High energy astrophysics

April 18 2013: X-ray Calibration

Jukka Nevalainen

University of Helsinki

Context

Whenever any astronomical instrument is used to measure anything, the interaction between the photons and the instrument imprints instrumental features into the data.

Instrumental effects must be removed from the data to do science.

To do this, one needs to know the instrument properties accurately i.e. to **calibrate** the instrument.

Nothing is perfect → Interpretation of data is **IS ALWAYS** affected by calibration uncertainties at some level.

When deriving science from data, one should estimate and take into account the effects of the calibration uncertainties.

Table of contents

1) Problem: Instrumental effects in the X-ray data

1.1 Components

1.1.1 Mirror effective area

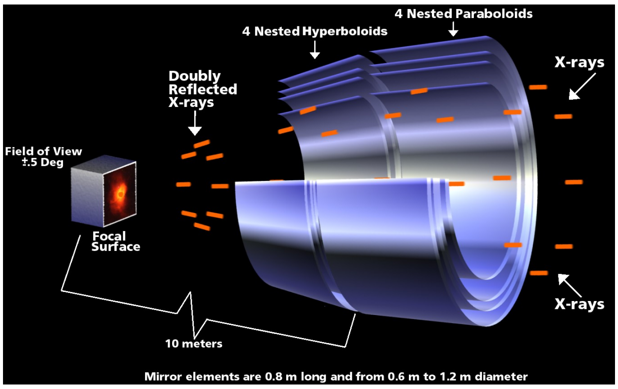

Fig. 1: X-ray satellites collect photons with their mirrors.(**Chandra X-ray Center)**

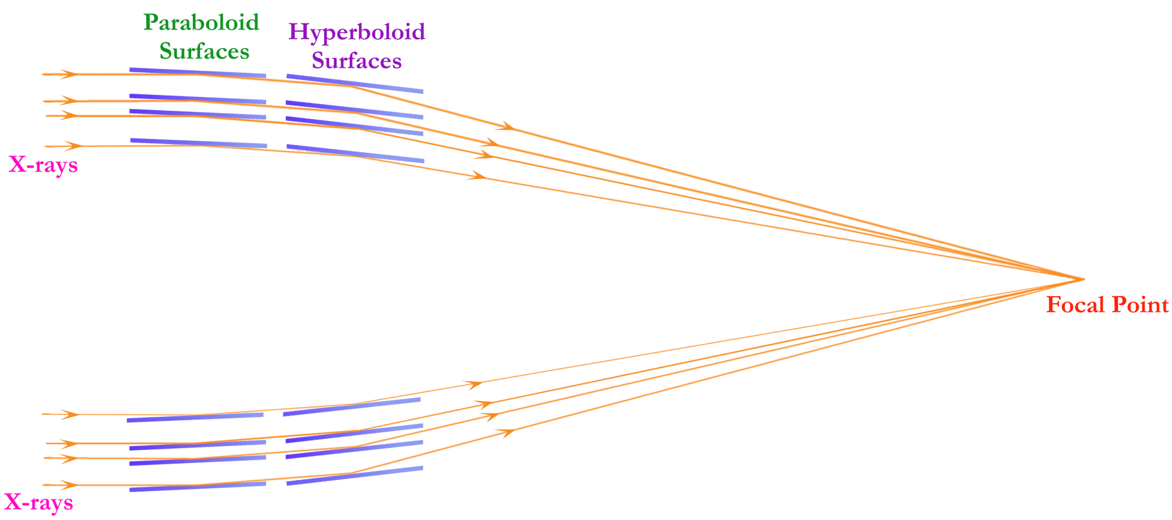

Fig. 2: Photons are focused to the detector at the focal plane**(Chandra X-ray Center)**

The photon collecting power of a mirror as a function of energy = mirror effective ***area*** (emission models give photons s^{-1} keV^{-1} ***cm***$^{-2}$).

Depending on the materials used for building the mirrors, the effective area has a particular shape (**Fig. 3**).

To do science, the effective area of a given instrument must be known accurately.

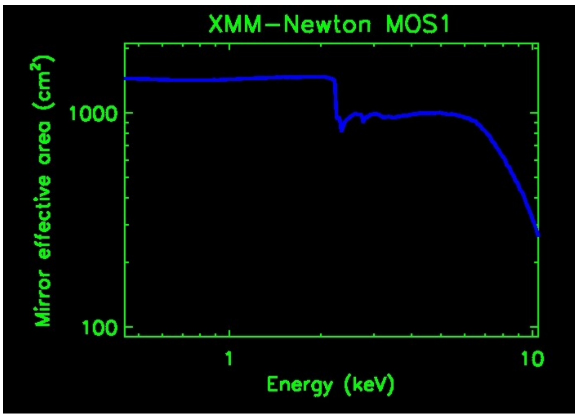

Fig. 3: Mirror effective area of MOS1

1.1 Components

1.1.2 Filter transmission

X-ray detectors are usually sensitive to IR, optical and UV photons as well. If the X-ray source has e.g. a high optical flux, of if an unwanted bright optical source is located within the field-of-view, the X-ray data could be contaminated in many ways.

Thus the photons collected by the mirror must be filtered. E.g. XMM-Newton has thin, medium and thick filters.

Unfortunately the materials in the filter also remove part of the X-ray photons. The transmission of a given filter describes the fraction of X-ray photons surviving the filter, compared to the intrinsic flux (**Fig. 4**).

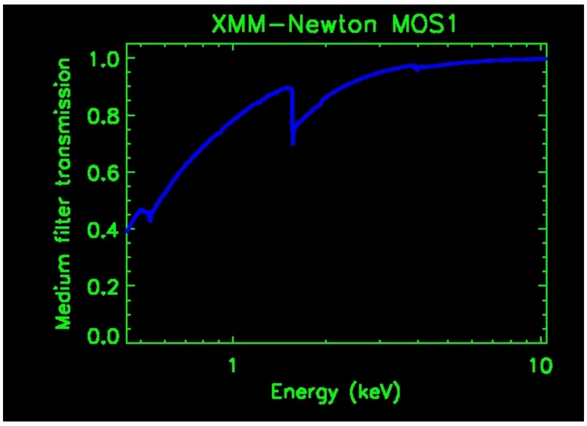

Fig. 4: Filter transmission of MOS1

1.1.3 Quantum efficiency

The mirror-collected, filtered photons then enter the detector in the focal plane (**Fig. 5**)

Detectors convert the photon events into particle bursts, which are then measured (energy, arrival time, position) (**Fig. 6**).

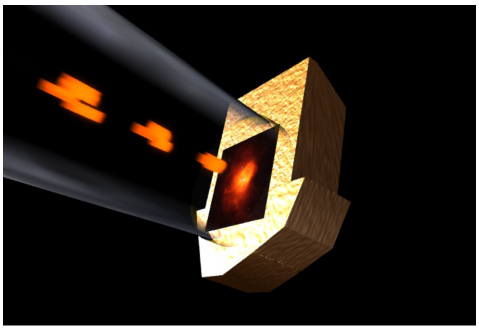

Fig.5 (Chandra X-ray Center)

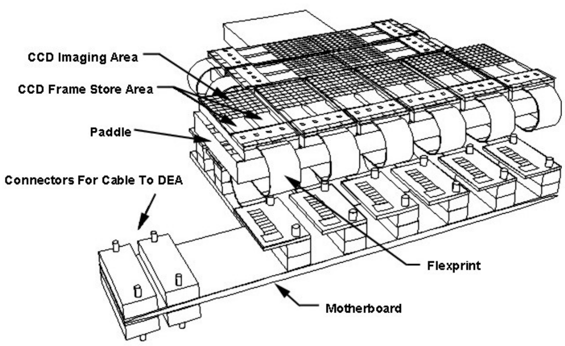

Fig. 6. Chandra/ACIS detector(Chandra X-ray Center)

The capability of a given instrument to convert the photons into counts is described by the quantum efficiency. Quantum efficiency = number of converted photons / number of intrinsic photons (**Fig. 7).**

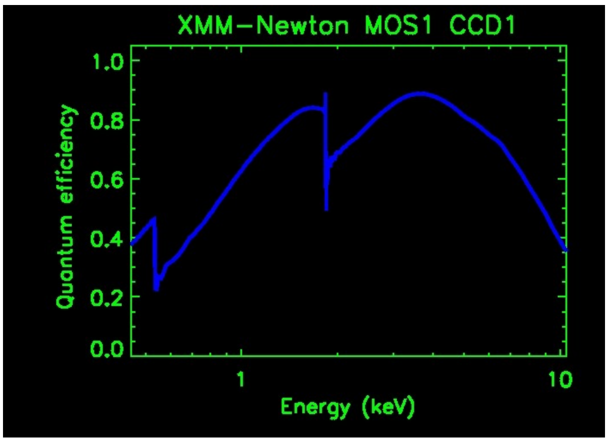

Fig. 7: Quantum efficiency of MOS1

1.1 Components

1.1.4 Total effective area

Total effective area (often called as ”effective area”) descibes the combined instrumental effects affecting the photon on its way towards the detector. Total effective area (**Fig. 11)** = mirror effective area (**Fig. 8)** $\times$ filter transmission (**Fig. 9)** $\times$ quantum efficiency (**Fig. 10)**.

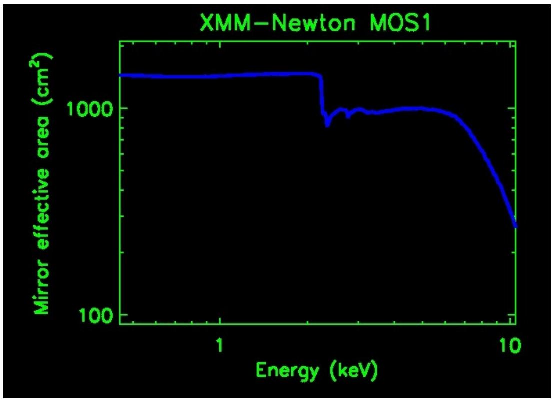

Fig.8: Mirror effective area of MOS1

×

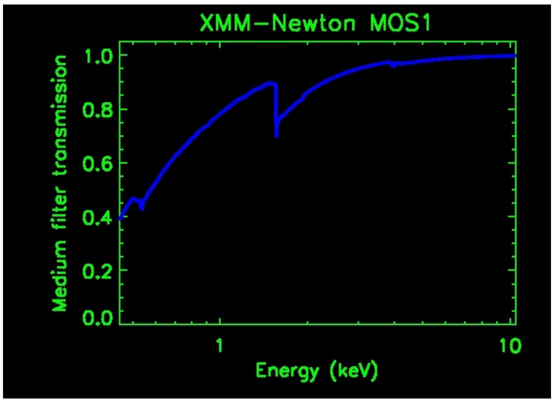

Fig. 9: Filter transmission of MOS1

×

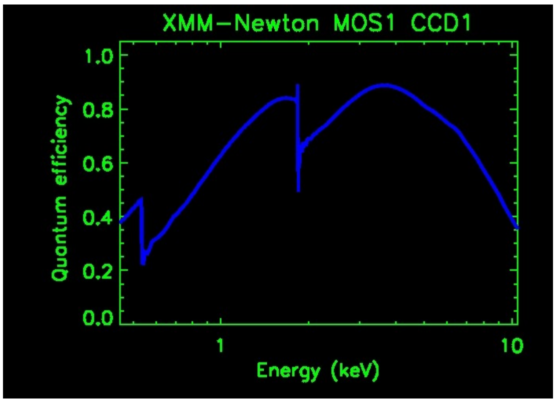

Fig.10: Quantum efficiency of MOS1

=

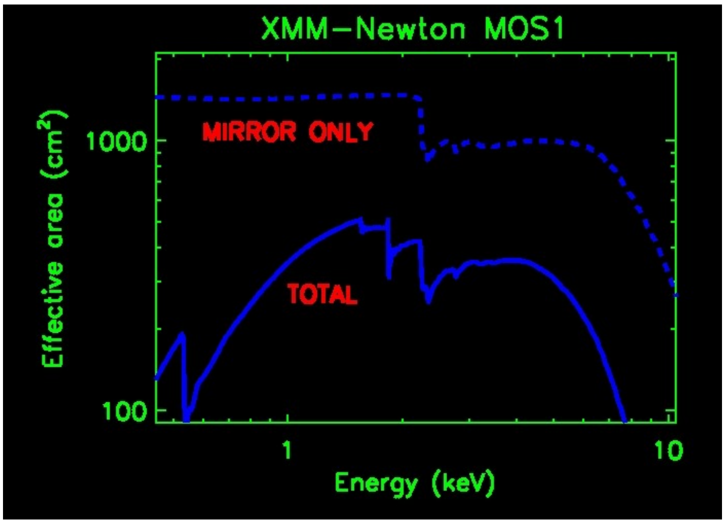

Fig. 11: Total effective area of MOS1

1.1.5 Other components

There are several other effects which will not be discussed in detail here, including energy redistribution, point spread function and vignetting.

1.2 Data analysis

The instrumental effects discussed above are collected into (total) effective area vector $A_{eff}(E)$ (arf – file in XSPEC).

When analysing the X-ray spectra obtained by an X-ray instrument, one first chooses the emission model with a particular choise of parameter values M(E).

To enable a comparison of a (theoretical) model M(E) with (observational) data D(E), one has to multiply the emission model with the effective area.

The product P(E) is the model prediction when the above instrumental effects are included (**Eq. 1**).

$$P(E) \approx M(E) \times A_{eff}(E) \qquad \textbf{Eq. 1}$$

However, due to the finite energy resolution of an instrument, a fraction of the photons are detected in an energy channel C which does not correspond to the true photon energy.

The probability that an incoming photon of energy E will be detected in energy channel C is descibed by the response matrix $R_D(C,E)$ (rmf-file in XSPEC) .

Thus the model prediction of the photon energy distribution P(E) is convolved with the response matrix $R_D(C,E)$ to yield the model prediction in detector channels P(C), (Eq.2).

$$\begin{gathered} P(C) = P(E) \otimes R_D(C,E) \Leftrightarrow \\ P(C) = M(E) \times A_{eff}(E) \otimes R_D(C,E) \end{gathered} \qquad \textbf{Eq. 2}$$

We do not study the response matrix calibration here. Thus, let's assume an infinite energy resolution, i.e. a diagonal response matrix, so that each input photon ends up in the corresponding channel. Thus (see **Figs. 12-14**)

$$\begin{gathered} P(C) = M(E) \times A_{eff}(E) \otimes R_D(C,E) \Rightarrow \\ P(E) = M(E) \times A_{eff}(E) \end{gathered} \qquad \textbf{Eq. 3}$$

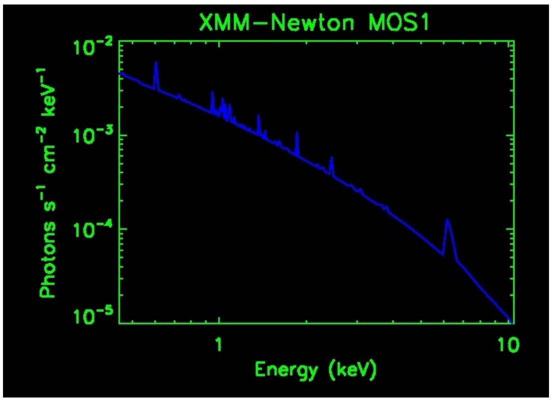

Fig. 12: Emission model M(E)

×

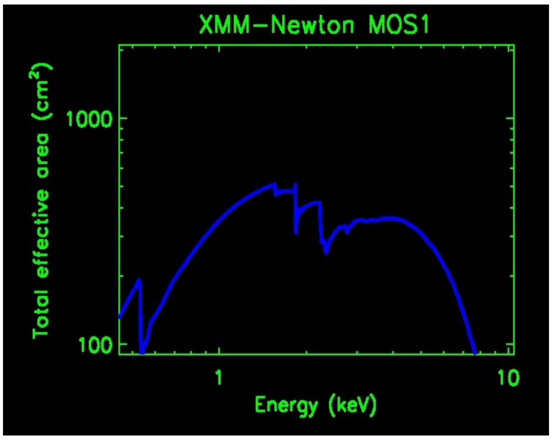

Fig. 13: Total effective area $A_{eff}(E)$

=

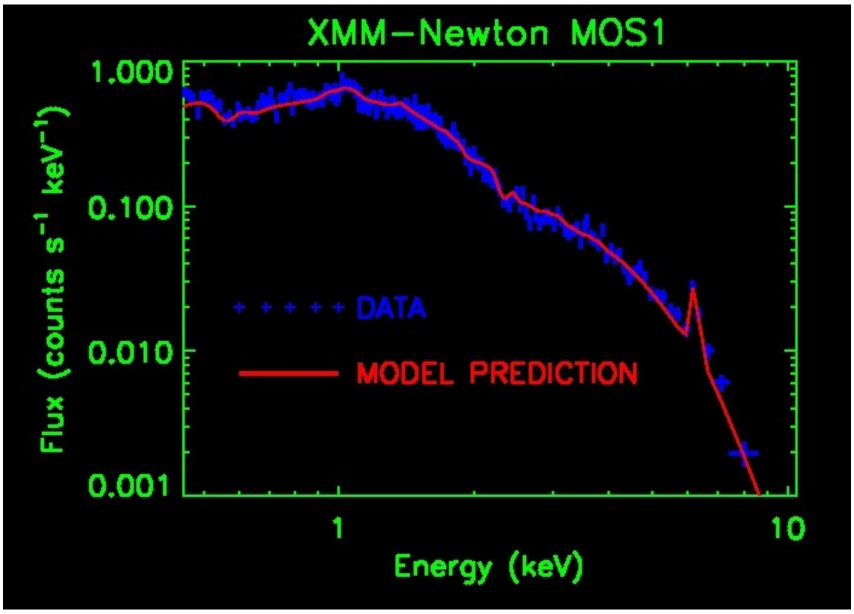

Fig. 14: Model prediction P(C)

By varying the model parameters one looks for the best match between the data D(E) and the model prediction P(E).

$$Min(\chi^2) = Min \sum \left[\frac{D(E) - P(E)}{\sigma(E)} \right]^2 \approx Min \sum \left[\frac{D(E) - (M(E) \times A_{eff}(E))}{\sigma(E)} \right]^2 \quad \textbf{Eq. 4}$$

The best fit model yields the physical intrperation of the emission (e.g. temperature) .

If there are significant calibration uncertainties i.e. the effective area $A_{eff}(E)$ is not correct.

Thus the best-fitting model M(E) is not correct either.

Often calibration uncertainties are at low level (~10%) so that the combination of inaccurate calibration and inaccurate model can produce an acceptable fit to data. Thus the quality of the fit (residuals) can be good and the problems are hidden.

Also, the model with incorrect parameters can still make physical sense.

Thus, an extra effort is needed to get rid of the systematics due to calibration uncertanties.

2) Solution: Learn to know your instrument i.e *calibrate* it

2.1 Ground calibration

Before the launch, some of the units of the mirror, filter units and CCD are tested in laboratory.

Typically an electron accelerator is used to generate a monochromatic synchrotron beam.

The beam is directed towards the studied unit and the effect is measured.

Since synchrotron physics is well known, the energy and flux are very accurately known a priori.

The synchrotron frequency is varied to scan the full spectral band.

Thus this measurement characterises the different components of the total effective area accurately.

So what is the ”calibration uncertainty” problem?

In practise, it is very difficult to assemble the full satellite instrumentation in the laboratory. The different components are typically tested separately. Thus, the effect of the combination of all possible photon paths is not tested.

One can only test a few modules of the instrumentation in a feasible time in lab. The properties of the other modules are assumed equal.

One can test the instrument response only at a limited number of frequencies. The response at non-tested frequencies is interpolated.

When mirrors are assembled into the spacecraft, possible uncertainties in the alignment may cause differences from the lab test.

Often things change after the launch. For example, in Suzaku and Chandra satellites there is a leakage of hydrocarbon contaminate, which did not show up in the lab.

2.2 In-flight calibration

2.2.1 On-board calibration source

In order to monitor the changes of the intrument properties in orbit, all X-ray satellites carry some kind of standard candle i.e a source whose emission properties are known very accurately a priori.

In case of XMM-Newton satellite, a 55Fe source is used which emits Al-K (1.5 keV) and Mn-K (5.9 keV) lines.

This is measured frequently with the detectors to keep track of energy scale and energy resolution.

This is limited because mirror effective area and filter transmissions do not affect this measurement.

Calibration source may vary. Who calibrates the calibrator?

2.2.2 Simultaenous observations of bright (variable) sources with many satellites

The idea is to observe bright X-ray sources (AGN...) with several satellites. One should get the same results if all the satellites are accurately calibrated.

Possible differences can be analysed to get information about which component of which satellite has significant calibration uncertainties and by how much.

But: bright sources like AGN are very variable, so one has to make simultaneous observations to make sure that the intrinsic emission entering the different satellites is equal.

The requirement of simultaneity renders the observations difficult and rare.

2.2.3 Observations of standard candles: general

BASIC LOGIC: If an astronomical source is very accurately known, and constant, different instruments should yield the same, a priori known X-ray properties (e.g. temperatures and fluxes) for the source at different observation epochs.

If they don't, there is a problem with the calibration of the instrument which gives ”wrong” results.

Crab nebula has often been used for this.

It was recently found that Crab is not constant as previously thought. Lesson learned: It is difficult to know anything accurately a priori.

3) Clusters of galaxies as standard candles

3.1 Introduction

OPTICAL	**X-RAYS**
100 – 1000 galaxies	Intergalactic gas: mass ~ 10^{14} $M_\odot$, T ~ $10^{7\text{-}8}$ K
Size ~ Mpc, mass ~ 10^{13} $M_\odot$	Dark matter ~ 10^{15} $M_\odot$

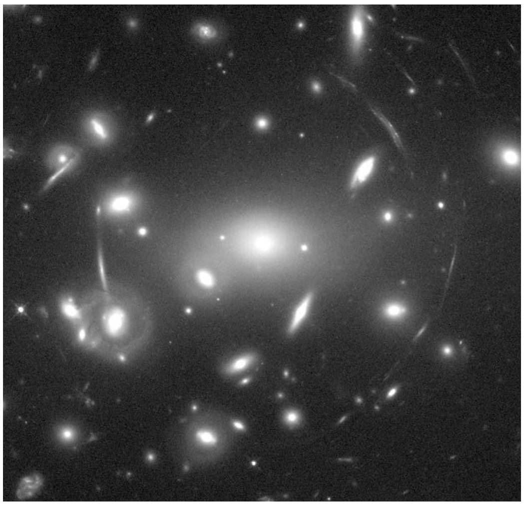

Fig. 15: A2218 observed by HST

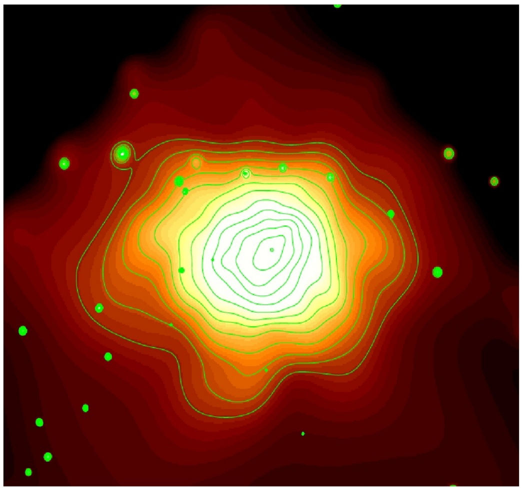

Fig. 16: A2218 observed by Chandra

Clusters of galaxies are suitable for X-ray calibration because

1) they have hard X-ray spectra (up to T ~ 10^8 K) → the hottest ones emit significantly at the highest energies (~ 10 keV)

2) nearest clusters are very bright (10^{-12}-10^{-11} erg s^{-1} cm^{-2}) → good statistics

3) physics is well understood for the relaxed clusters (bremsstrahlung continuum + collisionally excited line emission) → modelling accurate

4) clusters are stable in human time scales ➔ You don't need to observe them simultaneously with different satellites ➔You can build a large sample of objects observed with many satellites ➔ Large sample is necessary when examining systematic effects such as calibration uncertainties

3.2 Bremmsstrahlung

Radiation emitted by a charged particle accelerated in a Coulomb field of another charge

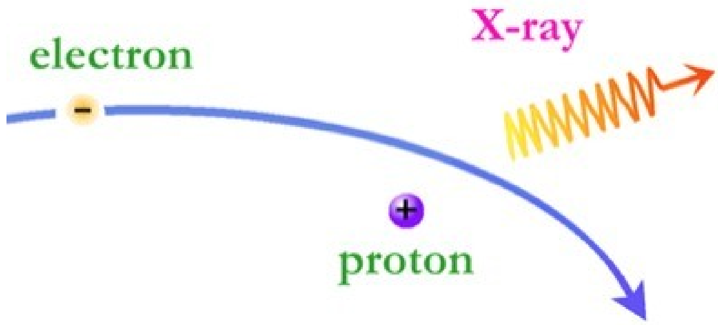

Fig. 17: Bremsstrahlung

Bremsstrahlung emissivity:

$$\epsilon_E \equiv \frac{dW}{dV\,dt\,dE} = 1.6\,x\,10^{-20}\,\bar{g}_{ff}\,Z^2\,n_e\,n_i\,T^{-1/2}\,e^{-E/kT}\left(erg\,s^{-1}\,cm^{-3}\,keV^{-1}\right) \qquad \textbf{Eq. 5}$$

Assuming pure fully ionised hydrogen (Z=1, $n_e = n_i$) and $\bar{g}_{ff} = 1.2$ →

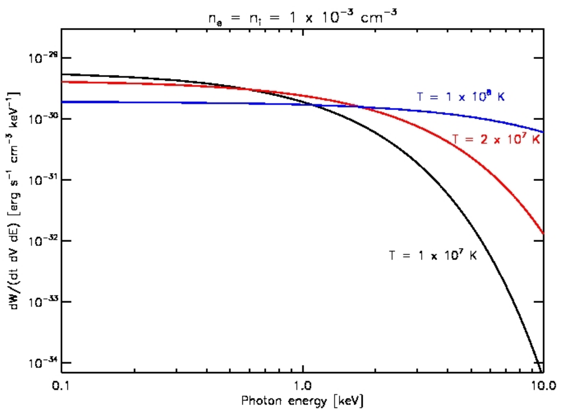

Fig. 18: temperature determines the shape

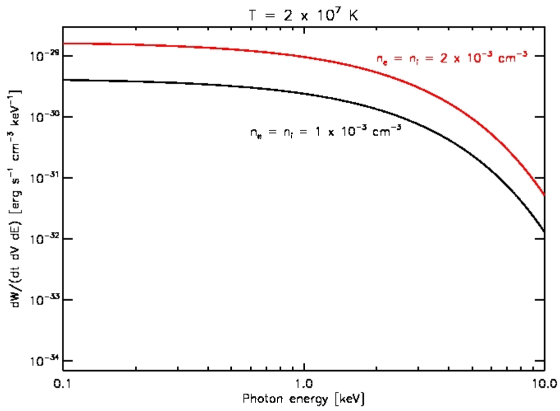

Fig. 19: density determines the normalization

3.3 Fe XXVI/XXV line ratio as a thermometer

Fe XXVI/XXV line flux ratio increases with higher ionisation temperature.

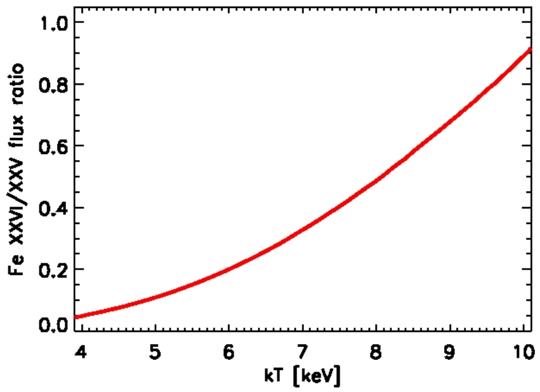

Fig. 20: Fe XXVI/XXV line flux ratio as a function of temperature.

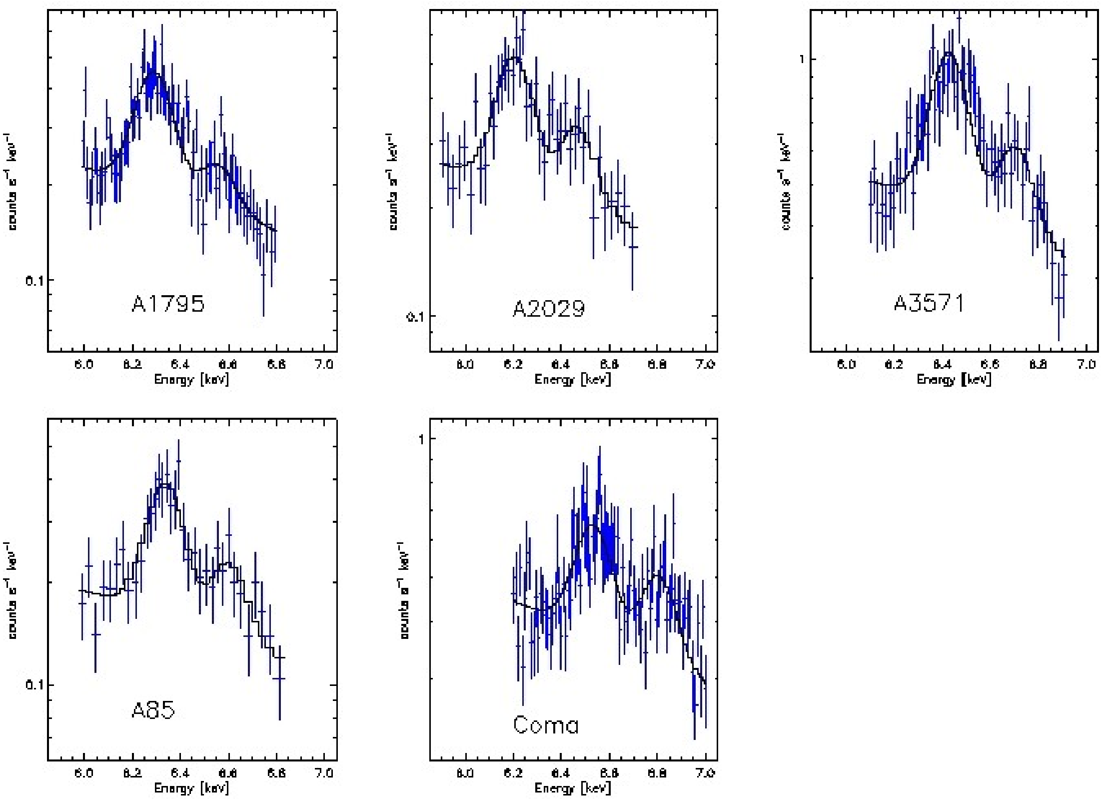

Fig. 21: Fe XXVI is measurable for the hottest clusters with the energy resolution of XMM-Newton/ EPIC and Chandra/ACIS instruments

4) IACHEC

In the following I will present calibration work done by IACHEC = International Astronomical Consortium for High Energy Calibration http://web.mit.edu/iachec/

IACHEC aims to provide standards for high energy calibration and supervise cross calibration between different missions.

This goal is reached through working groups, where IACHEC members cooperate to define calibration standards and procedures.

5) Physics as a calibration tool

5.1 Introduction

As discussed in Section 1.2, incorrect calibration yields incorrect physics.

Sometimes the derived physics does not make sense. This is an indicator of calibration uncertainties.

Usually we are not that ”lucky”. Rather, the biased emission parameters still make physical sense.

However, if one instrument yields systematically different physical properties while all other instruments yield consistent values, then there is probably something wrong with the instrument that gives the discrepant results.

5.2 Galaxy cluster temperatures

The bremsstrahlung temperature measurement is driven by the shape of the exponential cut-off (see Section 3.2).

If the shape of the effective area, as implemented in the calibration infromation, is too steep or too shallow, the derived temperature is too hard or too soft ➔ **cluster temperatures are useful for calibration of the shape (i.e. energy dependence) of the effective area.**

Fe XXVI/XXV line ratio measurement is independent of the effective area uncertainties, because the measurement is done in a very narrow band ➔ if bremsstrahlung and ionisation temperatures agree, effective area is correctly calibrated (assuming no deviations from the ionisation equilibrium and Maxwellian electron velocity distribution).

Because Fe XXVI/XXV ratio covers a narrow band, the exposure time must be quite high to get enough photons for meaningful statistics. This is a limitation for the method.

Within IACHEC we conducted a calibration accuracy test for X-ray satellites/instruments XMM-Newton/EPIC , Chandra/ACIS and BeppoSAX/MECS (Nevalainen et al., 2010, A&A, 523, 22)

We used a sample of 11 galaxy clusters

We performed X-ray spectroscopy on the data obtained with different instruments from the same regions of a given cluster.

We compared the bremsstrahlung and Fe XXVI/XXV temperatures obtained with different instruments.

We found that XMM-Newton pn and MOS instruments and BeppoSAX MECS yielded consistent temperatures obtained with a model consisting of a bremsstrahlung continuum and collisionally excited line emission temperatures in the 2-7 keV band (**Figs. 22 and 23**).

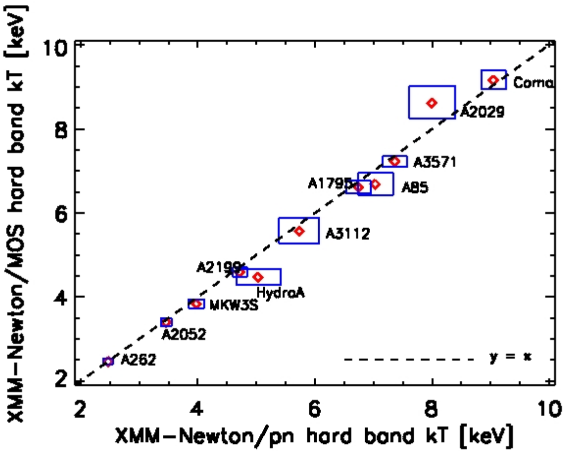

Fig. 22: hard band best-fit kT for pn and MOS

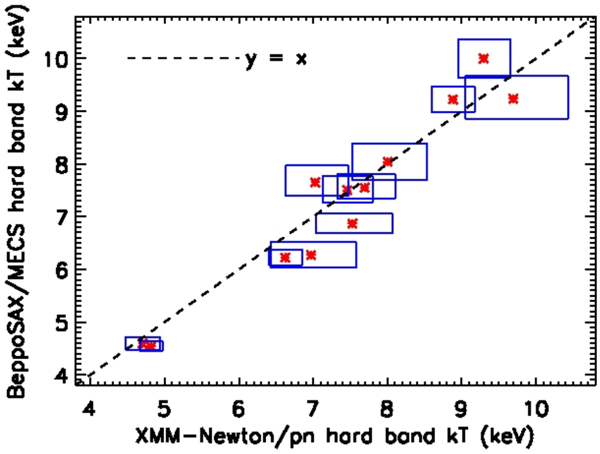

Fig. 23: hard band best-fit kT for pn and MECS

Also, the bremsstrahlung temperatures agreed with the continuum temperature for XMM-Newton (**Fig. 24).**

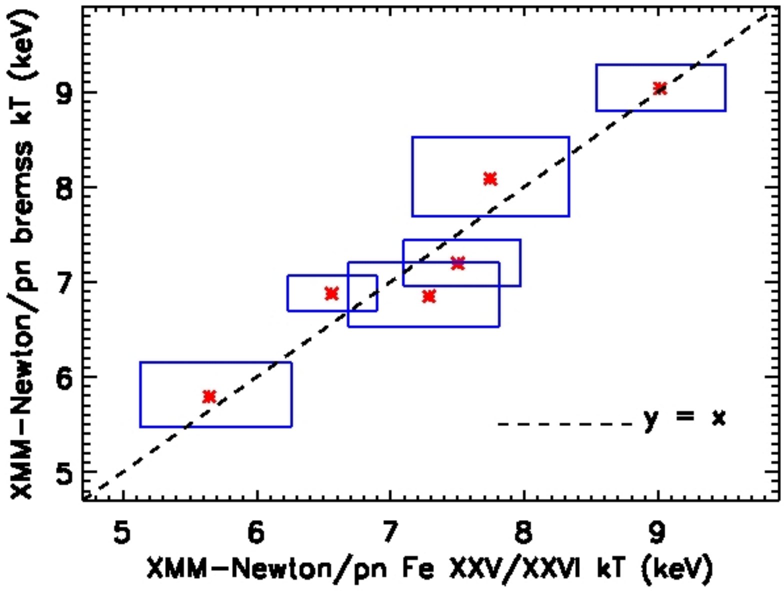

Fig. 24: Bremsstrahlung and Fe XXVI/XXV kT for pn

Thus, the consistence in the 2-7 keV band between several instruments and two independent temperature measurement methods indicates that calibration is close to absolute in XMM-Newton and BeppoSAX.

However, when using the public calibration database CALDB 3.4, the Chandra data yielded systematically lower cluster temperatures in the 2-7 keV band **(Fig. 25)**.

This indicated problems with Chandra calibration using CALDB 3.4.

Chandra team re-analysed the ground calibration data and found a problem with modelling of the hydrocarbonate contaminate in the mirrors.

Contaminate was re-measured shell-by-shell. The more detailed modelling changed the reflectivity ➔

Now with the new calibration information (CALDB 4.2.0) the Chandra temperatures agree with those obtained with XMM-Newton **(Fig. 26)** and BeppoSAX, and using the bremsstrahlung or Fe XXVI/XXV for temperature measurement i.e. effective area shape correctly calibrated.

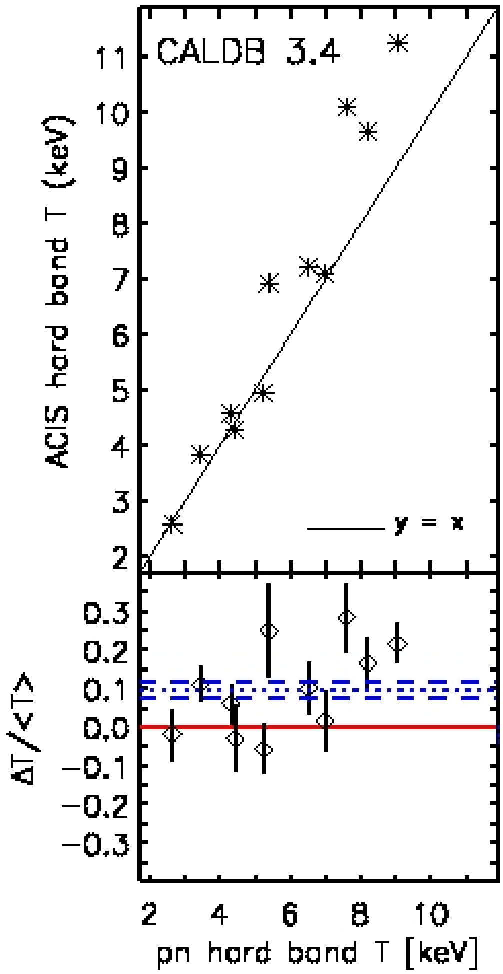

Fig. 25: pn and ACIS (old) kT

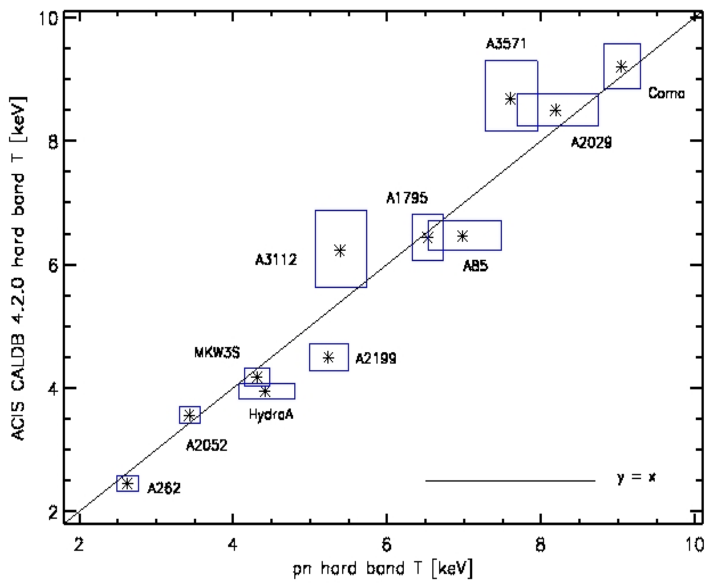

Fig. 26: pn and ACIS (current) kT

Lessons learned:

Galaxy cluster temperatures are very useful for X-ray calibration

Temperature consistence/inconsistence can found evidence about which instrument has problems. But it cannot say what is wrong with the calibration. The physical reason must be found. **DO NOT FUDGE!**

6) Stack residuals as a calibration tool

6.1 The method

The temperature comparison is useful, but it does not quantify in detail the calibration uncertainties of the effective are as a function of energy. Stack residuals will help here.

We select one instrument as a reference instrument (**ref**) , against which the other instruments (**I**) are compared.

Thus the stack residuals method quantifies a ***cross-calibration*** uncertainty.

First we fit the data of the reference instrument and obtain a best-fit model $M(E)_{\mathbf{ref}}$

Then we use the reference model $M(E)_{\mathbf{ref}}$ to produce a prediction in the studied instrument **I** by convolving (multiplying) the model with the instrument response (effective area) of **I**:

$$P_{I,ref}(E) \approx M_{ref}(E) \times A_{eff,I}(E) \quad \textbf{Eq. 6}$$

If the effective area of the reference instrument $A_{eff,\mathbf{ref}}$ (E) is accurately calibrated, then model $M(E)_{\mathbf{ref}}$ is correct

If also the effective area of the studied instrument $A_{eff,\mathbf{I}}$ (E) is accurately calibrated, the reference model $M(E)_{\mathbf{ref}}$, multiplied with $A_{eff,\mathbf{I}}$ (E) yields a prediction $P_{\mathbf{I},\mathbf{ref}}(E)$ which is consistent with the data obtained with the studied instrument, $D_{\mathbf{I}}(E)$

We thus divide the data $D_{\mathbf{I}}(E)$ by the prediction $P_{\mathbf{I},\mathbf{ref}}(E)$, and thus obtain residuals $R_{\mathbf{I}/\mathbf{ref}}(E)$ which should be unity at each energy, if both instruments are accurately calibrated. Deviations from 1 indicate cross-calibration uncertainties and quantify their energy dependence.

$$R_{I/ref} = \frac{D_I(E)}{M(E)_{ref} \times A_{eff,I}} \quad \textbf{Eq. 7}$$

The method assumes that the best-fit model to the data of the reference instrument is a perfect description of the data.

*Usually this is not the case. To fix this, we add a second term to the residual equation (**Eq.7**) , which corrects the model if it deviates from the data. Not discusssed in detail here.*

We repeat the analysis to a sample of objects and thus obtain a distribution of $R_{\mathbf{I}/\mathbf{ref}}(E)$ values at each energy. We calculate the median of the sample and the mean absolute deviation at each energy to quantify the information. These are the stack residuals.

6.2 Some results

Within IACHEC we analysed stack residuals for EPIC-pn and EPIC-MOS instruments of XMM-Newton satellite using calibration from Jan 2013 **(Fig. 27)**.

The objects are the galaxy cluster sample from Nevalainen et al. (2010) .

2-7 keV band effective area shape calibration OK (as already indicated by the temperature consistence).

2-7 keV band MOS effective area normalisation underestimated or pn overestimated by 5-10% (as indicated by the flux comparison)

Below 2 keV weird behaviour (as indicated by the discrepant 0.5-2.0 keV band temperatures): steep 1.0-2.0 keV feature.

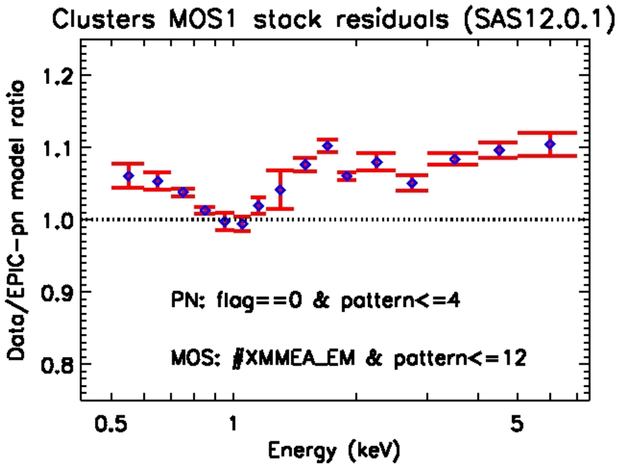

Fig. 27: MOS1/pn cluster stack residuals

We repeated the analysis of clusters of galaxies to compare XMM-Newton/pn and Chandra/ACIS instruments **(Fig. 28)**.

2-7 keV band effective area shape calibration OK (as already indicated by the temperature consistence).

2-7 keV band ACIS effective area normalisation underesti-mated or pn overestimated by ~10% (as indicated by the flux comparison).

Below 2 keV weird behaviour (as indicated by the discrepant 0.5-2.0 keV band temperatures): steep 1.0-2.0 keV feature.

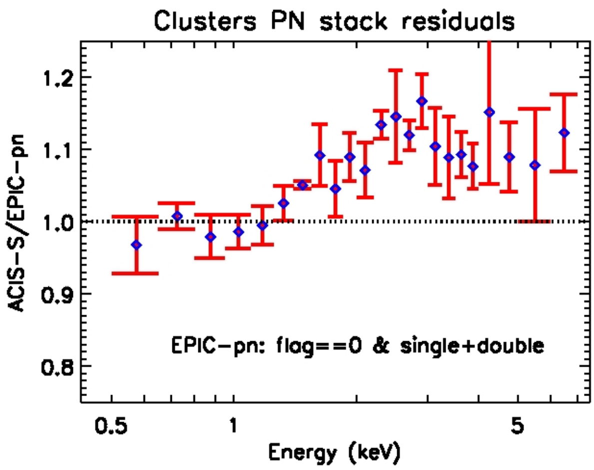

Fig. 28: ACIS/pn cluster stack residuals

7) Nothing is perfect: estimate and propagate the calibration uncertainties

The calibration work has yielded the level of calibration uncertainties between several satellites

It it sometimes possible to prove that one instrument is worse calibrated than the others

It is very difficult to prove that any instrument is absolutely correctly calibrated

As long as the absolute calibration is not reached, the calibration uncertainties should be propagated to the scientific results derived from the data.

There are a few possibilities to propagate the calibration uncertainties into the scientific results

Easy way: Use the published estimates of the effect of the calibration uncertainties to your parameters. For example, is you are interested in the galaxy cluster temperatures derived using XMM-Newton or Chandra, you can use the estimates from Nevalainen et al. (2010) for the temperature uncertainties.

Hard way: If you are interested in other objects than clusters, observed with XMM-Newton or Chandra, you could use the estimates for the effective area uncertainties from e.g. Nevalainen et al. (2010) and vary your effective area in the spectral fits.

Conclusions

Interpretation of data is **IS ALWAYS** affected by calibration uncertainties at some level

Instrumental effects must be removed from the data to do science

Calibration uncertainties can be evaluated by comparing results from standard candles obtained with different satellites

The calibration uncertainties should be propagated to scientific results