle (E - \langle E \rangle)^2 \rangle = \\ &= \langle E^2 \rangle - \langle E \rangle^2 \end{aligned} \tag{44}$$

Taking another derivative, you find

$$\langle E^2 \rangle = \frac{1}{\mathcal{Z}} \frac{\partial^2 \mathcal{Z}}{\partial \beta^2} \tag{45}$$

so that

$$\langle \delta E^2 \rangle = \frac{1}{\mathcal{Z}} \frac{\partial^2 \mathcal{Z}}{\partial \beta^2} - \frac{1}{\mathcal{Z}^2} \left(\frac{\partial \mathcal{Z}}{\partial \beta} \right)^2 \tag{46}$$

which can be rewritten

$$\langle \delta E^2 \rangle = -\frac{\partial \langle E \rangle}{\partial \beta}, \tag{47}$$

or in terms of temperature,

$$\langle \delta E^2 \rangle = kT^2 \frac{\partial \langle E \rangle}{\partial T}. \tag{48}$$

Since the specific heat at constant volume, c_V, is defined as $c_V \equiv \partial \langle E \rangle / \partial T$, this can also be written as

$$\langle \delta E^2 \rangle = kT^2 c_V. \tag{49}$$

This is a general thermodynamic relation, valid regardless of the precise nature of the system, provided it is in equilibrium.

Now we have to calculate c_V for our system. As a model, we will assume a pure insulating crystal. In that case, the only contribution to the specific heat is due to elastic vibrations of the ions in the crystal lattice. As you remember, at high temperatures such that classical physics applies, c_V is a

constant because the average energy per degree of freedom for each particle is equal to $\frac{1}{2}kT$. Each vibrating particle has six degrees of freedom, so for N particles

$$c_V = \frac{\partial\langle E\rangle}{\partial T} = \frac{\partial(N \cdot 3kT)}{\partial T} = 3Nk. \tag{50}$$

But at low temperatures, the quantized nature of the vibrations becomes evident and with decreasing temperature, eventually you start noticing that more and more oscillators are in the ground state, and that the amount of heat absorbed or given up per degree of temperature rapidly declines with decreasing temperature. This happens when kT becomes comparable to or goes below the energies $\hbar\omega_0$ associated with the highest eigenfrequencies ω_0 of the vibrations of the solid. The associated temperature is referred to as the Debye temperature, Θ:

$$\Theta \equiv \hbar\omega_0/k. \tag{51}$$

Well below the Debye temperature, the heat capacity must rapidly decrease with decreasing temperature. It is this fact that makes the calorimeter an attractive spectrometer at low temperatures: the specific heat can be made very small at low temperatures, and hence the thermodynamic fluctuations in the energy can be made very small.

The standard calculation for the specific heat associated with the lattice vibrations of a solid at low temperatures goes as follows (e.g., Peierls 1955). First, you write down the coupled equations of motion for the particles in the crystal, and you apply boundary conditions to find the normal modes. One usually applies periodic boundary conditions, in which case the allowed values of the wavevector $\mathbf{f}$ form a lattice, with a density of $V/(2\pi)^3$, where V is the volume of the crystal. The eigenfrequencies are $\omega(\mathbf{f}, s)$. For a lattice containing r particles in the unit cell, there are in principle $3r$ solutions with the same $\mathbf{f}$ to the boundary value problem, each with a different frequency $\omega(\mathbf{f}, s)$. In the case of a simple lattice with only one particle in the unit cell, the three solutions belonging to the same $\mathbf{f}$ are two transverse and one longitudinal wave, in which case s simply labels the three possible polarizations. This remains true for more complicated unit cells, for very-long-wavelength (low-frequency) vibrations, which correspond to relative motions of entire, undistorted unit cells with respect to each other.

You can write the total energy (kinetic plus potential) of the crystal in terms of the canonical coordinates of the normal modes. You then interpret these coordinates as operators and apply the commutation rules to them. The resulting expression for the total energy now looks like a sum over discrete excitations of a set of independent harmonic oscillators:

$$E = \sum_{\mathbf{f}} \sum_{s} \left(n(\mathbf{f}, s) + \frac{1}{2} \right) \hbar\omega(\mathbf{f}, s), \tag{52}$$

with the oscillator of eigenfrequency $\omega(\mathbf{f}, s)$ excited into its n'th excited state. You can also interpret this expression as saying that there are $n(\mathbf{f}, s)$ discrete quanta of vibration, phonons, with energy $\hbar\omega(\mathbf{f}, s)$.

By calculating the average energy $\langle E \rangle$ for this set of phonons, we get the expression for c_V. The partition function, $\mathcal{Z}_1$, for a single harmonic oscillator is

$$\mathcal{Z}_1 = \sum_{n=0}^{\infty} e^{-\beta(n+\frac{1}{2})\hbar\omega} = \\ = \frac{e^{\frac{1}{2}\beta\hbar\omega}}{e^{\beta\hbar\omega} - 1} \tag{53}$$

so that the average energy, $\langle E_1 \rangle$, for one oscillator is

$$\langle E_1 \rangle = -\frac{1}{\mathcal{Z}_1}\frac{\partial \mathcal{Z}_1}{\partial \beta} = \frac{1}{2}\hbar\omega + \frac{\hbar\omega}{e^{\beta\hbar\omega} - 1} \tag{54}$$

and for the entire crystal

$$\langle E \rangle = \sum_{\mathbf{f},s}\left\{\frac{1}{2}\hbar\omega(\mathbf{f},s) + \frac{\hbar\omega(\mathbf{f},s)}{e^{\beta\hbar\omega(\mathbf{f},s)} - 1}\right\}. \tag{55}$$

The first term is the familiar zero point energy of a harmonic oscillator summed over all modes, the second term is the thermal energy E_{T} of the crystal (kinetic plus potential energy of vibration):

$$E_{\mathrm{T}} = \sum_{\mathbf{f},s} \frac{\hbar\omega(\mathbf{f},s)}{e^{\beta\hbar\omega(\mathbf{f},s)} - 1}. \tag{56}$$

The evaluation of this sum is difficult, because it requires knowing $\omega(\mathbf{f}, s)$ for every value of $\mathbf{f}, s$. You can get an approximate solution, valid at very low temperatures, as follows (Peierls 1955). First, you argue that at very low T only very low frequency modes are excited, and those are the long-wavelength sound wave like excitations we mentioned above. Assume that the sound velocity c_{s} is constant at very-low-frequency, $\omega(\mathbf{f}, s) = c_{\mathrm{s}}(\theta, \phi)f$, where θ, ϕ indicate the direction of $\mathbf{f}$ (for each of the three polarizations $s = 1, 2, 3$). Replace the summation in the expression for E_{T} by an integral over $\mathbf{f}$-space:

$$E_{\mathrm{T}} = \frac{V}{(2\pi)^3}\sum_{s=1}^{3}\int_0^{\infty} f^3 df \int\int d\Omega \frac{\hbar c_{\mathrm{s}}(\theta,\phi)}{e^{\beta\hbar c_s f} - 1} \tag{57}$$

where $d\Omega$ is an element of solid angle, and the integral has been extended up to infinity, because at low temperatures only vibrations with low frequencies are excited, and the high-frequency part of the normal mode spectrum does not contribute. Change variables to $x = \beta\hbar c_{\mathrm{s}}(\theta, \phi)f$:

$$E_{\mathrm{T}} = \frac{V(kT)^4}{(2\pi\hbar)^3}\sum_{s=1}^{3}\int\int \frac{d\Omega}{c_{\mathrm{s}}^3}\int_0^{\infty} dx \frac{x^3}{e^x - 1}. \tag{58}$$

The integral over x is equal to $\pi^4/15$. The integral and sum of c_s^{-3} still require knowing the elastic constants of the crystal. Peierls introduces a mean effective sound velocity by

$$\sum_{s=1}^{3} \int\int \frac{d\Omega}{c_s^3} \equiv \frac{12\pi}{c_{\rm eff}^3} \tag{59}$$

and with this

$$E_{\rm T} = \frac{\pi^2 (kT)^4 V}{10\hbar^3 c_{\rm eff}^3}. \tag{60}$$

Since $E_{\rm T}$ depends on temperature as T^4, the specific heat depends on T as T^3, indeed a steep function of T, and c_V goes to zero as T goes to zero.

To do a practical calculation, we need an estimate for $c_{\rm eff}$. Debye has suggested a way of expressing this parameter in terms of the empirically easier to characterize parameter Θ. In transforming Eq.(56) to an integral over **f**-space, he also took c_s to be constant, even for high frequencies, and independent of polarization. The number of independent normal modes in a crystal can never be larger than three times the number of particles in the system. So, given that three polarizations are associated with each allowed **f**, the integral over **f**-space that replaces the sum over normal modes should only be extended up to a value of f such that the volume in **f**-space contains just Nr **f**-space lattice points, with N the number of unit cells (and r the number of ions in the unit cell). The corresponding radius f_0 of this sphere in **f**-space is thus found from the condition

$$\frac{4\pi}{3} f_0^3 \frac{V}{(2\pi)^3} = Nr, \tag{61}$$

and f_0 is then the highest frequency in the problem. With the definition of the Debye temperature, Θ,

$$\Theta = \frac{\hbar\omega_0}{k} = \frac{\hbar c_s f_0}{k} = \\ = \frac{\hbar c_s}{k} \left(\frac{6\pi^2 Nr}{V} \right)^{\frac{1}{3}}. \tag{62}$$

Expressing c_s in terms of Θ, and substituting for $c_{\rm eff}$ in Eq.(60), we get for the thermal energy at low temperature

$$E_{\rm T} = \frac{3\pi^4}{5} Nrk T \left(\frac{T}{\Theta} \right)^3, \tag{63}$$

and for the specific heat

$$c_V = \frac{12\pi^4}{5} kNr \left(\frac{T}{\Theta} \right)^3. \tag{64}$$

The Debye temperature is an empirical parameter, characteristic of the crystal material, to be determined experimentally from a fit to measured values of c_V as a function of T.

For the r.m.s. fluctuation in the energy of the crystal we finally obtain

$$\langle \delta E^2 \rangle^{\frac{1}{2}} = \left(\frac{12\pi^4}{5} \frac{k^2 N r}{\Theta^3} \right)^{\frac{1}{2}} T^{\frac{5}{2}}. \tag{65}$$

To get a very rough idea what this could work out to in practice, consider a piece of Si, of volume 1 mm^3, at a temperature of 0.1 Kelvin. The density of Si is 2.33 g cm^{-3}, its Debye temperature is $\Theta = 640$ K. It is customary to quote the FWHM energy resolution, which is equal to 2.35 times the r.m.s. resolution. In this sense, the energy resolution is 4.2 eV! And notice there is no fundamental reason it couldn't be smaller—just lower the number of particles and/or the temperature.

Let us take a short break, and examine the expression for the r.m.s. fluctuation in the energy of the quantized harmonic oscillator. With very little work, we can demonstrate a result of profound significance for bosonic quantum systems, too beautiful not to highlight here. From Eq.(47) we had $\langle \delta E^2 \rangle = -\partial \langle E \rangle / \partial \beta$, which you can write as an equation for the mean square fluctuation $\langle \delta n^2 \rangle$ in the average occupation number $\langle n \rangle$ of a harmonic oscillator:

$$\langle \delta n^2 \rangle = -\frac{1}{\hbar\omega} \frac{\partial \langle n \rangle}{\partial \beta}. \tag{66}$$

Substitute

$$\langle n \rangle = \frac{1}{e^{\beta\hbar\omega} - 1} \tag{67}$$

and you get

$$\begin{aligned} \langle \delta n^2 \rangle &= -\frac{1}{\hbar\omega} \frac{\partial}{\partial \beta} \left(\frac{1}{e^{\beta\hbar\omega} - 1} \right) = \\ &= \frac{e^{\beta\hbar\omega}}{(e^{\beta\hbar\omega} - 1)^2} = \\ &= \frac{e^{\beta\hbar\omega} - 1}{(e^{\beta\hbar\omega} - 1)^2} + \frac{1}{(e^{\beta\hbar\omega} - 1)^2} = \\ &= \langle n \rangle + \langle n \rangle^2. \end{aligned} \tag{68}$$

This seemingly innocent result has a profound implication. In the classical regime, we have $\hbar\omega \ll kT$, so $\langle n \rangle \approx kT/\hbar\omega \gg 1$, and in that case evidently $\langle \delta n^2 \rangle \approx \langle n \rangle^2$. The fractional fluctuation in the number of excitations is then $\langle \delta n^2 \rangle / \langle n \rangle^2 \approx 1$, which is the result for the fractional mean square fluctuation in the energy density of a random superposition of plane waves (e.g., Longair 1984, p. 245). In the quantum regime, however, the other term dominates, and the expression for the mean square fluctuations reduces to the

usual Poisson counting statistics expression for discrete particles. In general, therefore, the thermodynamic properties of vibrations in a crystal indicate that the vibrations exhibit both a wave- and a particle-like character!

Since this is just a property of the quantized harmonic oscillator, the same is true for a box filled with pure radiation in thermodynamic equilibrium. What produces the result is that the harmonic oscillator in thermodynamic equilibrium has a Planck spectrum (Eq.(67)). Einstein used Eq.(48) in 1909 to point out that the empirical fact that radiation in thermodynamic equilibrium has a Planck spectrum implies that, thermodynamically, radiation behaves as if it has both a wave- and a particle-like character (Pais 1982; Longair 1984). In fact, in the 1905 paper on the photoelectic effect (as it is usually referred to), he had already used a different thermodynamic relation to show that the Wien spectrum led to the inference that 'monochromatic radiation of low density behaves thermodynamically as though it consisted of a number of independent energy quanta of magnitude $[h\nu]$'. Equation (68) holds for the Bose-Einstein distribution in general, and, even more miraculously, in 1925, before the advent of wave mechanics, Einstein showed that the thermodynamic fluctuation properties of a Bose gas of material particles also implied that the 'particles' exhibit both particle- and wave-like characteristics!

The above digression is actually relevant to the theory of the microcalorimeter. The qualitative argument is sometimes made that you can estimate the amplitude of the thermodynamic energy fluctuations in the absorber as follows. There are of order $N \sim c_V/k$ modes in the crystal, the typical mode has occupation number $\langle n \rangle \sim 1$ and variance $\langle \delta n^2 \rangle \sim 1$, and each mode carries of order kT energy. Hence,

$$\langle \delta E^2 \rangle \sim N \cdot \langle \delta n^2 \rangle \cdot (kT)^2 \sim \frac{c_V}{k} \cdot 1 \cdot k^2 T^2 = \\ = kT^2 c_V . \tag{69}$$

This is somewhat misleading. At high T, the average occupation number of each mode is large, but the variance in the number is also large, and equal to $\langle n \rangle^2$, so the squared fractional fluctuation is equal to unity, and the correct estimate is

$$\langle \delta E^2 \rangle \sim N \cdot \frac{\langle \delta n^2 \rangle}{\langle n \rangle^2} \cdot (kT)^2 \sim \frac{c_V}{k} \cdot 1 \cdot k^2 T^2 = \\ = kT^2 c_V . \tag{70}$$

But at low T, $\langle \delta n^2 \rangle \approx \langle n \rangle$, and the average energy per mode is *not* kT, but rather $\sim \hbar\omega_i \exp(-\beta\hbar\omega_i)$ (with ω_i the eigenfrequency of the mode), and you don't get the right answer this way because the oscillations are quantized with most oscillators in the ground state, and because the particle oscillations are not statistically independent.

This also brings us to the question of the ultimate energy resolution of an ideal microcalorimeter. At very low T, or with a very small device (small number of modes), there are only very few phonons present in equilibrium, and the fractional (quantum) fluctuations are large. But Eq.(49) still holds, that is just thermodynamics (provided the time scale for thermalization of the phonon population is short compared to other timescales in the problem). Since c_V is now very small, the fluctuations only carry tiny amounts of energy, and the resolving power is still high (and given by Eq.(49)). Presumably, however, at some point you will start to notice the effect of the statistical spread in the number of phonons excited by the absorption of a single photon, and how fast a given phonon population evolves (e.g., relaxes to a Planck distribution). The resolving power averaged over a large number of identical photons will still be given by thermodynamics (if the thermalization is sufficiently rapid), but for a single photon of unknown energy (and that is what you build the spectrometer for, to determine uniquely the energy of each photon), the resolution will degrade.

7.3 An alternative derivation

The thermodynamic fluctuations we worked out in the previous section provide a useful measure of the energy resolution of the microcalorimeter. In this section we will work out an alternative derivation of the energy resolution that is actually physically more correct, in that it explicitly models how the actual signal of the temperature pulse induced by the absorption of a photon is detected against the noise in the absorber. This treatment is also the starting point for the calculation of the properties of actual devices (including amplifier noise, etc.). As it turns out, the derivation yields a surprise: the 'thermodynamic limit' we derived above is not a real limit on the energy resolution!

We start again from the fluctuation formula Eq.(49). We argue that the energy fluctuations $\langle \delta E^2 \rangle$ correspond to temperature fluctuations:

$$\langle \delta T_{\rm n}^2 \rangle = \frac{\langle \delta E^2 \rangle}{c_V^2} = \frac{kT^2}{c_V} \tag{71}$$

where the subscript 'n' indicates that $\delta T_{\rm n}$ is the noise fluctuation in the temperature due to the spontaneous energy fluctuations. This is a somewhat questionable step, as mentioned earlier—the fluctuation formula for $\langle \delta E^2 \rangle$ only holds in thermodynamic equilibrium, but we will assume again that the equality Eq.(71) is good to order unity. For a complete treatment, thermodynamics doesn't work, and you would have to resort to solving the Boltzmann equation to study the energy transport and thermalization processes!

When analyzing the signal pulse shape and the noise properties, it will be convenient to work in frequency space, rather than in the time domain. The principles of this type of analysis, common in detector physics, can be found

in various texts; I have used Houghton & Smith (1966, Ch. 5). The calculation goes as follows: we will derive the power spectral density of the temperature noise due to the thermodynamic fluctuations, and the power spectral density of the temperature signal induced by the absorption of an X-ray photon. The ratio, at each frequency, of these two spectra, is a 'monochromatic' signal-to-noise ratio. Integrating this signal-to-noise ratio over all frequencies, we obtain the total signal to noise ratio, and the ultimate energy resolution obtainable in the presence of noise.

We will analyze the noise first. We start by writing down the energy conservation law for the absorber:

$$\frac{d(\delta E)}{dt} = c_V \frac{d(\delta T)}{dt} = W - G\delta T \tag{72}$$

where δT is the difference in temperature of the absorber and the heat reservoir, and W is any external source of power. The term $G\delta T$ describes the conduction of heat from the absorber to the heat reservoir; G is the thermal conductance of the link (in units Watt Kelvin^{-1}).

Eq.(72) is most conveniently solved in frequency space. But if we apply the equation to the fluctuations $\delta T_{\rm n}$ caused by a source of noise power $W_{\rm n}$, we have to be careful with the Fourier transforms; the usual definition of the transform, $\widehat{\delta T_{\rm n}}(\omega)$, of $\delta T_{\rm n}(t)$,

$$\widehat{\delta T_{\rm n}}(\omega) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} {\rm d}t\ e^{-i\omega t} \delta T_{\rm n}(t), \tag{73}$$

does not converge. Instead, assume that $\delta T_{\rm n}(t)$ is non-zero only for a finite time interval, $t = -a \rightarrow +a$, where a is chosen large, but finite. Then *define* the function (see for example Bennett 1956)

$$\widehat{\delta T_{\rm n}}(\omega, a) \equiv \frac{1}{\sqrt{2\pi}} \int_{-a}^{a} {\rm d}t\ e^{-i\omega t} \delta T_{\rm n}(t). \tag{74}$$

Then, for $-a < t < a$,

$$\delta T_{\rm n}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} {\rm d}\omega\ \widehat{\delta T_{\rm n}}(\omega, a) e^{i\omega t}. \tag{75}$$

The definite integral of $\delta T_{\rm n}^2(t)$ is

$$\begin{aligned}
\int_{-a}^{a} {\rm d}t\ \delta T_{\rm n}^2(t) &= \int_{-a}^{a} {\rm d}t\ \delta T_{\rm n}(t) \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} {\rm d}\omega\ \widehat{\delta T_{\rm n}}(\omega, a) e^{i\omega t} = \\
&= \int_{-\infty}^{\infty} {\rm d}\omega\ \widehat{\delta T_{\rm n}}(\omega, a) \cdot \frac{1}{\sqrt{2\pi}} \int_{-a}^{a} {\rm d}t\ \delta T_{\rm n}(t) e^{i\omega t} = \\
&= \int_{-\infty}^{\infty} {\rm d}\omega\ \widehat{\delta T_{\rm n}}(\omega, a) \cdot \widehat{\delta T_{\rm n}}^{*}(\omega, a) = \\
&= \int_{-\infty}^{\infty} {\rm d}\omega\ \left|\widehat{\delta T_{\rm n}}(\omega, a)\right|^2,
\end{aligned} \tag{76}$$

so the variance of the temperature noise can be written as the integral over a spectral density,

$$\langle \delta T_{\rm n}^2 \rangle = \frac{1}{2a} \int_{-a}^{a} {\rm d}t \; \delta T_{\rm n}^2(t) = \int_{-\infty}^{\infty} {\rm d}\omega \; \frac{\left| \widehat{\delta T_{\rm n}}(\omega, a) \right|^2}{2a}. \tag{77}$$

You can find a logically rigorous definition of the spectral density in Kittel (1958, Ch. 28).

Now we are ready to solve Eq. (72) for the noise; in terms of the transforms $\widehat{\delta T_{\rm n}}(\omega, a)$ and $\widehat{W}_{\rm n}(\omega, a)$, we get

$$\widehat{\delta T_{\rm n}}(\omega, a) = \frac{\widehat{W}_{\rm n}(\omega, a)}{G + i\omega c_V}, \tag{78}$$

from which we find

$$\begin{aligned} \left| \widehat{\delta T_{\rm n}}(\omega, a) \right|^2 &= \frac{\left| \widehat{W}_{\rm n}(\omega, a) \right|^2}{G^2 + \omega^2 c_V^2} = \\ &= \frac{\left| \widehat{W}_{\rm n}(\omega, a) \right|^2}{G^2 (1 + \omega^2 \tau^2)} \end{aligned} \tag{79}$$

with $\tau \equiv c_V / G$. The interpretation of τ is obvious from solving Eq.(72) in the absence of an external power source: the equation

$$c_V \frac{{\rm d}(\delta T)}{{\rm d}t} = -G \delta T \tag{80}$$

has the solution

$$\delta T(t) = \delta T_0 \; e^{-t/\tau} \tag{81}$$

with $\tau = c_V / G$, so τ is the thermal relaxation time of the system. Eq.(79) says the same thing: if the power source W varies rapidly, on time scales short compared to τ (frequencies high compared to $1/\tau$), the amplitude of the temperature change goes down due to the finite relaxation time.

Now we argue that $\left| \widehat{W}_{\rm n}(\omega, a) \right|^2$ is independent of frequency, that is, that the power associated with the thermodynamic fluctuations has a white noise spectrum. Now integrate $\left| \widehat{\delta T_{\rm n}}(\omega, a) \right|^2 / 2a$ over all frequencies, and set the result, $\langle \delta T_{\rm n}^2 \rangle$, equal to kT^2 / c_V:

$$\frac{\left| \widehat{W}_{\rm n}(\omega, a) \right|^2}{2a} = \frac{kT^2 G}{\pi}. \tag{82}$$

This result is usually quoted in terms of a spectral density of the noise power, W_f, defined such that

$$\langle W_{\rm n}^2 \rangle \equiv \int_0^{\infty} {\rm d}f \; W_f \tag{83}$$

with f the linear frequency ($\omega = 2\pi f$); with that definition

$$W_f = 4kT^2G. \tag{84}$$

For the spectral density of the temperature fluctuations we get

$$\frac{\left|\widehat{\delta T_{\rm n}}(\omega, a)\right|^2}{2a} = \frac{kT^2}{\pi G(1+\omega^2\tau^2)}. \tag{85}$$

Now we need to derive the spectral density of the temperature signal due to the absorption of a photon. Assume that the photon is absorbed at $t = 0$, by an ideal detector. The temperature pulse shape will look like a sharp spike of amplitude $\delta T_{\rm X,0} = E_\gamma/cV$, with E_γ the photon energy, at $t = 0$, followed by exponential decay as the heat leaks away to the reservoir:

$$\delta T_{\rm X}(t) = \delta T_{\rm X,0}\, e^{-t/\tau}. \tag{86}$$

Its Fourier transform, in the sense of Eq.(74) is

$$\begin{aligned}\widehat{\delta T_{\rm X}}(\omega, a) &= \frac{\delta T_{\rm X,0}}{\sqrt{2\pi}} \frac{1 - e^{-(i\omega+1/\tau)a}}{i\omega + 1/\tau} \cong \\ &\cong \frac{\delta T_{\rm X,0}}{\sqrt{2\pi}} \frac{1}{i\omega + 1/\tau}\end{aligned} \tag{87}$$

provided $a \gg \tau$. We have

$$\left|\widehat{\delta T_{\rm X}}(\omega, a)\right|^2 = \frac{\delta T_{\rm X,0}^2}{2\pi} \frac{\tau^2}{1+\omega^2\tau^2}. \tag{88}$$

When we take the ratio of Eq.(88) to the spectral density of the noise temperature fluctuations, we get the spectral density of the square of the signal-to-noise ratio, $\rho^2(\omega)$ (dimension: Hz^{-1}):

$$\rho^2(\omega) = \frac{\delta T_{\rm X,0}^2 G\tau^2}{2kT^2}. \tag{89}$$

And this is the surprise: because the shape of the spectral densities of the noise and signal temperature variations are identical (actually, not surprising: signal and noise satisfy the same energy conservation equation), the spectral density of the signal to noise ratio is constant. When you integrate $\rho^2(\omega)$ over frequency, you get the total (square of the) signal to noise ratio, in our case equal to the square of the resolving power $\mathcal{R}$, and the integral diverges, which implies that the resolving power is infinite!

You can understand this remarkable result in another way, in the time domain. Superimposed on the randomly fluctuating temperature, you see a sharp spike right when a photon is absorbed. You make a naive measurement of the photon energy by measuring the amplitude of the temperature rise right at the start of the pulse. This number has a statistical fluctuation in

it due to the presence of the noise. But as the pulse decays, you can of course make another measurement, and another, and another, correcting the measurements for the known decay of the pulse with time constant τ. The noise averages out, and in fact goes to zero in the limit that you can take infinitely many samples infinitely closely spaced in time. And this is allowed, because we have said that the detector was ideal: it reacts instantaneously to changes in temperature of the absorber (or: it has an infinite frequency bandwidth), so noise samples taken over infinitesimal time intervals are all statistically independent.

Had you designed the circuitry such that you integrated $\rho^2(\omega)$ only up to the cutoff frequency $1/\tau$, you would have found

$$\mathcal{R}^2 = \frac{\delta T_{\mathrm{X},0}^2 G \tau^2}{2kT^2} \int_{-1/\tau}^{1/\tau} \mathrm{d}\omega = \\ = \frac{E_\gamma^2}{kT^2 c_V}, \tag{90}$$

and this is precisely the result we derived in the previous section. It is now clear how 'imperfections' in the detector can be incorporated into the calculation. For instance, the presence of another source of noise can be incorporated by adding its spectral density in quadrature to the density of the thermodynamic fluctuations. If this other noise source also has a white power spectrum, the effect will be to just lower the resolving power by a constant number. If there is another time constant in the problem, for instance, associated with a finite time for thermalization of the photon energy, this changes the boundaries of the integral in Eq.(90), again changing the resolving power by a constant factor.

To obtain a *really* complete description of the performance of the microcalorimeter, we need to write down the coupled equations for the thermal and electrical properties of the system, which take the place of the single thermal balance equation Eq.(72). The coupling between these equations of course arises from terms which describe the ohmic heating of the detector, the change in electrical resistance of the thermometer in response to changes in the temperature, etc. The resulting change in the properties of the detector with respect to the simple properties we derived from Eq. (72) is referred to as 'electrothermal feedback'. Moseley, Mather & McCammon (1984) provide such a description in terms of complex impedance theory; Labov et al. (1997) present a derivation based on the differential equations. It is instructive to solve the equations for the simplified case of a constant voltage on the thermometer, assuming that the circuit only contains this thermometer, of resistance R, and the bias power source. Then $W = I^2R$ in Eq.(72), and $d(IR)/dt \equiv 0$. Taking the time derivative of Eq. (72), and neglecting small terms, you find

$$c_V \frac{\mathrm{d}^2(\delta T)}{\mathrm{d}t^2} + \left(G + \frac{W_0}{T_0}\alpha\right)\frac{\mathrm{d}(\delta T)}{\mathrm{d}t} = 0 \tag{91}$$

that is, the same equation as before, but with an effective time constant

$$\frac{1}{\tau_{\rm eff}} = \frac{1}{\tau_{\rm th}}\left(1 + \frac{W_0}{GT_0}\alpha\right) \tag{92}$$

with $\tau_{\rm th} = c_V/G$ the 'thermal time constant' which characterizes the 'bare' thermal problem. Here, $\alpha \equiv {\rm d}(\ln R)/{\rm d}(\ln T)$, which measures the steepness of the thermometer response to changes in temperature. This suggests integrating up to a frequency $1/\tau_{\rm eff}$, in which case the spectral resolving power will exceed the 'thermodynamic limit' by a factor $\left(1 + (W_0/GT_0)\alpha\right)^{1/2}$: a more sensitive thermometer (bigger α) is better.

In general, therefore, the uncertainty in the determination of the photon energy will be approximately

$$\left\langle \delta E^2 \right\rangle = \xi k T^2 c_V \tag{93}$$

with ξ a pure number that depends on the strength of additional noise sources, other relaxation time constants (inclusing electrical), and, implicitly, on the thermometer response and operating conditions of the entire detector system.

Finally, traditionally in detector physics, one does not calculate with the spectral density of the temperature, but of course with that of the (square of the) physical power delivered by the various components of the detector system. You can redo the above calculation in these terms, remembering that the absorption of a photon in an ideal detector delivers a power spike in the form of a delta function (which has a flat power spectrum, just like the noise), normalized such that the time integral over the spike is equal to the photon energy. In this connection, one often uses the concept of the 'noise equivalent power' (NEP), defined such that the NEP is the amount of power delivered to the detector at frequency f equal to the noise power at that frequency in a 1 Hz bandwidth. For the thermodynamic, or phonon, noise, we had the spectral density of the noise power $W_f = 4kT^2G$, so for this noise source,

$${\rm NEP}_{\rm phonons} \equiv W_f^{\frac{1}{2}} = (4kT^2G)^{\frac{1}{2}} \tag{94}$$

with dimension $\mathrm{W}/\sqrt{\mathrm{Hz}}$. The energy resolution in these terms is

$$\left\langle \delta E^2 \right\rangle = \tau^2 \int_0^{f_{\rm max}} {\rm d}f \ \left|{\rm NEP}_{\rm phonons}(f)\right|^2 . \tag{95}$$

7.4 The microcalorimeter on *Astro-E*

In this section, we will look at some of the features of the operation of a real first-generation X-ray microcalorimeter, the X-ray Spectrometer (XRS) on *Astro-E*. Information on *Astro-E*, a joint Japan-US project, can be found on `http://heasarc.gsfc.nasa.gov/docs/frames/astroe_about.html`; a recent conference paper can be found on `http://wwwvms.mppmu.mpg.de/ltd7/`

`contribute/by_field/C.htm` (Stahle et al. 1997). It carries five telescopes, four with a CCD focal-plane detector, one with the microcalorimeter; there is also a hard X-ray instrument. The microcalorimeter has been developed in collaboration by NASA/Goddard Space Flight Center and the University of Wisconsin (PI: Richard Kelley, GSFC). Launch is in early 2000.

The microcalorimeter consists of 6×6 separate pixels, covering $3' \times 3'$ on the sky (the telescope angular response is of the order of 2 arcmin HEW). The X-ray absorbing element is HgTe, attached to a Si substrate with an implanted thermistor as thermometer. The thermal time constant for the conduction of heat between the absorber and the heat bath is approximately 80μs, so that the detector can count of order 10 counts/sec per pixel without (difficult to discriminate) overlapping pulses. For bright sources, neutral density filters can be inserted in the X-ray beam to reduce count rates.

The microcalorimeter is inside a dewar. The instrument itself is cooled to 65 mK by an Adiabatic Demagnetization Refrigerator, located inside a liquid He dewar, which in turn is located inside a solid Ne dewar. The lifetime is determined by the cryogenics, and is expected to be of order two years. During this initial phase of the mission, priority will be given to microcalorimeter observations.

The energy resolution is approximately 12 eV, while 7 eV has been seen in individual array elements. The microcalorimeter itself has good quantum efficiency (basically X-ray absorption optical depth) up to 10 keV. The low end of the bandpass is determined by the X-ray transmission of a set of filters that prevent radiation from space from heating the instrument, which limits the band to roughly $E \gtrsim 400$ eV.

The performance of the XRS in terms of spectral resolving power, and the spectral diagnostics accessible with it, was plotted in Fig. 8. The XRS will have most of its impact at high energies, specifically in the Fe K band. Novel (for astrophysics) diagnostics that will become available at Fe K are the unambiguous determination of the (distribution of) ionization stages of fluorescing Fe, as long as it has been ionized up into the L-shell, from a direct measurement of the energies of the Fe K fluorescent lines; direct detection of the Compton recoil spectrum of Fe K photons scattered by cold electrons, and possibly even detection of Raman scattered Fe K photons. The powerful hot plasma diagnostics derived from spectroscopy of the dielectronic satellites may just be accessible, provided the spectrum is bright enough.

8 The 21st Century

Astrophysical considerations suggest that after *AXAF*, *XMM*, and *Astro-E*, extending the spectroscopic capability to cosmologically interesting redshifts is the most important thing to do, rather than, for instance, increasing spectral resolving power for its own sake.

X-ray line emission arises predominantly in optically thin plasmas (although significant optical depths may occur in resonance lines), and the emission line power is relatively easy to calculate as a function of the local variables. In addition, the X-ray band is so wide that it simultaneously covers the K transitions of all ionization stages of all the elements from C up to Fe. This makes, for instance, absolute abundance determinations easy, because they don't have to rely on model-dependent corrections for the abundances of unseen ionization stages (as is the case in the optical and UV bands). The importance of X-ray spectroscopy to nucleosynthesis, stellar evolution, and formation and evolution of galaxies is obvious. As another example, detailed information on the physical conditions in clusters at significant redshifts (distribution of temperature, density, and abundances) has direct implications for the evolution of large-scale structure and the cosmological parameters. Galactic astrophysics will of course also benefit. To name just one example, detailed photospheric spectroscopy of hot neutron stars will finally constrain the mass-radius relation for these objects, and hopefully will uniquely constrain the equation of state at supranuclear densities.

Obviously, what is required is a spectrometer, preferably imaging, with resolving power no less than *AXAF*, *XMM*, and *Astro-E* (so at least $\mathcal{R} \gtrsim$ few hundred), and a huge effective area. For the study of redshifted objects, it is crucial that these characteristics extend to low photon energies (the oxygen Lyα line, to name an example, shifts to ~ 300 eV at $z \sim 1$). These requirements naturally suggest lightweight, medium angular resolution optics, with as large a combined area as you can afford, with an imaging spectrometer at the focal plane. The natural (or only) candidate for the latter is, of course, a microcalorimeter array, with significantly improved energy resolution.

As you can see from Fig. 8, the $\sim$ 12 eV characteristic of the *Astro-E* spectrometer is not good enough for detailed spectroscopy at low energies. Hence, work is in progress to construct $\Delta E \sim 2$ eV microcalorimeters. The most important innovation is the use of an extremely sensitive thermometer, based on a superconducting element that is held at a temperature in the middle of the transition between the superconducting and normal states. Hence, its resistance is extremely sensitive to changes in temperature. Thermometers of this type are referred to as Transition Edge Sensors (TES). Recently, a resolution of 4 eV was already demonstrated with such a device by a group at the National Institute of Standards and Technology (NIST, Boulder, Colo.; see *Physics Today*, July 1998, p. 19).

NASA is currently funding work on the *Constellation X* project, a fleet of six identical observatories, with a combined effective area of order 15,000 cm^2 at 1 keV and 6,000 cm^2 at 6 keV. The proposed instrument package includes $\Delta E = 2$ eV microcalorimeter arrays, reflection grating spectrometers, and hard X-ray telescopes that will be sensitive up to at least 40 keV (1500 cm^2 at 40 keV).

The grating spectrometers complement the microcalorimeters at low energies, where their sensitivity declines due to declining resolving power and strong X-ray absorption in the necessary long-wavelength radiation filters. The design of the spectrometers is similar to those on *XMM*, but the gratings could be of a novel design: lightweight thin Si films, etched along crystal planes to achieve an extremely smooth and accurate groove shape. In order to increase the graze angles and allow a higher resolving power (Eq. 31) without loss of reflectivity, multi-layer coatings may be applied to the gratings.

A prime motivation for including sensitive hard X-ray telescopes is the study of AGN spectra. The study of these spectra with spectrometers that cut off at 10 keV has shown that it is important to have the ability to characterize the hard continuum spectrum, to allow for an accurate extrapolation to lower energies, where the continuum is strongly modulated by absorption and the presence of emission line complexes. In addition, one hopes to study heavily absorbed type II objects (Seyfert 2 galaxies and, if they are ever found, the analogous 'type 2 QSO's') with much higher sensitivity. And there are, of course, numerous other novel applications.

The European *XEUS* project (*X-ray Early Universe Spectrometer*) pursues similar goals. The design currently revolves around a truly giant (10 m^2 aperture) single X-ray telescope, which would be assembled in sections over the years at the space station. Operations can begin as soon as the core of this assembly has been placed in orbit. A telescope with such a large area needs to have a very long focal length, in order to keep graze angles small and reflectivity high up to high energies; the current figure is $F \sim 50$ m! With such an extreme focal length, building a 'monolithic' telescope becomes more expensive than simply putting focusing optics and focal plane instrument on two separate satellites that track each other, and this is the current plan. The focal-plane satellite will be designed such that its instrumentation can be exchanged. *XEUS* might start out with CCD detectors, and then move on to microcalorimeter arrays.

Acknowledgements

I would like to express my gratitude to the following people, for conversations, reading of the draft manuscript, or permission to use their data: Duane Liedahl (LLNL), Steve Kahn, Andy Rasmussen, Masao Sako, Joshua Spodek, and Jean Cottam (Columbia University), Bert Brinkman, Theo Gunsing, Jan van Rooijen, Jelle Kaastra, Rolf Mewe, and Piet de Korte (SRON), John Houck, Dan Dewey, John Davis, and the entire HETGS group (MIT), and Steve Holt (GSFC). Finally, I would like to thank the organizers of the School, Johan Bleeker and Jan van Paradijs, for their invitation to teach the lectures, and Jan van Paradijs for his most generous hospitality during the School.

References

Beiersdorfer, P., Schneider, M.B., Bitter, M. & von Goeler, S. (1992): Rev. Sci. Instrum. 63(10), 5029

Bennet, W.R. (1956): Proc. IRE 44, 609; reprinted in M. S. Gupta: *Electrical Noise: Fundamentals and Sources*, (IEEE Press: New York) (1977)

Beuermann, K.P., Bräuniger, H. & Trümper, J. (1978): Appl. Optics 17, 2304

Bixler, J.V. et al. (1991): Proc. SPIE 1549, 420

Born, M. & Wolf, E. (1959): *Principles of Optics*, (Pergamon: London).

Bowyer, S. & Malina, R.F. (1996): *Astrophysics in the Extreme Ultraviolet*, Proc. IAU Coll. 152, Berkeley, March 1995 (Kluwer:Dordrecht)

Brinkman, A.C. et al. (1980): Appl. Optics 19, 1601

Brinkman, A.C. et al. (1987): Astrophys. Letters 26, 73

Brinkman, A.C. et al. (1997): Proc. SPIE, 3113, 181

Canizares, C.R. et al. (1987): Astrophys. Letters, 26, 87

Canizares, C.R. (1990): in *Imaging X-ray Astronomy*, M. Elvis (Ed.), (Cambridge UP: Cambridge), p. 123

Church, E.L., Jenkinson, H.A. & Zavada, J.M. (1977): Opt. Engineering 16, 360

Compton, A.H. & Allison, S.K. (1935): *X-rays in Theory and Experiment*, Second Edition (van Nostrand: New York).

Davis, J.E. (1997): on `http://space.mit.edu/HETG/LRF/scatter.html`

Decaux, V., Beiersdorfer, P., Osterheld, A., Chen, M. & Kahn, S.M. (1995): ApJ 443, 464

Debye, P. (1914): Ann. der Physik 43, 49

Decaux, V., Beiersdorfer, P., Kahn, S.M. & Jacobs, V.L. (1997): ApJ 482, 1076

Deiker, S. et al. (1997): in *Proceedings of the VIIth International Conference on Low Temperature Detectors*, July 1997, Munich, Germany.

Dewey, D. et al. (1997): Proc. SPIE 3113, 144

Done, C., Madejski, G.M. & Smith, D.A. (1996): ApJ 463, L63

Evans, A. (1986): MNRAS 223, 219

Fabian, A. et al. (1994): in *New Horizons in X-ray Astronomy*, F. Makino & T. Ohashi, Eds (Tokyo: Universal Academic Press), p. 573

Fraser, G.W. (1989): *X-ray Detectors in Astronomy*, (Cambridge UP: Cambridge)

Fukazawa, Y. et al. (1994): PASJ 46, L141

Gabriel, A.H. & Jordan, C. (1969): MNRAS 145, 241

Hamann, F. et al. (1998): ApJ 496, 761

Henke, B.L., Gullikson, E.M. & Davis, J.C. (1993): *Atomic Data and Nuclear Data Tables*, Vol. 54

Hettrick, M.C. & Bowyer, S. (1983): Appl. Optics 22, 3921

Houghton, J.T. & Smith, S.D. (1966): *Infra-red Physics*, (Oxford University Press: Oxford)

Illarionov, A., Kallman, T., McCray, R. & Ross, R. (1979): ApJ 228, 279

Iwasawa, K., Yaqoob, T., Awaki, H. & Ogasaka, Y. (1994): PASJ 46, L167

Iwasawa, K. et al. (1997): MNRAS 285, 683

Jark, W. & Nevière, M. (1987): Appl. Optics 26, 943

Kaastra, J.S., Roos, N. & Mewe, R. (1995): A&A 300, 25

Kahn, S.M. & Hettrick, M.C. (1985): in *A cosmic X-ray spectroscopy mission : proceedings of a workshop held in Lyngby, Denmark on 24-26 June 1985*, ESA SP 239 (ESA: Paris), p. 237

Kahn, S.M. et al. (1996): Proc. SPIE 2808, 450
Kitamoto, S. et al. (1994): PASJ 46, L105
Kittel, C. (1958): *Elementary Statistical Physics*, (Wiley:New York).
Kawashima, K. & Kitamoto, S. (1996): PASJ, 48, L113
Klos, R.A. (1985): Proc. SPIE 597, 135
Kunieda, H. et al. (1994): in *New Horizon in X-ray Astronomy*, F. Makino & T. Ohashi (Eds) (Universal Academy Press: Tokyo), p. 317
Labov, S. et al. (1997): in *Proceedings 7th International Workshop on Low Temperature Detectors LTD-7*, July 27-August 2, 1997, Munich (MPI für Physik: Munich) (`http://wwwvms.mppmu.mpg.de/ltd7/welcome.html`).
Lee, P.A., Citrin, P.H., Eisenberger, P. & Kincaid, B.M. (1981): Rev. Mod. Phys. 53, 769
Liedahl, D.A., Kahn, S.M., Osterheld, A.L. & Goldstein, W.H. (1990): ApJ 350, L37
Liedahl, D.A., Kahn, S.M., Osterheld, A.L. & Goldstein, W.H. (1992): ApJ 391, 306
Liedahl, D.A. & Paerels, F. (1996): ApJ 468, L33
Longair, M.S. (1984): *Theoretical Concepts in Physics*, (Cambrdige University Press: Cambridge).
Markert, T.H. et al. (1994): Proc. SPIE 2280, 168
Marshall, H.L. et al. (1997): Proc. SPIE 3113, 160
Martin, P.G. (1970): MNRAS 149, 221
Mather, J.C. (1982): Appl. Optics 21, 1125
Mathur, S., Elvis, M. & Wilkes, B. (1995): ApJ 452, 230
Matt, G. (1996): `astro-ph/9612002`.
McCray, R. (1984): Physica Scripta T7, 73
Mewe, R., Gronenschild, E.H.B.M. & van den Oord, G.H.J. (1985): A&AS 62, 197
Mewe, R. (1991): A&AR 3, 127
Michette, A.G. (1986): *Optical Systems for Soft X-rays* (Plenum Press: New York).
Moseley, S.H., Mather, J.C. & McCammon, D. (1984): J. Appl. Phys. 56, 1257
Nelson, C.S. et al. (1994): Proc. SPIE 2280, 191
Nussbaumer, H., Schmid, H.M. & Vogel, M. (1989): A&A 211, L27
Owens, A. (1997): ApJ 476, 924
Paerels, F. et al. (1994): Proc. SPIE 2283, 107
Paerels, F. (1997): at `http://astro1.nevis.columbia.edu/xmm/documents/index.html`.
Paerels, F., Kahn, S.M. & Wolkovitch, D.N. (1998): ApJ 496, 473
Pais, A. (1982): *Subtle is the Lord...*, (Oxford University Press: Oxford).
Peierls, R.E. (1955): *Quantum Theory of Solids*, (Oxford University Press: Oxford).
Petit, R. (1980): *Electromagnetic Theory of Gratings* (Ed.) (Springer: Berlin)
Predehl, P. et al. (1997): Proc. SPIE 3113, 172
Rando, N. et al. (1996): Nucl. Instr. Meth., A 370, 85
Rasmussen, A.R. et al. (1998): Proc. SPIE 3444.
Ross, R., Weaver, R. & McCray, R. (1978): ApJ 219, 294
Rybicki, G.B. & Lightman, A.P. (1979): *Radiative Processes in Astrophysics* (Wiley: New York)

Sanders, W.T., Edgar, R.J. & Liedahl, D.A. (1996): in *Röntgenstrahlung from the Universe*, H.U. Zimmermann, J. Trümper, & H. Yorke (Eds), MPE Report nr. 263, p. 339

Schattenburg, M.L. et al. (1994): Proc. SPIE 2280, 181

Schnopper, H.W. et al. (1977): Appl.Optics 16, 1088

Sevier, K.D. (1979): *Atomic Data and Nuclear Data Tables* 24, 323.

Seward, F.D. et al. (1982): Appl. Optics 21, 2012

Snow, T.P. & Witt, A.N. (1996): ApJ 468, L65

Spodek, J., Rasmussen, A., Cottam, J., Kahn, S.M. & Paerels, F. (1998): *in preparation*

Stahle, C.K. et al. (1997): *Proceedings 7th International Workshop on Low Temperature Detectors LTD-7*, July 27-August 2, 1997, Munich (MPI für Physik: Munich) (`http://wwwvms.mppmu.mpg.de/ltd7/welcome.html`).

Stroke, G.W. (1967): in *Handbuch der Physik*, Vol. XXIX, S. Flügge (Ed.) (Springer: Berlin), pp. 426-754

Sunyaev, R.A. & Churazov, E.M. (1996): Astronomy Letters 22, 723

Tanaka, Y., Inoue, H. & Holt, S.S. (1994): PASJ 46, L37

Turner, M.J.L. et al. (1996): in *The Next Generation of X-ray Observatories: Workshop Proceedings*, M.J.L. Turner & M.G. Watson (Eds), Leicester X-ray Astronomy Group Special Report, XRA97/02, p. 165 (`http://ledas-www.star.le.ac.uk/ngxo`)

Ueno, S. et al. (1994): PASJ 46, L71

van den Heuvel, E.P.J. (1994): in *New Horizon of X-ray Astronomy - First Results from ASCA*, F. Makino & T. Ohashi (Eds), (Universal Academy Press: Tokyo).

van Kerkwijk, M.H. et al. (1992): Nat 355, 703

Weisskopf, M.C. et al. (1987): Astrophys. Letters 26, nrs. 1 & 2

Woo, J.W. (1995): ApJ 447, L129

Zarnecki, J.C. & Culhane, J.L. (1978): MNRAS 178, 57P