X-ray data analysis

Andrea Marinucci

Università degli Studi
Roma Tre
marinucci@fis.uniroma3.it

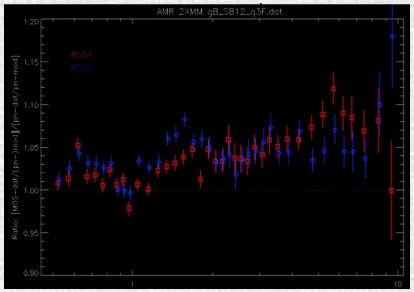
Goal of these lectures

X-ray data analysis

why?

what?

how?



How?

This is intentended to be a general reference for topics relevant to spectroscopy of low-resolution (i.e. CCD) spectra:

- How do we fit spectra?
– [and, by the way, what does it min “fitting a spectrum”?]

- What files do we need? what are they?

- How do we turn the fitting wheel?

- How do we deal with calibration uncertainties?

If I make things too messy, no panic! Look at (e.g.):
http://heasarc.gsfc.nasa.gov/docs/xanadu/xspec/manual/XspecSpectralFitting.html

This section is taken from past lectures by Dr. Matteo Guainazzi

How?

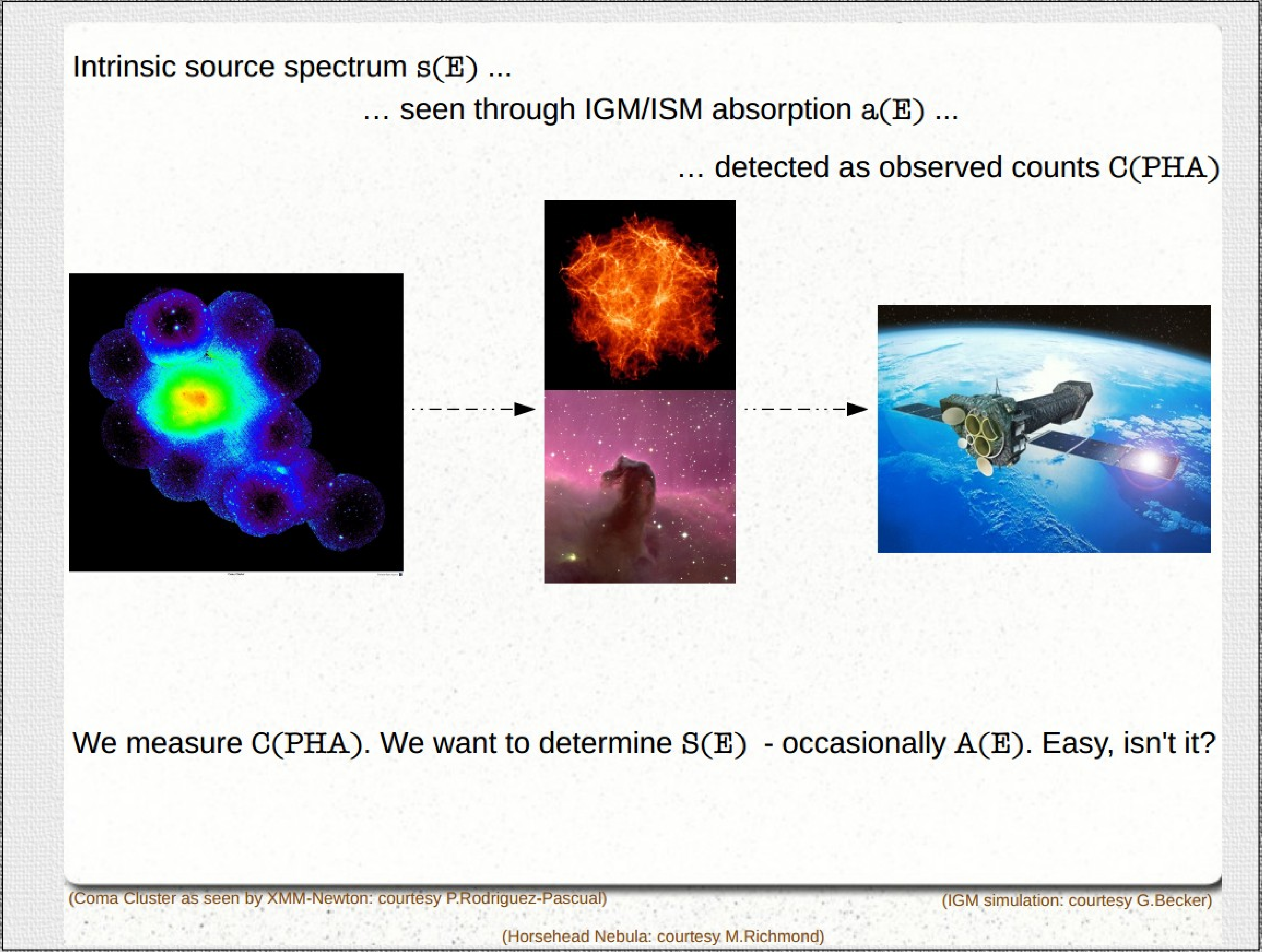
This talk is primarily intended for users of CCD spectra:

  - ASCA/SIS
  - Chandra/ACIS
  - Swift/XRT
  - Suzaku/XIS
  - XMM-Newton/EPIC (-MOS and -pn)

- However, some basic principles can be applied to instruments with even lower resolution:

  - ROSAT/PSPC, ASCA/GIS, BeppoSAX, RXTE, Suzaku/HXD, NuSTAR/FPMA-B

The ultimate goal of data analysis..



Intrinsic source spectrum $\mathtt{s(E)}$ ...

... seen through IGM/ISM absorption $\mathtt{a(E)}$ ...

... detected as observed counts $\mathtt{C(PHA)}$

We measure $\mathtt{C(PHA)}$. We want to determine $\mathtt{S(E)}$ - occasionally $\mathtt{A(E)}$. Easy, isn't it?

(Coma Cluster as seen by XMM-Newton: courtesy P.Rodriguez-Pascual)

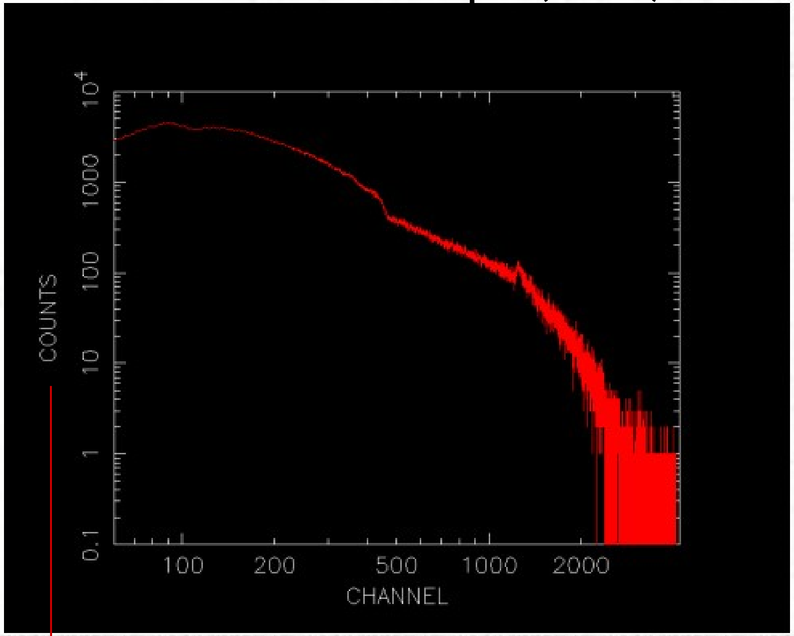
(IGM simulation: courtesy G.Becker)

(Horsehead Nebula: courtesy M.Richmond)

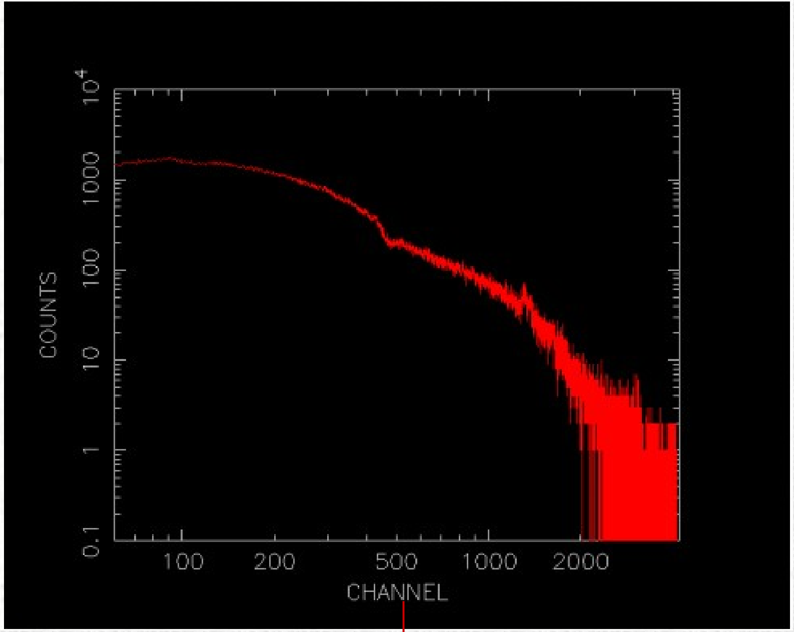
When all candles be out, all cats are gray..

CCD spectra extracted by dmextract, xmm/evselect, or xselect look like this:

Ark 120 – EPIC-pn (AGN)



Coma – EPIC-pn (Galaxy Cluster)



These are “COUNTS per bin”, not flux!

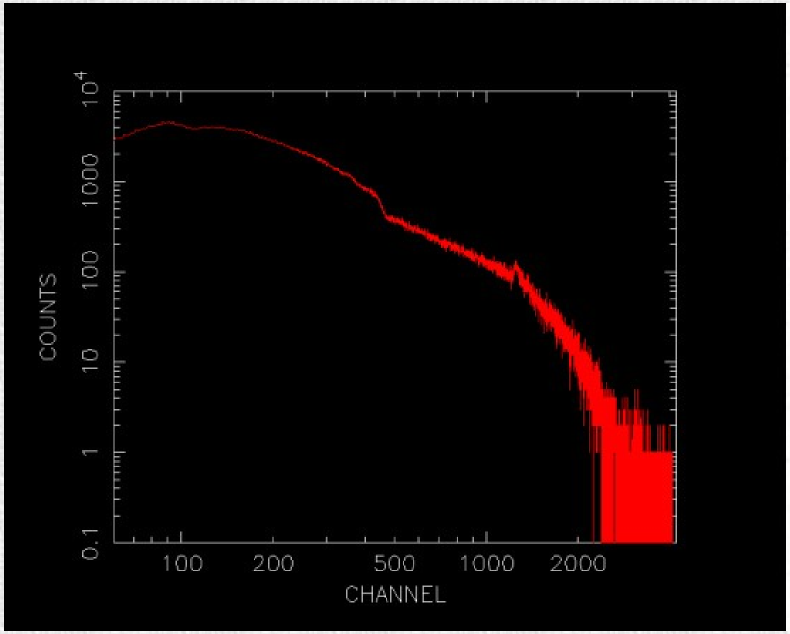
These are “CHANNELs”, not energy!

First problem: spectral extractors produce spectra in instrumental quantities

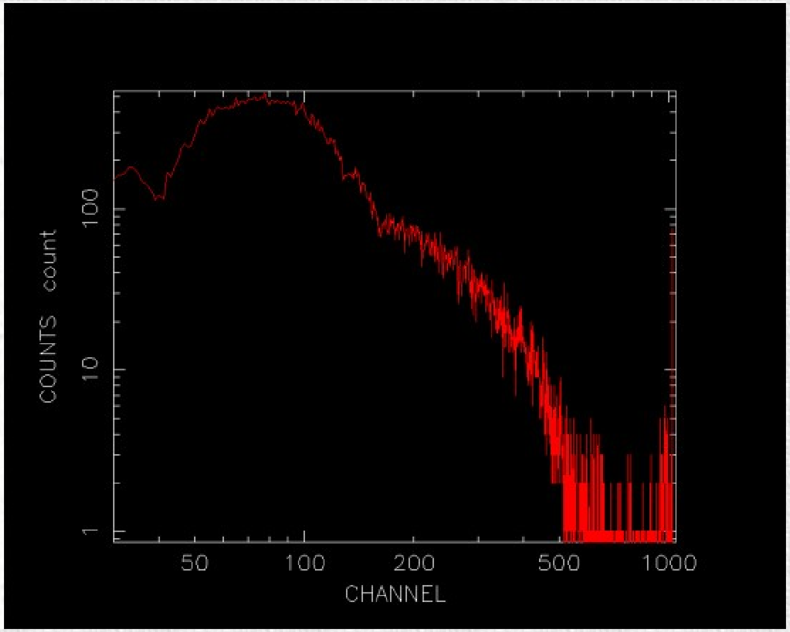
When all candles be out, all cats are gray..

And now, something completely different..

Ark 120 – EPIC-pn (AGN)

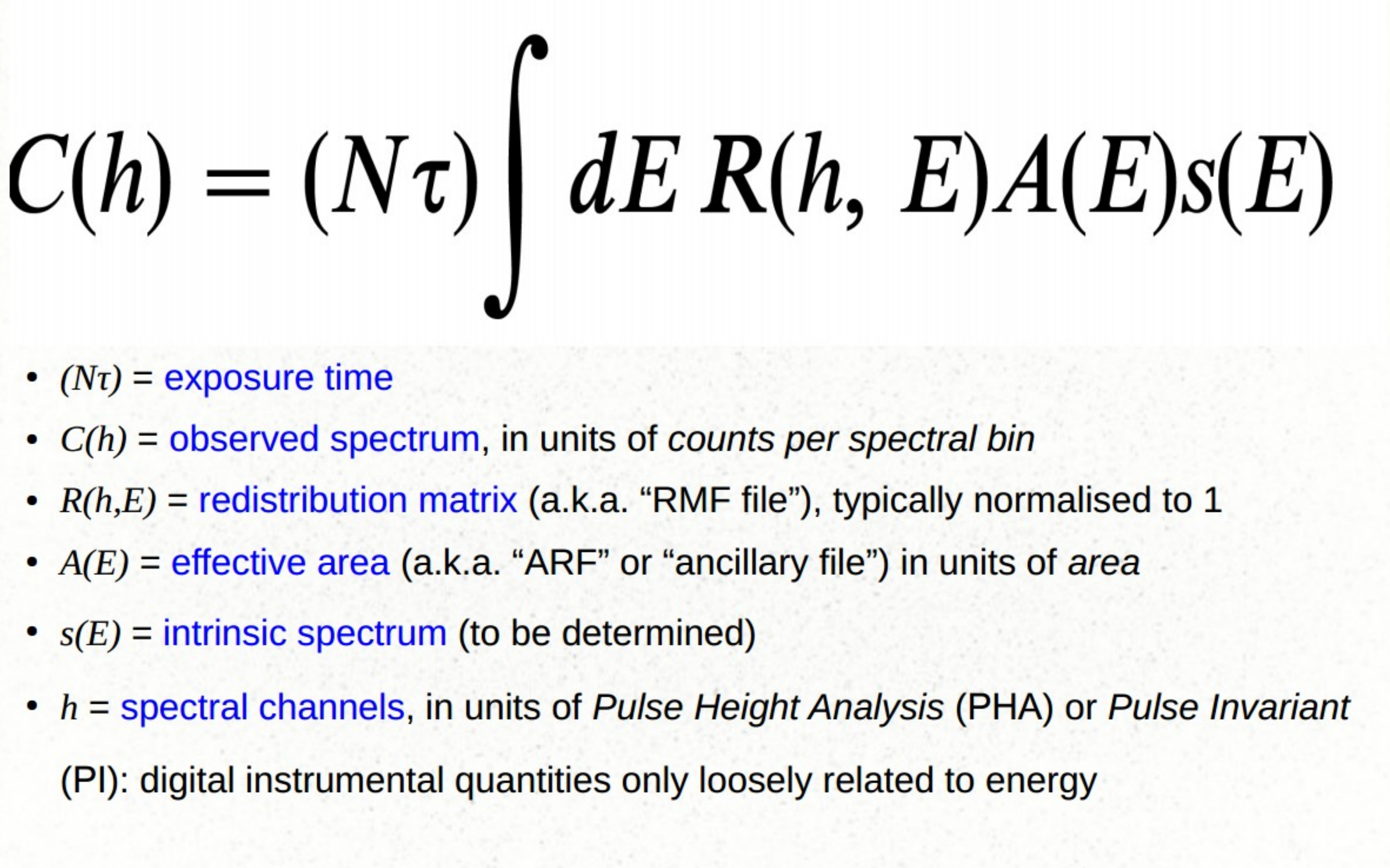


Ark 120 – SIS (AGN)



Second problem: the shape of the count spectra is dominated by the transfer function of the telescope+detector: we must “decode” it

The spectral equation



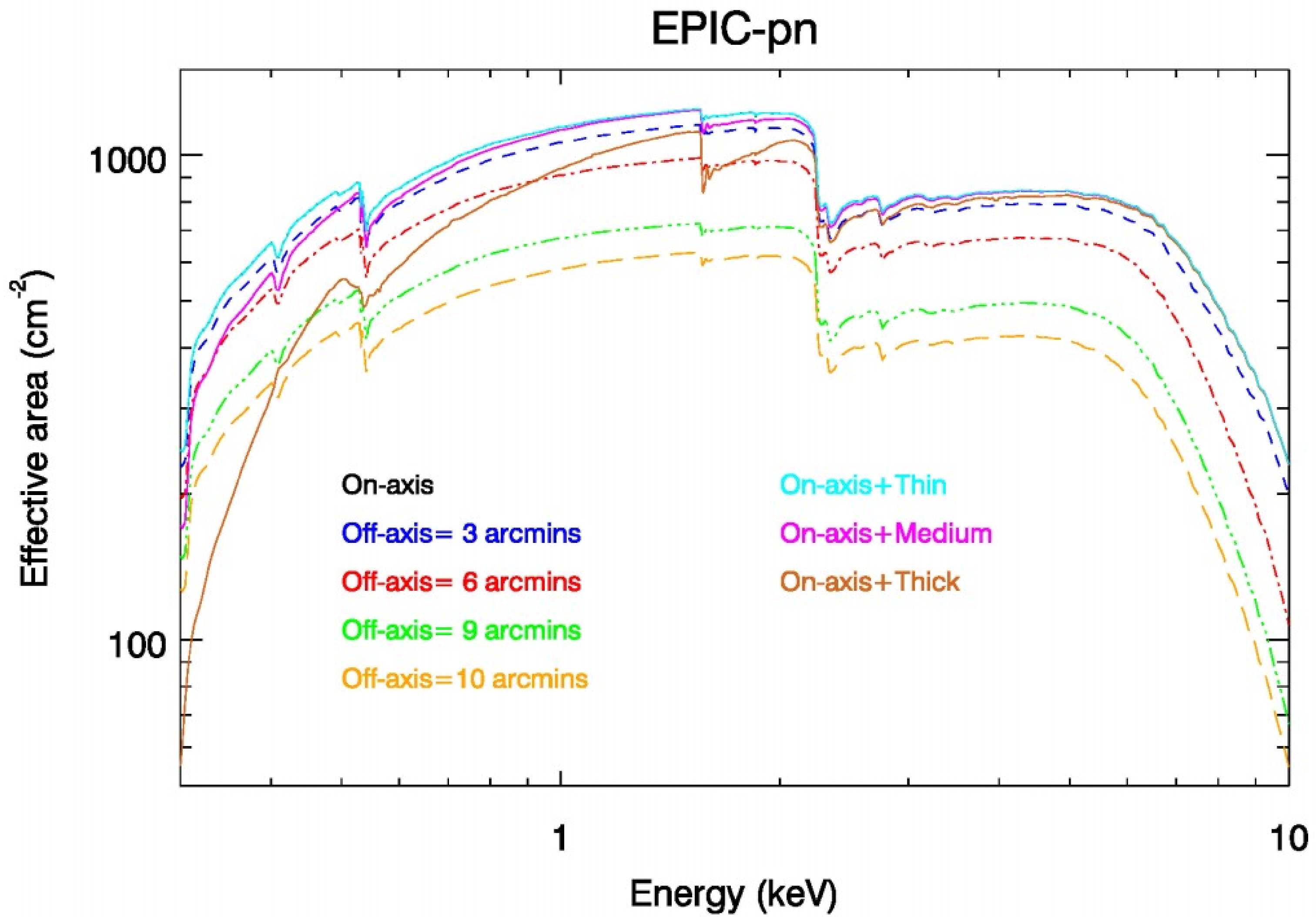
$$C(h) = (N\tau)\int dE\, R(h, E)A(E)s(E)$$

- $(N\tau)$ = exposure time
- $C(h)$ = observed spectrum, in units of *counts per spectral bin*
- $R(h,E)$ = redistribution matrix (a.k.a. “RMF file”), typically normalised to 1
- $A(E)$ = effective area (a.k.a. “ARF” or “ancillary file”) in units of *area*
- $s(E)$ = intrinsic spectrum (to be determined)
- h = spectral channels, in units of *Pulse Height Analysis* (PHA) or *Pulse Invariant* (PI): digital instrumental quantities only loosely related to energy

Davis, 2001, ApJ, 562, 575

We would need to invert this equation to get s(E).
However, in general this is not possible. Why?

The effective area A(E)

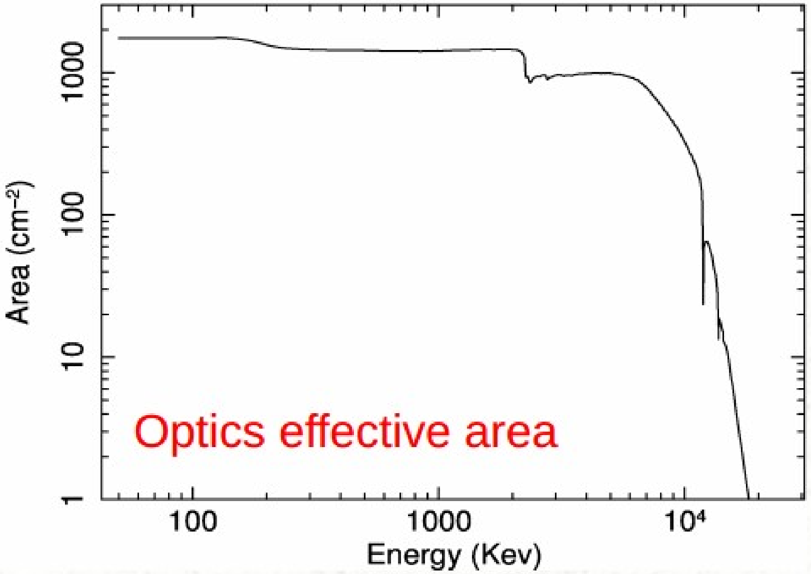


Measure (conventionally expressed in units of “*area*”) of the collecting power of telescope+filter+detector. It depends on energy and position (“off-axis”)

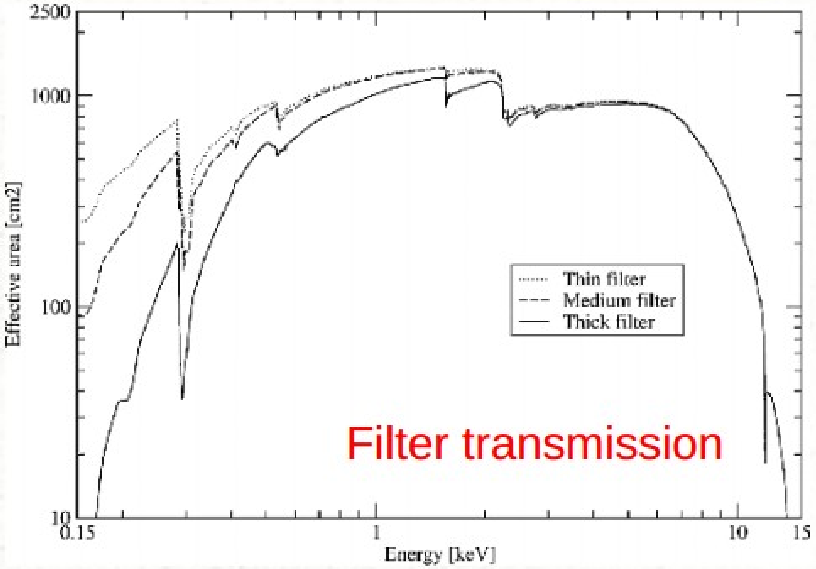
[Beware: not all observatories carry “*optical photon blocking filters*”]

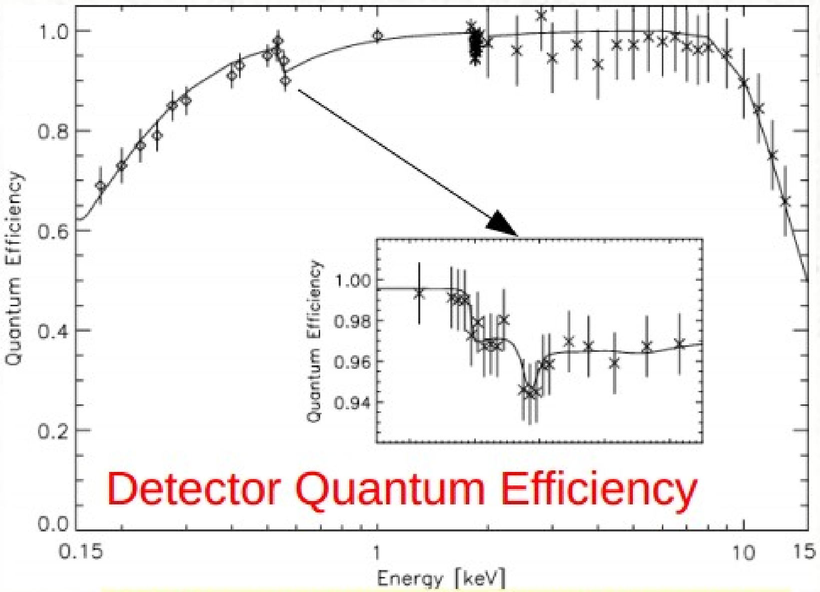
Components of the effective area

How many photons are focused by the mirrors

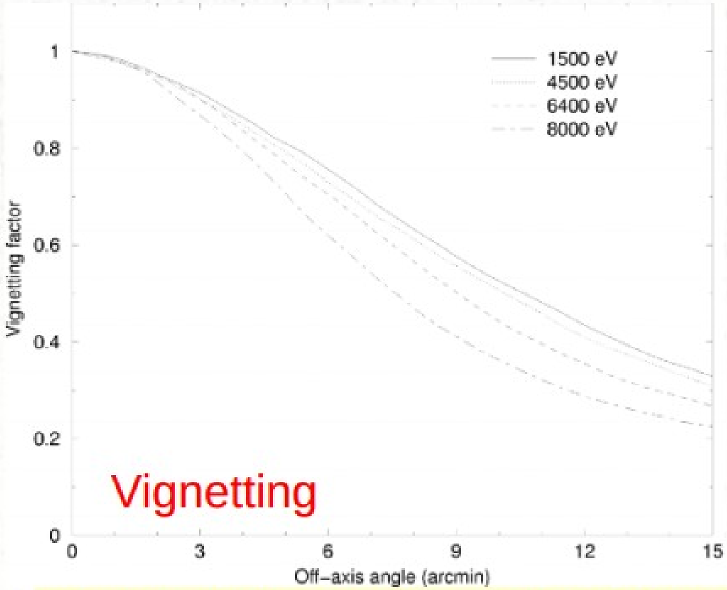


How many photons are transmitted by the filters





How many photons are detected by the camera

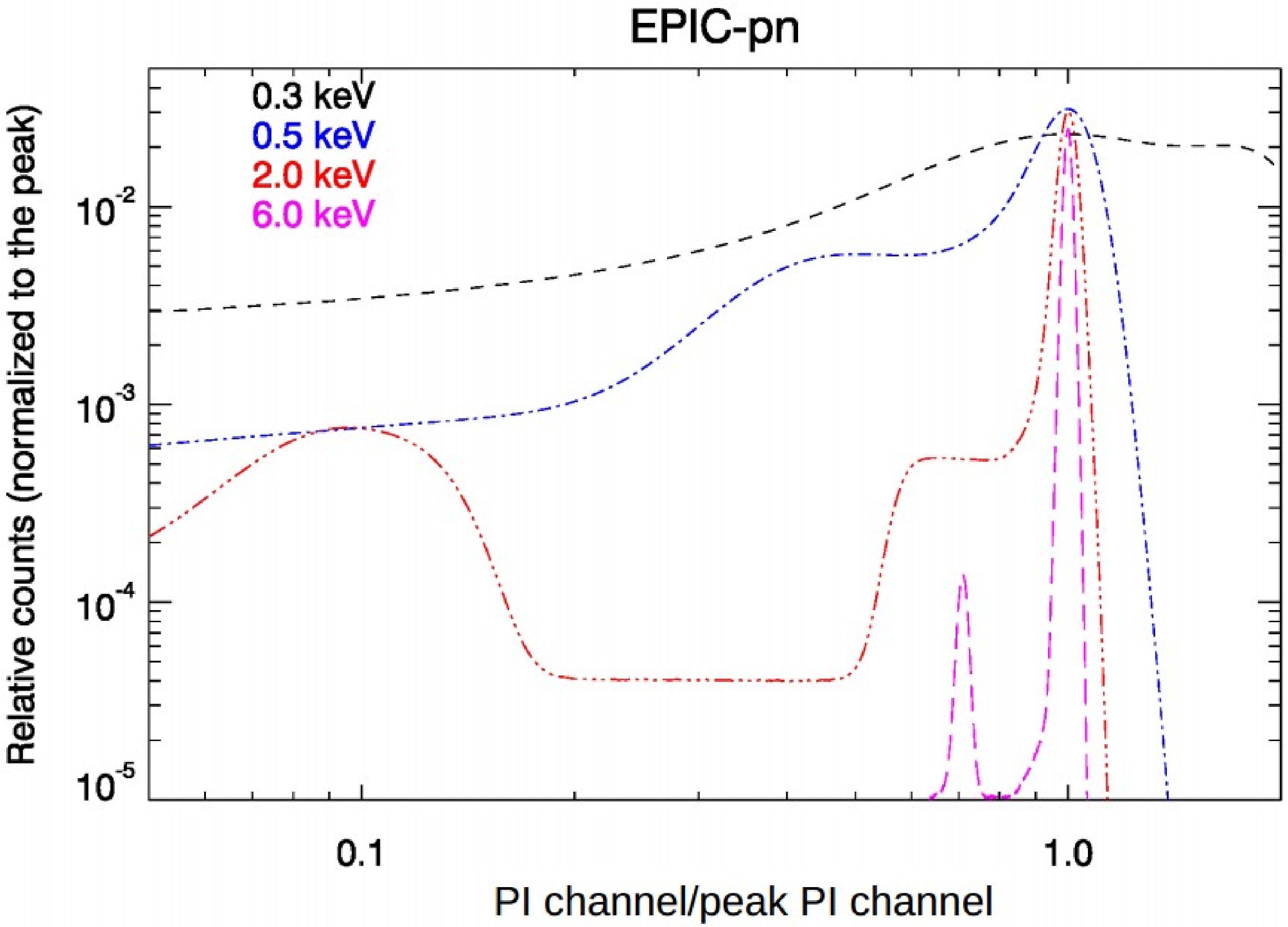


How many photons are intercepted by shadowing

Strüder et al., 2001, A&A, 365, L18

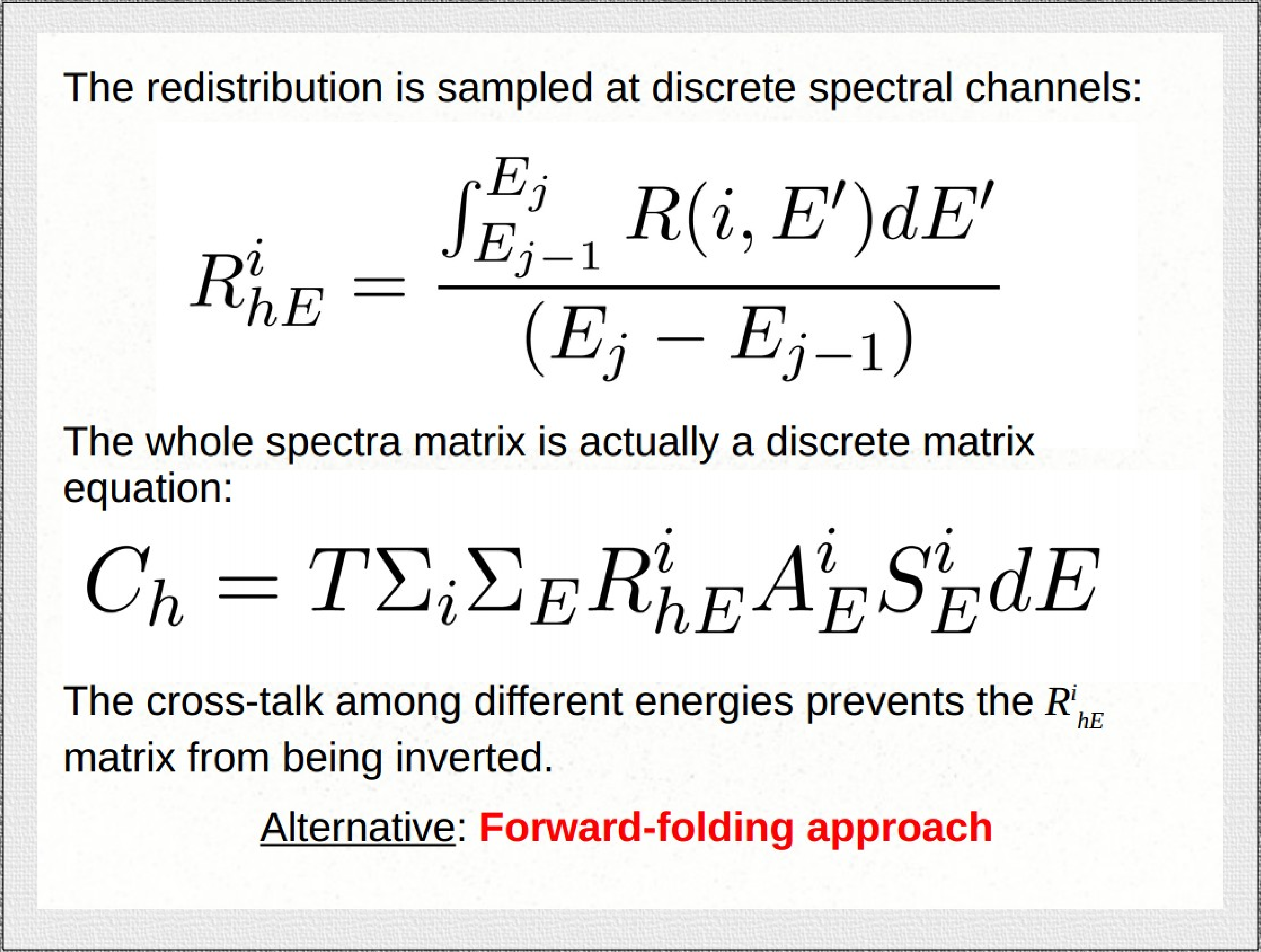
XMM-Newton User's Handbook

Redistribution matrix R(E)



Response of the detector to a monochromatic line. Highly dependent on the energy. The width of the core defines the instrument resolution:
$\boldsymbol{\sigma_{PHA} = [n^2+fE]^{0.5}}$ (n^2 → noise term)

Inverting the spectral equation?



The redistribution is sampled at discrete spectral channels:

$$R^i_{hE} = \frac{\int_{E_{j-1}}^{E_j} R(i, E')dE'}{(E_j - E_{j-1})}$$

The whole spectra matrix is actually a discrete matrix equation:

$$C_h = T\Sigma_i \Sigma_E R^i_{hE} A^i_E S^i_E dE$$

The cross-talk among different energies prevents the R^i_{hE} matrix from being inverted.

Alternative: **Forward-folding approach**

Forward-folding approach

1. Assume a model with its defining parameters

2. Define a set of parameter values

3. Convolve the model with the instrument response

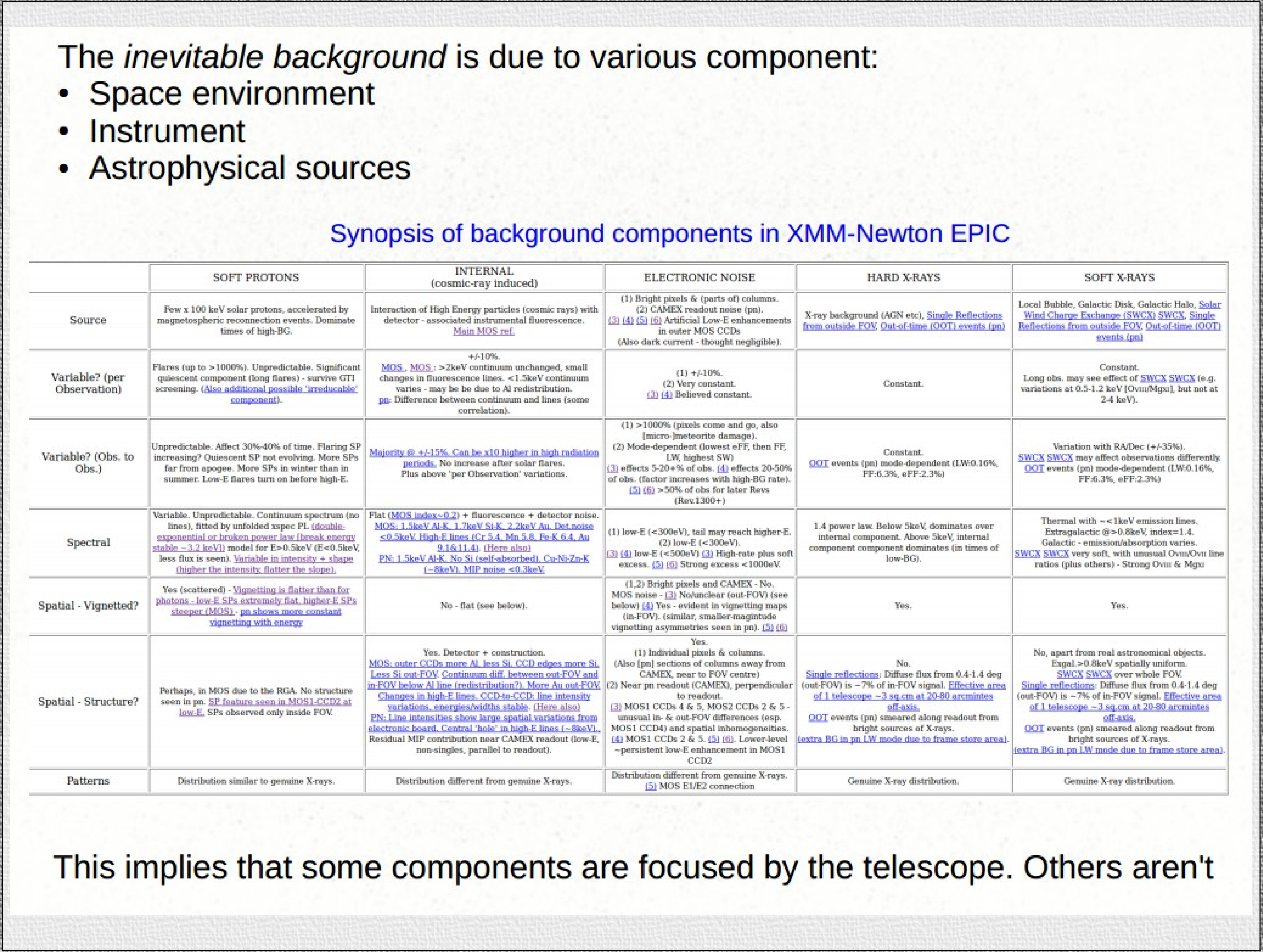
4. Compare the (dis)agreement between the observed spectrum and the folded model through a goodness-of-fit statistical test

5. Change the parameter values to minimize the goodness-of-fit test = **fit**

6. Once the best fit is found, calculate the confidence intervals on the best-fit parameters

Spectral packages are looping machines through the steps above
(+ some fancy cosmetic features)

Background spectra



The *inevitable background* is due to various component:

- Space environment
- Instrument
- Astrophysical sources

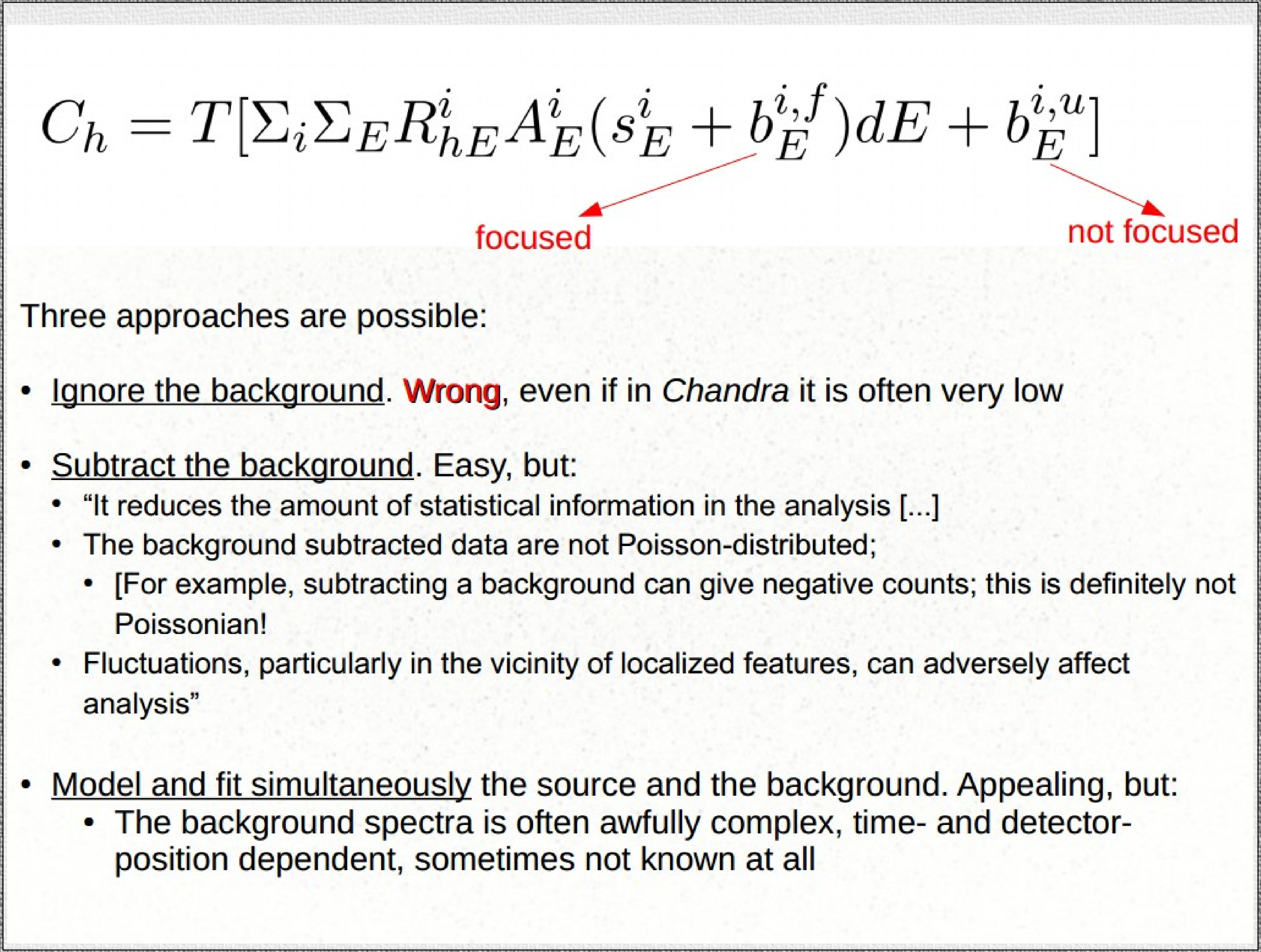
Synopsis of background components in XMM-Newton EPIC

	SOFT PROTONS	INTERNAL (cosmic-ray induced)	ELECTRONIC NOISE	HARD X-RAYS	SOFT X-RAYS
Source	Few x 100 keV solar protons, accelerated by magnetospheric reconnection events. Dominate times of high-BG.	Interaction of High Energy particles (cosmic rays) with detector - associated instrumental fluorescence. Main MOS ref.	(1) Bright pixels & (parts of) columns. (2) CAMEX readout noise (pn). (3) (4) (5) (6) Artificial Low-E enhancements in outer MOS CCDs (Also dark current - thought negligible).	X-ray background (AGN etc), Single Reflections from outside FOV, Out-of-time (OOT) events (pn)	Local Bubble, Galactic Disk, Galactic Halo, Solar Wind Charge Exchange (SWCX) SWCX, Single Reflections from outside FOV, Out-of-time (OOT) events (pn)
Variable? (per Observation)	Flares (up to >1000%). Unpredictable. Significant quiescent component (long flares) - survive GTI screening. (Also additional possible 'irreducable' component).	+/-10%. MOS , MOS : >2keV continuum unchanged, small changes in fluorescence lines. <1.5keV continuum varies - may be be due to Al redistribution. pn: Difference between continuum and lines (some correlation).	(1) +/-10%. (2) Very constant. (3) (4) Believed constant.	Constant.	Constant. Long obs. may see effect of SWCX SWCX (e.g. variations at 0.5-1.2 keV [Oviii/Mgxi], but not at 2-4 keV).
Variable? (Obs. to Obs.)	Unpredictable. Affect 30%-40% of time. Flaring SP increasing? Quiescent SP not evolving. More SPs far from apogee. More SPs in winter than in summer. Low-E flares turn on before high-E.	Majority @ +/-15%. Can be x10 higher in high radiation periods. No increase after solar flares. Plus above 'per Observation' variations.	(1) >1000% (pixels come and go, also [micro-]meteorite damage). (2) Mode-dependent (lowest eFF, then FF, LW, highest SW) (3) effects 5-20+% of obs. (4) effects 20-50% of obs. (factor increases with high-BG rate). (5) (6) >50% of obs for later Revs (Rev.1300+)	Constant. OOT events (pn) mode-dependent (LW:0.16%, FF:6.3%, eFF:2.3%)	Variation with RA/Dec (+/-35%). SWCX SWCX may affect observations differently. OOT events (pn) mode-dependent (LW:0.16%, FF:6.3%, eFF:2.3%)
Spectral	Variable. Unpredictable. Continuum spectrum (no lines), fitted by unfolded xspec PL (double-exponential or broken power law [break energy stable ~3.2 keV]) model for E>0.5keV (E<0.5keV, less flux is seen). Variable in intensity + shape (higher the intensity, flatter the slope).	Flat (MOS index~0.2) + fluorescence + detector noise. MOS: 1.5keV Al-K, 1.7keV Si-K, 2.2keV Au. Det.noise <0.5keV. High-E lines (Cr 5.4, Mn 5.8, Fe-K 6.4, Au 9.1&11.4). (Here also) PN: 1.5keV Al-K. No Si (self-absorbed). Cu-Ni-Zn-K (~8keV). MIP noise <0.3keV.	(1) low-E (<300eV), tail may reach higher-E. (2) low-E (<300eV). (3) (4) low-E (<500eV) (3) High-rate plus soft excess. (5) (6) Strong excess <1000eV.	1.4 power law. Below 5keV, dominates over internal component. Above 5keV, internal component component dominates (in times of low-BG).	Thermal with ~<1keV emission lines. Extragalactic @>0.8keV, index=1.4. Galactic - emission/absorption varies. SWCX SWCX very soft, with unusual Oviii/Ovii line ratios (plus others) - Strong Oviii & Mgxi
Spatial - Vignetted?	Yes (scattered) - Vignetting is flatter than for photons - low-E SPs extremely flat, higher-E SPs steeper (MOS) - pn shows more constant vignetting with energy	No - flat (see below).	(1,2) Bright pixels and CAMEX - No. MOS noise - (3) No/unclear (out-FOV) (see below) (4) Yes - evident in vignetting maps (in-FOV). (similar, smaller-magintude vignetting asymmetries seen in pn). (5) (6)	Yes.	Yes.
Spatial - Structure?	Perhaps, in MOS due to the RGA. No structure seen in pn. SP feature seen in MOS1-CCD2 at low-E. SPs observed only inside FOV.	Yes. Detector + construction. MOS: outer CCDs more Al, less Si. CCD edges more Si. Less Si out-FOV. Continuum diff. between out-FOV and in-FOV below Al line (redistribution?). More Au out-FOV. Changes in high-E lines. CCD-to-CCD: line intensity variations, energies/widths stable. (Here also) PN: Line intensities show large spatial variations from electronic board. Central 'hole' in high-E lines (~8keV).. Residual MIP contribution near CAMEX readout (low-E, non-singles, parallel to readout).	Yes. (1) Individual pixels & columns. (Also [pn] sections of columns away from CAMEX, near to FOV centre) (2) Near pn readout (CAMEX), perpendicular to readout. (3) MOS1 CCDs 4 & 5, MOS2 CCDs 2 & 5 - unusual in- & out-FOV differences (esp. MOS1 CCD4) and spatial inhomogeneities. (4) MOS1 CCDs 2 & 5. (5) (6). Lower-level ~persistent low-E enhancement in MOS1 CCD2	No. Single reflections: Diffuse flux from 0.4-1.4 deg (out-FOV) is ~7% of in-FOV signal. Effective area of 1 telescope ~3 sq.cm at 20-80 arcmintes off-axis. OOT events (pn) smeared along readout from bright sources of X-rays. (extra BG in pn LW mode due to frame store area).	No, apart from real astronomical objects. Exgal.>0.8keV spatially uniform. SWCX SWCX over whole FOV. Single reflections: Diffuse flux from 0.4-1.4 deg (out-FOV) is ~7% of in-FOV signal. Effective area of 1 telescope ~3 sq.cm at 20-80 arcmintes off-axis. OOT events (pn) smeared along readout from bright sources of X-rays. (extra BG in pn LW mode due to frame store area).
Patterns	Distribution similar to genuine X-rays.	Distribution different from genuine X-rays.	Distribution different from genuine X-rays. (5) MOS E1/E2 connection	Genuine X-ray distribution.	Genuine X-ray distribution.

This implies that some components are focused by the telescope. Others aren't

(Courtesy A.Read: http://www.star.le.ac.uk/~amr30/BG/BGTable.html))

How to deal with background spectra



$$C_h = T[\Sigma_i \Sigma_E R^i_{hE} A^i_E (s^i_E + b^{i,f}_E) dE + b^{i,u}_E]$$

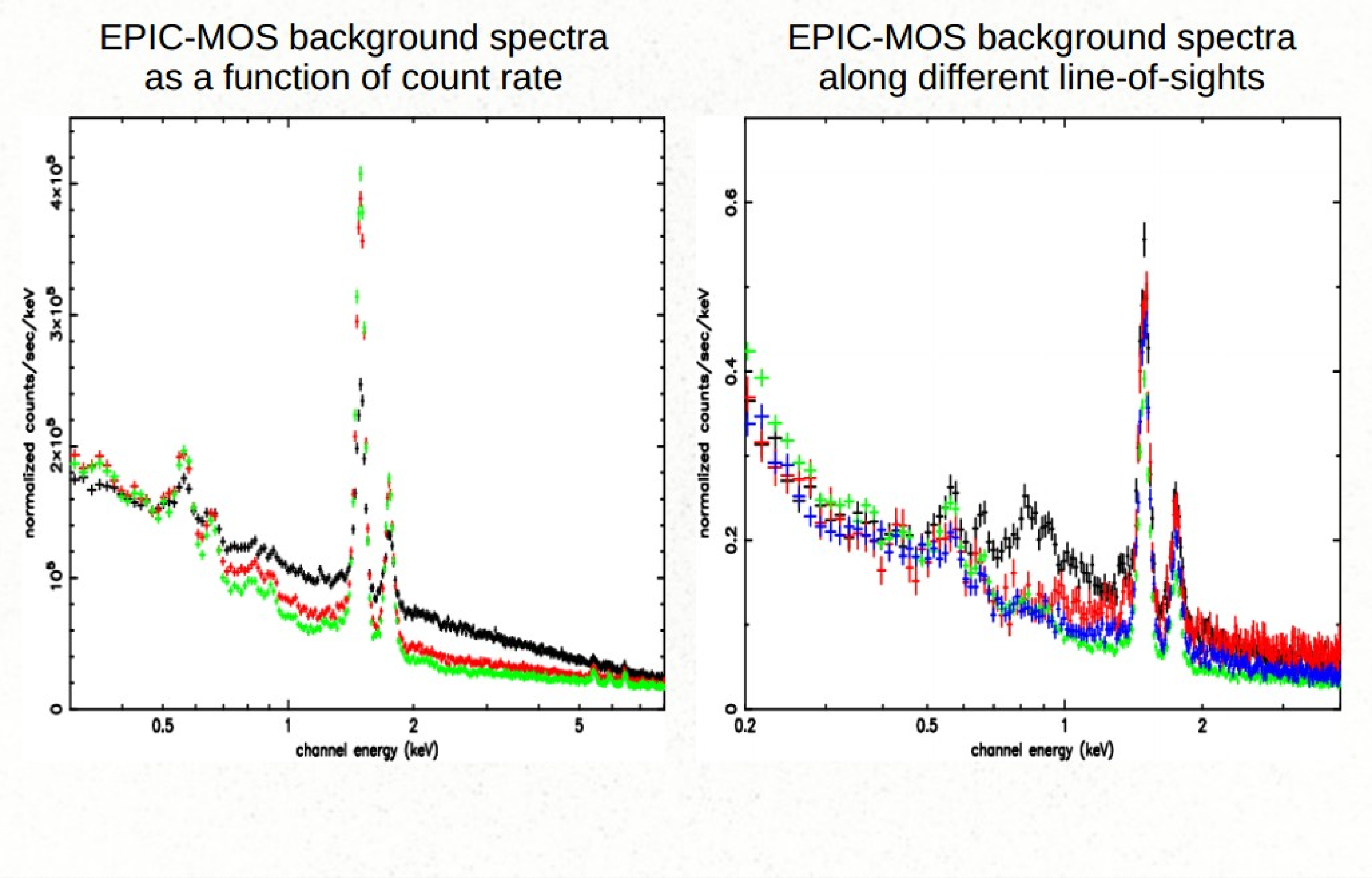
focused

not focused

Three approaches are possible:

- Ignore the background. **Wrong**, even if in *Chandra* it is often very low
- Subtract the background. Easy, but:
  - “It reduces the amount of statistical information in the analysis [...]
  - The background subtracted data are not Poisson-distributed;
    - [For example, subtracting a background can give negative counts; this is definitely not Poissonian!
  - Fluctuations, particularly in the vicinity of localized features, can adversely affect analysis”
- Model and fit simultaneously the source and the background. Appealing, but:
  - The background spectra is often awfully complex, time- and detector-position dependent, sometimes not known at all

How to deal with background spectra

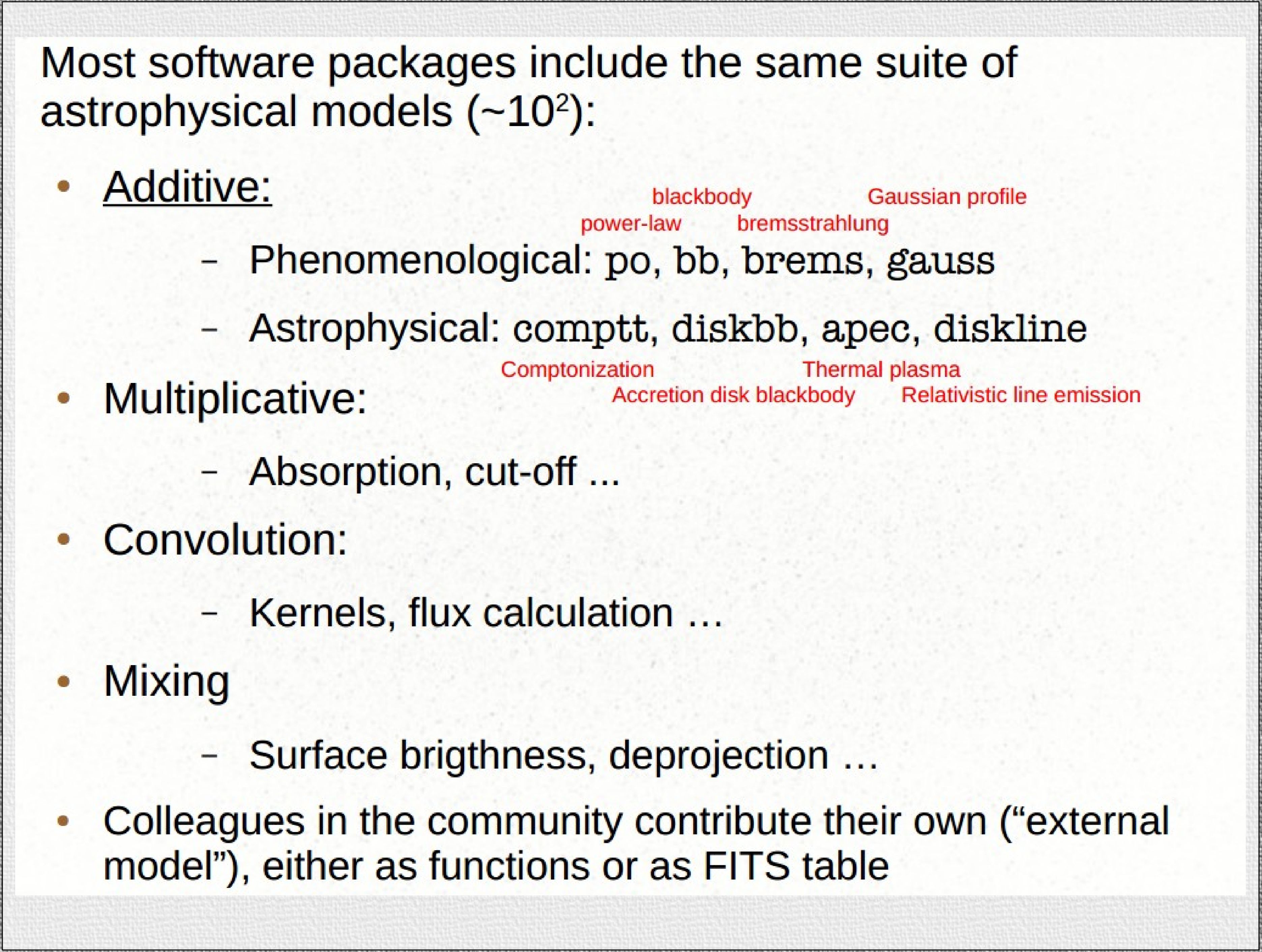


EPIC-MOS background spectra as a function of count rate

EPIC-MOS background spectra along different line-of-sights

(Carter & Read, 2007, A&A, 464, 1155)

Models



Most software packages include the same suite of astrophysical models (~10^2):

- Additive:

  blackbody Gaussian profile
  power-law bremsstrahlung

  - Phenomenological: po, bb, brems, gauss
  - Astrophysical: comptt, diskbb, apec, diskline

  Comptonization Thermal plasma
  Accretion disk blackbody Relativistic line emission

- Multiplicative:
  - Absorption, cut-off ...
- Convolution:
  - Kernels, flux calculation …
- Mixing
  - Surface brigthness, deprojection …
- Colleagues in the community contribute their own (“external model”), either as functions or as FITS table

Goodness-of-fit tests

- The most common goodness-of-fit statistic test is the “chi-squared” (χ^2):

$$X^2 = \sum \frac{(\text{observed - expected})^2}{\text{expected}}$$

  - It requires that the distribution of background-subtracted counts in each spectral channel is well approximated by a Gaussian (5-10 counts)
  - Different alternatives from the denumerator: *weight churazov* or *weight model*

- Alternatively, one can use the Cash (C-)statistics: $C = 2\sum_{i=1}^{n} s_i - N_i + N_i \ln(N_i/s_i).$

  – Applicable to data following the Poissonian statistics only i.e.: non-background subtracted spectra)

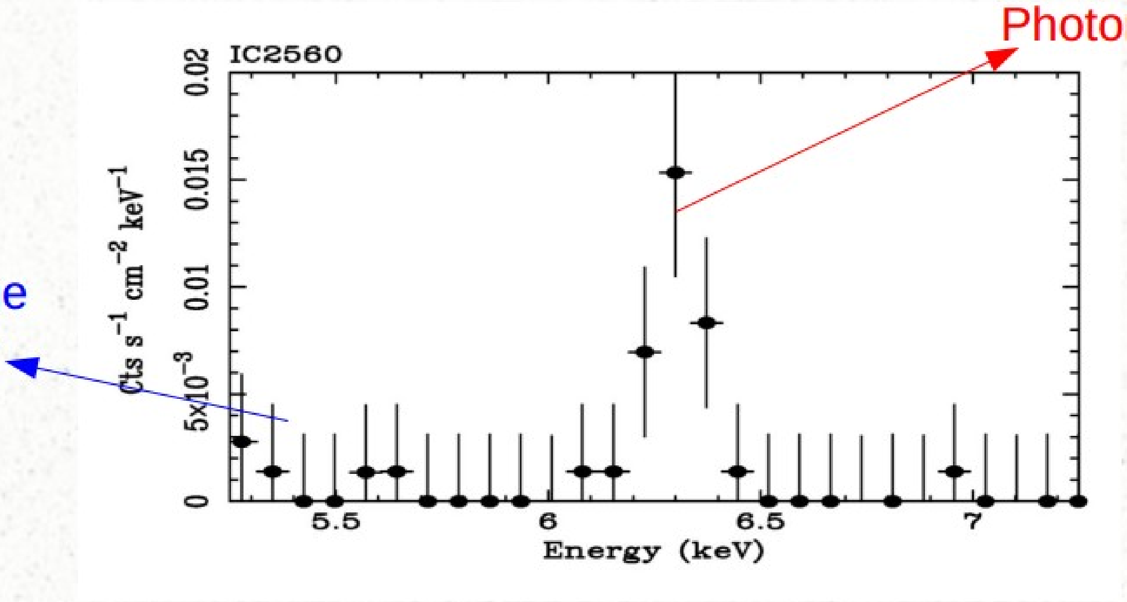
  – XSPEC implements a flavour (the “W-statistics”) which can be directly applied to background-subtracted spectra

  – It does not yield a metrics of the absolute quality of a a fit (one need to use Monte-Carlo simulations in this case)

(see Arnaud et al., “Fitting low count spectra”,https://astrophysics.gsfc.nasa.gov/XSPECwiki/low_count_spectra)
(Cash, 1976, ApJ, 52, 307)

Data rebinning

- Rebin you spectra is pure evil, and may lead to loss of scientific information:



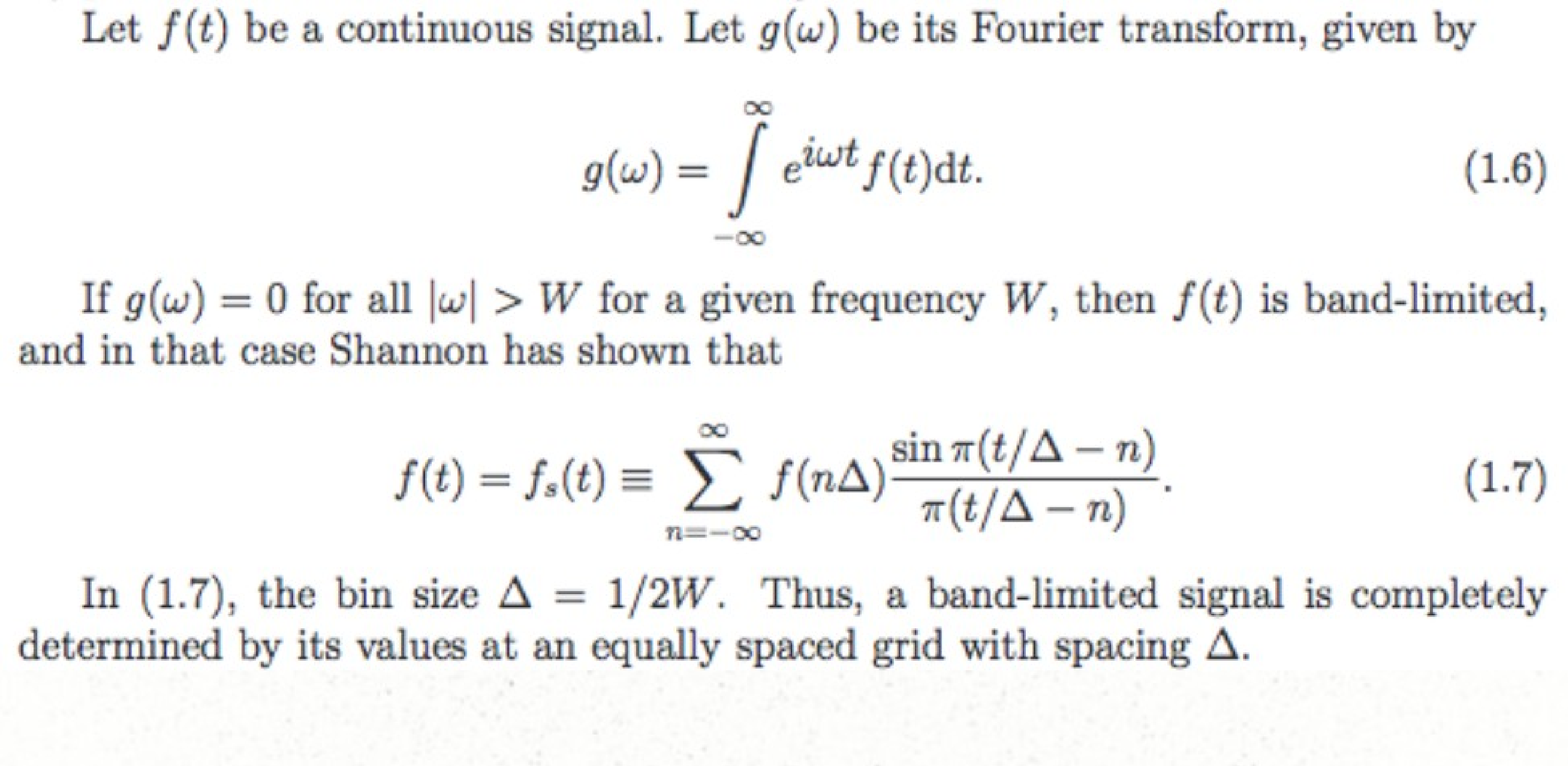
Photons in the line: 21

Photons in the continuum: 9

- However, a minimum level of spectral rebinning is required to avoid oversampling the intrinsic resolution of the instrument

(Guainazzi et al., 2005, MNRAS; 356, 295)

Shannon theorem



Let $f(t)$ be a continuous signal. Let $g(\omega)$ be its Fourier transform, given by

$$g(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} f(t) \mathrm{d}t. \tag{1.6}$$

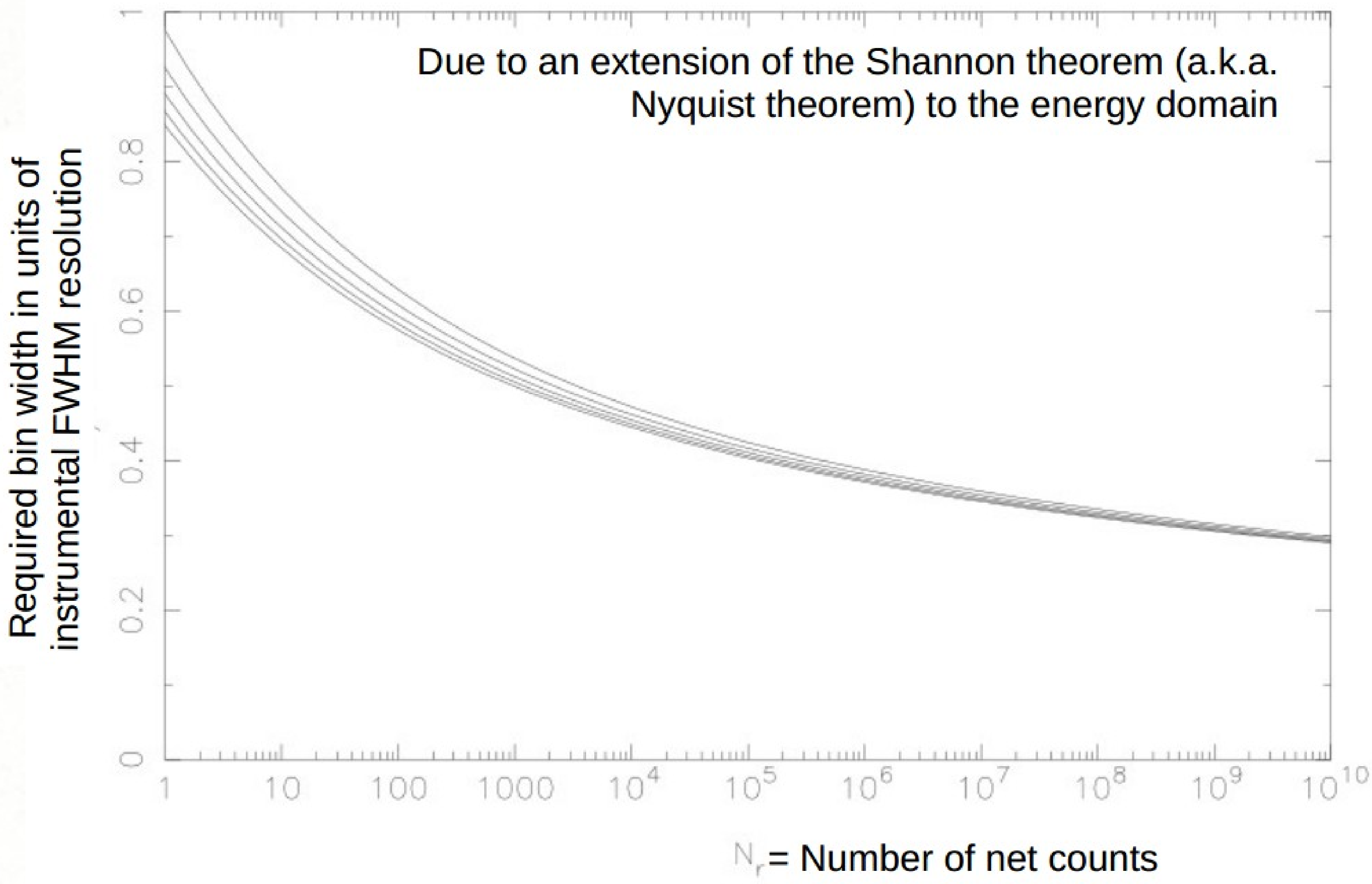
If $g(\omega) = 0$ for all $|\omega| > W$ for a given frequency W, then $f(t)$ is band-limited, and in that case Shannon has shown that

$$f(t) = f_s(t) \equiv \sum_{n=-\infty}^{\infty} f(n\Delta) \frac{\sin \pi(t/\Delta - n)}{\pi(t/\Delta - n)}. \tag{1.7}$$

In (1.7), the bin size $\Delta = 1/2W$. Thus, a band-limited signal is completely determined by its values at an equally spaced grid with spacing Δ.

(from J.Kaastra's and F.Verbunt's lecture notes on high-energy astrophysics, 2008)

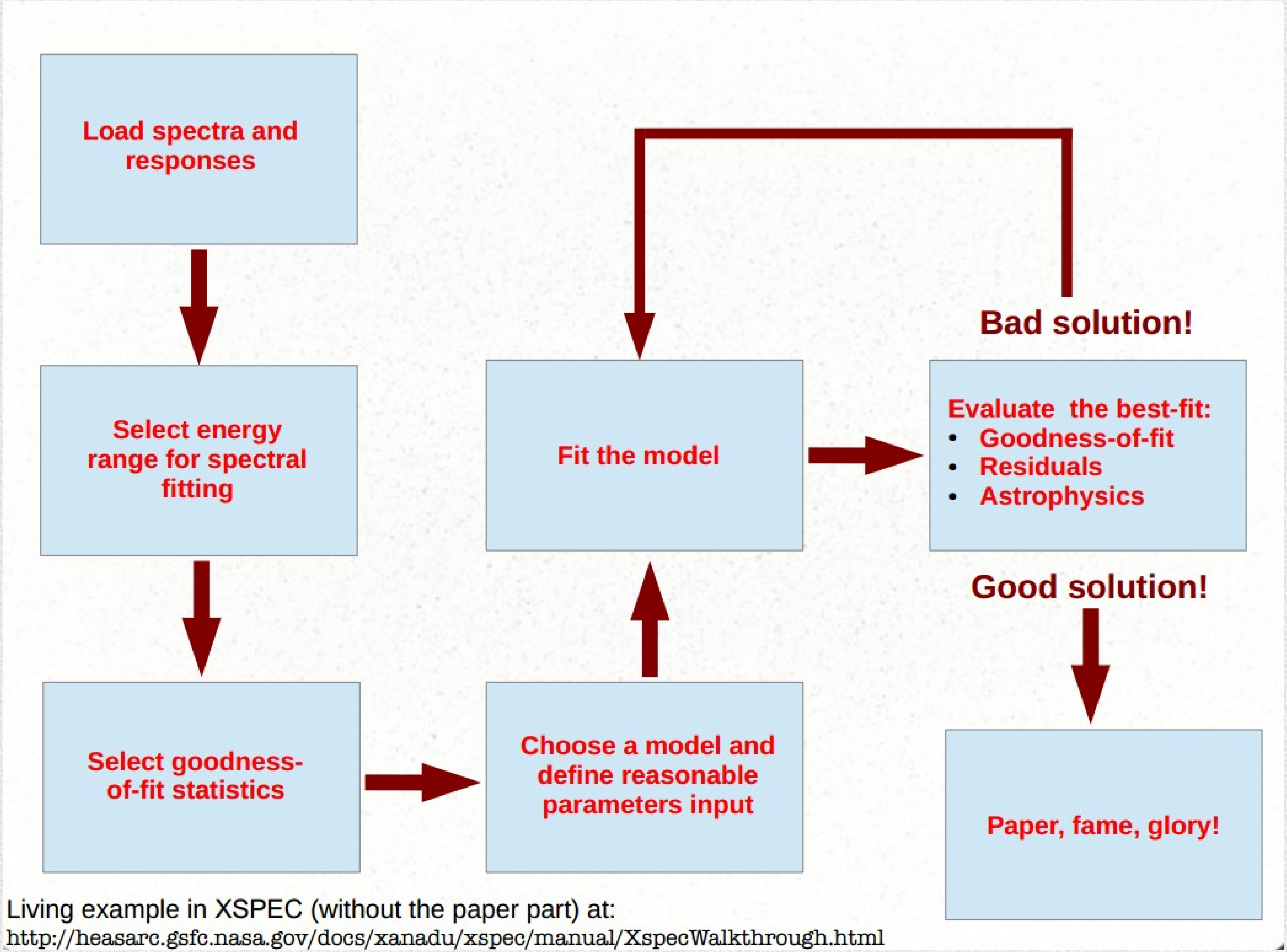
An ideal rebinning strategy



Due to an extension of the Shannon theorem (a.k.a. Nyquist theorem) to the energy domain

`specgroup` in SAS implements this, and many other spectral rebinning schemes

Forward-folding in action



Load spectra and responses

Select energy range for spectral fitting

Select goodness-of-fit statistics

Choose a model and define reasonable parameters input

Fit the model

Evaluate the best-fit:
- Goodness-of-fit
- Residuals
- Astrophysics

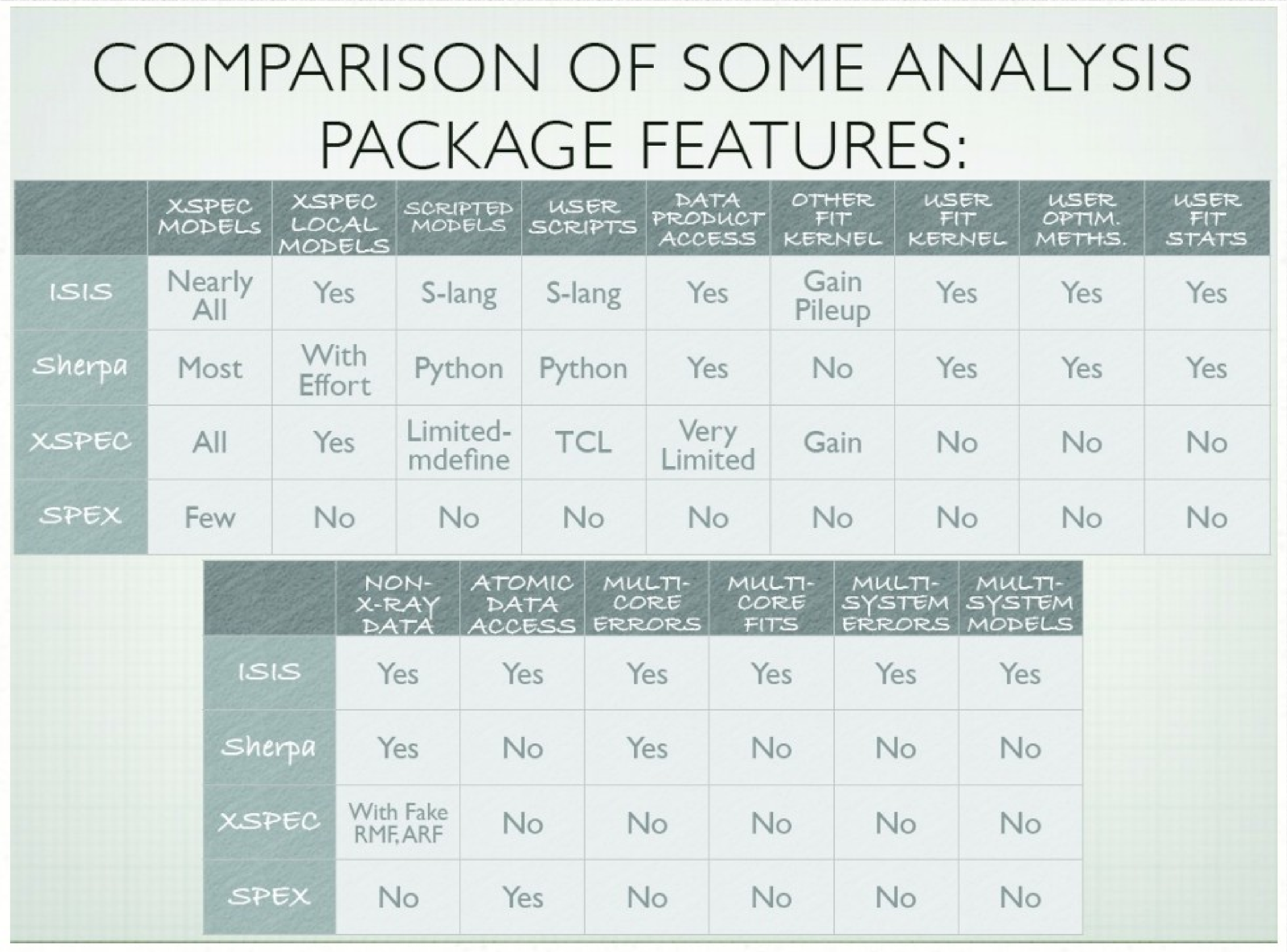
Bad solution!

Good solution!

Paper, fame, glory!

Living example in XSPEC (without the paper part) at:
http://heasarc.gsfc.nasa.gov/docs/xanadu/xspec/manual/XspecWalkthrough.html

Features of the existing X-Ray fitting packages



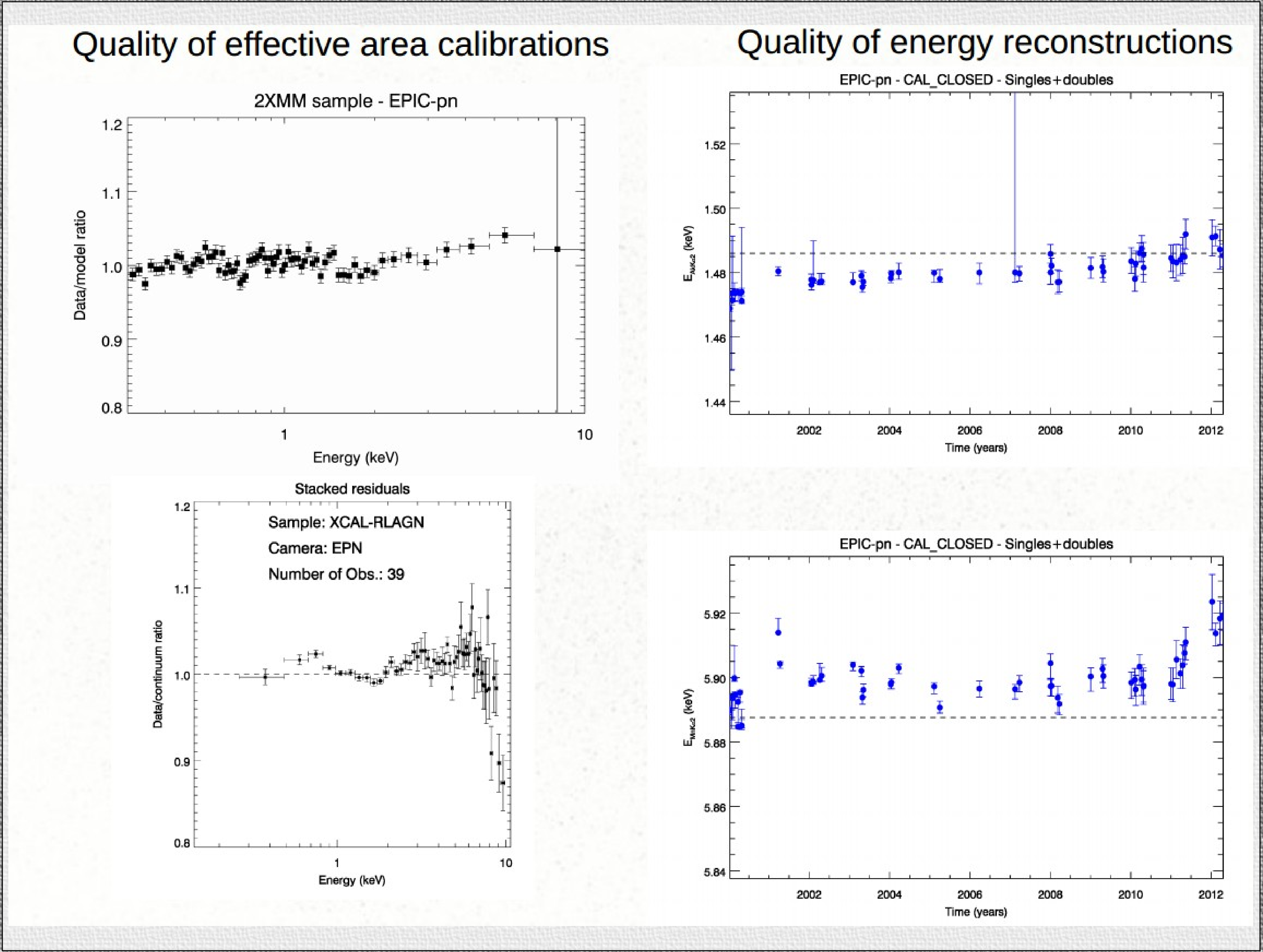
COMPARISON OF SOME ANALYSIS PACKAGE FEATURES:

	XSPEC MODELS	XSPEC LOCAL MODELS	SCRIPTED MODELS	USER SCRIPTS	DATA PRODUCT ACCESS	OTHER FIT KERNEL	USER FIT KERNEL	USER OPTIM. METHS.	USER FIT STATS
ISIS	Nearly All	Yes	S-lang	S-lang	Yes	Gain Pileup	Yes	Yes	Yes
Sherpa	Most	With Effort	Python	Python	Yes	No	Yes	Yes	Yes
XSPEC	All	Yes	Limited-mdefine	TCL	Very Limited	Gain	No	No	No
SPEX	Few	No	No	No	No	No	No	No	No

	NON-X-RAY DATA	ATOMIC DATA ACCESS	MULTI-CORE ERRORS	MULTI-CORE FITS	MULTI-SYSTEM ERRORS	MULTI-SYSTEM MODELS
ISIS	Yes	Yes	Yes	Yes	Yes	Yes
Sherpa	Yes	No	Yes	No	No	No
XSPEC	With Fake RMF, ARF	No	No	No	No	No
SPEX	No	Yes	No	No	No	No

Credit M. Nowak

Systematic errors (example from EPIC-pn)

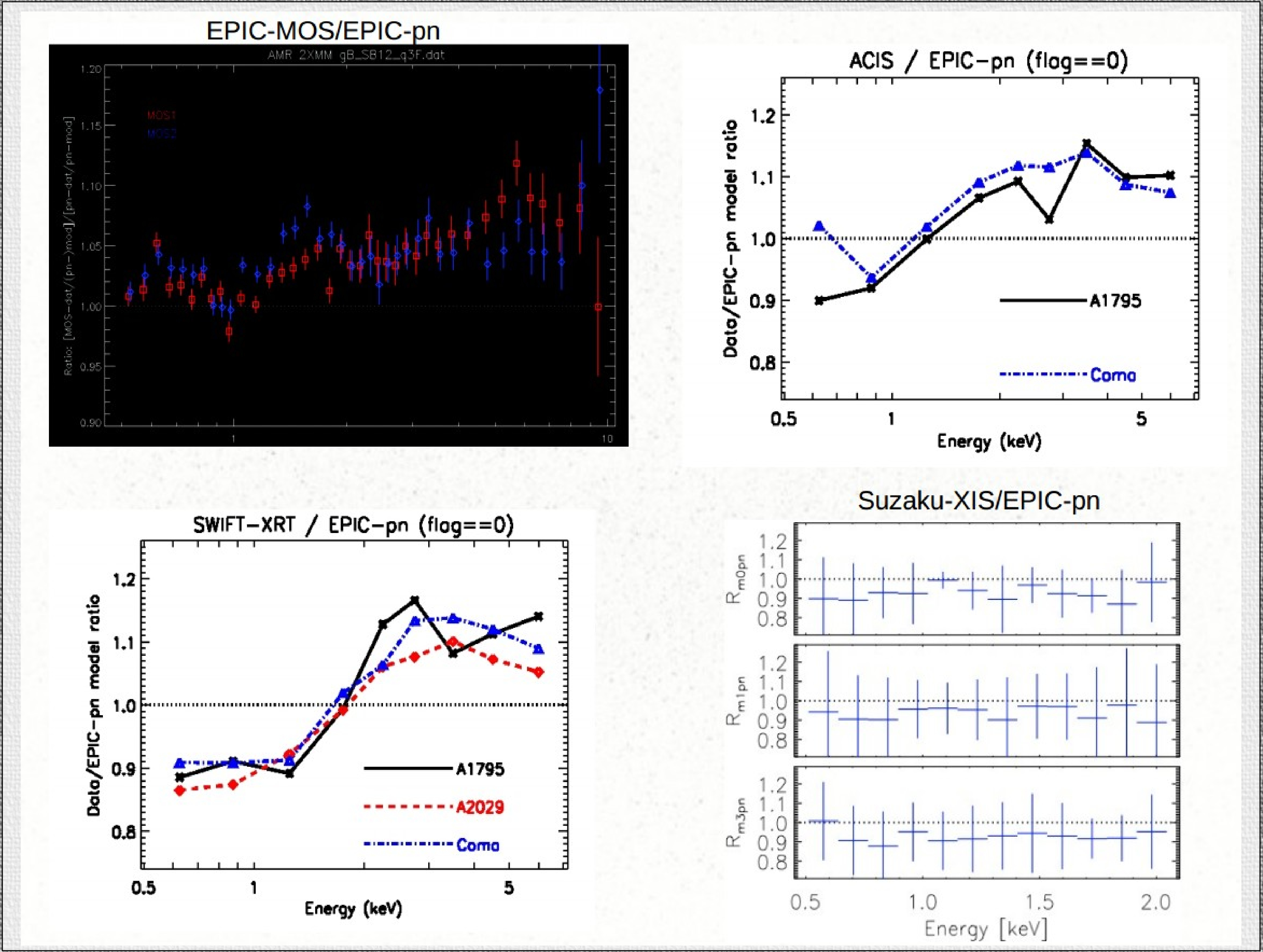


Quality of effective area calibrations

Quality of energy reconstructions

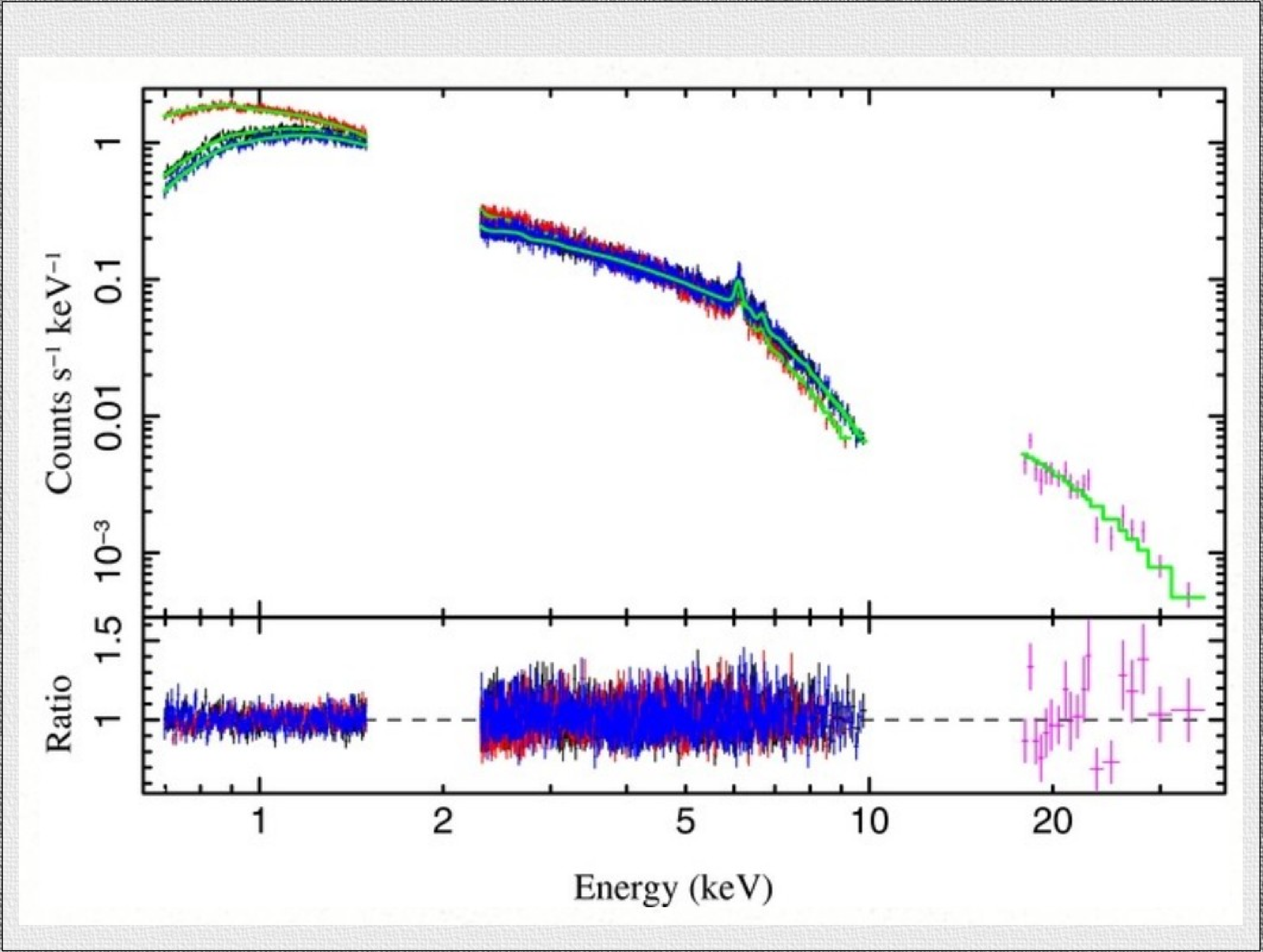
(Guainazzi et al., 2013, XMM-SOC-CAL-TN_0018)

Features of the existing X-Ray fitting packages



(Courtesy A.Read, J.Nevalainen (2x); Kettula et al., 2013, A&A, 552, 47)

An example of how not to deal with systematic errors



An example of how not to deal with systematic errors