Part 2. **OBSERVATIONS**

Lecture 6: Statistics of measurements

Do we measure the source or simply fluctuation in the noise:

- instrumental noise, spurious signals in the absence of any photons, (CCDs, PC, readout processes):
  i) *statistical* in nature – i.e. due to randomly arriving cosmic rays
  ii) *systematical* in nature – i.e. aging of detector

- statistical fluctuations - “noise” inherent randomness of certain types of events.

We measure the rate of arrival of photons in a limited time interval.

Poisson distribution:

Consider source of constant luminosity (not pulsating), that produces, on average, 100 counts every second. 10 ms between each count.

Poisson distribution:

Consider source of constant luminosity (not pulsating), that produces, on average, 100 counts every second. 10 ms between each count.

Each photon arrives at a time completely uncorrelated with each others (randomly)

Poisson distribution:

Consider source of constant luminosity (not pulsating), that produces, on average, 100 counts every second. 10 ms between each count.

Each photon arrives at a time completely uncorrelated with each others (randomly).

Average time of arrival is governed by fixed *probability* of an event occurring in some fixed interval of time.

Poisson distribution:

Consider source of constant luminosity (not pulsating), that produces, on average, 100 counts every second. 10 ms between each count.

Each photon arrives at a time completely uncorrelated with each others (randomly).

Average time of arrival is governed by fixed *probability* of an event occurring in some fixed interval of time.

A distribution of counts $N(x)$ can be obtained from many measurements: $N(105)$, $N(95)$, $N(87)$, $N(101)$ ... in 1 s intervals.
N – number of times
x – a given value occurs.

Poisson distribution:

For a random process such as photon arrival time, this distribution is well known theoretically as Poisson distribution:

$$P_x = \frac{m^x e^{-m}}{x!}$$

Poisson distribution:

For a random process such as photon arrival time, this distribution is well known theoretically as Poisson distribution:

$$P_x = \frac{m^x e^{-m}}{x!}$$

m – average (mean) number of events over a large number of tries,
x – integer number of events (counts).

if m=10.3 P(6 photons) = 0.056
if m=6 P(6 photons) = 0.161 but < 1

It is not particularly likely that one will detect the mean number.

Poisson distribution:

Is valid for discrete independent events that occur randomly (equal probability of occurrence per unit time) with a low probability of occurrence in a differential time integral *dt*.

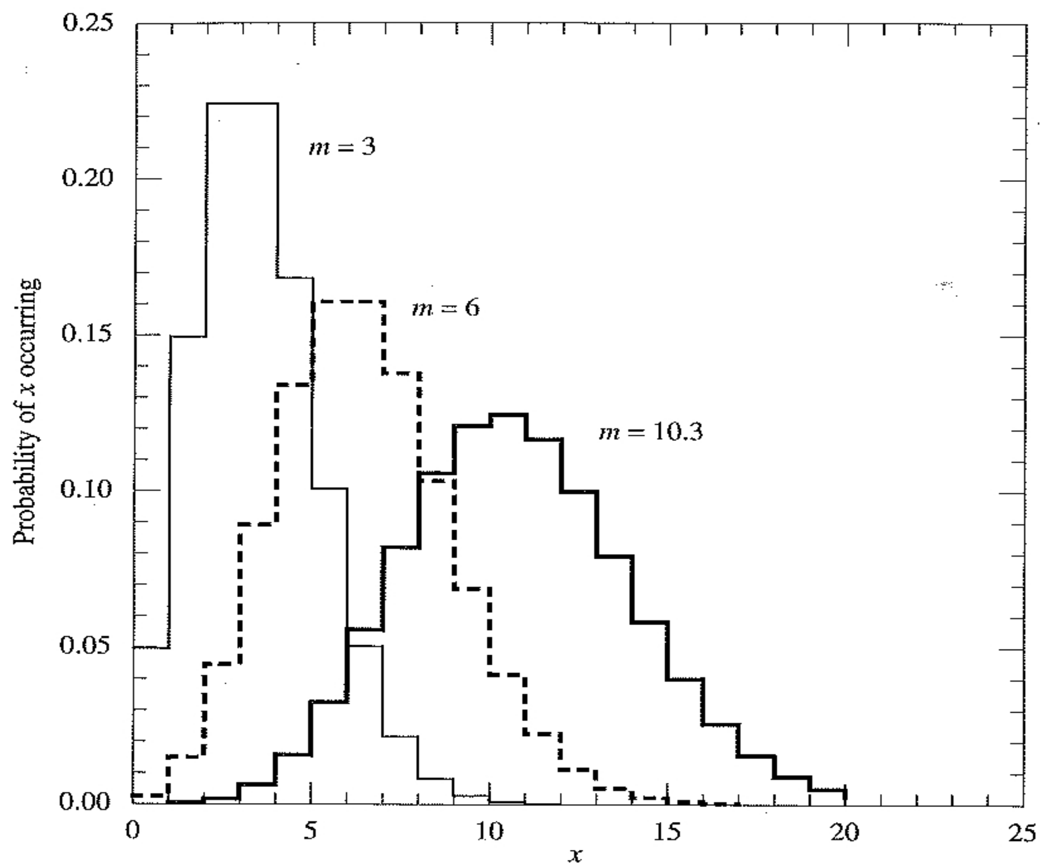


Figure 6.7. The Poisson distribution for small mean numbers, $m = 3.0$, 6.0 and 10.3. The ordinate gives the probability of the value x occurring, for the given mean value. Note the asymmetry of the histograms.

Poisson distribution:

$$\sum_{x=0}^{x=\infty} P_x = 1$$

- distribution i.s not symmetric,
- more symmetric for higher m.

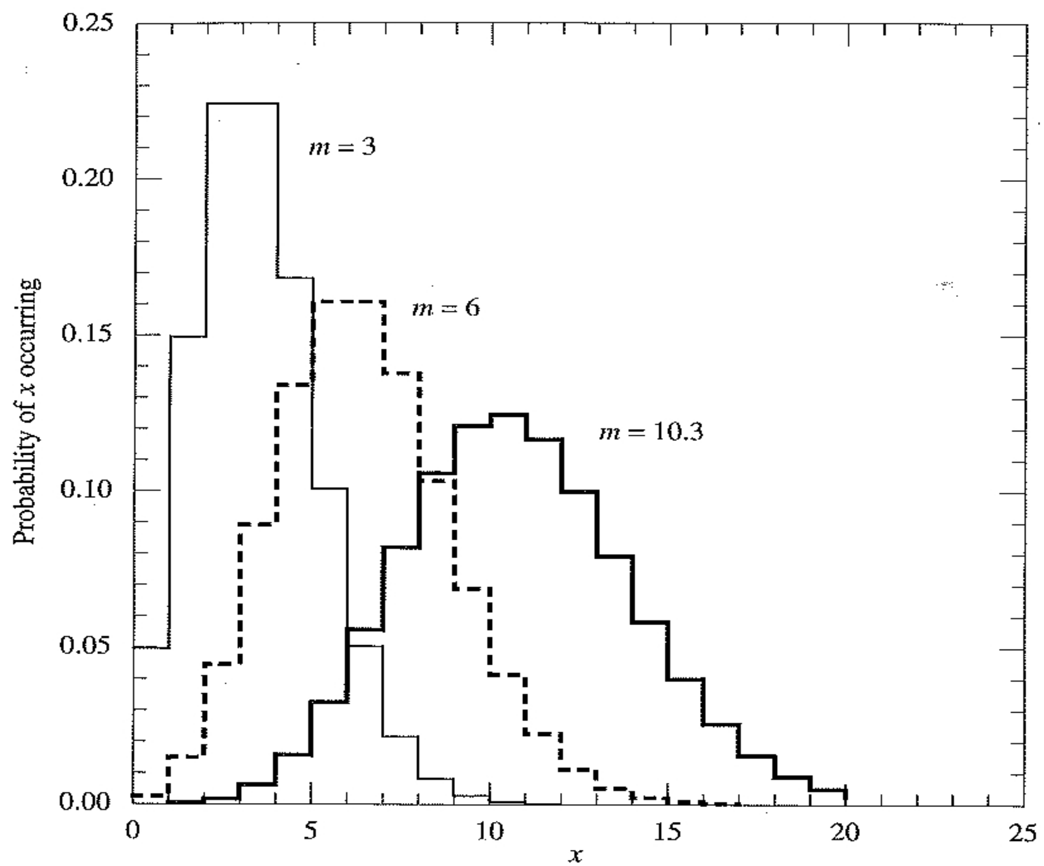


Figure 6.7. The Poisson distribution for small mean numbers, $m = 3.0$, 6.0 and 10.3. The ordinate gives the probability of the value x occurring, for the given mean value. Note the asymmetry of the histograms.

Poisson distribution:

-probability of obtaining non zero event is significant.

Table 6.1. *Sample values of Poisson function P_x*

x:	0	1	2	3	4	5	6	7[a]	8	9
$m = 1$	0.368	0.368	0.184	0.061	0.015	0.003	0.001	7E–5	9E–6	1E–6
$m = 2$	0.135	0.271	0.271	0.180	0.090	0.036	0.012	0.003	0.001	2E–4
$m = 3$	0.050	0.149	0.224	0.224	0.168	0.101	0.050	0.022	0.008	0.003
$m = 4$[b]	0.018	0.073	0.147	0.195	0.195	0.156	0.104	0.060	0.030	0.013
$m = 6$[c]	0.002	0.015	0.045	0.089	0.134	0.161	0.161	0.138	0.103	0.069
$m = 10$[d]	5E–5	5E–4	0.002	0.008	0.019	0.038	0.063	0.090	0.113	0.125

[a] The notation 7E-5 indicates 7×10^{-5}.
[b] The values of P_x for $m = 4$ at $x = 10$ and 11 are 0.005 and 0.002 respectively.
[c] The values of P_x for $m = 6$ at $x = 10$–14 are 0.041, 0.023, 0.011, 0.005, 0.002.
[d] The values of P_x for $m = 10$ at $x = 10$–18 are: 0.125, 0.114, 0.095, 0.073, 0.052, 0.035, 0.022, 0.013, 0.007.

Normal distribution, Gaussian:

Continuous and symmetrical distribution which gives the *differential probability dP* of finding the value x within the differential interval *dx*:

$$dP_x = \frac{1}{\sigma_w \sqrt{2\pi}} \exp\left[\frac{-(x-m)^2}{2\sigma_w^2}\right] dx$$

Normal distribution, Gaussian:

Continuous and symmetrical distribution which gives the *differential probability dP* of finding the value x within the differential interval *dx*:

$$dP_x = \frac{1}{\sigma_w \sqrt{2\pi}} \exp\left[\frac{-(x-m)^2}{2\sigma_w^2} \right] dx$$

m – the mean, which is the true value of the quantity being measured,

σ_w - width, the standard deviation of the distribution.

Two parameters instead of one in Poisson distribution.

Normal distribution, Gaussian:

Bell curve of probability, symmetric around *m*, can extend to negative values of *x*.

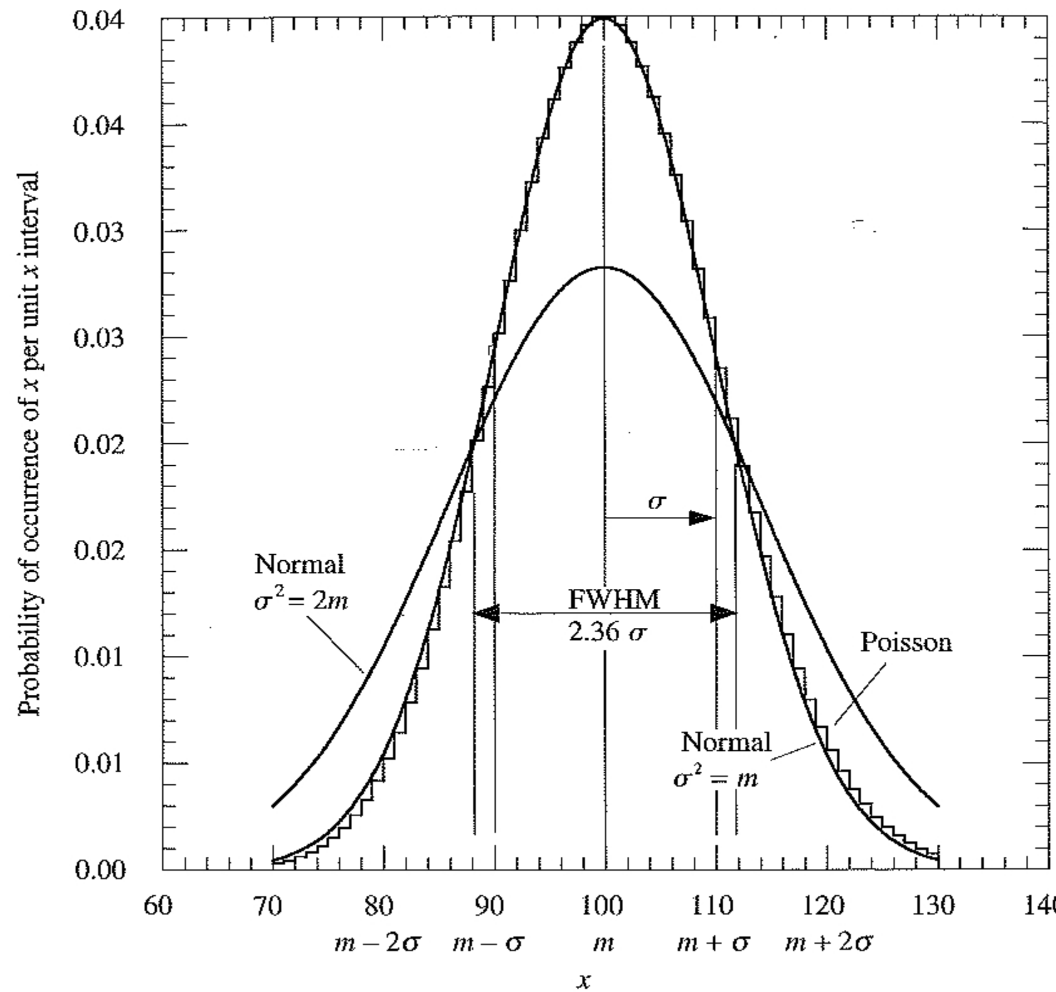


Figure 6.8. The Poisson (step curve) and normal distributions (smooth curves) for the mean value $m = 100$. The normal distribution is given for two values of the width parameter σ_w which is shown in the text to be equal to the standard deviation σ. The Poisson distribution approximates well the normal distribution if the latter has $\sigma = \sqrt{m}$. Note the slight asymmetry of the Poisson distribution relative to the normal distribution. The standard deviation and full width half maximum widths are shown for the higher normal peak; the two normal curves happen to cross at the FWHM point.

Normal distribution, Gaussian:

Bell curve of probability, symmetric around *m*, can extend to negative values of *x*.

$$(2\pi)^{-1/2}$$

is chosen so that this distribution is also normalized:

$$\int_{x=-\infty}^{x=\infty} dP_x = 1$$

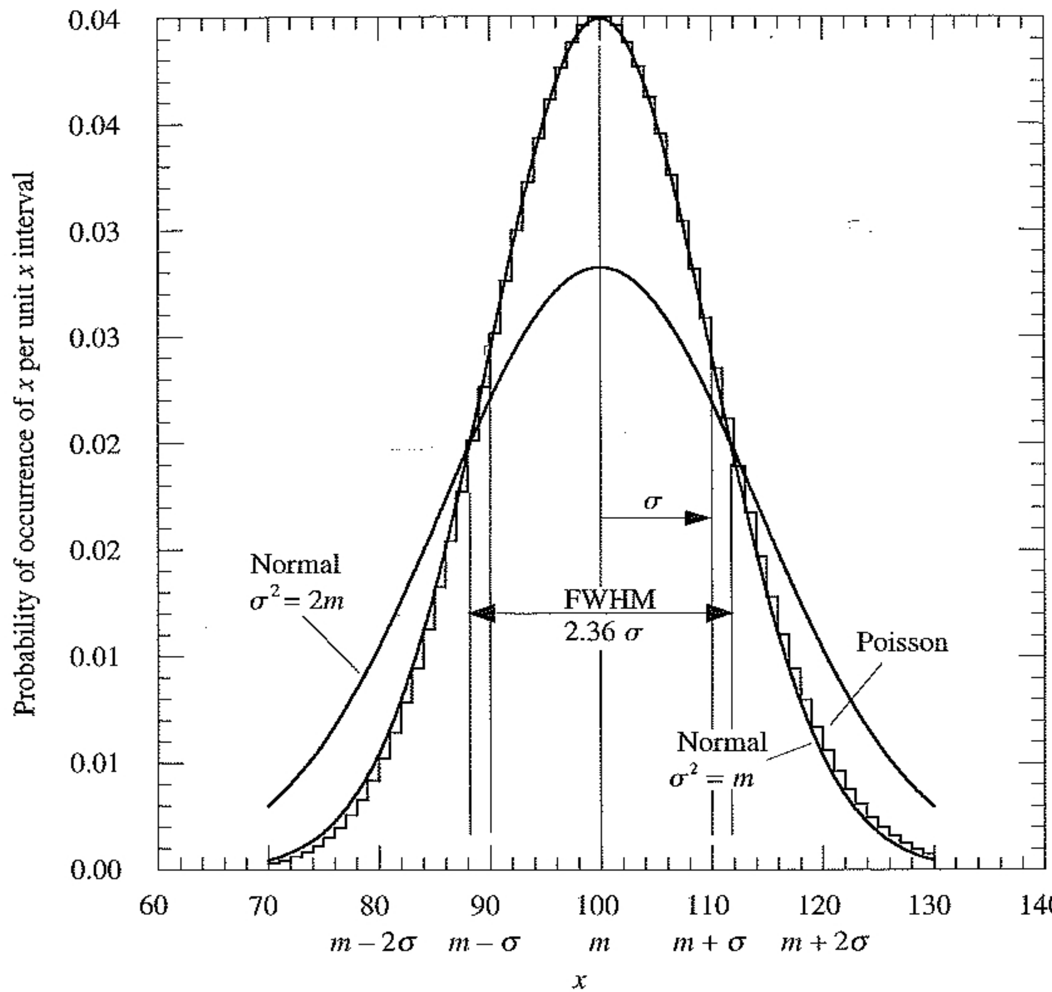


Figure 6.8. The Poisson (step curve) and normal distributions (smooth curves) for the mean value $m = 100$. The normal distribution is given for two values of the width parameter σ_w which is shown in the text to be equal to the standard deviation σ. The Poisson distribution approximates well the normal distribution if the latter has $\sigma = \sqrt{m}$. Note the slight asymmetry of the Poisson distribution relative to the normal distribution. The standard deviation and full width half maximum widths are shown for the higher normal peak; the two normal curves happen to cross at the FWHM point.

Normal distribution, Gaussian:

σ_w - a characteristic width:

$x = m \pm \sigma_w$

the function has fallen to:

$e^{-0.5} = 0.601$

of its maximum value.

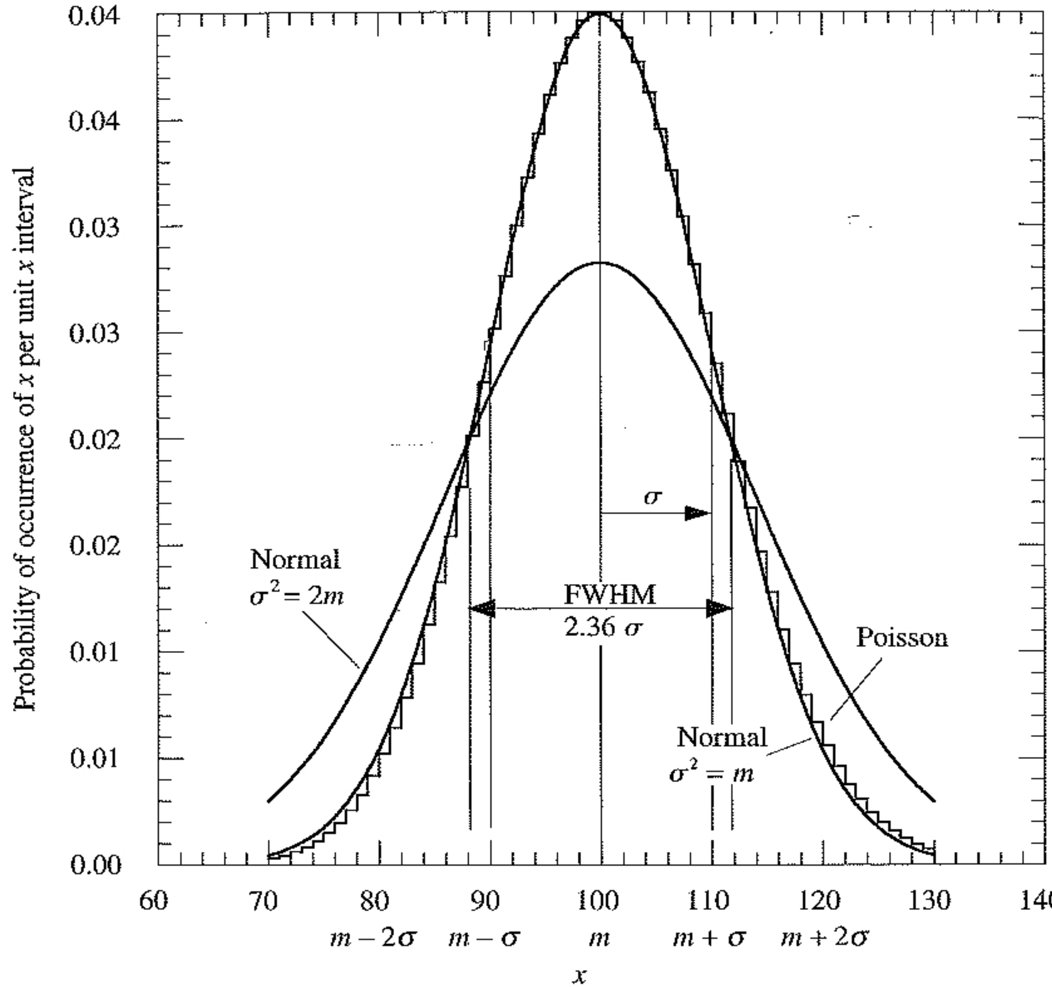


Figure 6.8. The Poisson (step curve) and normal distributions (smooth curves) for the mean value $m = 100$. The normal distribution is given for two values of the width parameter σ_w which is shown in the text to be equal to the standard deviation σ. The Poisson distribution approximates well the normal distribution if the latter has $\sigma = \sqrt{m}$. Note the slight asymmetry of the Poisson distribution relative to the normal distribution. The standard deviation and full width half maximum widths are shown for the higher normal peak; the two normal curves happen to cross at the FWHM point.

Normal distribution, Gaussian:

σ_w - a characteristic width:

$$x = m \pm \sigma_w$$

the function has fallen to:

$$e^{-0.5} = 0.601$$

of its maximum value.

For:

$$x - m = \sqrt{2}\,\sigma_w$$

$$e^{-1} = 0.37$$

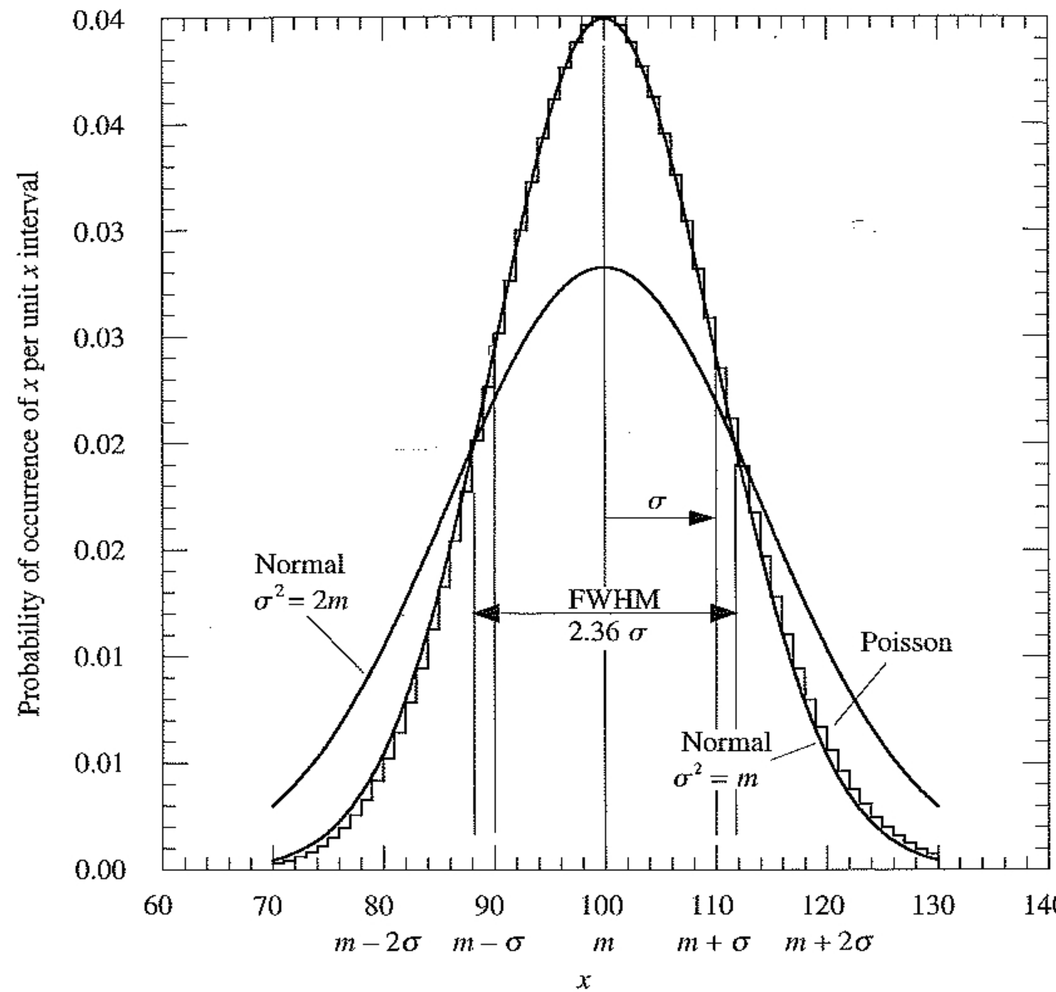


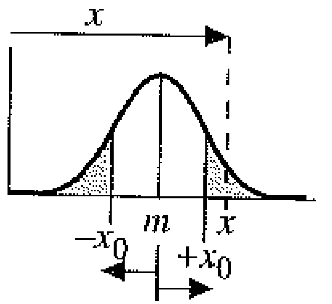
Figure 6.8. The Poisson (step curve) and normal distributions (smooth curves) for the mean value $m = 100$. The normal distribution is given for two values of the width parameter σ_w which is shown in the text to be equal to the standard deviation σ. The Poisson distribution approximates well the normal distribution if the latter has $\sigma = \sqrt{m}$. Note the slight asymmetry of the Poisson distribution relative to the normal distribution. The standard deviation and full width half maximum widths are shown for the higher normal peak; the two normal curves happen to cross at the FWHM point.

Normal distribution, Gaussian:

- 68% of the area falls in 1 σ_w
- 95.5 % falls in 2 σ_w
- 99.73 % falls in 3 σ_w

from integrating eq. of distribution.

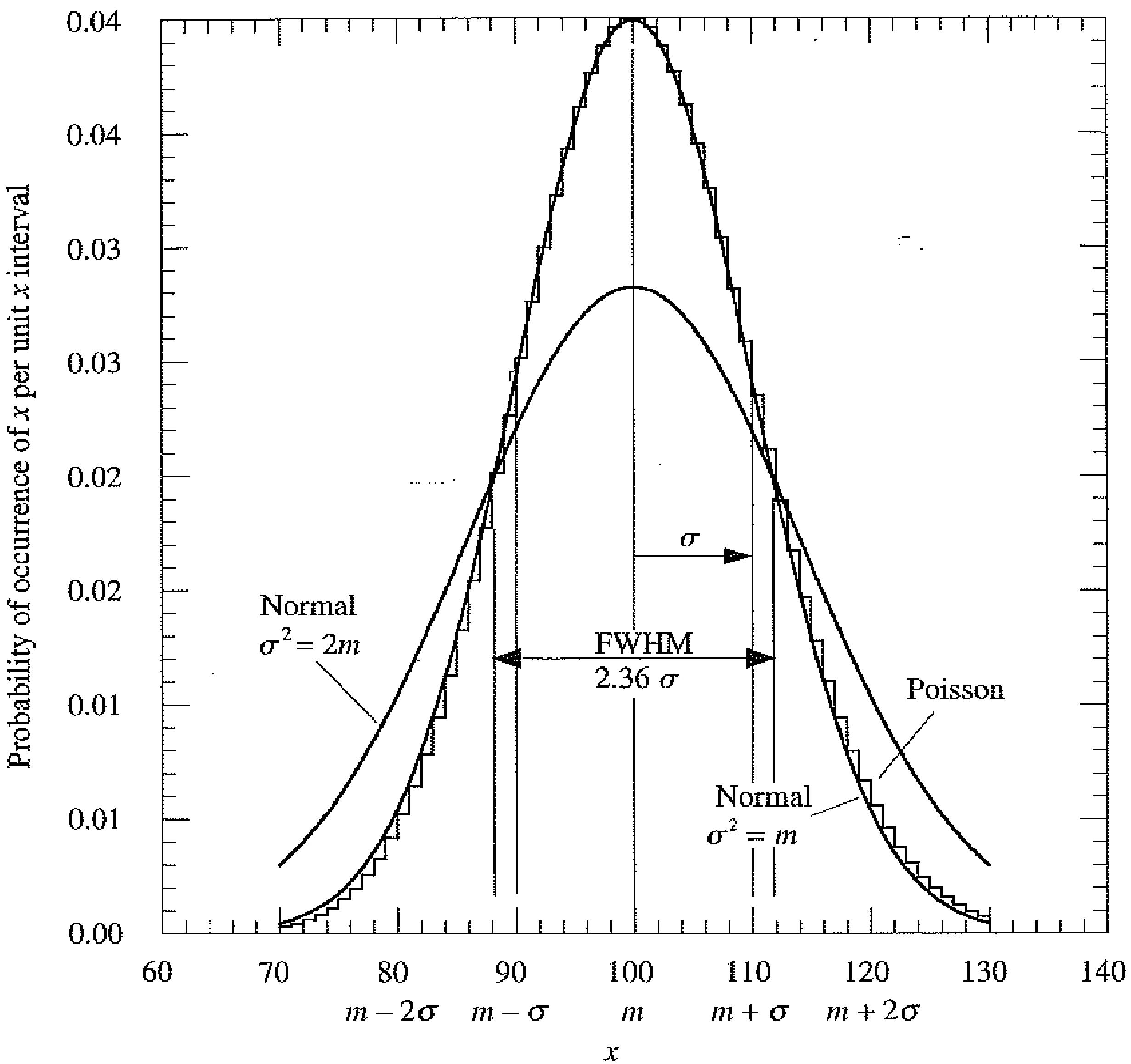
Table 6.2. *Normal distribution probabilities*



$\left(\frac{x_0}{\sigma}\right)^a$	Area (shaded) at $\lvert x - m \rvert > x_0$ [b]	$\left(\frac{x_0}{\sigma}\right)^a$	Area (shaded) at $\lvert x - m \rvert > x_0$ [b]
0	1.00	2.5	0.0124
0.5	0.617	3.0	0.00270
1.0	0.317	3.5	4.65×10^{-4}
1.2	0.230	4.0	6.34×10^{-5}
1.4	0.162	5.0	5.73×10^{-7}
1.6	0.110	6.0	2.0×10^{-9}
1.8	0.0719	7.0	2.6×10^{-12}
2.0	0.0455		

[a] Ratio of deviation x_0 to standard deviation σ. The standard deviation σ is equal to σ_w, the width parameter of the distribution.
[b] Probability of occurrence of deviation greater than $\pm x_0$.

Normal and Poisson distribution:



For large values of *m*, the Poisson distribution approaches in shape the central part of the normal distribution if the width parameter is set to:

$$\sigma_w = m^{1/2}$$

Normal distribution describes the arrival of random events for large m.

Variance and standard deviation:

The width of a *measured* distribution indicates the range of values obtained from a set of individual measurements of *x*. Formally *root-mean-square deviation*, i.e. *standard deviation,* σ, its square is called the *variance*:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{i=n} (x_i - m)^2$$

definition of variance.

n – number of independent measurements,

x_i – individual measurements,

m – mean value can be only obtained with an infinite amount of data !!!

Variance and standard deviation:

In practice the average value x_{av} of the *n* measured numbers may be the best approximation of *m* that is available:

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{i=n}(x_i - x_{av})^2$$

practical variance.

Practical variance equal theoretical for large *n*.

Variance and standard deviation:

In practice the average value x_{av} of the *n* measured numbers may be the best approximation of *m* that is available:

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{i=n}\left(x_i - x_{av}\right)^2$$

practical variance.

Practical variance equal theoretical for large *n*.

Variance can be evaluated for any given experimental distribution as Poisson or Normal.

Variance and standard deviation:

Variance of theoretical Poisson distribution:
n_j – occurrences of the same value x_j,
It is useful to rewrite variance in terms of the probability used in theoretical expression: Summation will be over *x*, than the trial number *i*.

$$\sigma^2 = \sum_{j=1}^{j=n} (x_j - m)^2 \frac{n_j}{n} = \sum_{x=-\infty}^{x=\infty} (x-m)^2 P_x , \qquad \left(x_{av} = \sum_{j=1}^{j=n} \frac{x_j}{n} \right)$$

Variance and standard deviation:

Variance of theoretical Poisson distribution:
n_j – occurrences of the same value $x_{j,}$
It is useful to rewrite variance in terms of the probability used in theoretical expression. Summation will be over *x*, than the trial number *i*.

$$\sigma^2 = \sum_{j=1}^{j=n} (x_j - m)^2 \frac{n_j}{n} = \sum_{x=-\infty}^{x=\infty} (x-m)^2 P_x, \qquad \left(x_{av} = \sum_{j=1}^{j=n} \frac{x_j}{n} \right)$$

Substituting the Poisson distribution:

$$\sigma^2 = \sum_{x=-\infty}^{x=\infty} \frac{(x-m)^2 m^x e^{-m}}{x!} = m,$$

$$\sigma_w = m^{1/2}$$

Example:

If 100 photons are expected to arrive at pixel of a CCD during exposure of 1s, standard deviation for a single measurement is:

$$\sigma = \sqrt{100} = 10$$

We can expect fluctuations ± 10 or even ± 30 about the 100 count mean.

Example:

If 100 photons are expected to arrive at pixel of a CCD during exposure of 1s, standard deviation for a single measurement is:

$$\sigma = \sqrt{100} = 10$$

We can expect fluctuations ± 10 or even ± 30 about the 100 count mean.
The uncertainty relative to the mean value is:

$$\sigma / m = 10/100 = 10\,\%$$

It is “a 10% measurement”. If in 100 s, one expect 10000 counts:

$$\sigma = \sqrt{10000} = 100, \quad \sigma / m = 100/10000 = 1\,\%$$

More counts leads to higher absolute fluctuations and uncertainty, but fractional uncertainty is reduced. Longer means better rates.

Variance and standard deviation:

Variance of Normal distribution is obtained through substitution of sum into integral form:

$$\sigma^2 = \int_{x=-\infty}^{x=\infty} (x-m)^2 \, dP_x$$

$$\sigma^2 = \frac{1}{\sigma_w \sqrt{2\pi}} \int_{x=-\infty}^{x=\infty} (x-m)^2 \exp\left[\frac{-(x-m)^2}{2\sigma_w^2}\right]$$

$$\sigma^2 = \sigma_w^2$$

Measurement significance:

For large number of events $\sigma = m^{1/2}$

The probability of exceeding 3 sigma is 0.27%.
If the pixel of CCD us expected to record 100 photons during the exposure time thus $\sigma = 10.$

If we measure 130 photons, we may ask, is it bright source or fluctuation? This is 3 sigma detection, but still there is one chance in 1/(0.0027)=370 that this would happen from statistical fluctuations.

One measures of 5 sigma error in one of measurements. The probability of a statistical fluctuations in one given trial is 6 x 10-7, but CCD has 4 million pixels, thus:

$$expectation\, value = 6 \times 10^{-7} \times 4 \times 10^{6} = 2.4$$

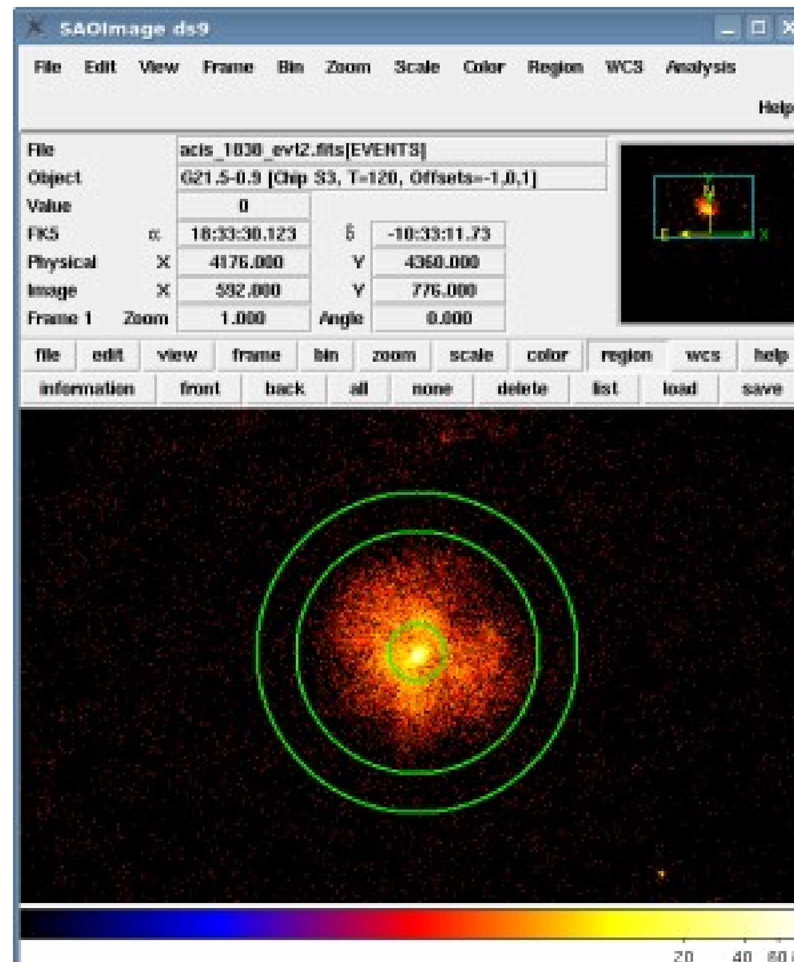
Background:

- Counts due to cosmic rays particles – anticoincidence logic.
- Counts due to diffuse X-ray background.

Commonly, two measurements will be made:

1) one with astrophysical source in the field of view,
2) one with offset from the source to measure bkgr only.

If the detector/telescope produces a sky image, we can make both measurements in a single exposure.



Propagation of errors:

After data are taken, one invariably manipulates then to obtain other quantities:

$$\frac{\text{accumulated number of counts}}{\text{accumulated time}} = \text{rate of photon arrival}$$

Propagation of errors:

After data are taken, one invariably manipulates then to obtain other quantities:

$$\frac{\text{accumulated number of counts}}{\text{accumulated time}} = \text{rate of photon arrival}$$

Assume, x and y be a length, each accurate to 1mm:

$$z = x + y, \qquad z = x - y$$

$$dz = dx + dy$$

$$|dz|_{max} = |dx|_{max} + |dy|_{max}$$

Maximum error is thus a sum of the individual maximum errors.

Propagation of errors:

Assume, x and y be a length, each accurate to 1mm:

$$z = x * y, \qquad z = x / y$$

Fractional error is the sum of the individual fractional errors:

or:

$$dz = x * dy + y * dx$$

$$|dz/z|_{max} = |dx/x|_{max} + |dy/y|_{max}$$

We assume to maximize the error.. In fact the measurements of *x* and *y* would most likely be uncorrelated.

Fractional errors are thus, on average, less than the maximum values found above.

Propagation of errors:

If x and y vary independently with normal distribution, characterized by standard deviation,

error of summation or subtraction:

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2$$

$$x = y \quad \Rightarrow \quad \sigma_z = \sqrt{2}\,\sigma_x$$

$$x > y \quad \Rightarrow \quad \sigma_z \approx \sigma_x$$

error in a product or quotient:

$$\frac{\sigma_z}{z}^2 = \frac{\sigma_x}{x}^2 + \frac{\sigma_y}{y}^2$$

Background subtraction:

S – expected number of counts detected in Δt time interval,

B – expected number of counts of bkgr in the same time interval.

ON Source – $S+B$,
OFF Source – B.

Background subtraction:

S – expected number of counts detected in *Δt* time interval,

B – expected number of counts of bkgr in the same time interval.

ON Source – *S*+*B*,
OFF Source – *B*.

Signal counts; equal exposures:

$$S = (S + B) - B$$

Two measurements are quite independent: different photons and different bkgr are involved. Thus the fluctuation will be uncorrelated.

$$\sigma_s^2 = \sigma_{s+b}^2 + \sigma_b^2$$

Background subtraction:

Two standard deviations obtained from the Poisson distribution:

$$\sigma_s^2 = S + B + B = S + 2\mathrm{B}$$

$$\sigma_s \ll S \quad \Rightarrow \quad high\ quality\ of\ measurement\ ,$$

$$S = 3\sigma_s \quad \Rightarrow \quad 3\sigma \quad result\ ,$$

$$S < 3\sigma_s \quad \Rightarrow \quad detection\ questionable.$$

<u>Background subtraction:</u>

Significance equals number of standard deviations or S/N:

$$\frac{S}{\sigma_s} = \frac{S}{\sqrt{S + 2\mathrm{B}}} \qquad \text{signal-to-noise ratio.}$$

Background subtraction:

Significance equals number of standard deviations or S/N:

$$\frac{S}{\sigma_s} = \frac{S}{\sqrt{S+2B}} \qquad \text{signal-to-noise ratio}.$$

The intensity of the source is best represented by the source event rate r_s (counts/s). With equal on-source and off-source accumulated time Δt :

$$r_s = \frac{S}{\Delta t} \qquad r_b = \frac{B}{\Delta t}$$

Low and hight background limits:

The low-background (B<<S) case gives:

$$\frac{S}{\sigma_s} \approx \frac{S}{\sqrt{S}} = \sqrt{S} = \sqrt{r_s \Delta t}$$

bkgr negligible

Significance increases as the square root of the number of counts.

To increase significance to 5 sigma – to increase duration time by a factor of $(5/2)^2=6.25$.

Low and hight background limits:

The low-background (B<<S) case gives:

$$\frac{S}{\sigma_s} \approx \frac{S}{\sqrt{S}} = \sqrt{S} = \sqrt{r_s \Delta t}$$

bkgr negligible

Significance increases as the square root of the number of counts.
To increase significance to 5 sigma – to increase duration time by a factor of $(5/2)^2=6.25$.

The high-background (B>> S) case gives:

$$\frac{S}{\sigma_s} \approx \frac{S}{\sqrt{2B}} = \frac{r_s \Delta t}{\sqrt{2 r_b \Delta t}} = \frac{r_s}{\sqrt{2 r_b}} \sqrt{\Delta t}$$

bkgr dominates.

It takes a lot of observing time to increase significance.

Low and hight background limits:

Let us compare the S/N ratios of two hypothetical detectors, one *high-B* and the other of *low-B*.

Comparison of two sensitivities:

$$\left(\frac{S}{\sigma_s}\right)_{B \gg S} = \sqrt{\frac{r_s}{2\mathrm{r}_b}}\left(\frac{S}{\sigma_s}\right)_{B \ll S}$$

Since $r_s << r_b$, the expression tells us that the significance is much less in the high-B case than for the low-B case for similar exposures.

<u>Bright and faint source observations:</u>

Focusing instruments are low-B systems.
3 X-rays photons in one resolution element of the focal plane could be highly significant since bkgr so low.

If the expected bkgr in the element is only 0.1 counts, the probability of this bkgr giving rise to the 3 x-rays is:

$$P_x = \frac{m^x e^{-m}}{x!} = \frac{0.1^3 e^{-0.1}}{3!} = 1.5 \times 10^{-4}$$

Focusing instruments – the best for faint sources.

Bright and faint source observations:

How does the significance of a detection in a given time depends on source intensity i.e., on the rate r_s.

When S>B as in focusing instruments, the statistical noise arises from the source itself.

$$S \ \ increases \ \ \Rightarrow \ \ Statistical \ \ noise \ \ increases$$

$$\frac{S}{\sigma_s} \ \ increases \, slowly \, with \ \ \sqrt{r_s}$$

($S/\sigma_s \approx \sqrt{r_s \Delta t}$ **for low-B)**

Bright and faint source observations:

How does the significance of a detection in a given time depends on source intensity i.e., on the rate r_s.

When S>B as in focusing instruments, the statistical noise arises from the source itself:

$$S \quad increases \quad \Rightarrow \quad Statistical \quad noise \quad increases$$

$$\frac{S}{\sigma_s} \quad increases\, slowly\, with \quad \sqrt{r_s}$$

($S/\sigma_s \approx \sqrt{r_s \Delta t}$ for low-B)

When S<B: $$\frac{S}{\sigma_s} \quad increases\, with \quad r_s$$

($S/\sigma_s \approx r_s / \sqrt{2 r_b} \cdot \sqrt{\Delta t}$ for high-B)

source with twice intensity will be measured with twice the significance, statistical noise depends only on the bkgr rate.

Bright and faint source observations:

$$\left(\frac{S}{\sigma_s}\right)_{B\gg S} = \sqrt{\frac{r_s}{2r_b}}\left(\frac{S}{\sigma_s}\right)_{B\ll S}$$

They can differ by:
-effective area
-different energies control.

For example non-focusing high-B X-ray detector, PC & MC has large collecting area A, is sensitive up to 60 keV.

Focusing low-B reaches only about 8 keV. A_eff << A_coll.

For bright sources, a high-B large area system can yield a higher significance (S/N) in a given time that can a low-B system.

The Rossi X-ray Timing Explorer (RXTE)
low E and ang. resolution, great timing accuracy.

<u>Bright and faint source observations:</u>

$$\left(\frac{S}{\sigma_s}\right)_{B \gg S} = \sqrt{\frac{r_s}{2r_b}} \left(\frac{S}{\sigma_s}\right)_{B \ll S}$$

They can differ by:
-effective area
-different energies control.

$$\frac{r_s}{r_b} \quad increases\ but\ still\ well \ll 1$$

The sensitivity of high-B detector moves toward the sensitivity of low-B detector.

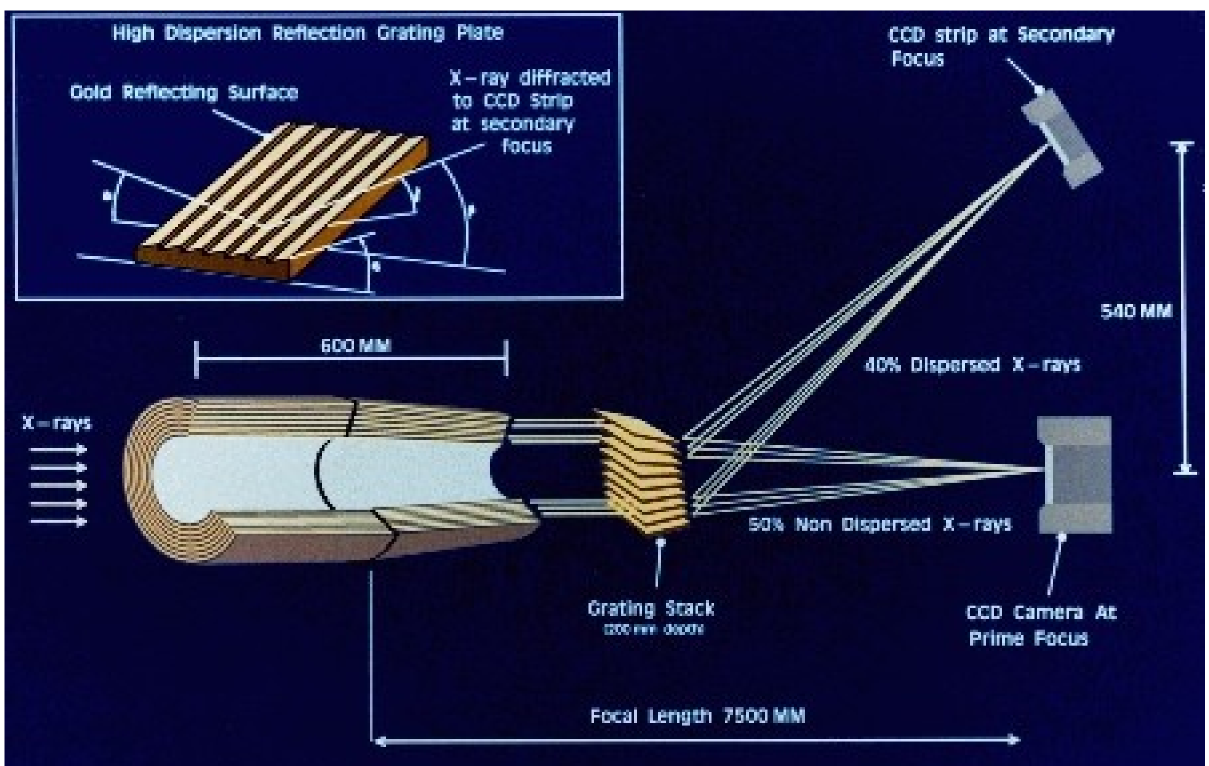
The adventure of the low-B detector decreases as the source brightens.

When source becomes so bright in the high-B detector that it exceeds its high bkgr, the weak-bkgr limit applies to both detectors.

Homework: find A_{eff} (E), A_{eff} (off-axis angle), PSF
for following gratings + mirrors + detector:

EINSTAIN, EXOSAT, ASCA, CHANDRA (LEG, HEG), XMM RGS, SUZAKU.

Next lecture on Nov. 29th 2011.



EINSTEIN

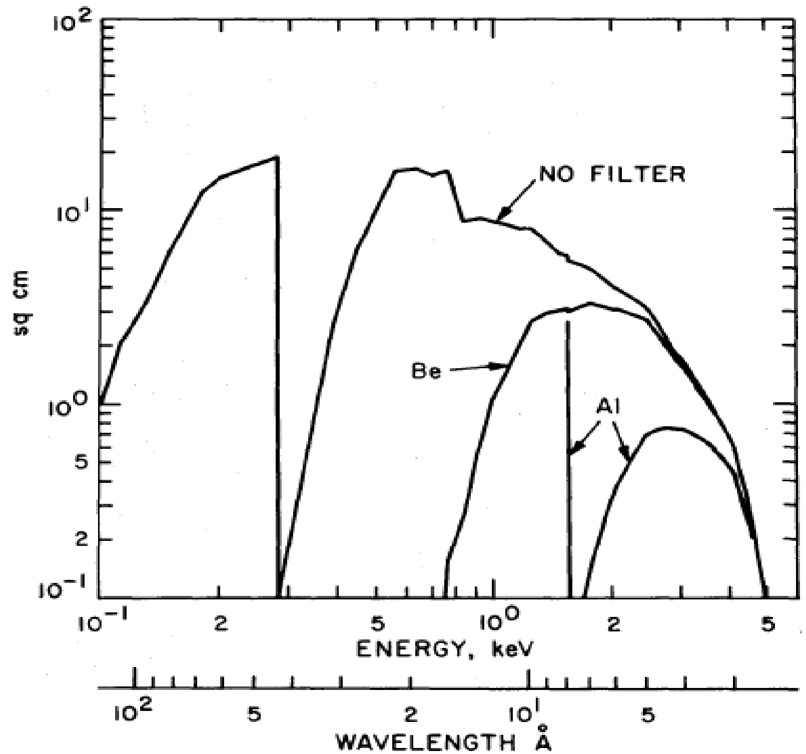
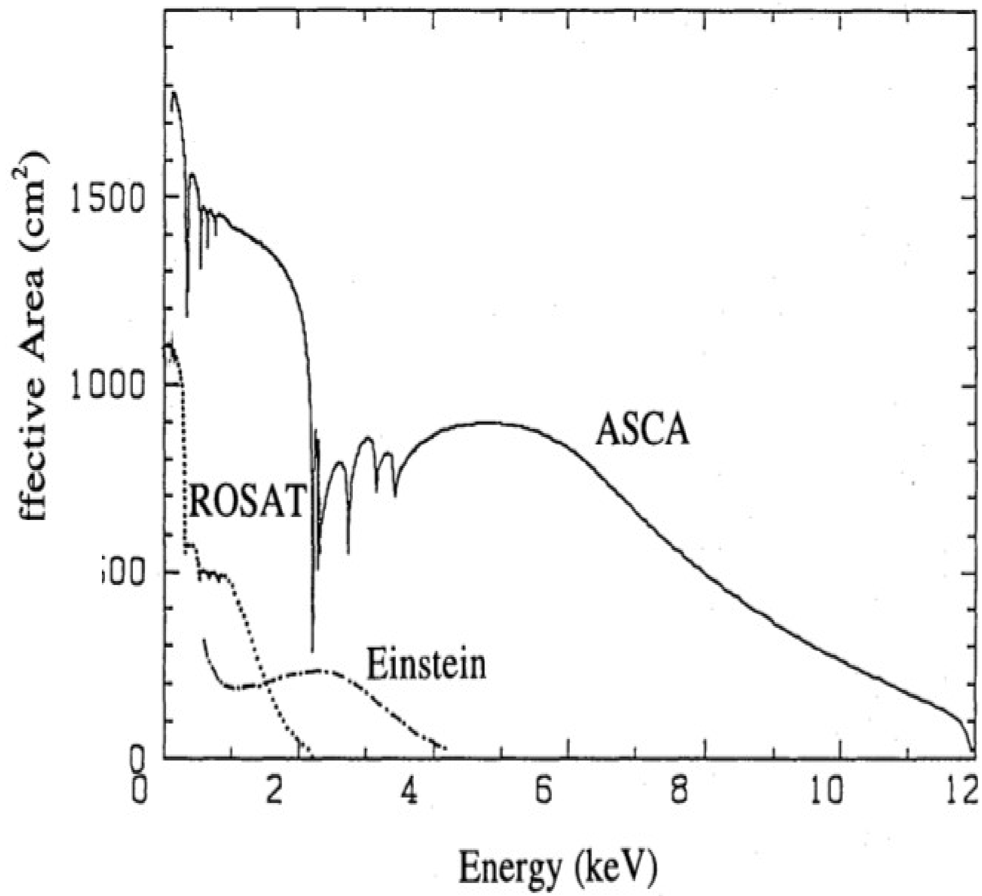


FIG. 5.—Effective area HRI No. 2 with no filters, with Be filter, and with Al filter. Photons which scatter from the mirror with angles greater than 6″ have not been included; thus this efficiency is for 12″ pixels.



EINSTEIN

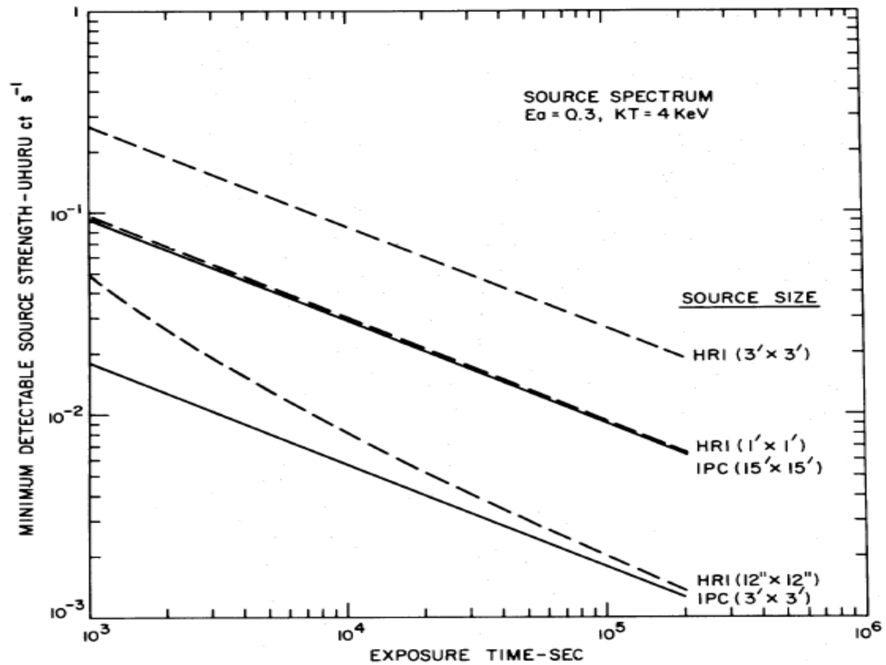


FIG. 2.—Sensitivity of *Einstein* imaging instruments

Effective Area of one XRT + SIS (within 12mm diameter)

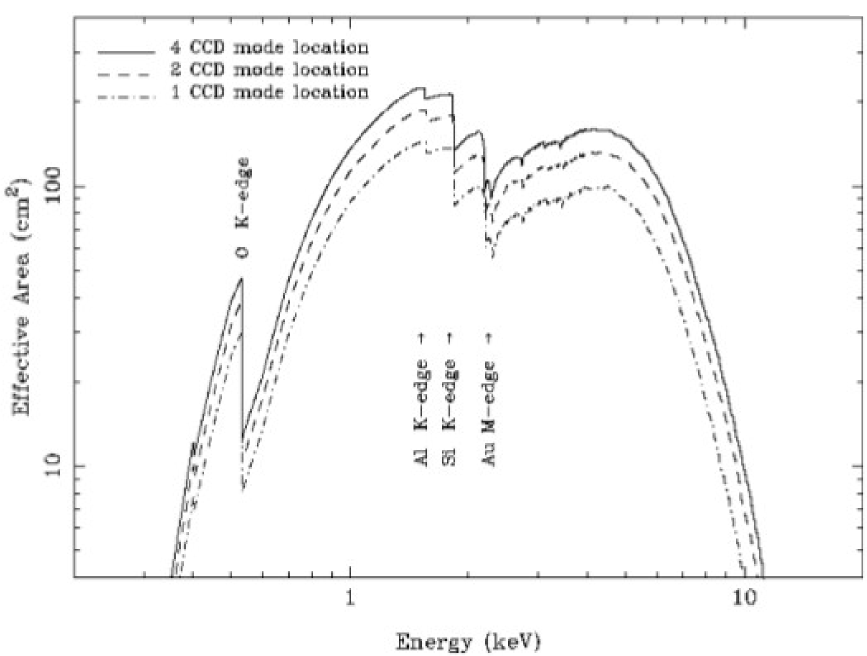


Figure 3.10: Dependence on energy of the effective area of a single SIS combined with an XRT. Dependence on the source locations is also shown.

ASCA

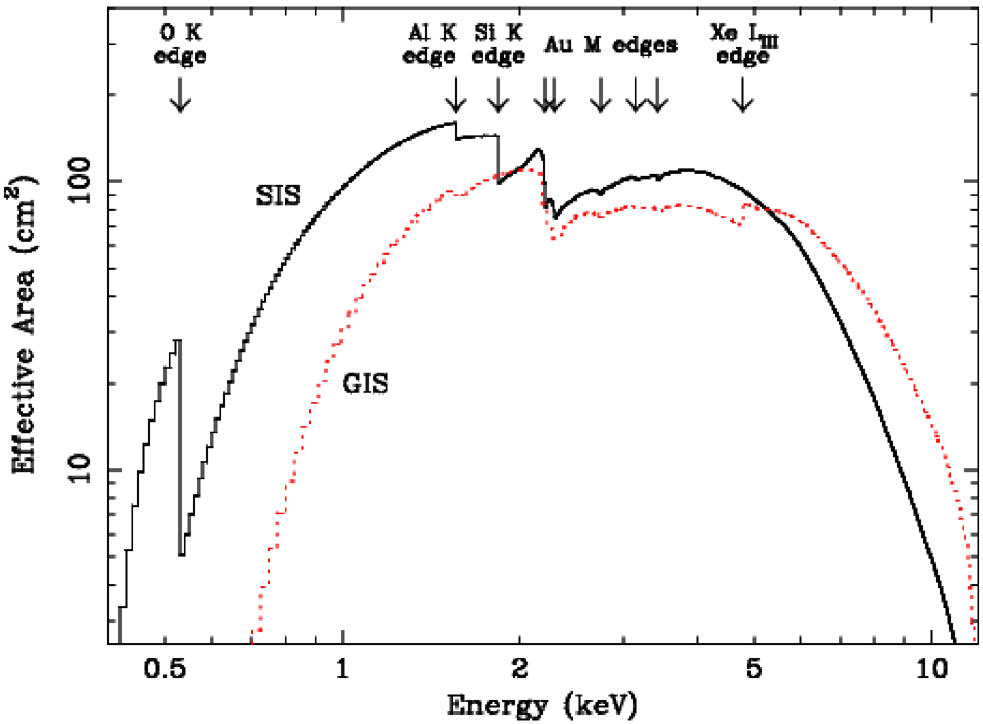


Figure 2.3 Effective area curves for the SIS (solid line) and GIS (dotted line) when the pairs of like detectors are averaged together. In practice, the effective area is a function of position in the field of view, and also time dependent in the case of the SIS, but these curves (for NGC 5548, *ASCA* sequence number 76029010) are typical. The sharp features correspond to absorption arising in either the mirrors or the detectors; the associated atomic shells are labeled.

ASCA

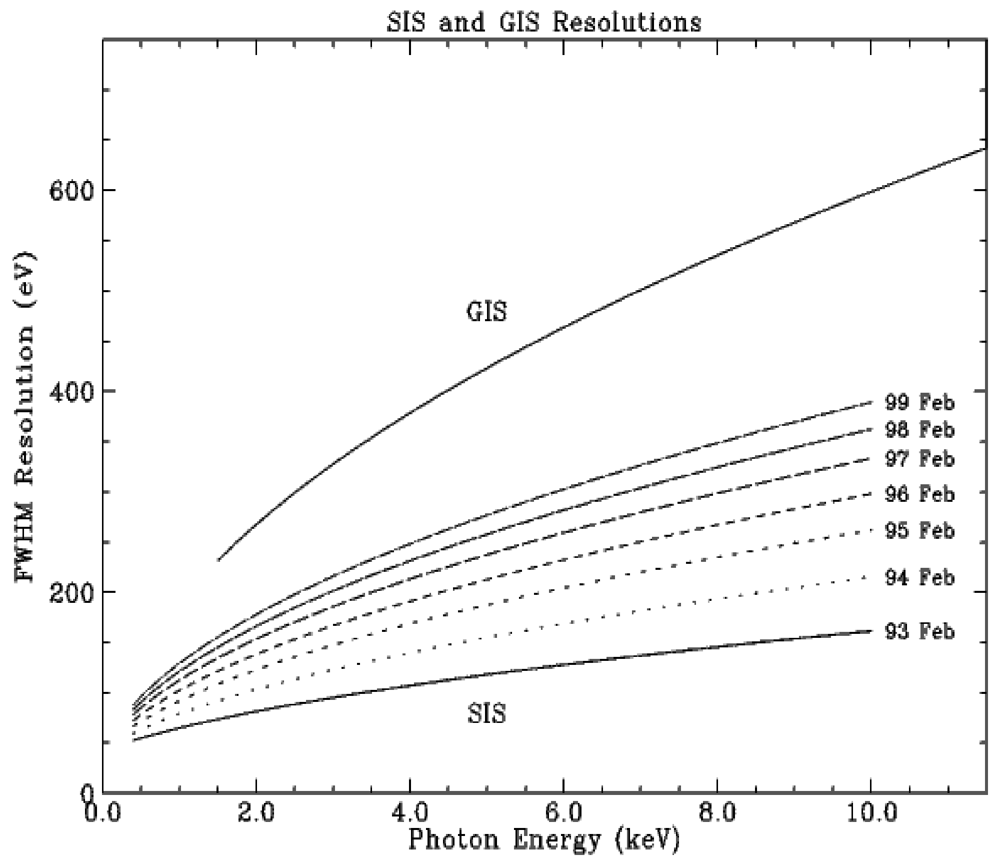


Figure 2.6 The progression of the SIS energy resolution over the lifetime of the satellite (*ASCA* AO7 Specs, 1998, figure 8.4b, courteously provided by the *ASCA* Guest Observer Facility, NASA/GSFC).

ASCA

:

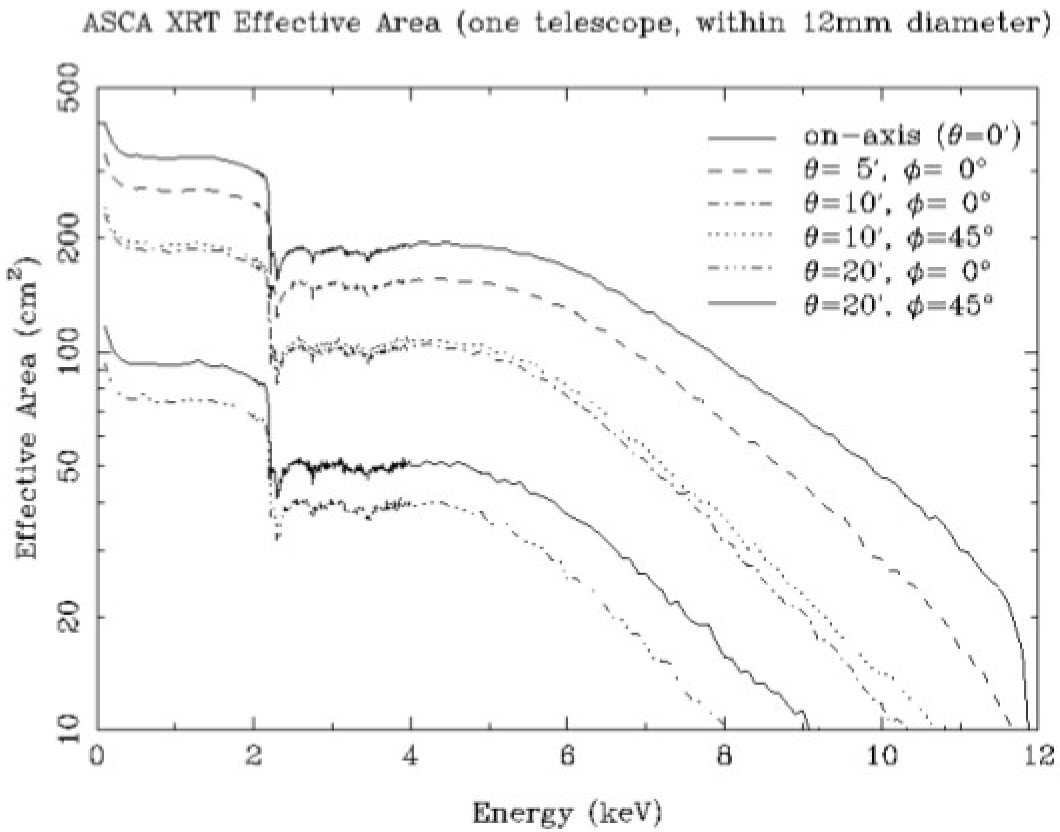


Figure 3.3: Effective area of a single XRT as a function of the energy of the incident X-ray. Dependence on the off-axis angle θ and azimuthal angle ϕ is also shown.

ASCA

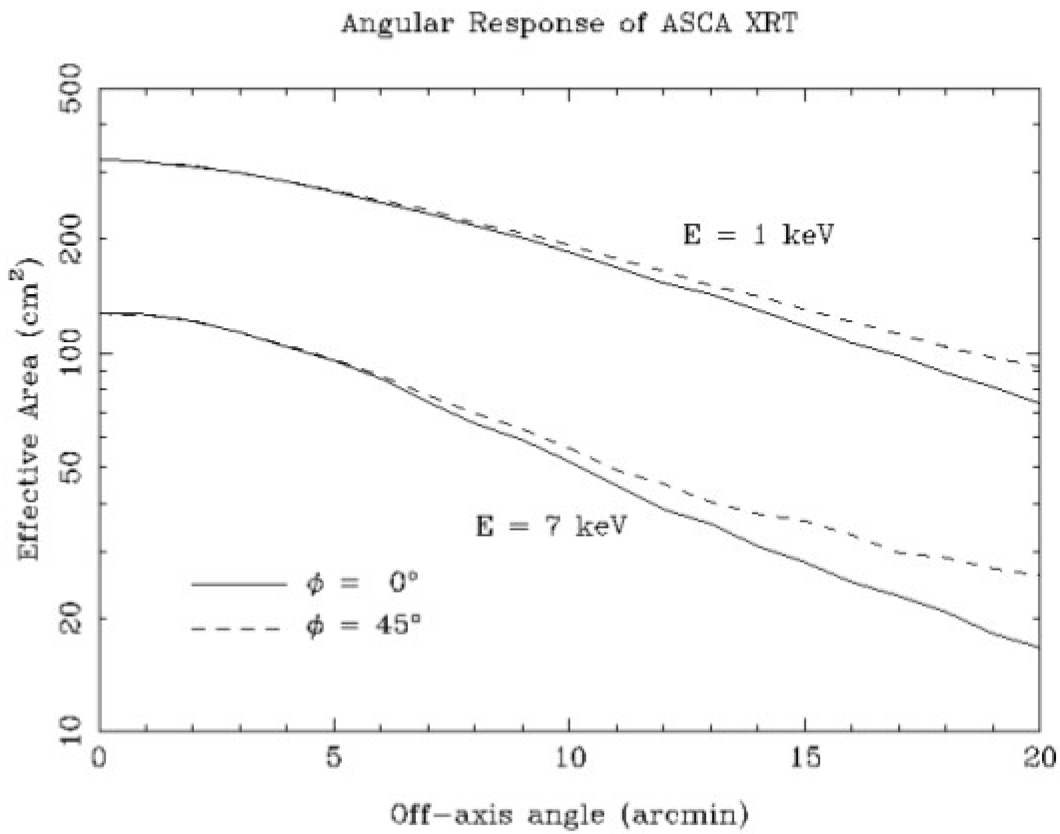


Figure 3.4: Effective area of a single XRT as a function of the off-axis angle. The data for incident X-rays of 1.0 and 7.0 keV are shown.

ASCA

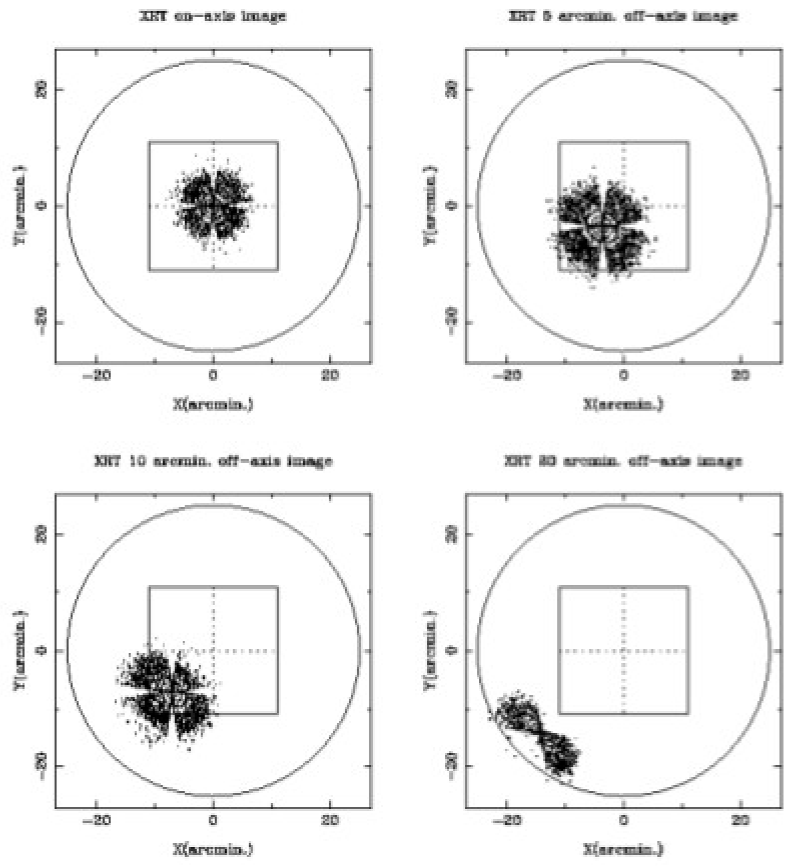


Figure 3.5: Point spread function of the ASCA XRT.

ASCA

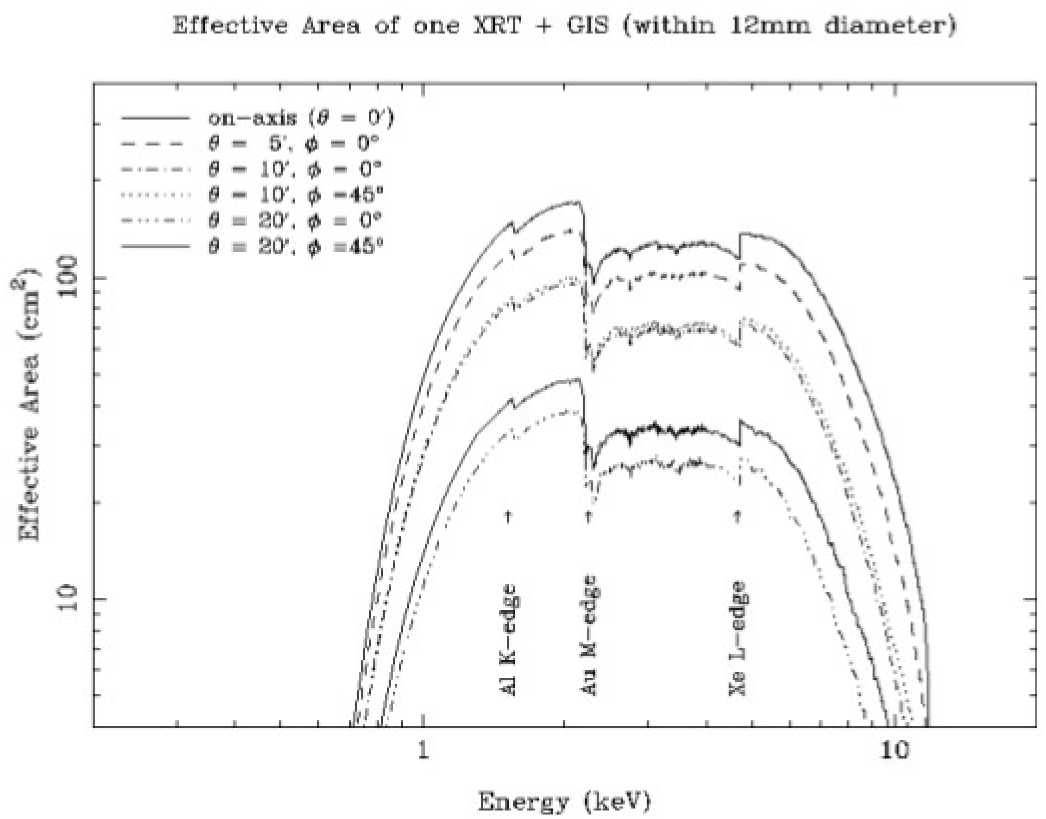


Figure 3.8: Dependence of the effective area of a single GIS combined with and XRT. Dependence on the off-axis angle θ and azimuthal angle ϕ is also shown.

ASCA

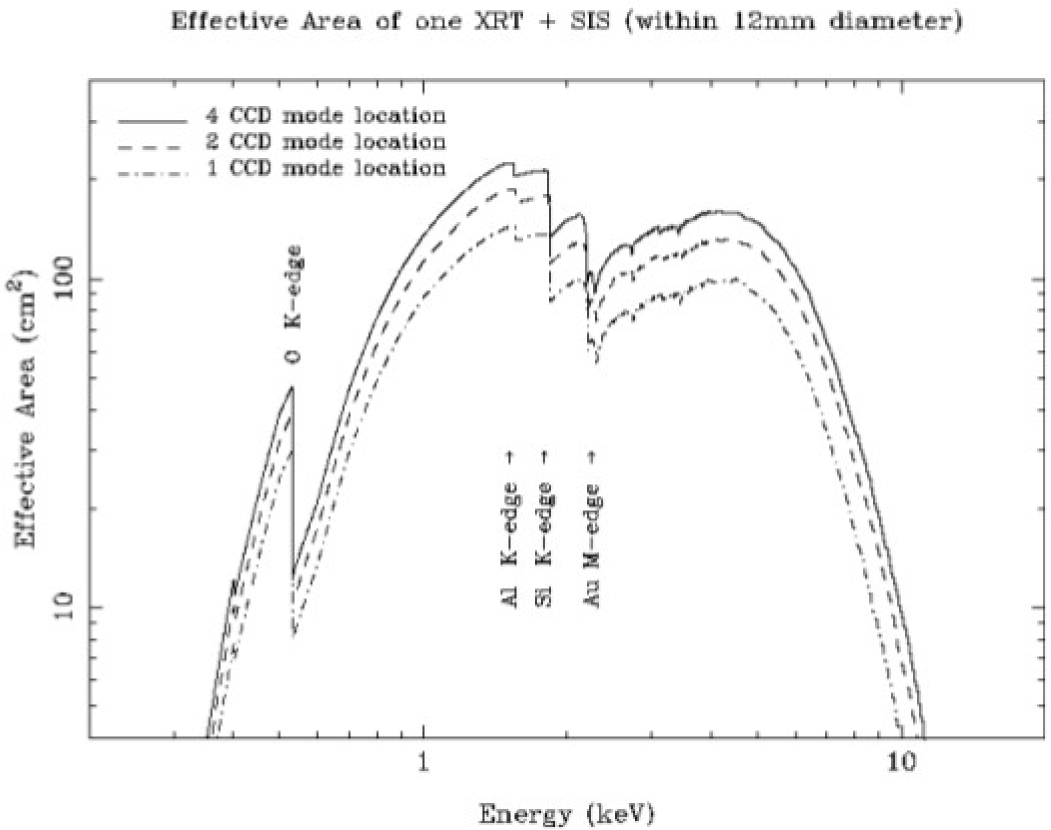


Figure 3.10: Dependence on energy of the effective area of a single SIS combined with an XRT. Dependence on the source locations is also shown.

Principles of ranking the lecture:

- to be here

- to participate into discussions

- to make a homework

- hand – on sessions with the use of the computer.....

- exam – very simple :::)))

wi-fi password: a w sercu maj