Accreting Binary Pulsars

A test bed for matter under extreme conditions

Mauro Orlandini
INAF/IASF Bologna

CSV Rev: 0.2 (2015/09/09)

Portrait of an Accreting Binary Pulsars

- Discovery of pulsed emission at 4.8 sec in Cen X–3

THE ASTROPHYSICAL JOURNAL, 167:L67–L73, 1971 July 15

DISCOVERY OF PERIODIC X-RAY PULSATIONS IN CENTAURUS X-3 FROM *UHURU*

R. GIACCONI, H. GURSKY, E. KELLOGG, E. SCHREIER, AND H. TANANBAUM
American Science & Engineering, Inc., Cambridge, Massachusetts 02142
Received 1971 May 17

ABSTRACT

A search for X-ray sources exhibiting pulsating characteristics similar to the ones recently discovered in Cyg X-1 by Oda *et al.*, and confirmed by Holt *et al.*, has revealed the existence of periodic pulsations in the X-ray emission from Cen X-3.

The pulsations are of large amplitude, comprising at least 70 percent of the total flux from the source. The period is approximately 5 s. Due to the amplitude of the effect and the long period, we are able to establish that the period and phase of the pulsations are maintained for several hours. In addition, two very remarkable events are observed to occur during a 1-day interval. The first is a change in the intensity of the source by a factor of 10 occurring in approximately 1 hour. The second is a decrease and then an increase in the period of the pulsations, corresponding to about 0.02 and 0.04 percent, each change of period occurring in a time of about 1 hour. The magnitude of the changes, and the short time interval in which they take place, are quite different from any observed in radio pulsars. In addition, we see evidence for changes of the period by about 1 percent occurring over a time span of 3 months. Present data are inconclusive about the presence of faster pulsations from the source. The X-ray spectrum shows evidence of substantial absorption below 3.8 keV.

Portrait of an Accreting Binary Pulsars

- Discovery of pulsed emission at 4.8 sec in Cen X–3

THE ASTROPHYSICAL JOURNAL, 167:L67–L73, 1971 July 15

DISCOVERY OF PERIODIC X-RAY PULSATIONS IN CENTAURUS X-3 FROM *UHURU*

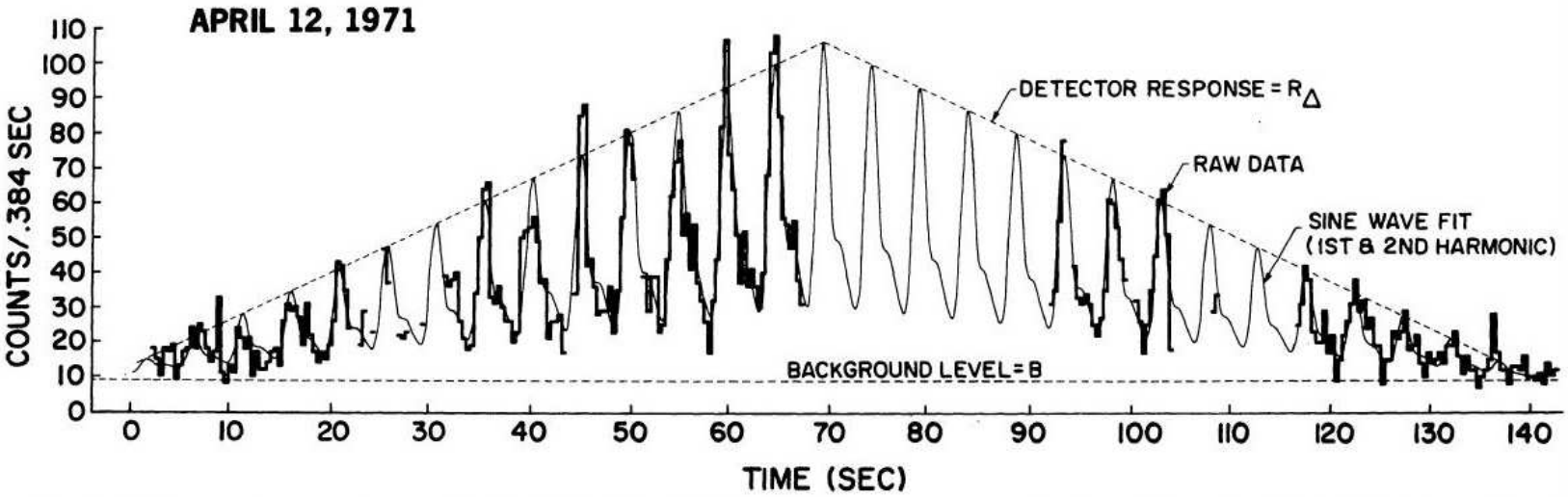


FIG. 2.—Histogram shows counts accumulated in 0.384-s intervals in the 2–6-keV energy range as a function of time on 1971 April 12. Missing portions of the data are due to quick-look transmission dropout. The spacecraft spin rate during this observation was about 0°.07 s^{-1}. The sinusoidal function fit to the data is shown as the continuous curve and is given analytically by the function $f = B + R_\Delta[A_0 \sin(\omega t + \phi_1) + A_1 \sin(2\omega t + \phi_2) + A_2 \sin(3\omega t + \phi_3) + C]$.

Portrait of an Accreting Binary Pulsars

- Pulsed emission due to rotation

  Not disruption by centrifugal force implies

$$\frac{GM}{R^2} \geq \Omega^2$$

Portrait of an Accreting Binary Pulsars

- Pulsed emission due to rotation

  Not disruption by centrifugal force implies

$$\frac{GM}{R^2} \geq \Omega^2$$

$$G\,\frac{M}{R^3}$$

Portrait of an Accreting Binary Pulsars

- Pulsed emission due to rotation

  Not disruption by centrifugal force implies

$$\frac{GM}{R^2} \geq \Omega^2$$

$$G\,\frac{M}{R^3} \simeq G\langle\rho\rangle$$

Portrait of an Accreting Binary Pulsars

- Pulsed emission due to rotation

  Not disruption by centrifugal force implies

$$\frac{GM}{R^2} \geq \Omega^2$$

$$G\,\frac{M}{R^3} \simeq G\langle\rho\rangle \geq \Omega^2$$

Portrait of an Accreting Binary Pulsars

- Pulsed emission due to rotation

Not disruption by centrifugal force implies

$$\frac{GM}{R^2} \geq \Omega^2$$

$$G\frac{M}{R^3} \simeq G\langle\rho\rangle \geq \Omega^2$$

$$\Omega = \frac{2\pi}{4.8} = 1.3\,\text{Hz}$$

Portrait of an Accreting Binary Pulsars

- Pulsed emission due to rotation

Not disruption by centrifugal force implies

$$\frac{GM}{R^2} \geq \Omega^2$$

$$G\frac{M}{R^3} \simeq G\langle\rho\rangle \geq \Omega^2$$

$$\Omega = \frac{2\pi}{4.8} = 1.3\,\text{Hz} \Rightarrow \langle\rho\rangle \geq 10^7\ \text{g/cm}^3$$

Portrait of an Accreting Binary Pulsar

- Gravitational field of a Neutron Star

  From the Newton law

$$F = G \frac{m\,M}{R^2}$$

Portrait of an Accreting Binary Pulsar

- Gravitational field of a Neutron Star

From the Newton law

$$F = G\,\frac{m\,M}{R^2}$$

$$\frac{F_{\mathrm{NS}}}{F_{\mathrm{Earth}}} =$$

Gravitational field of a Neutron Star

From the Newton law

$$F = G \frac{m\,M}{R^2}$$

$$\frac{F_{\mathrm{NS}}}{F_{\mathrm{Earth}}} = \frac{M_{\mathrm{NS}}}{M_{\mathrm{Earth}}}$$

Portrait of an Accreting Binary Pulsar

- Gravitational field of a Neutron Star

From the Newton law

$$F = G\,\frac{m\,M}{R^2}$$

$$\frac{F_{\mathrm{NS}}}{F_{\mathrm{Earth}}} = \frac{M_{\mathrm{NS}}}{M_{\mathrm{Earth}}} \times \left(\frac{R_{\mathrm{Earth}}}{R_{\mathrm{NS}}}\right)^2$$

Portrait of an Accreting Binary Pulsar

- Gravitational field of a Neutron Star

From the Newton law

$$F = G\,\frac{m\,M}{R^2}$$

$$\frac{F_{\mathrm{NS}}}{F_{\mathrm{Earth}}} = \frac{M_{\mathrm{NS}}}{M_{\mathrm{Earth}}} \times \left(\frac{R_{\mathrm{Earth}}}{R_{\mathrm{NS}}}\right)^2 \simeq 10^{10}$$

- Magnetic field of a Neutron Star

  From the conservation of the magnetic flux for a solar type star, with $B \sim 100$ gauss, we have

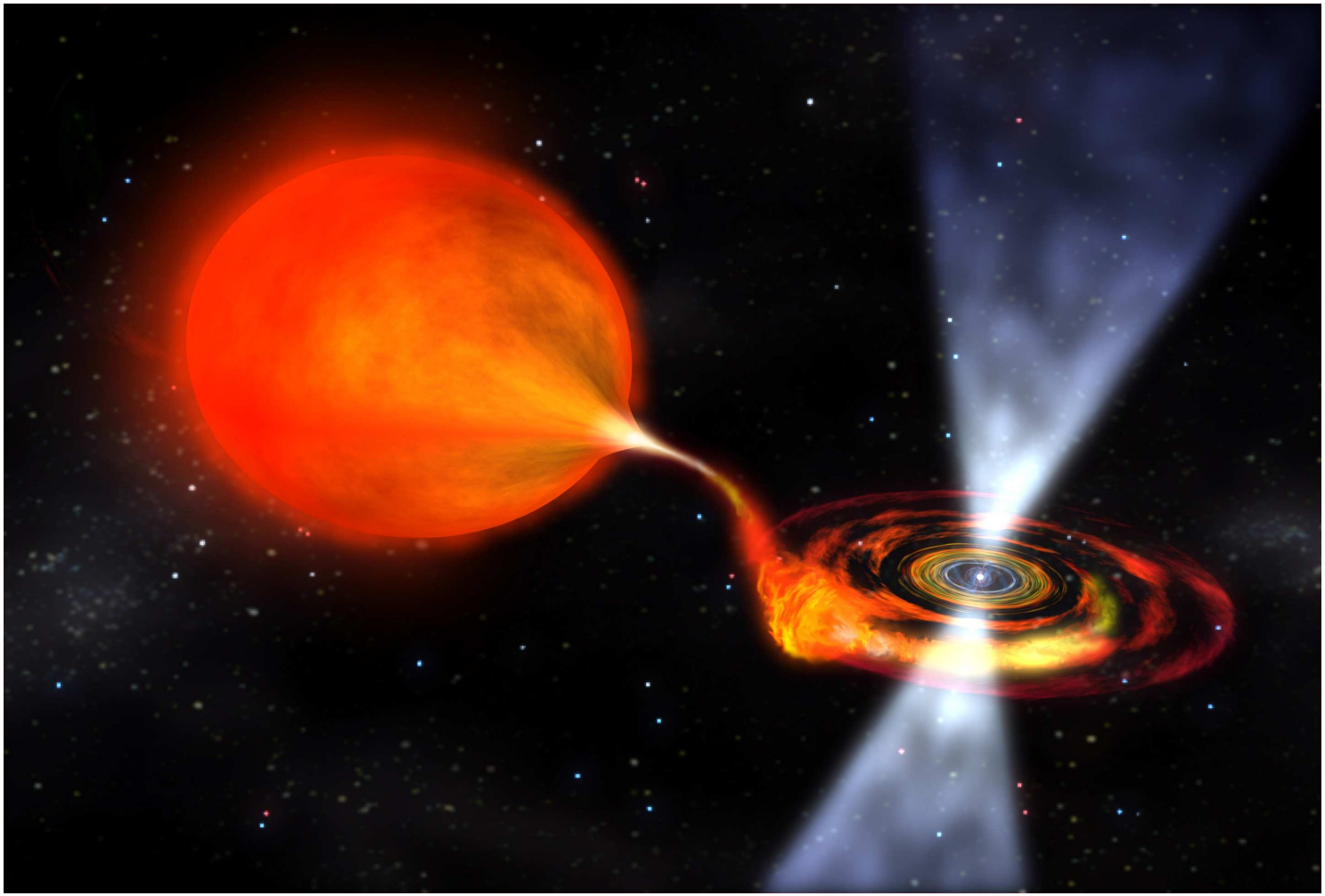
Portrait of an Accreting Binary Pulsar

- Magnetic field of a Neutron Star

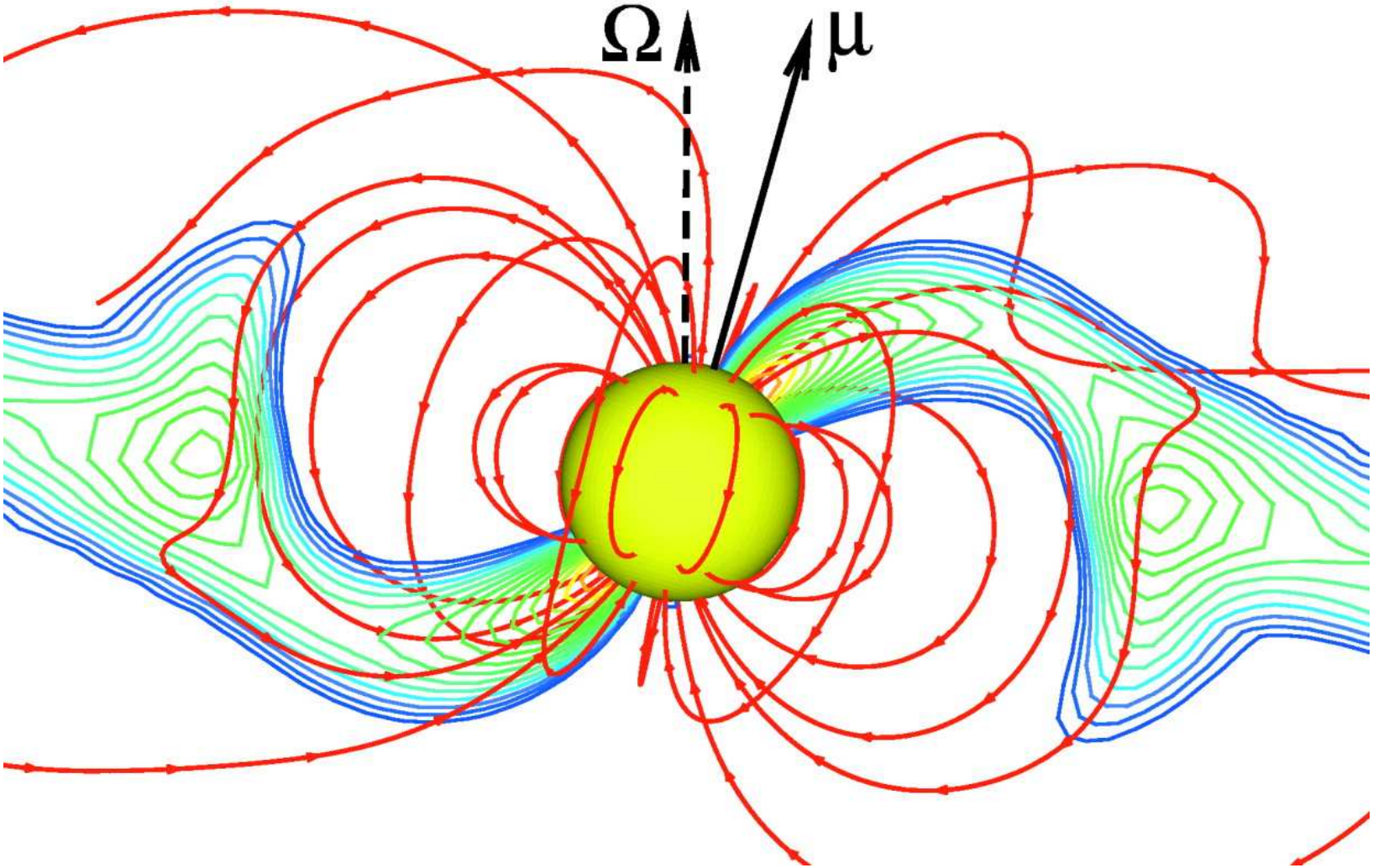
From the conservation of the magnetic flux for a solar type star, with $B \sim 100$ gauss, we have

$$100 \times \left(\frac{R_{\text{Sun}}}{R_{\text{SN}}} \right)^2 \simeq 10^{12} \text{ gauss}$$

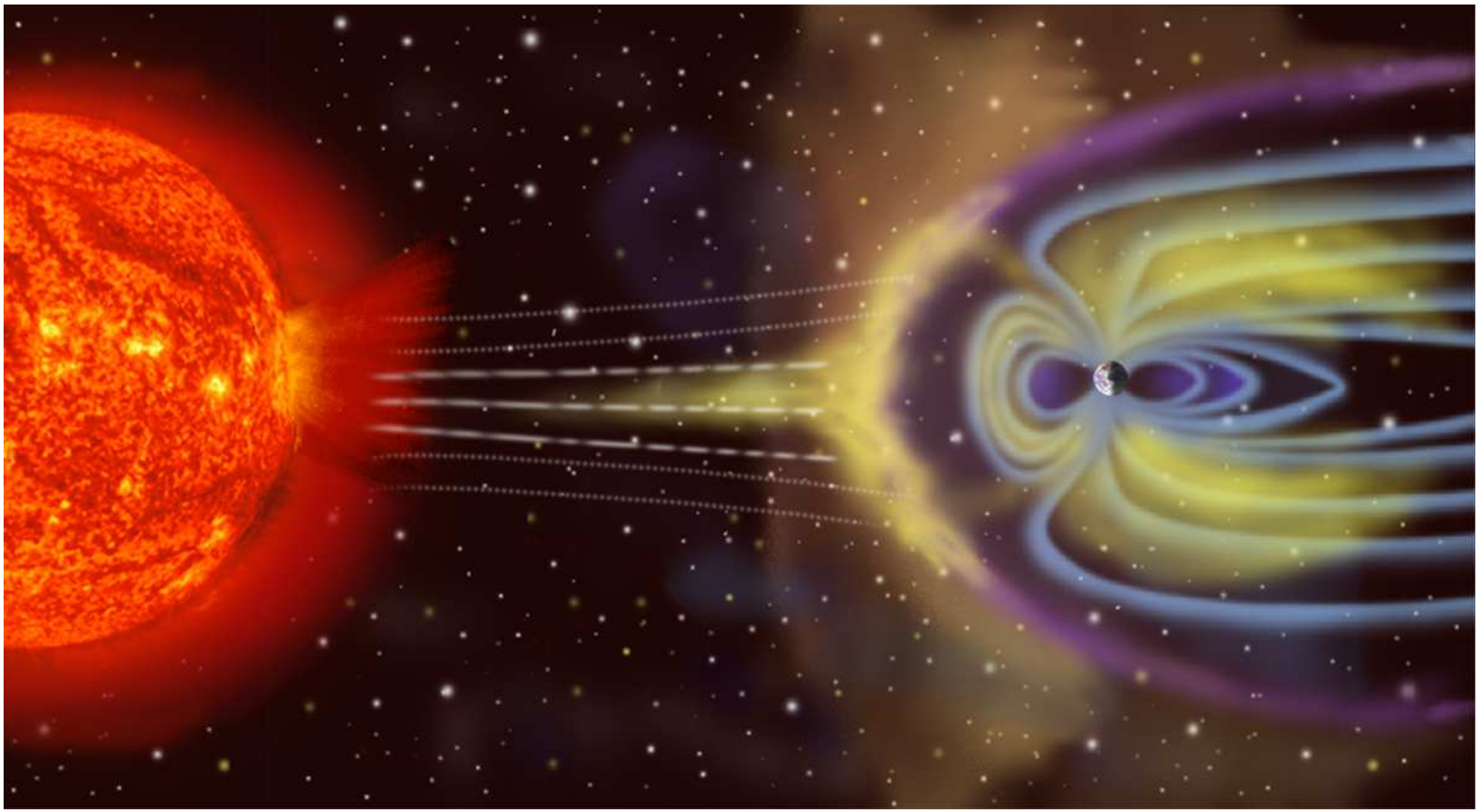
Why AXPs Are Important



Why AXPs Are Important



Why AXPs Are Important



Part I
Theory

Flux of Matter Falling from Stellar Wind and/or Roche Lobe Overflow

Flux of Matter Falling from Stellar Wind and/or Roche Lobe Overflow

|

Interaction with the Magnetic Field at the Magnetospheric Radius r_{MAG}

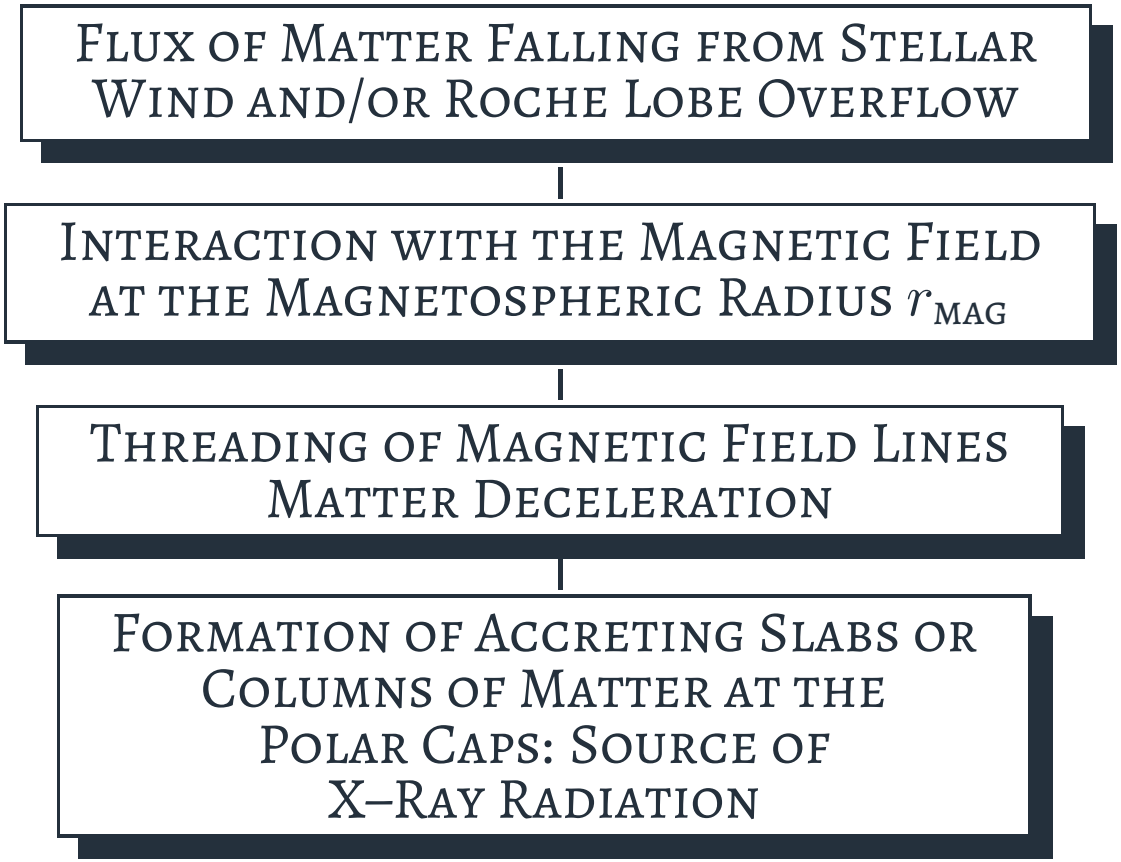
Flux of Matter Falling from Stellar Wind and/or Roche Lobe Overflow

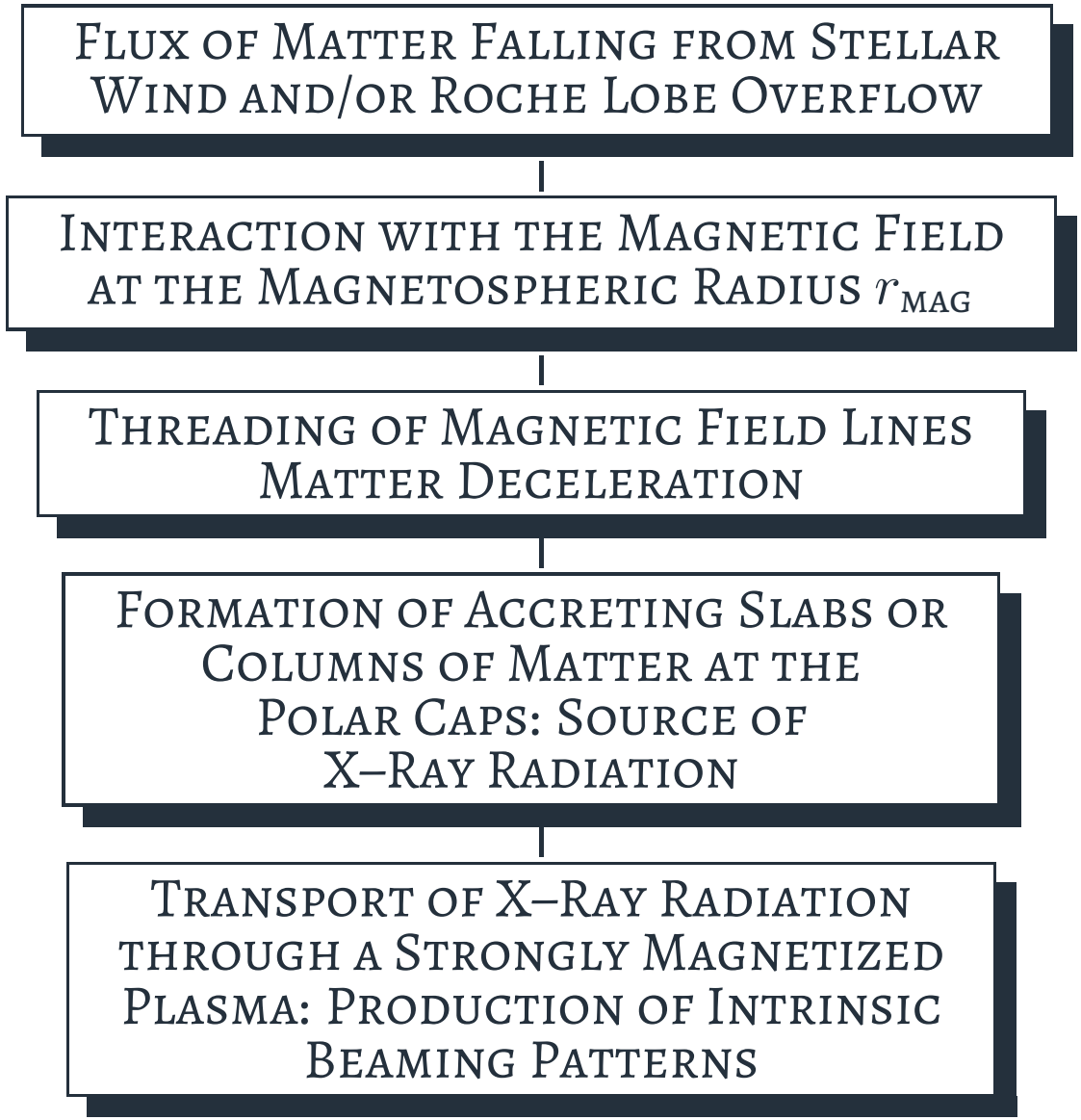
↓

Interaction with the Magnetic Field at the Magnetospheric Radius r_{MAG}

↓

Threading of Magnetic Field Lines Matter Deceleration

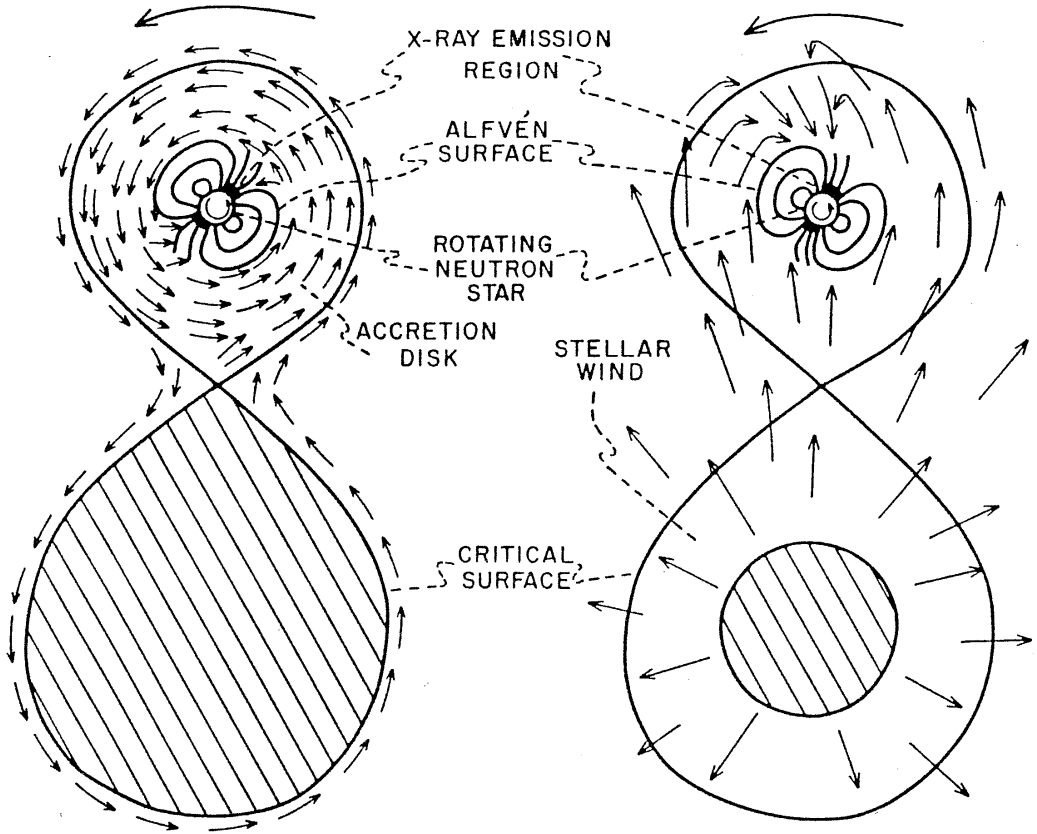




Region I: Outside the B field interaction

Two types of AXPs according to their accretion regime

- ❏ disk-fed AXPs;
- ❏ wind-fed AXPs.



Region I: Outside the B field interaction

The Accretion Radius

All the matter swept by the NS inside a distance called "accretion radius" will be captured and accreted.

$$r_a = \frac{2GM_x}{v_{\rm rel}^2 + c_s^2} \approx \frac{2GM_x}{v_{\rm orb}^2 + v_{\rm win}^2}$$

where M_x is the neutron star mass, c_s is the sound speed (negligible because the wind matter is supersonic), and $v_{\rm orb}$ and $v_{\rm win}$ are the orbital and wind velocity, respectively.
We expect that the characteristic time scales in this "block" be dynamical, of the order of 100–1000 s.

Region I: Outside the B field interaction

The Corotation Radius

In order for matter to be accreted it is necessary that the neutron star does not rotate so fast that plasma is expelled because of the centrifugal force. The distance at which there is balance between these two forces is called corotation radius, defined as

$$r_c = \left(\frac{GM_x}{\Omega_p^2}\right)^{1/3} = 1.5 \times 10^8 \; P^{2/3} m^{1/3} \, \mathrm{cm}$$

Region II: Influence of the B field

The Magnetospheric Radius

At some distance from the neutron star surface, called magnetospheric radius, the magnetic field of the neutron star becomes the main interaction which drives the motion of the captured matter toward the stellar surface. At this distance matter is halted by the very strong magnetic field of the neutron star and accretion can occur only if matter can penetrate the shock layer by means of magneto-hydrodynamical instabilities.
From its definition, the magnetospheric radius will depend on the magnetic field strength and the ram pressure of the accreted matter

Region II: Influence of the B field

$$r_m = \begin{cases} 5.1 \times 10^8 \, \xi^{2/7} \mu_{30}^{4/7} m^{-1/7} \dot{M}_{16}^{\ -2/7} \ \mathrm{cm} \\ \\ 2.9 \times 10^8 \, \xi^{2/7} \mu_{30}^{4/7} m^{-1/7} r_6^{-2/7} \epsilon_{0.1}^{2/7} L_{37}^{-2/7} \ \mathrm{cm} \end{cases}$$

where $\xi \lesssim 1$, μ_{30} is the dipolar magnetic moment in units of 10^{30} G cm^3, $\dot{M}_{16}$ the mass accretion rate onto the neutron star in units of 10^{16} g/s, r_6 the neutron star radius in units of 10^6 cm, $\epsilon_{0.1}$ the accretion efficiency in units of 0.1, and L_{37} the X–ray luminosity in units of 10^{37} erg/s.

Because the main physical processes occurring in this “block” are magneto-hydrodynamical instabilities, the characteristic time scales will be 0.1–10 s.

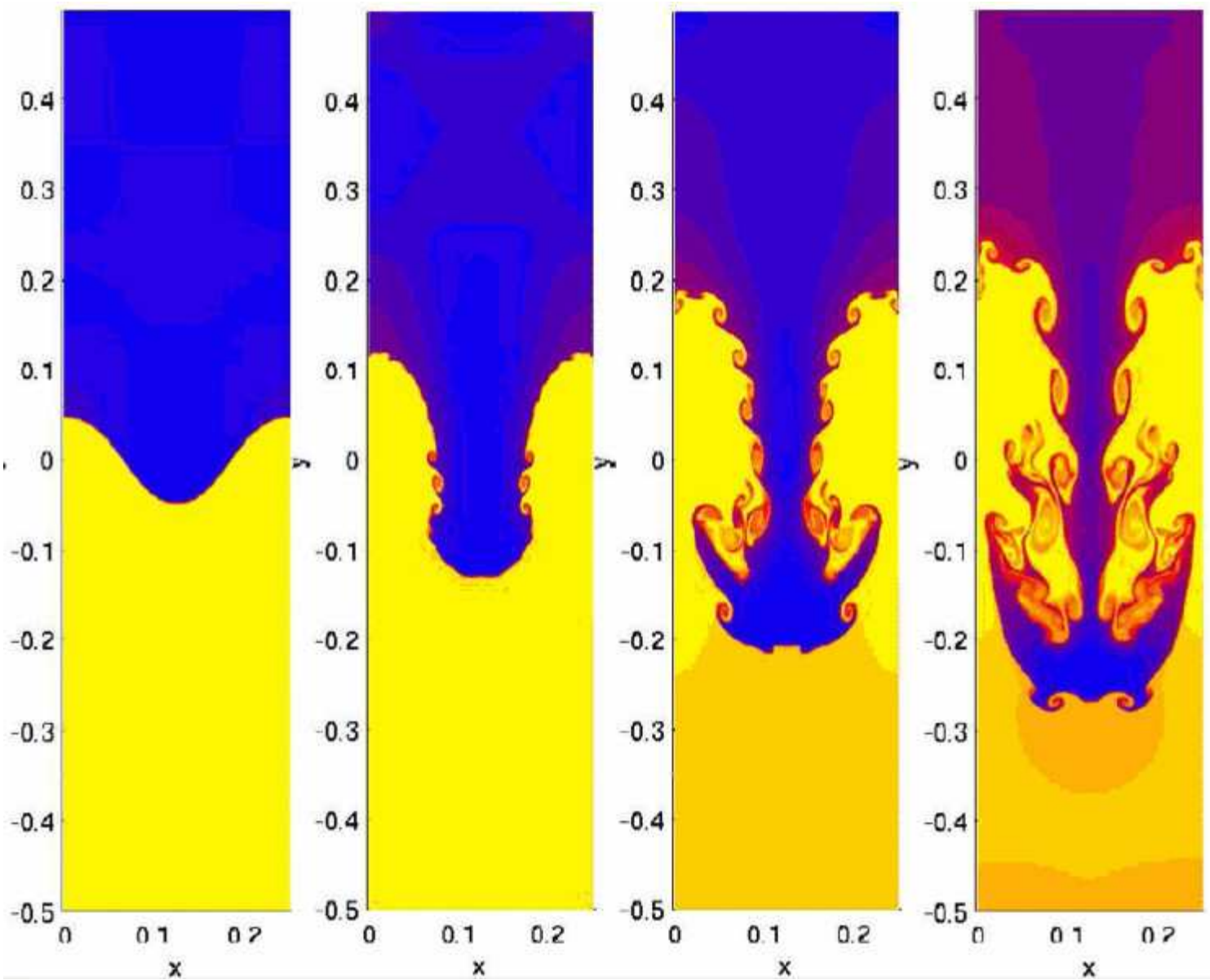
Region III: Plasma Penetration

According to fastness of rotation, different kind of instabilities will determine plasma penetration:

- Slow rotators: gravity-driven interchange Rayleigh-Taylor instability
- Fast rotators: shear Kelvin-Helmholtz instability

fastness parameter: $\omega_k \equiv \Omega_p/\Omega_k(r_m)$, where Ω_k is the angular velocity of matter orbiting into Keplerian orbits.

Region III: Plasma Penetration



Region III: Plasma Penetration

Region III: Inhibition of Accretion

By comparing the two length scales, magnetospheric and corotation radii, it is possible to distinguish two accretion regimes:

- $r_c \gtrsim r_m$: the centrifugal force is smaller than the magnetic force, and therefore matter can be accreted.
- $r_c \lesssim r_m$: centrifugal force inhibits matter from being channeled and is swept away. This is the so-called *propeller* regime.

The Propeller Effect

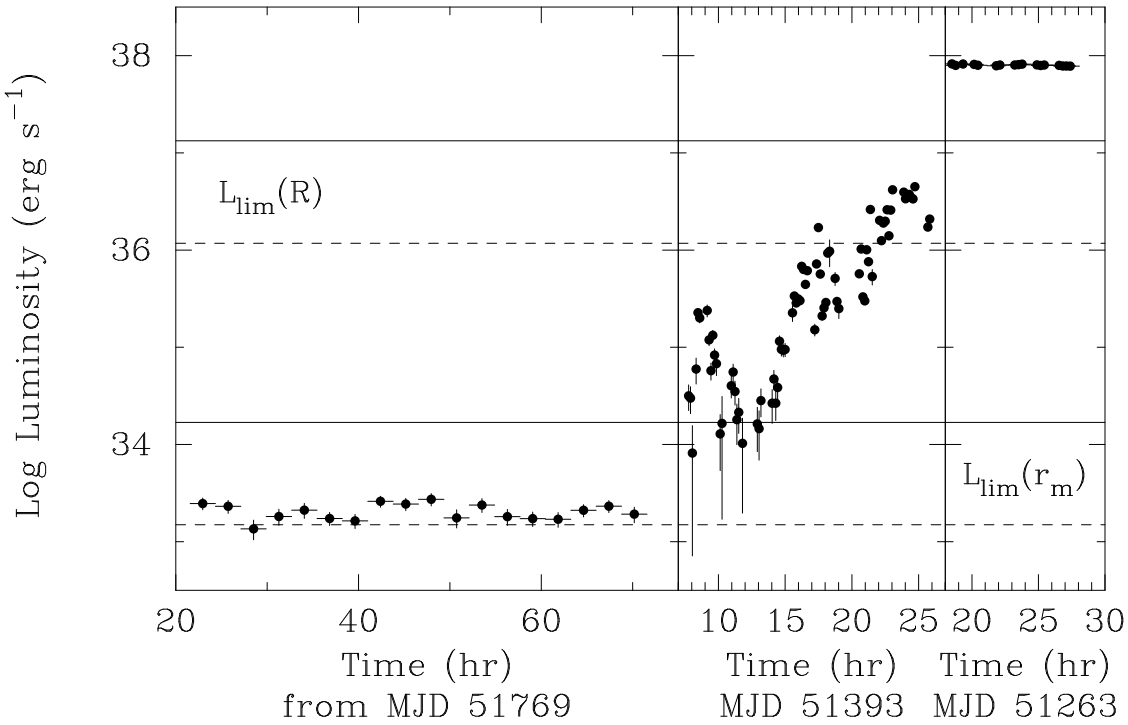
At low accretion rates pulsars enter a centrifugally inhibited state, the so-called *propeller state*. **BUT** four pulsars show pulsation at low luminosity.

Source	$P_{\rm spin}$ (s)	$L_{\rm X}^{\ddagger}$	Energy (keV)	Mission	Ref.
1A 0535+26	103.5	1.3	2–10	BeppoSAX	Orlandini *et al.* (2004)
4U 1145–619	290	5.9	0.5–2	Einstein	Mereghetti *et al.* (2006)
SAX J2103.5+4545	351	1.2	0.5–10	Chandra	Reig *et al.* (2014)
1A 1118–615	409	1.8	0.5–10	Chandra	Rutledge *et al.* (2007)

$^{\ddagger}\ 10^{33}$ erg/s

The Propeller Effect

At low accretion rates pulsars enter a centrifugally inhibited state, the so-called *propeller state*. **BUT** four pulsars show pulsation at low luminosity.



BeppoSAX observation of 4U0115+65: from quiescence to outburst (Campana *et al.* 2001)

Region IV: Radiation Production

Once matter has penetrated the magnetosphere, it will follow the magnetic field lines up to the magnetic polar caps of the NS, where it will be decelerated.

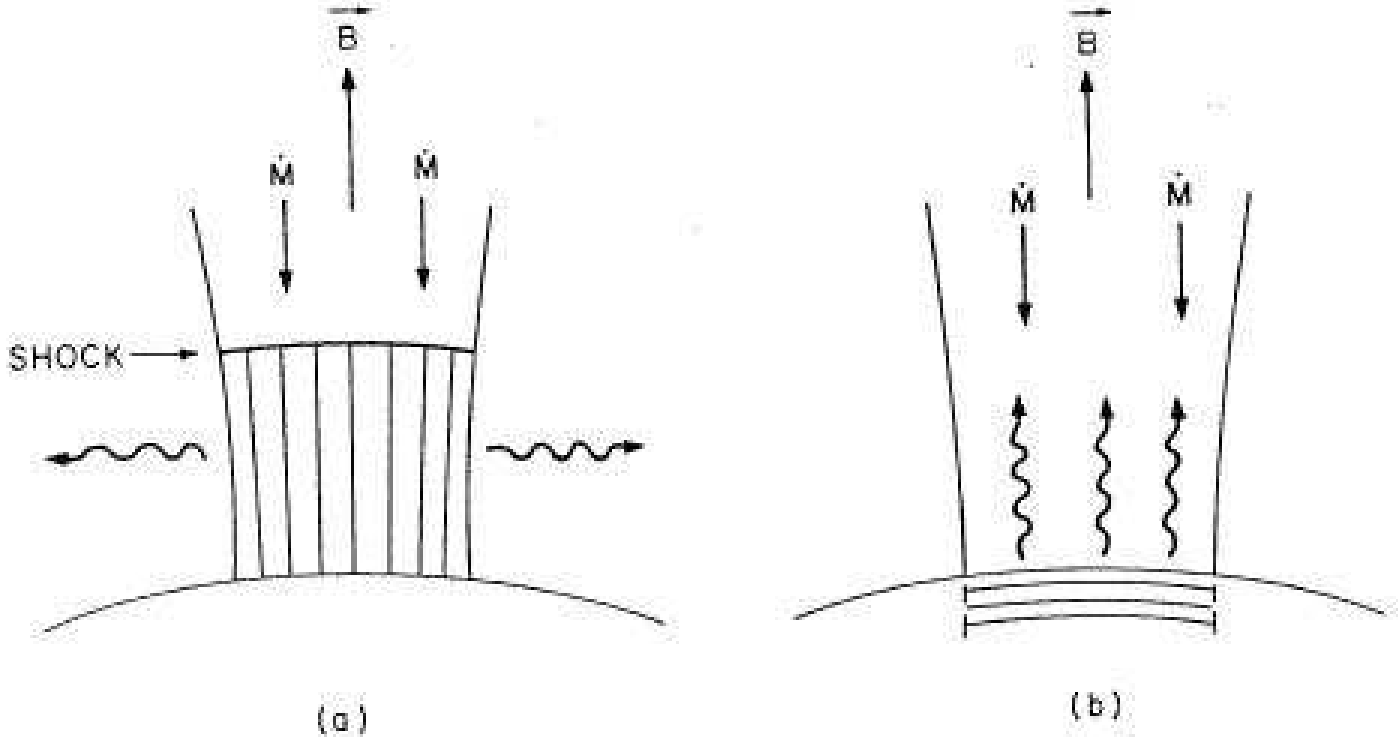
If the amount of matter falling on the polar caps is high enough that a X–ray luminosity greater than about 10^{37} erg/s is reached, then a radiative shock will form.

In this case an accretion column just above the polar cap will form; this accretion column will be optically thick to X–rays, therefore radiation will be emitted mainly *sideways*. Radiation is emitted mainly in a direction perpendicular to the magnetic field lines and we call this pattern “fan beam emission pattern”.

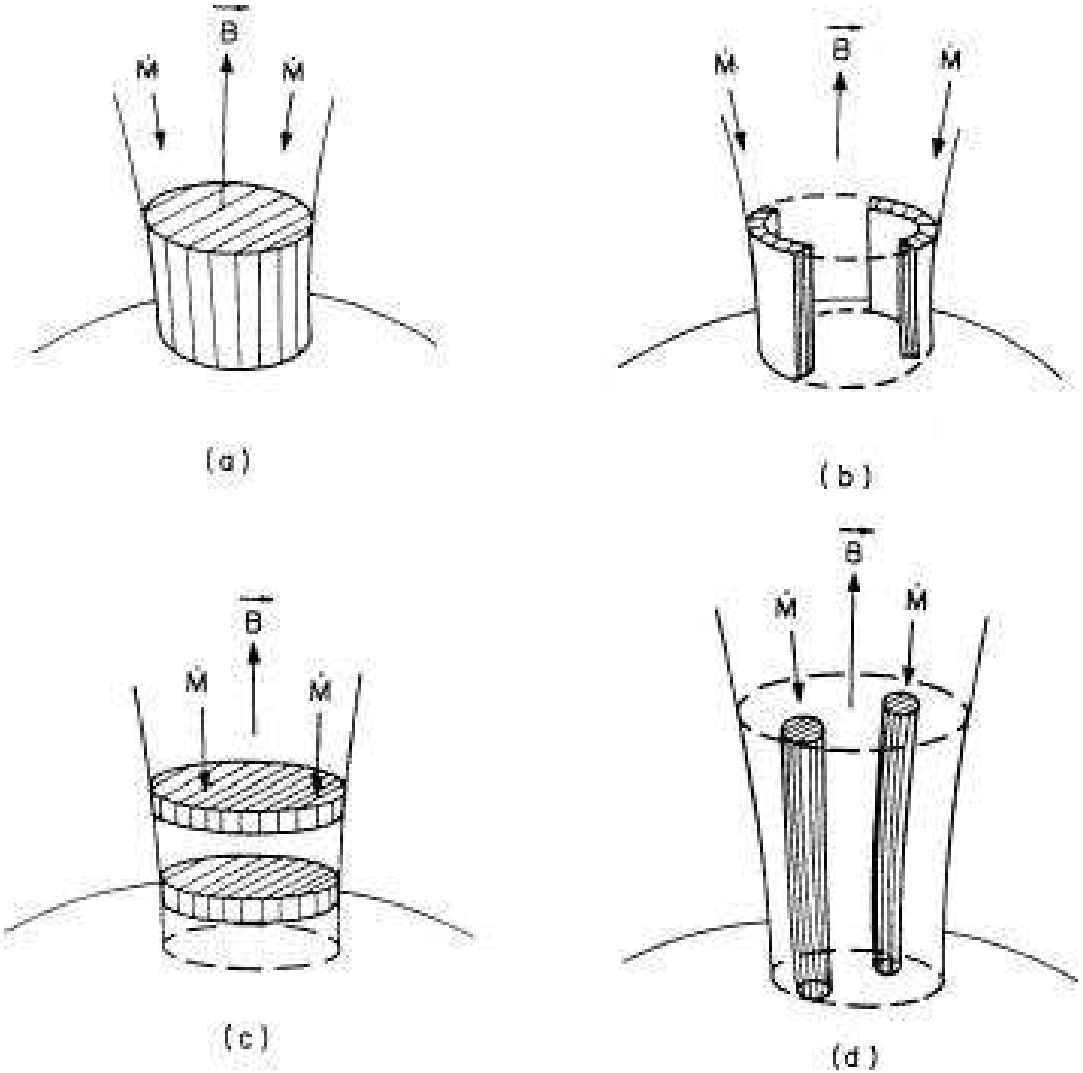
Region IV: Radiation Production

On the other hand, if the X–ray luminosity is lower than about 10^{37} erg/s, then the radiative shock will not form and matter will be able to reach the neutron star surface. In this case we will have the formation of an emitting "slab" and radiation will be emitted mainly in a direction parallel to the magnetic filed lines. We call this pattern "pencil beam emission pattern".
Because the main physical processes occurring in this "box" are Compton heating and cooling, bremsstrahlung and Coulomb interactions, the characteristic time scale will be $\lesssim 0.001$ s.

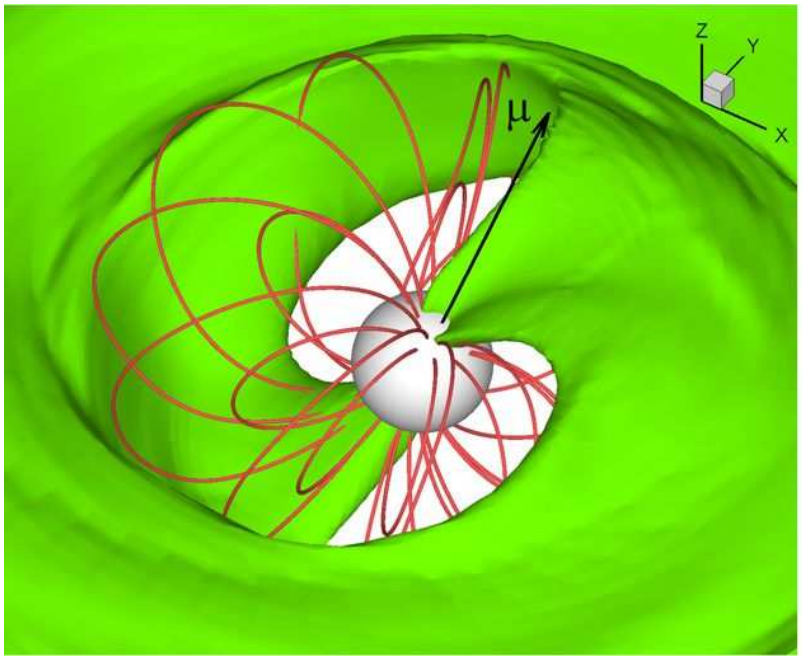
Emitting Region Structure

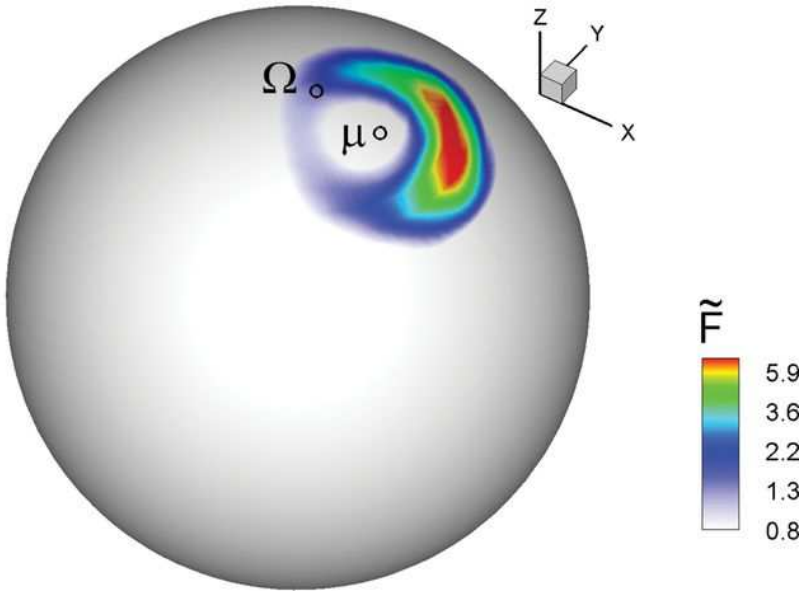


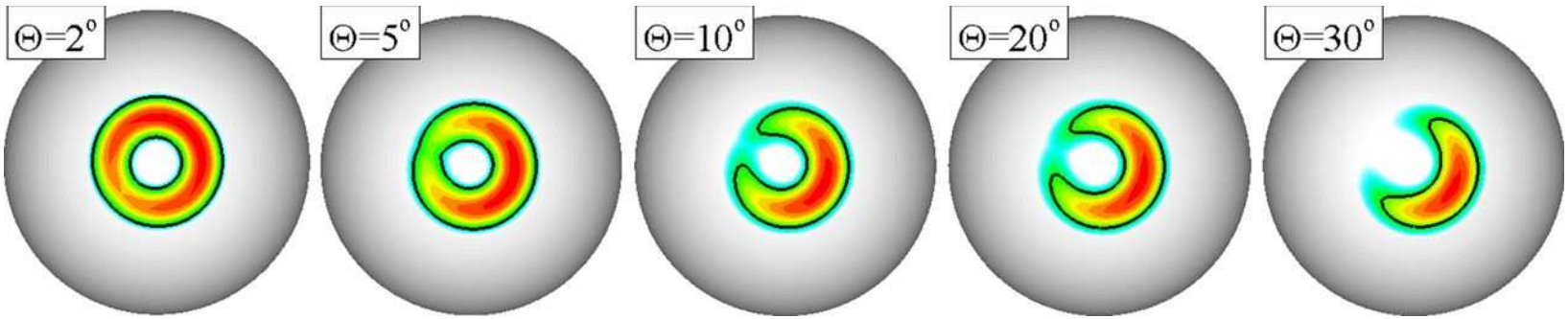
Emitting Region Structure



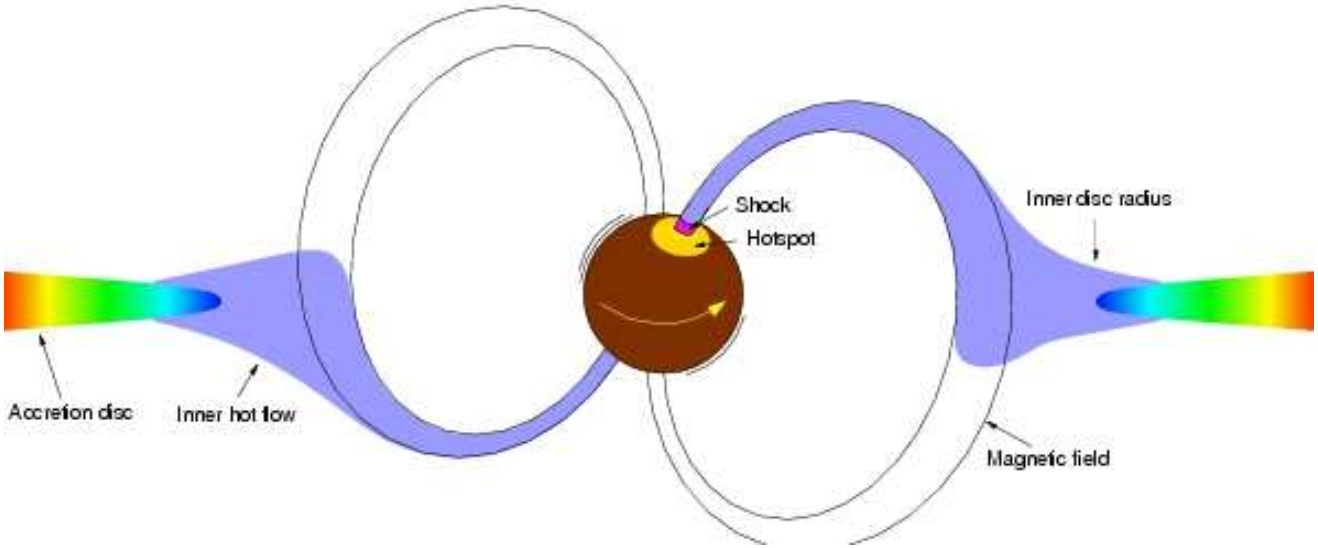
Emitting Region Structure







Time Scales



	Physics	Time Scale
$r > r_{\text{mag}}$	Dynamics	100–1000 s
$r \simeq r_{\text{mag}}$	MHD Instability	0.1–10 s
$r < r_{\text{mag}}$	Compton H&C	$<$0.001 s
	Bremsstrahlung	
	Coulomb	

Region V: Emergent Spectrum

The electrons present in the magnetosphere will have an helicoidal motion along the magnetic field lines, with gyromagnetic (Larmor) frequency given by

$$\omega_c = \frac{eB}{\gamma mc}$$

where γ is the Lorenz factor. For the magnetic field strength B expected in the NS magnetosphere, the motion of the electron in the direction perpendicular to B is quantized in the so-called Landau levels.

Region V: Emergent Spectrum

In the non-relativistic case, the energy associated to each level is given by

$$\hbar\omega_n = n\,\hbar\omega_c$$

where ω_c is the Larmor gyrofrequency.
As an aside, from this Eq. we have that $E_n = 11.6 \cdot B_{12}$ keV, where B_{12} is the magnetic field strength in units of 10^{12} gauss. Therefore we expect to observe cyclotron features in the hard ($E > 10$ keV) energy range.

Region V: Emergent Spectrum

When relativistic corrections are taken into account a slight anharmonicity is introduced in the Landau levels. Indeed, we have

$$\hbar\omega_n = mc^2 \frac{\sqrt{mc^2 + 2n\hbar\omega_c \sin^2\theta} - 1}{\sin^2\theta}$$

where θ is the angle between the line of sight and B.

Cyclotron Resonance Features

Another consequence of the existence of the Landau levels is that an electromagnetic wave propagating in such a plasma will have well defined polarization normal modes, *i.e.* the medium will be birifringent.
If we introduce the complex refraction index N, with its real part the geometric refraction index and with its imaginary part the absorption coefficient, then the dispersion relation in the non-relativistic case can be written as a bi-quadratic equation in N. The solution for N will have the form

$$N_1^2 \propto \frac{1}{\omega - \omega_c} \qquad N_2^2 \propto \frac{1}{\omega + \omega_c} \quad .$$

N_1 extraordinary mode (right-handed circularly polarized)
N_2 ordinary mode (left-handed circularly polarized)

Cyclotron Resonance Features

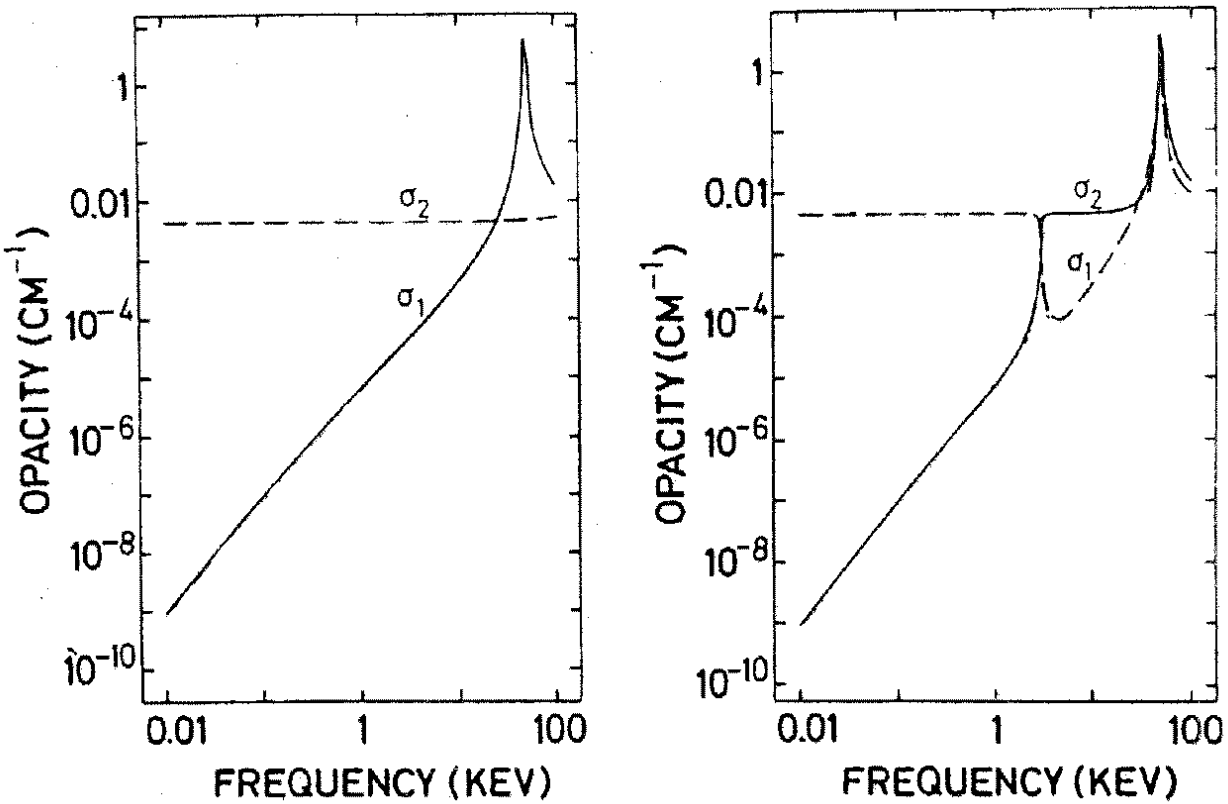
By introducing the complex refractive index is straightforward to obtain the cyclotron ***absorption*** cross section. By means of the optical theorem we obtain

$$\sigma_c \propto \alpha_f \, |e_1|^2 \begin{cases} \delta(\omega - \omega_c) & \text{Not relativistic} \\ \dfrac{W^2}{(\omega - \omega_c) + W^2} & \text{Relativistic} \end{cases}$$

where $\alpha_f = e^2/\hbar c$ is the fine structure constant, and $\vec{e_1}$ is the polarization versor of the extraordinary wave.

Cyclotron Resonance Features

By introducing the complex refractive index is straightforward to obtain the cyclotron ***absorption*** cross section. By means of the optical theorem we obtain



Cyclotron Resonance Features

Once the electron absorbs a photon it (almost) immediately de-excitates on a time scale

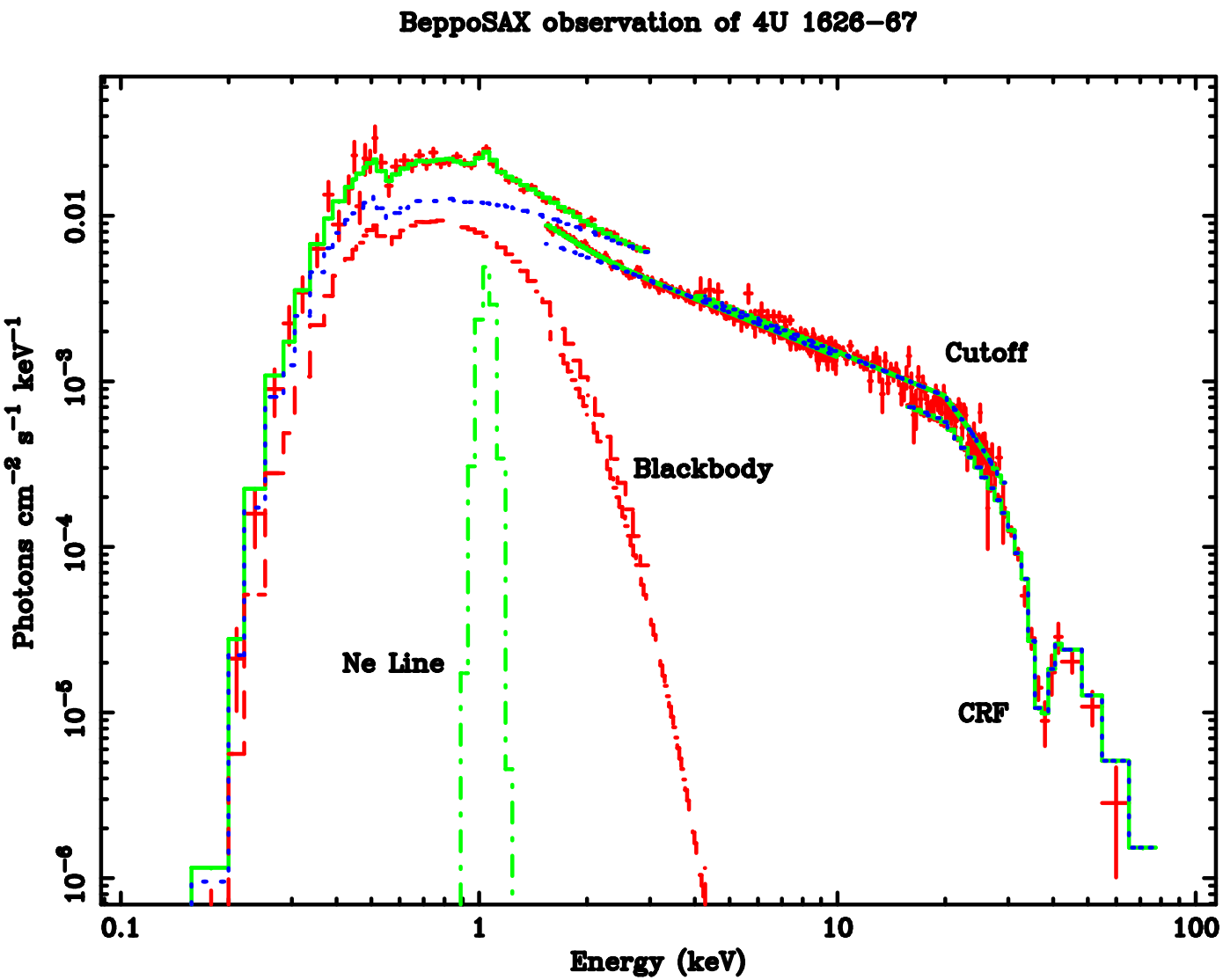
$$t_{\mathrm{r}} \sim 2.6 \times 10^{-16}\, B_{12}^{-1}\ \mathrm{s}$$

This means that an absorbed photon is immediately re-emitted. The absorption-emission process is therefore equivalent to a scattering.

Photons with frequency close to ω_c will be scattered out of the line of sight, creating a drop in their number.

The cyclotron features are therefore **NOT** due to absorption processes, but are due to scattering of photons resonant with the magnetospheric electrons (as it occurs for the Fraunhofer lines in the Solar spectrum).

Cyclotron Resonance Features



Observed CRSFs

AXP with CRSFs

	Source	$P_{\rm spin}$ (sec)	E_c (keV)
1	Swift J1626.6-5156	15.37	10
2	XMMU J054134.7-682550	61.6	10 (2)
3	EXO 2030+375	41.8	11
4	GRO J1948+32	18.7	12.5
5	4U 0115+634	3.61	14 (5)
6	4U 1907+09	438	19 (2)
7	GS 1843+009	29.5	20
8	IGR J18179-1621	11.82	22?
9	4U 1538-52	529	22 (2)
10	2S 0114+650	10008	22 (2)
11	Vela X-1	283	25 (2)
12	V 0332+53	4.4	26 (3)
13	1H 2138+579	66.2	28
14	4U 0352+309	835	29
15	Cen X-3	4.84	30

AXP with CRSFs

	Source	$P_{\rm spin}$ (sec)	E_c (keV)
16	IGR J16493-4348	1069	30
17	RX J0520.5-6932	8.03533	31.5
18	XTE J0658-073	160.7	33
19	GX 1+4	138.17	34?
20	XTE J1946+274	15.8	36
21	OAO 1657-415	38	36?
22	4U 1626-67	7.7	37
23	Her X-1	1.24	41
24	MAXI J1409-619	506	44 (3)
25	GX 301-2	696	45
26	1A 0535+262	104	46 (2)
27	GX 304-1	272	51
28	1A 1118-615	405	55
29	GRO J1008-57	93.5	76
30	LMC X-4	13.5	100?

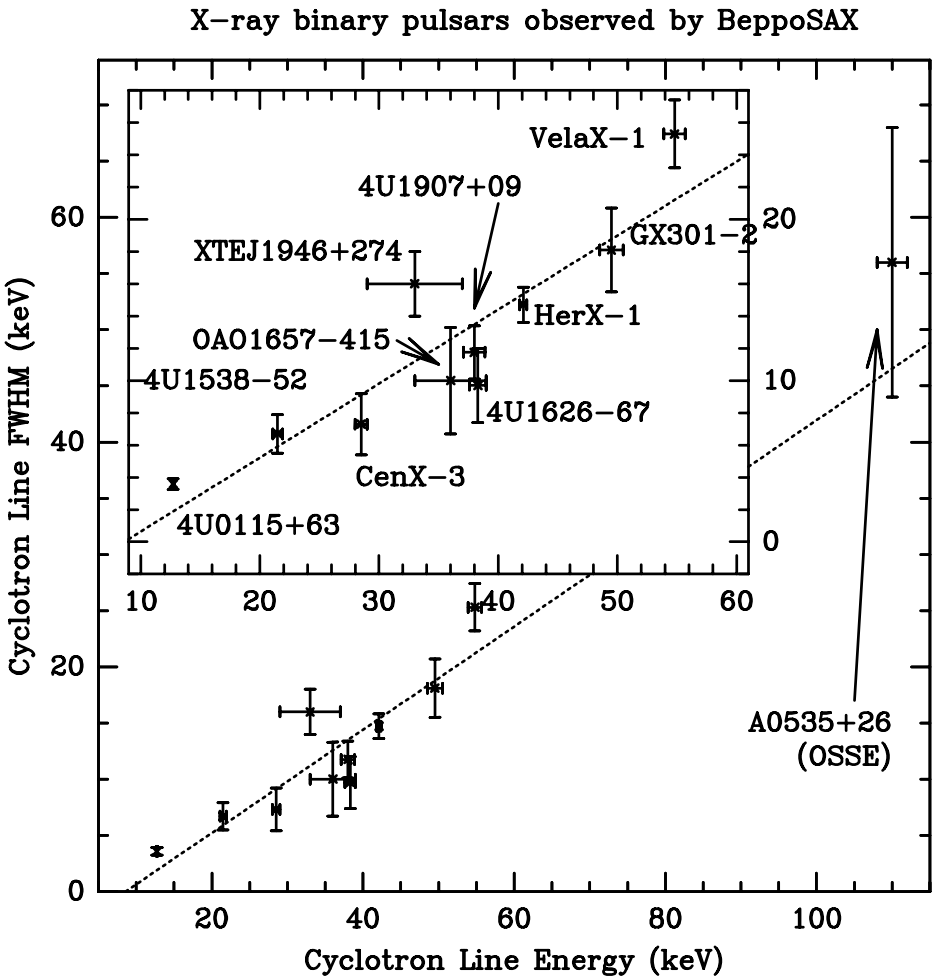
Doppler Broadening of the CRSF

Up to now, we worked neglecting both relativistic corrections and thermal motions (cold plasma approximation). The release of the latter condition allows an electron to absorb waves not only of frequency $\omega = \omega_c$, but in the interval $\omega_c \pm \Delta\omega_D$, where the Doppler width is given by

$$\Delta\omega_D = \omega_c \sqrt{\frac{2kT}{mc^2}} \, |\cos\theta|$$

where T is the electron temperature (we assumed a Maxwell-Boltzmann distribution for the electrons).

Doppler Broadening of the CRSF



The Continuum Spectrum

The main physical process responsible for the continuum emission in AXPs is Compton scattering.

We will not enter into the details of the problem of repeated scatterings in a finite, thermal medium. Let us only summarize that an input photon of energy E_i will emerge from a cloud of non-relativistic electrons (at a temperature T) with an average energy $E_f \sim E_i\, e^y$ (this is valid in the regime $E_f \ll 4kT$).

The Comptonization parameter y therefore gives a measure of the photon energy variation in traversing the plasma, and is given by

$$y = \begin{cases} \dfrac{4kT}{mc^2}\,\max\left(\tau,\tau^2\right) & \text{Nonrelativistic} \\ \left(\dfrac{4kT}{mc^2}\right)^2\,\max\left(\tau,\tau^2\right) & \text{Relativistic} \end{cases}$$

where $\max\left(\tau,\tau^2\right)$ is nothing else but the average number of scattering suffered by the photons (τ is the optical depth of the medium).

Note that if $E_i < 4kT$ then photons can increase their energy at the expense of the electrons: this is *inverse* Compton scattering.

The Continuum Spectrum

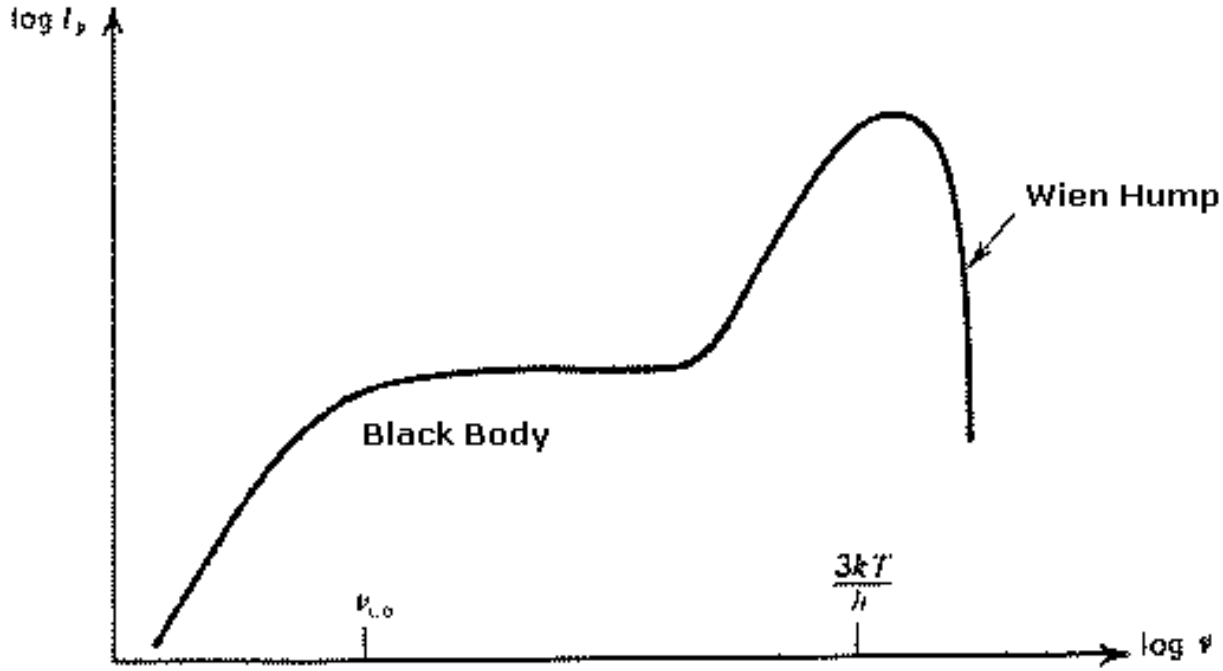
The detailed description of the spectrum of the emergent photons requires the solution of the Kompaneets equation, but it is possible to obtain qualitative information for special cases:

- $y \ll 1$
  In this case only coherent scattering is important, and the emergent spectrum will be a blackbody spectrum or a "modified" blackbody spectrum according whether the photon frequency is lower or greater than the frequency at which scattering and absorption coefficients are equal.

The Continuum Spectrum

The detailed description of the spectrum of the emergent photons requires the solution of the Kompaneets equation, but it is possible to obtain qualitative information for special cases:

- $y \ll 1$



The Continuum Spectrum

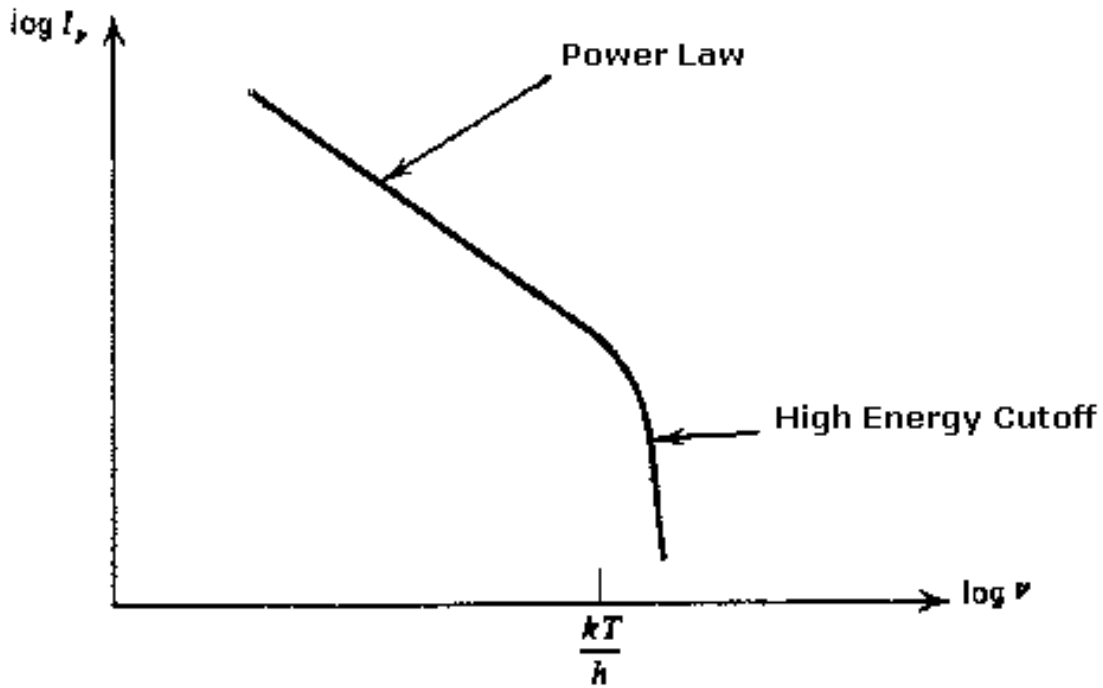
The detailed description of the spectrum of the emergent photons requires the solution of the Kompaneets equation, but it is possible to obtain qualitative information for special cases:

- $y \gg 1$
  Inverse Compton scattering can be important. If we define a frequency ω_{co} such that $y(\omega_{co}) = 1$, then for $\omega \gg \omega_{co}$ the inverse Compton scattering is saturated and the emergent spectrum will show a Wien hump, due to low-energy photons up-scattered up to $\hbar\omega \sim 3kT$. In the case in which there is not saturation a detailed analysis of the Kompaneets equation shows that the spectrum will have the form of a power law modified by a high energy cutoff.

The Continuum Spectrum

The detailed description of the spectrum of the emergent photons requires the solution of the Kompaneets equation, but it is possible to obtain qualitative information for special cases:

- $y \gg 1$



The Continuum Spectrum

From the analysis of the HEAO-1/A2 spectra of AXPs White et al. found an empirical law that was able to fit their energy spectra

$$\mathrm{POHI}(E) = \begin{cases} E^{-\alpha} & E < E_{\mathrm{cut}} \\ E^{-\alpha} \exp\left(-\dfrac{E - E_{\mathrm{cut}}}{E_f}\right) & E > E_{\mathrm{cut}} \end{cases}$$

It is evident that this model tries to simulate the unsaturated inverse Compton process shown before.

But this model suffers the problem of a too abrupt break around the cutoff energy E_{cut}, so smoother cutoffs were introduced.

- Fermi-Dirac cutoff

$$\mathrm{FDCO}(E) = \frac{1}{1 + \exp\left(\dfrac{E - E_{\mathrm{cut}}}{E_f}\right)}$$

called Fermi-Dirac cutoff because of it resemblance with the Fermi-Dirac distribution function. It is important to stress that the FDCO model does not have any physical meaning: it only gives a better description of the break in the AXP spectra.

Makishima and Mihara were the first to note that in the AXPs showing CRFs there was a correlation between the cutoff energy E_{cut} and the CRF energy E_c, namely $E_c \simeq (1.2 - 2.5) \cdot E_{\mathrm{cut}}$.

Mihara introduced the so-called NPEX (Negative Positive EXponential) model

$$\mathrm{NPEX}(E) = (AE^{-\alpha} + BE^{+\beta}) \exp\left(-\frac{E}{kT}\right) \quad .$$

It mimics the saturated inverse Compton spectrum shown before if $\beta = 2$.

Furthermore, because the (non relativistic) energy variation of a photon during Compton scattering is

$$\frac{\Delta E}{E} = \frac{4kT - E}{mc^2}$$

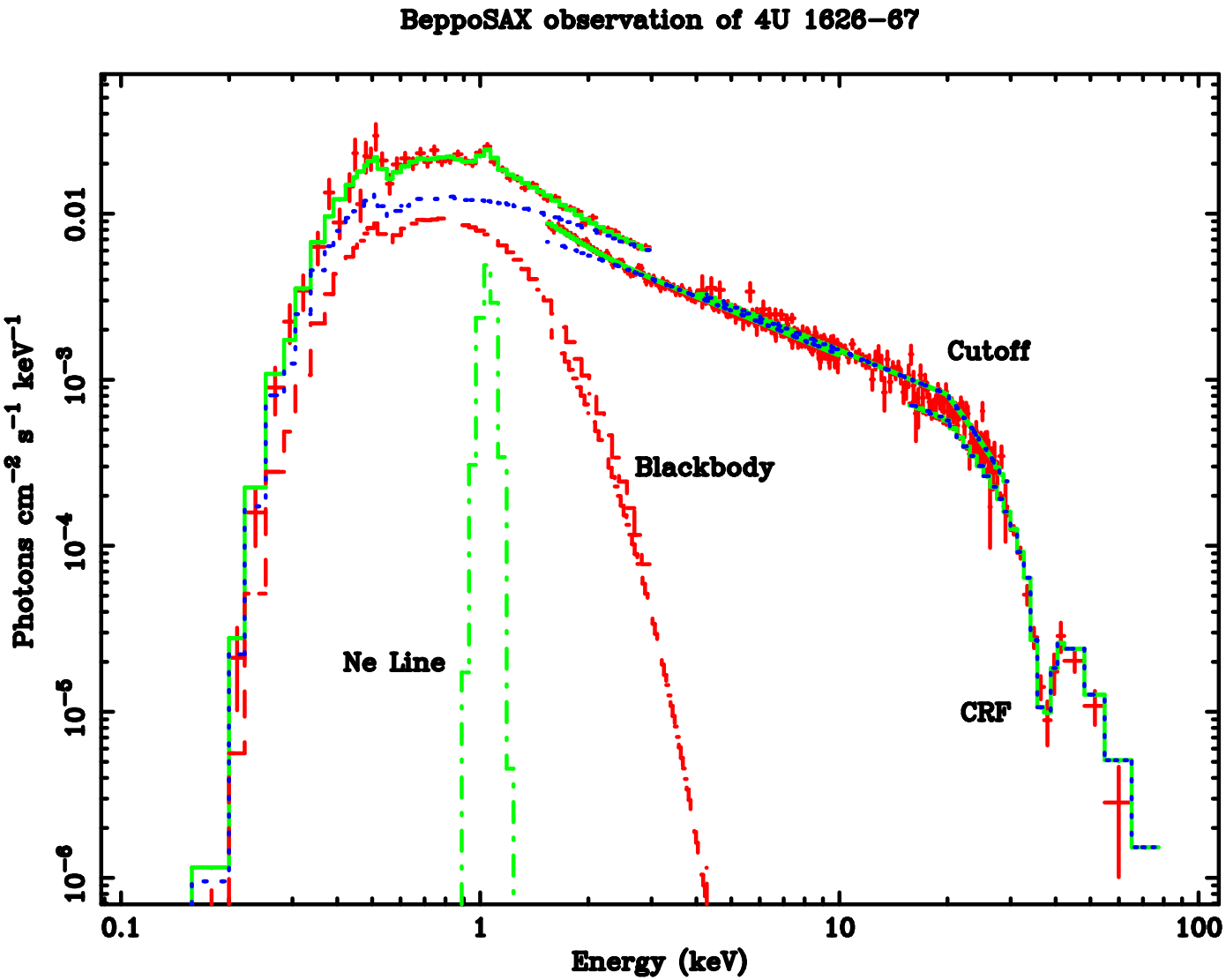
then when $E = E_c$ the medium is optically thick and therefore $E_c \sim 4kT$.

The Continuum Spectrum

From an observational point of view, AXP continua are described in terms of

- a black-body component with temperature of a few hundred eV;
- a power law of photon index $\sim$1 up to 10 keV;
- a high energy ($>$10 keV) cutoff that makes the spectrum rapidly drop above 40–50 keV.

The Continuum Spectrum



The Continuum Spectrum

From an observational point of view, AXP continua are described in terms of

- a black-body component with temperature of a few hundred eV;
- a power law of photon index $\sim$1 up to 10 keV;
- a high energy ($>$10 keV) cutoff that makes the spectrum rapidly drop above 40–50 keV.

Problem: an *ad hoc* source of soft photons is required in order to fit the low energy part of the spectrum.

The Continuum Spectrum

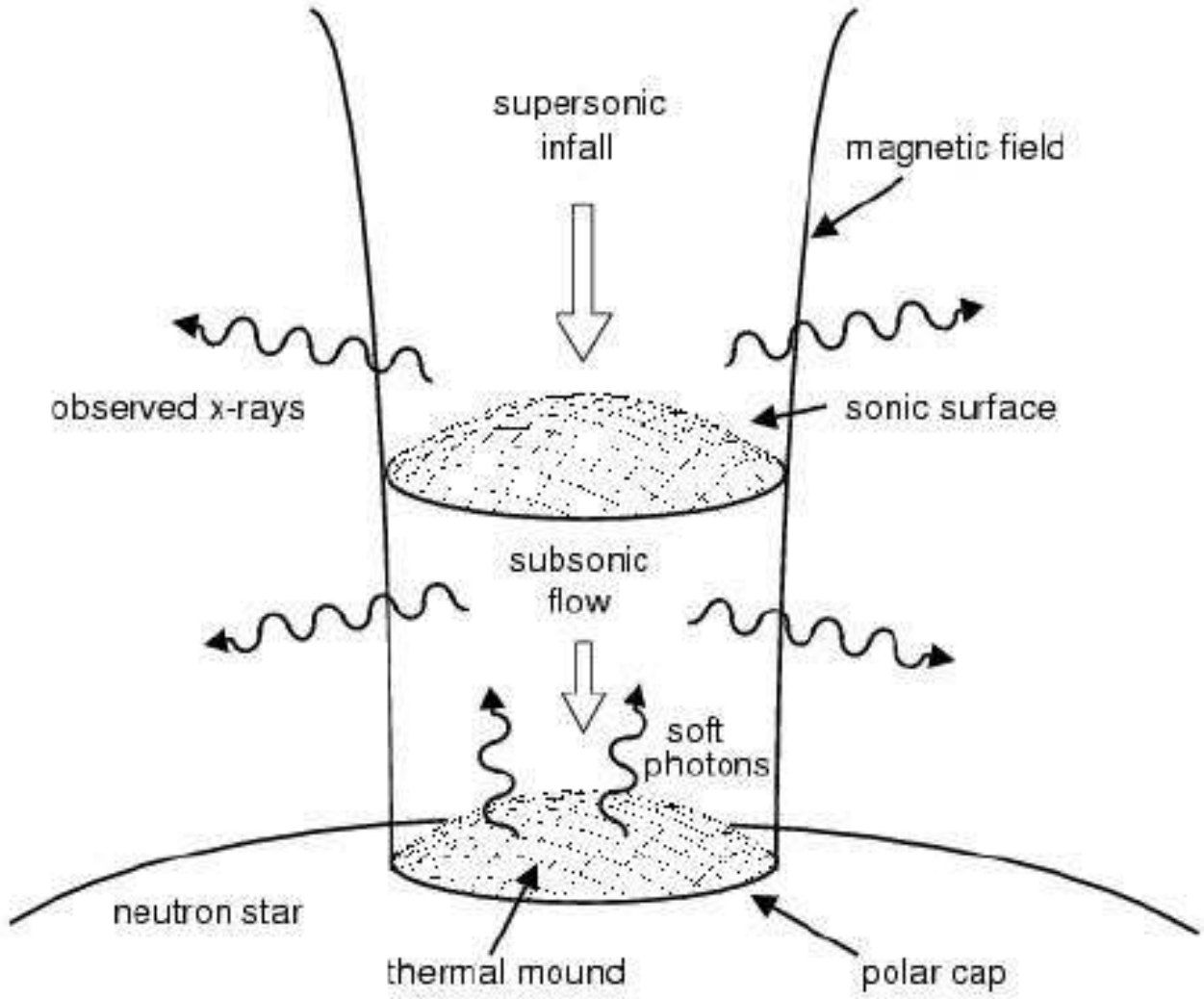
Becker & Wolff naturally derived the source of these soft photons by taking into account the effect of *bulk* (or dynamical) Comptonization in the accretion column.

In the case of a *thermal* Comptonization process, photons gain energy via second-order Fermi acceleration because of the incoherent, stochastic motion of the plasma.

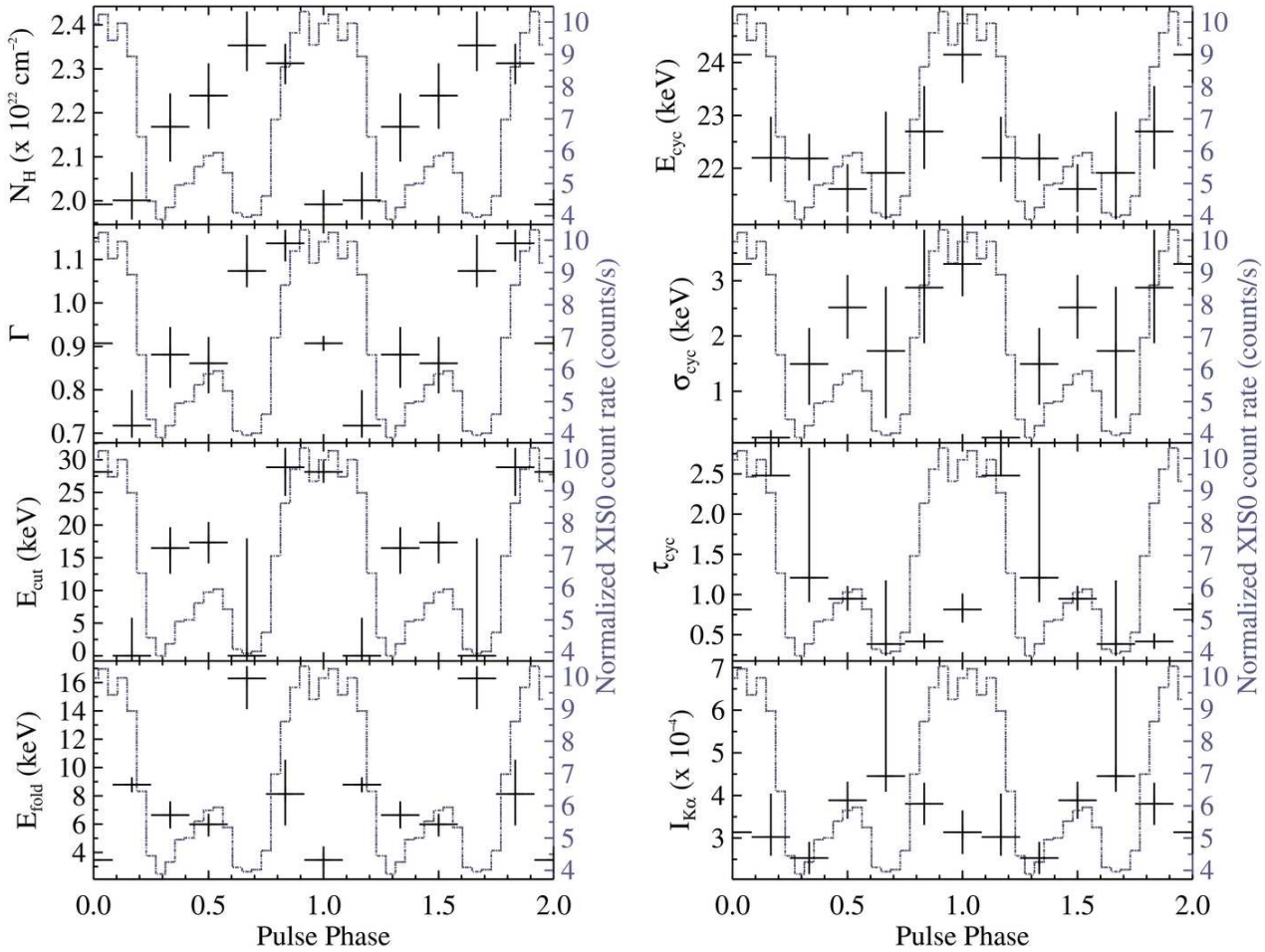
In the accretion column the infalling electrons possess a preferred motion: this implies that photons gain energy through first-order Fermi acceleration.

The soft component comes out naturally from soft photons produced by the "thermal mound" at the base of the accretion column.

The Continuum Spectrum

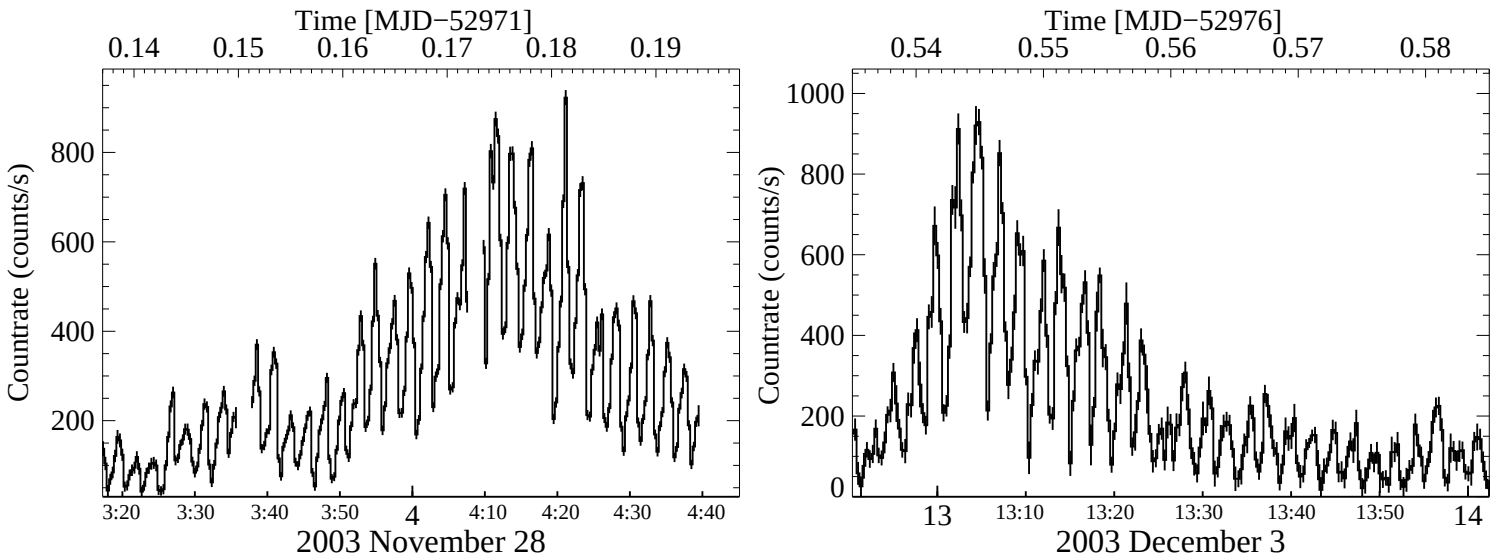


Pulse Phase and Time Resolved Spectroscopy



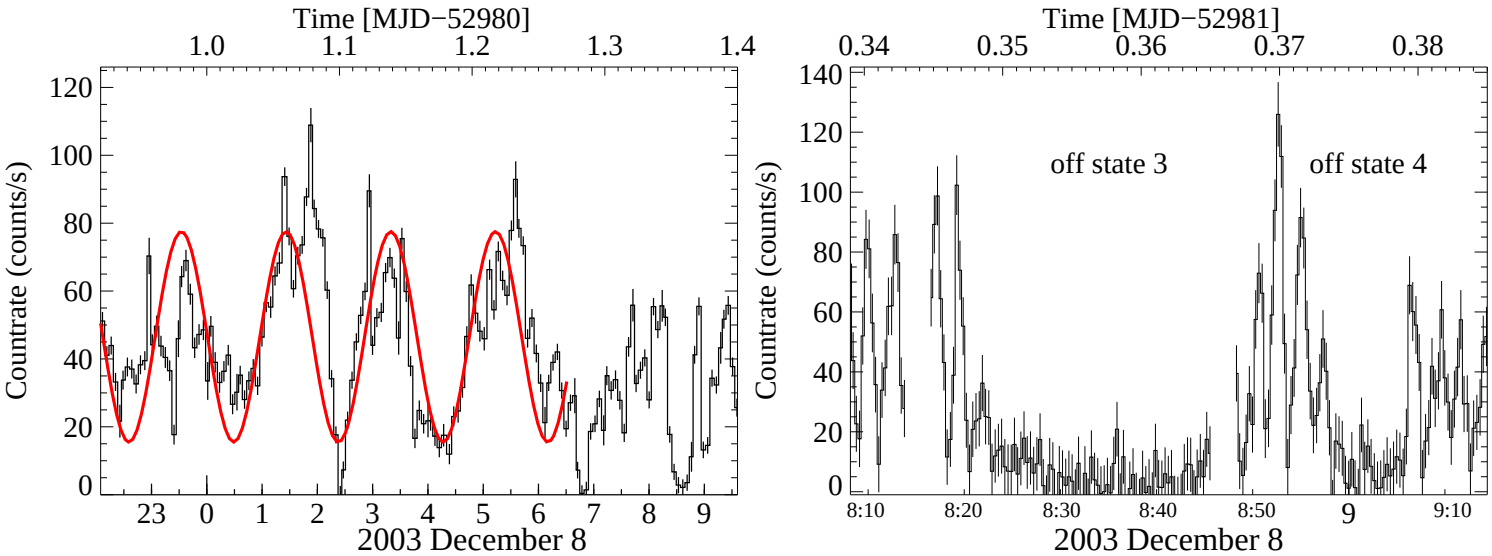
Suzaku PPS of 4U 1538–52 (Hemphill *et al.* 2014. ApJ 792, 14)

Pulse Phase and Time Resolved Spectroscopy



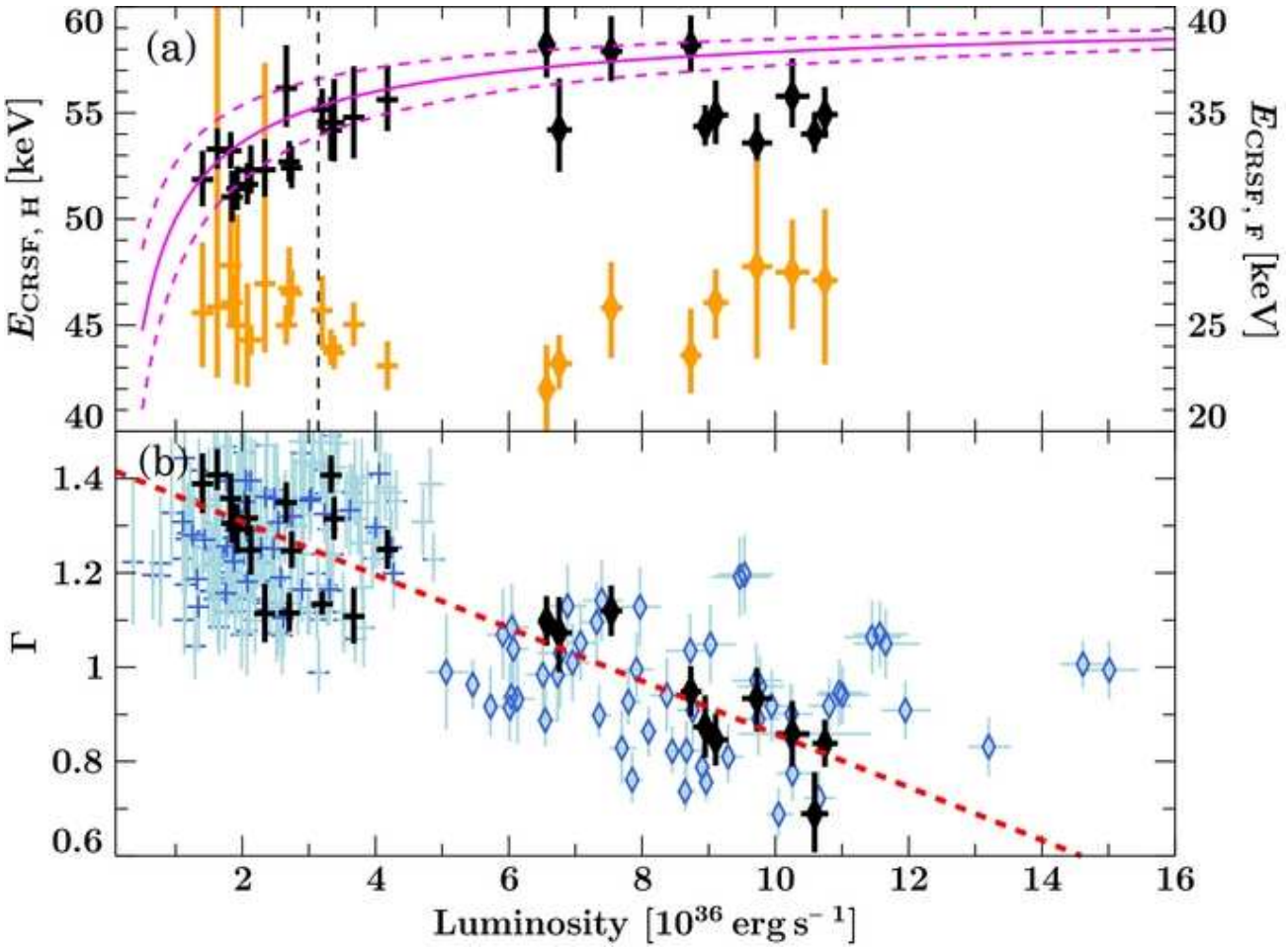
INTEGRAL ISGRI 20–40 keV light curve of Vela X–1 (Kreykenbohm *et al.* 2008. A&A 492, 511)

Pulse Phase and Time Resolved Spectroscopy



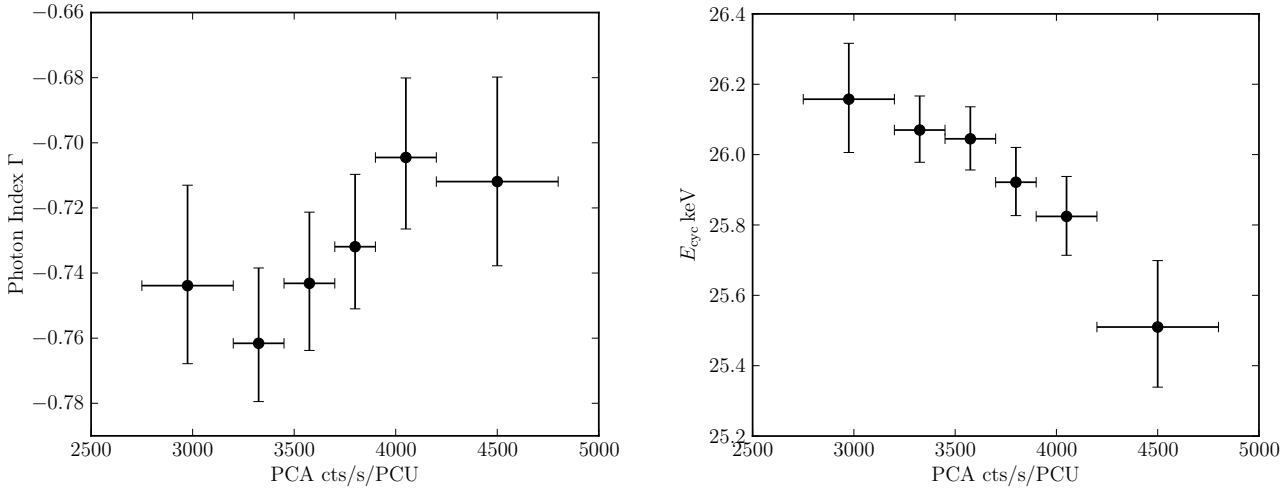
INTEGRAL ISGRI 20–40 keV light curve of Vela X–1 (Kreykenbohm *et al.* 2008. A&A 492, 511)

Pulse Phase and Time Resolved Spectroscopy



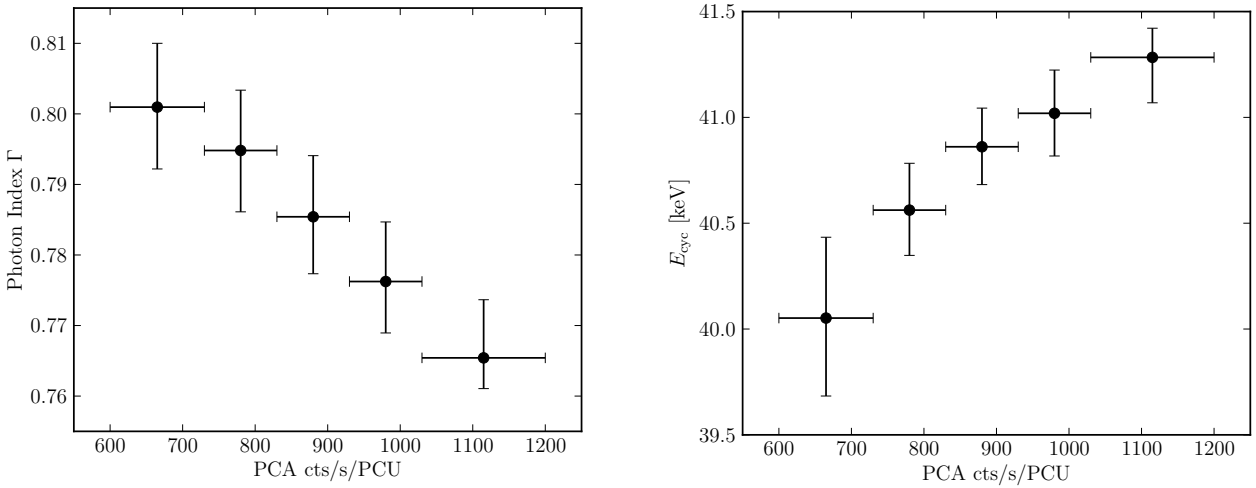
NuSTAR observations of Vela X–1 (Fürst *et al.* 2014, ApJ 780, 133).

Pulse Phase and Time Resolved Spectroscopy



RXTE observation of V0332+53: Continuum (left) and CRF energy (right) vs Pulse Amplitude (Klochkov *et al.* 2011)

Pulse Phase and Time Resolved Spectroscopy

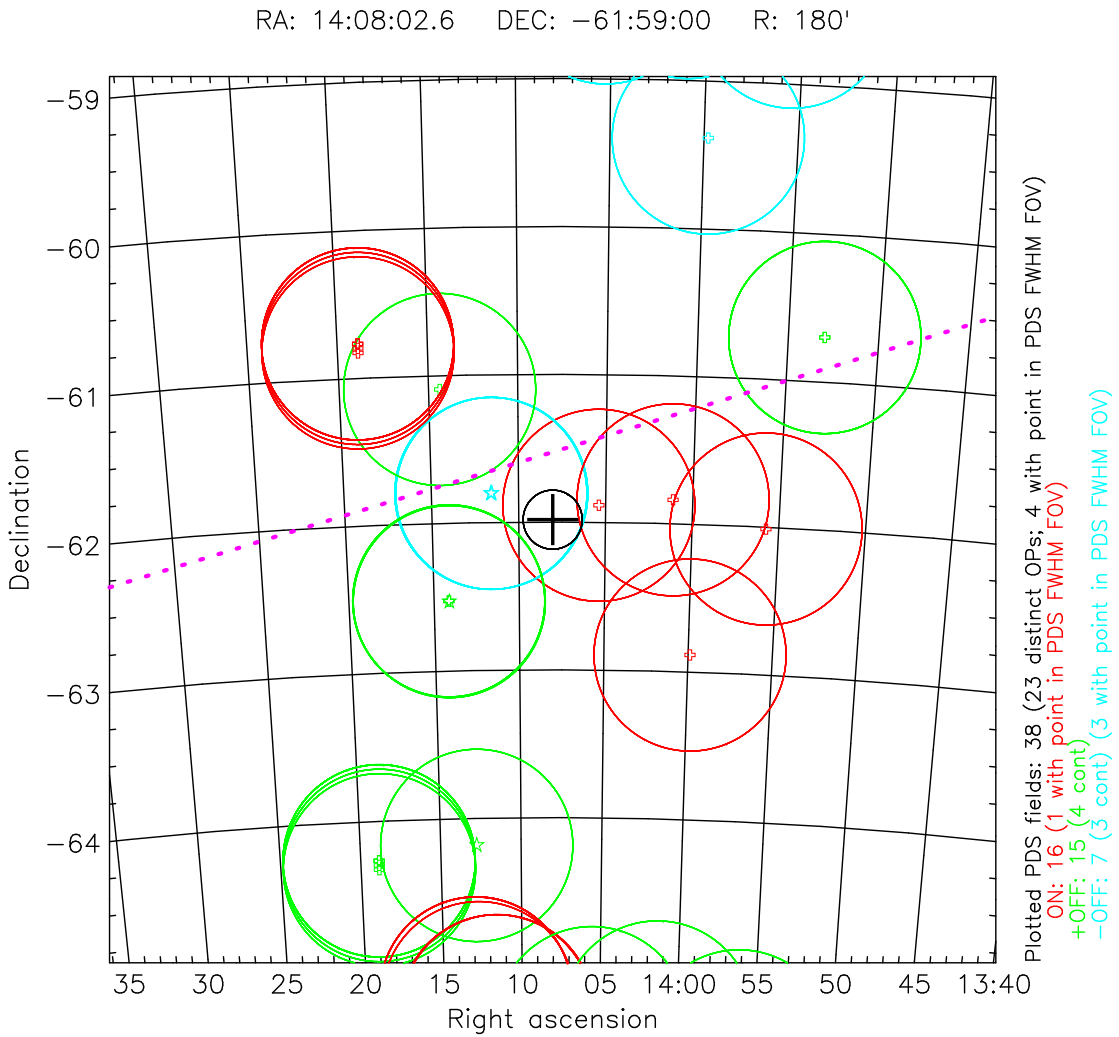


RXTE observation of Her X–1: Continuum (left) and CRF energy (right) vs Pulse Amplitude (Klochkov *et al.* 2011).

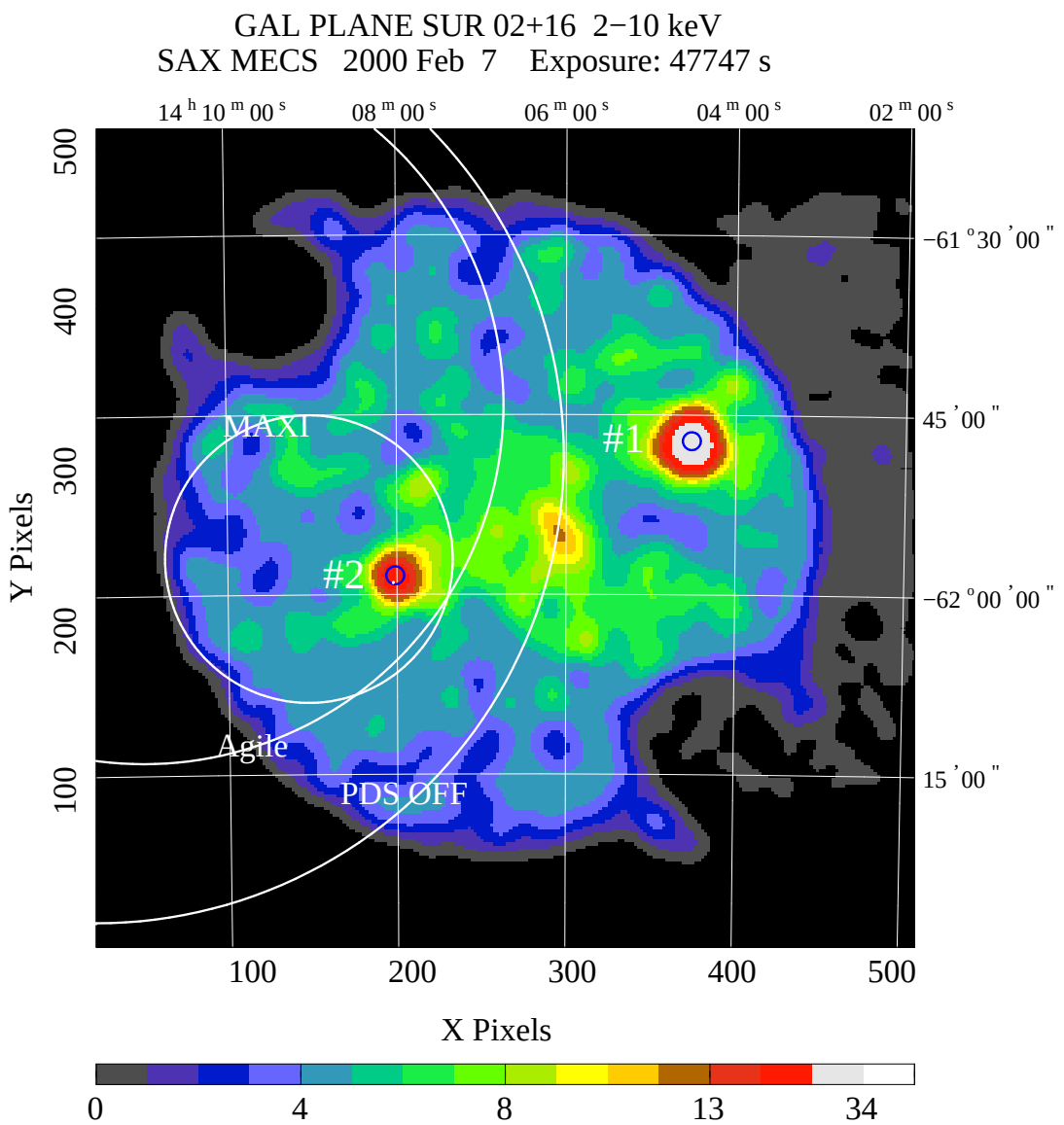
Warning: the correlation depends on the continuum adopted (see 4U 0115+63; Müller *et al.* 2013. A&A 551, A6).

Part II
A Real CRSF: the case of MAXI J1409−619

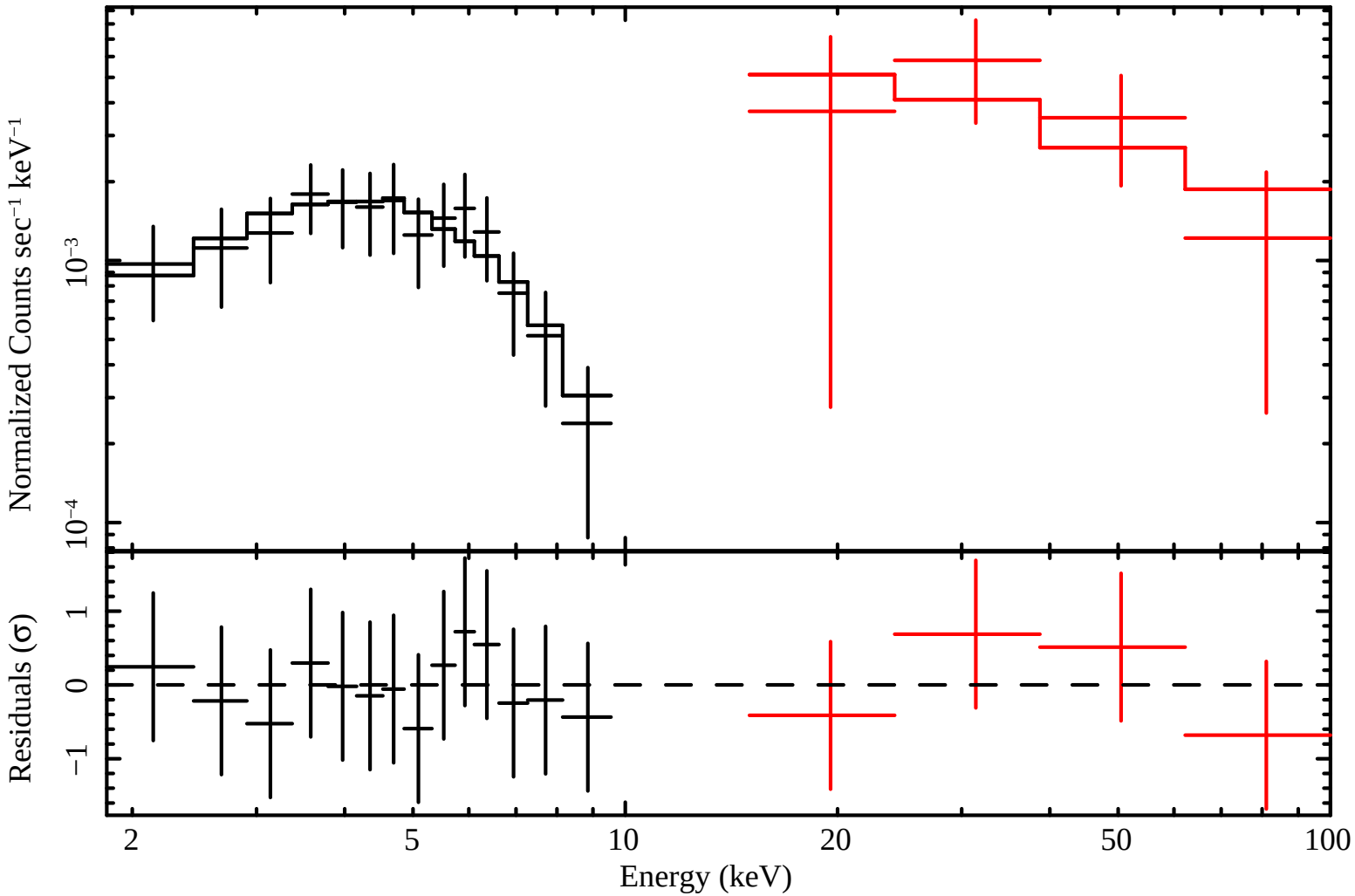
The PDS Spectrum



The PDS Spectrum

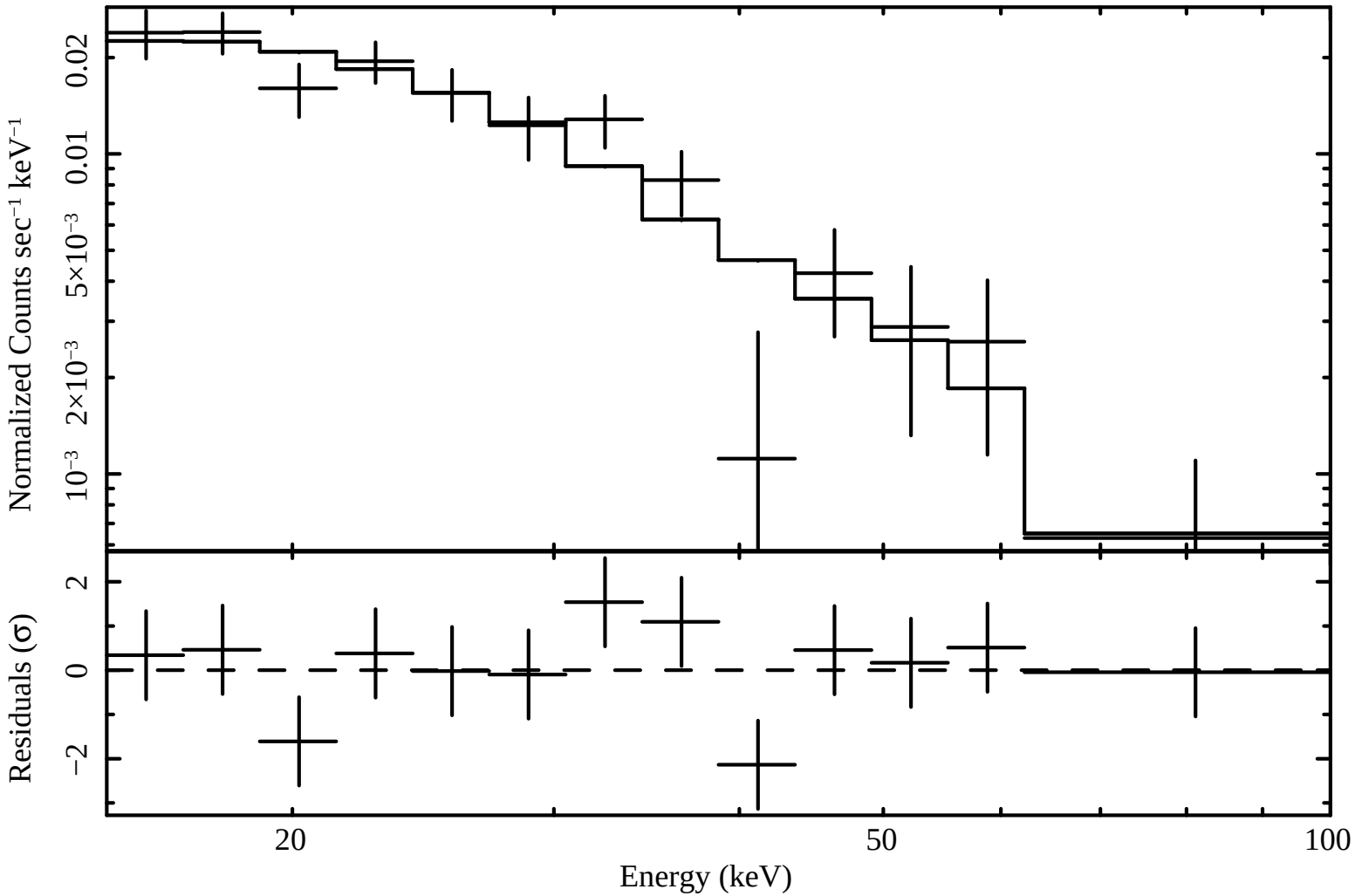


The PDS Spectrum



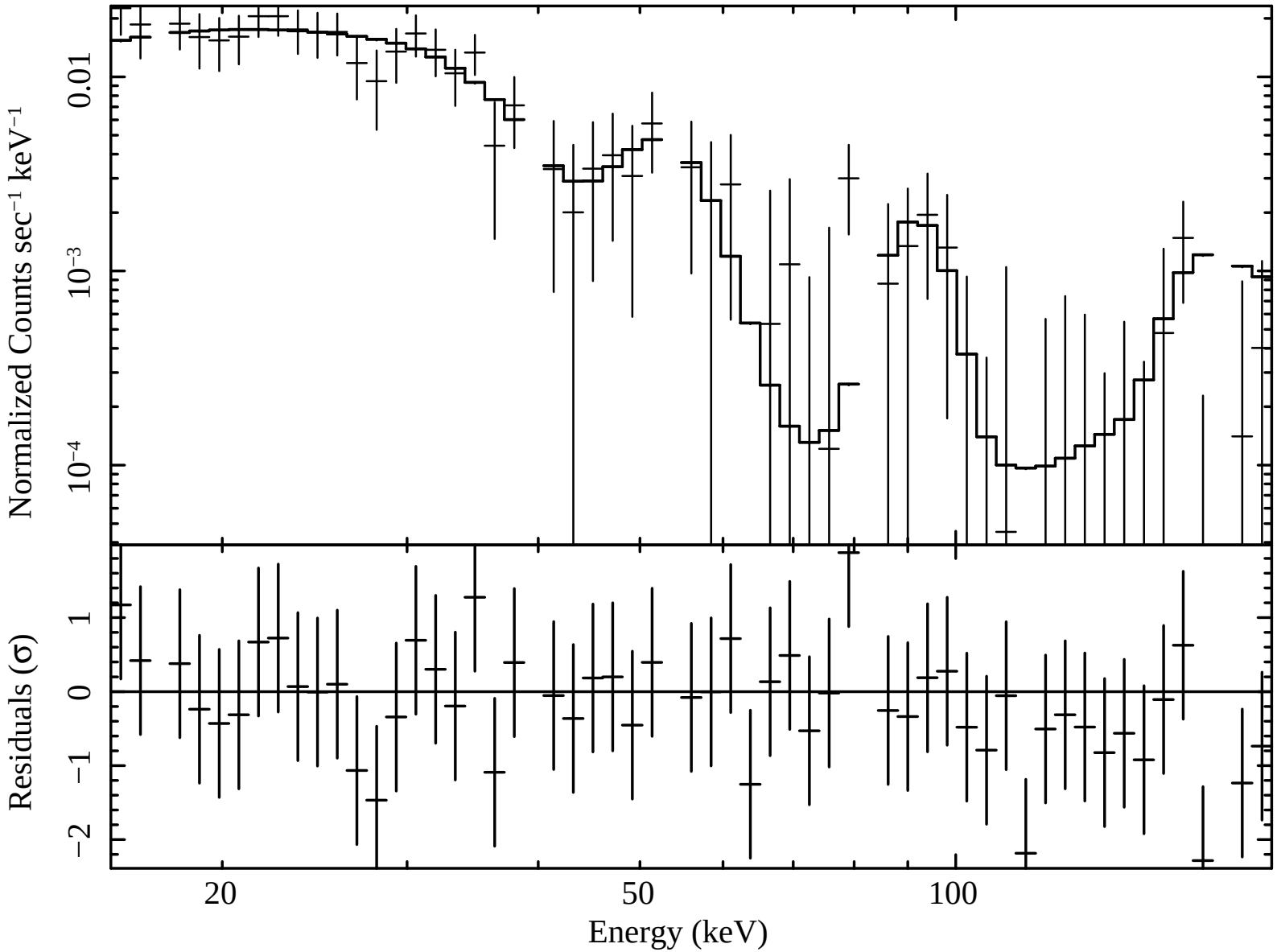
Power Law fit: $\alpha = 0.87^{+0.29}_{-0.19}$, $N_H = 2.8^{+3.4}_{-2.2} \times 10^{22}\,\mathrm{cm}^{-2}$, $\chi^2_\nu = 0.18$ for 15 dof.

The PDS Spectrum



Power Law fit: $\alpha = 2.2 \pm 0.3$, $\chi^2_\nu = 1.31$ for 11 dof.

The PDS Spectrum



Rule I: Be careful with the rebinning

- Customary to rebin data to have at least 20 counts in each bin.

Rule I: Be careful with the rebinning

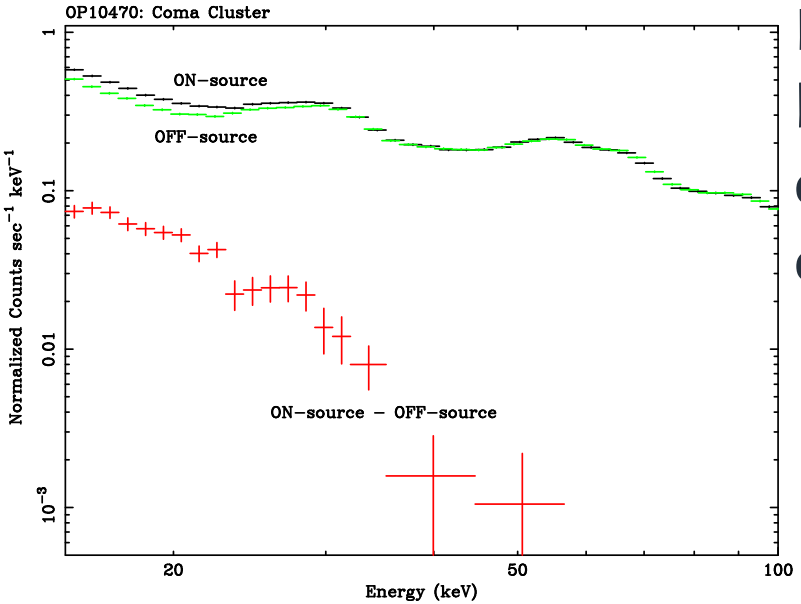
- Customary to rebin data to have at least 20 counts in each bin.
  Why? Because the χ^2 distribution with k dof is the distribution of a sum of the squares of k *independent* ***standard normal*** random variables.

Rule I: Be careful with the rebinning

- Customary to rebin data to have at least 20 counts in each bin.
- Correct for spectra coming from not-background dominated instruments.
  Not necessary when the net source spectrum is obtained by subtracting two high counts spectra.

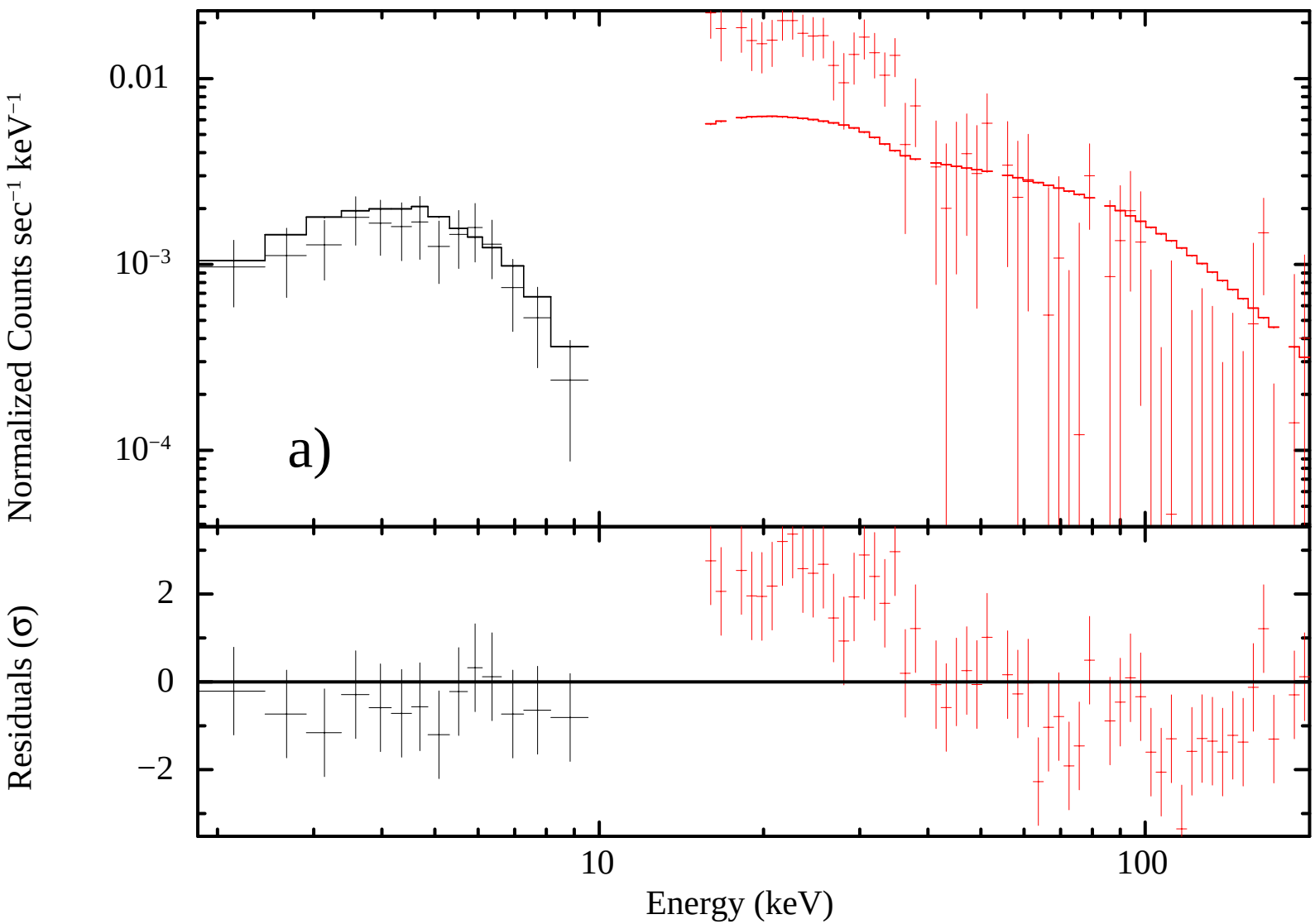
Rule I: Be careful with the rebinning

- Customary to rebin data to have at least 20 counts in each bin.
- Correct for spectra coming from not-background dominated instruments.
  Not necessary when the net source spectrum is obtained by subtracting two high counts spectra.



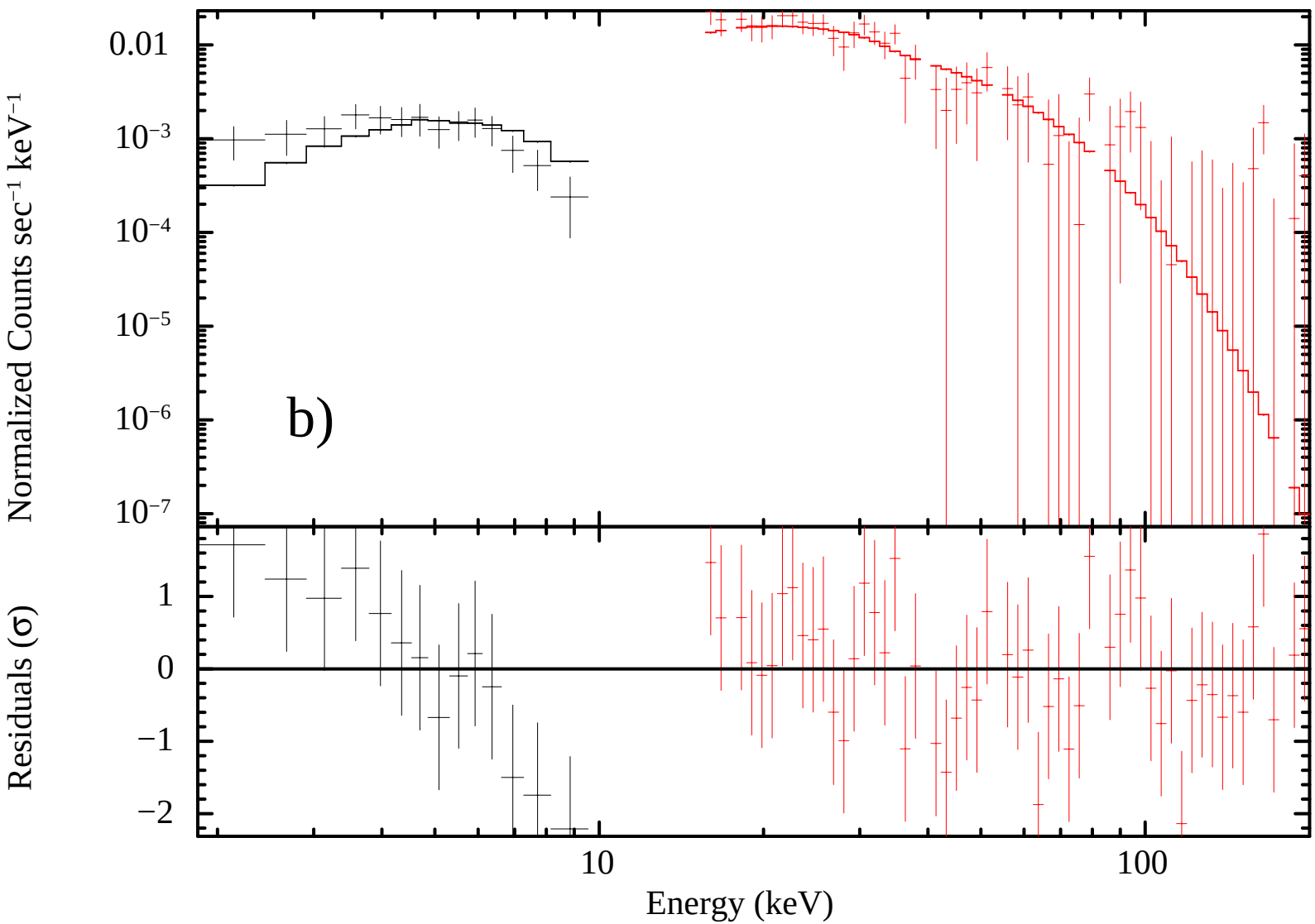
Normally distributed because they result from the difference of two normally distributed χ^2 variables.

The MECS+PDS Spectrum



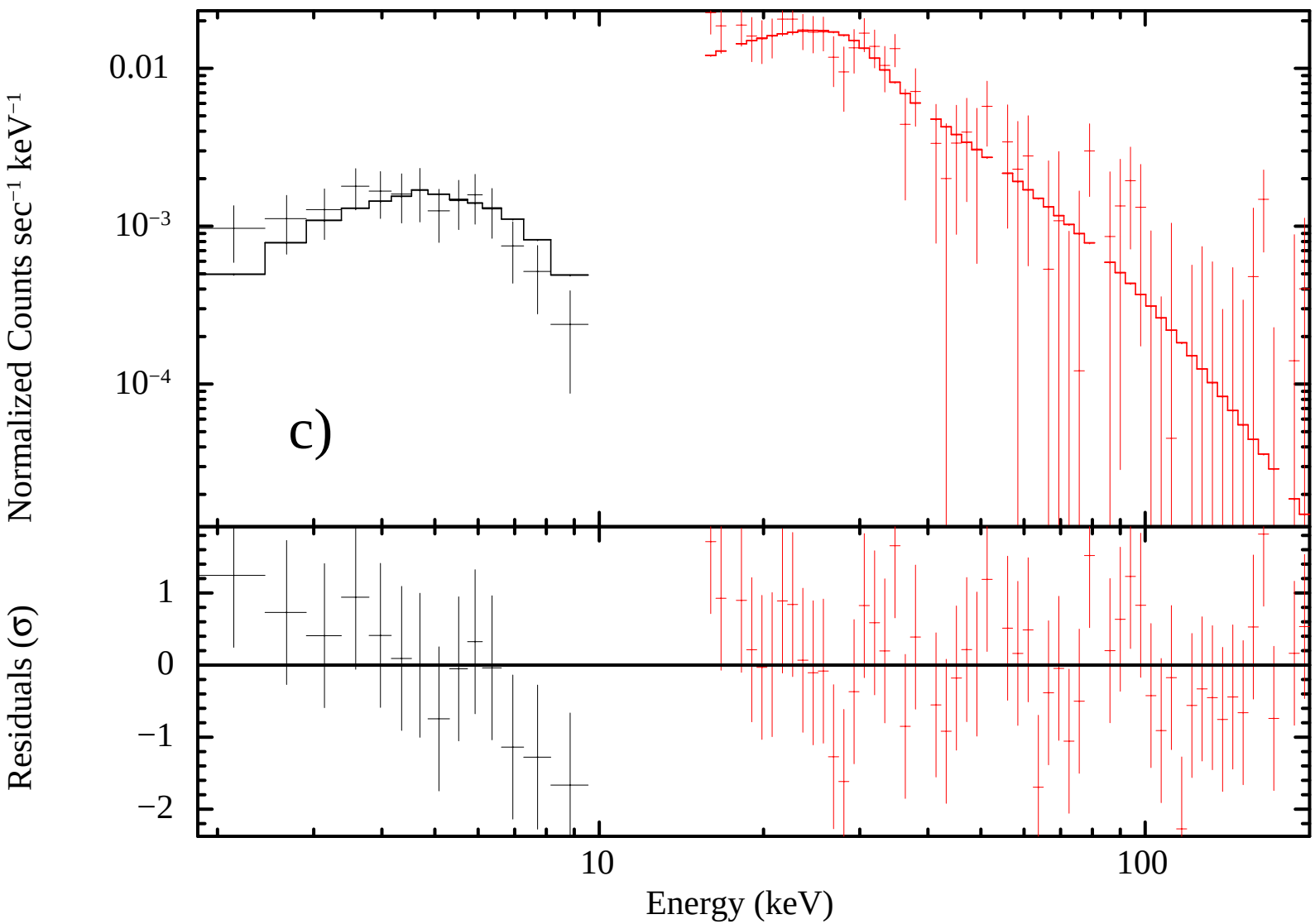
Power Law

The MECS+PDS Spectrum



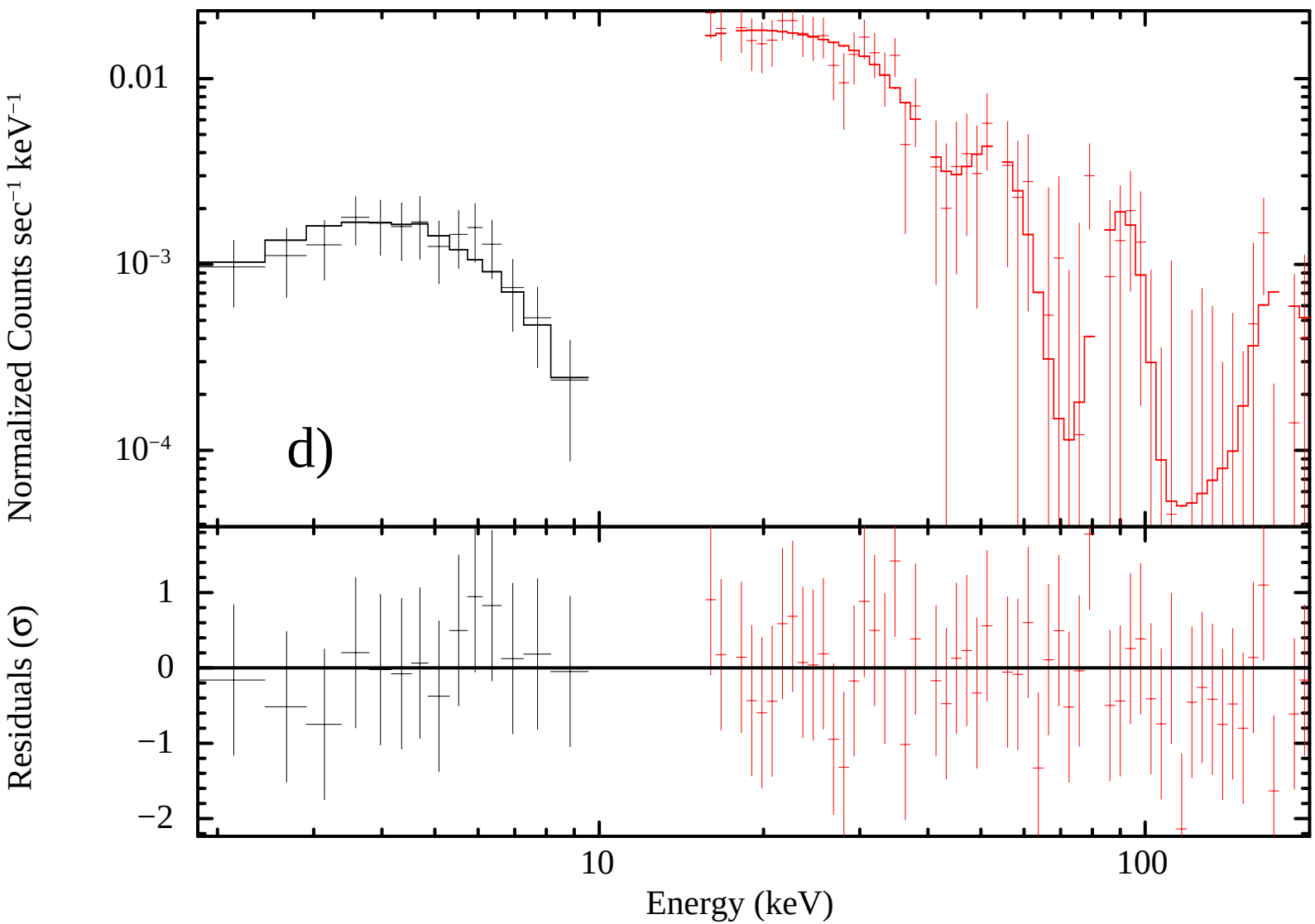
Cutoff Power Law

The MECS+PDS Spectrum



Broken Power Law

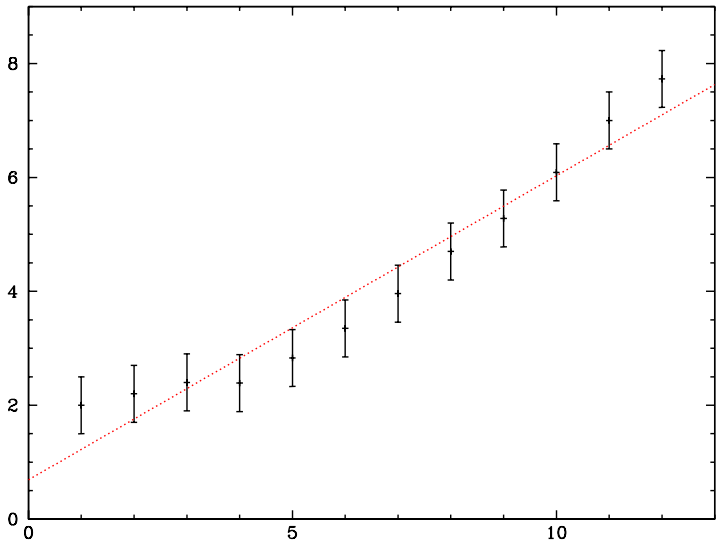
The MECS+PDS Spectrum



Power Law + 3 CRSFs

Other tests in support: the run test

The run test, also known as Wald–Wolfowitz test, works on the *signs* of the deviations, that is on the form of the residuals.

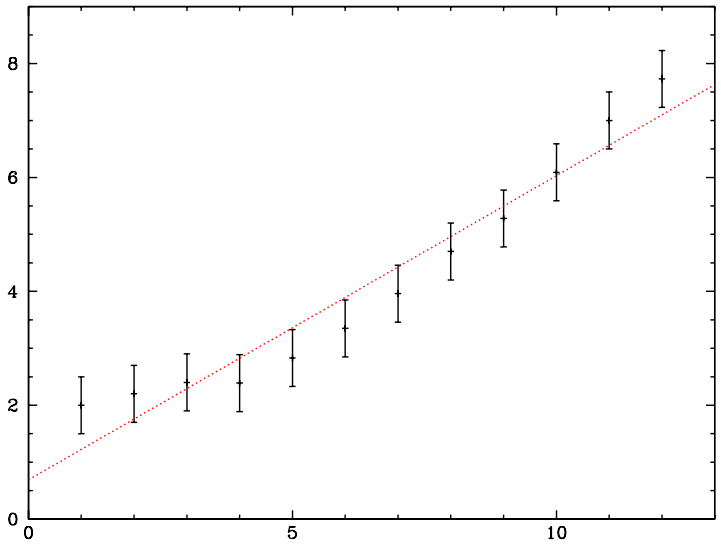


When fitting the 12 data points with a straight line the χ^2_ν is exactly 1 (likely due to error overestimation).

For a "good" fit we should expect that the number of points "above" the fitting line should not group together, but should be intermixed with points "below" the fitting line (and this should be more true as the number of data points increases).

Other tests in support: the run test

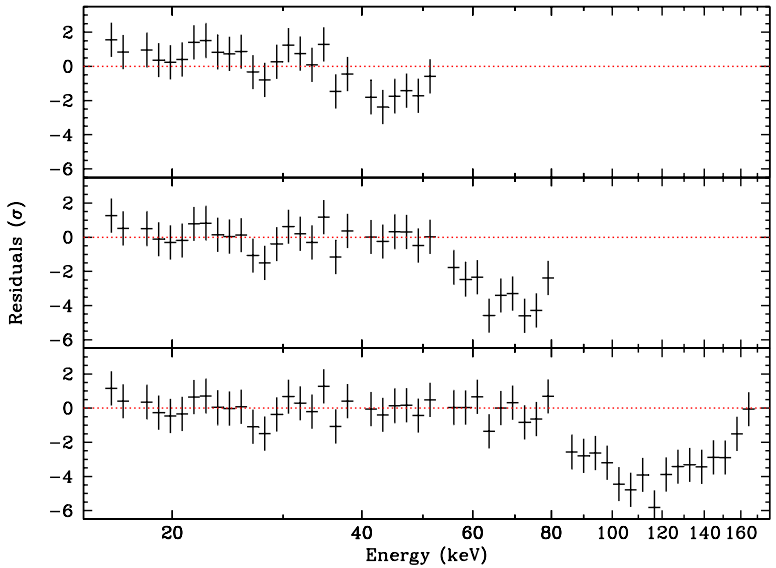
The run test, also known as Wald–Wolfowitz test, works on the *signs* of the deviations, that is on the form of the residuals.



We have 12 points, 6 points "above" the fitting line (let us call them N_+) and 6 points "below" the fitting line (N_-). The number of runs N_r is only 3, suspiciously small. Indeed the probability of obtaining $N_r \leq 3$ is 1.3%: the structure observed in the residuals is not due to random fluctuations.

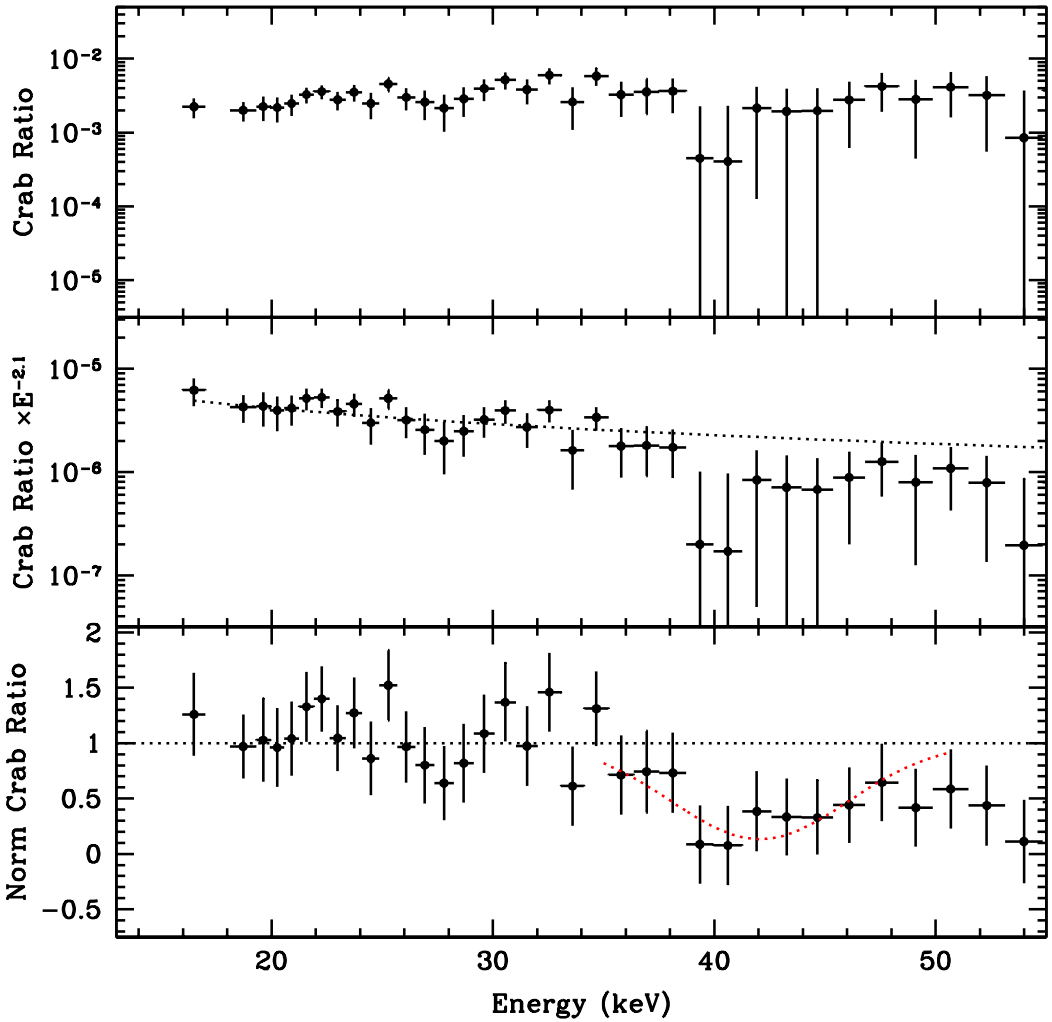
Other tests in support: the run test

The run test, also known as Wald–Wolfowitz test, works on the *signs* of the deviations, that is on the form of the residuals.

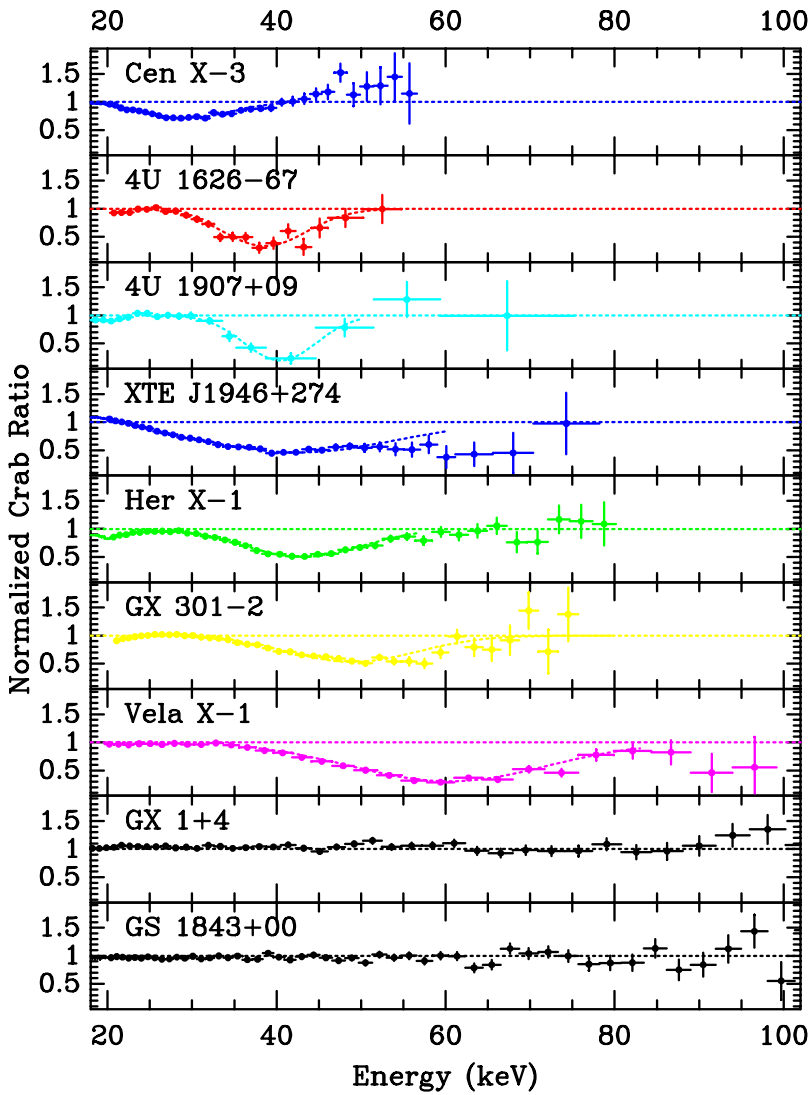


Up to 34 keV the residuals are consistent with random fluctuations at the 84%: $N_+ = 12; N_- = 9; N_r = 13.$ On the other hand, in the 34–50 keV band we see a clear structure in the residuals. In this case we have $N_+ = 2$, $N_- = 14, N_r = 2$, so we we can reject the null hypothesis of randomness at the 99% confidence level.

The Normalized Crab Ratio



The Normalized Crab Ratio



Rule II: Significance supported by other tests

The evaluation of the statistical significance of a CRSF must be supported by other tests:

- In the presence of structures in the fit residuals the run test is able to discriminate whether the structure is due to random fluctuations or not.

Rule II: Significance supported by other tests

The evaluation of the statistical significance of a CRSF must be supported by other tests:

- In the presence of structures in the fit residuals the run test is able to discriminate whether the structure is due to random fluctuations or not.
- A Crab Ratio can discriminate whether the feature is of instrumental origin or is intrinsic to the source.

Testing Statistical Hypothesis

1. Analyze the problem: identify the hypothesis and the alternative hypothesis;

Testing Statistical Hypothesis

1. Analyze the problem: identify the hypothesis and the alternative hypothesis;
2. Choose a test statistics;

Testing Statistical Hypothesis

1. Analyze the problem: identify the hypothesis and the alternative hypothesis;
2. Choose a test statistics;
3. Compute the test statistics;

Testing Statistical Hypothesis

1. Analyze the problem: identify the hypothesis and the alternative hypothesis;
2. Choose a test statistics;
3. Compute the test statistics;
4. Determine the frequency distribution of the test statistic under the hypothesis;

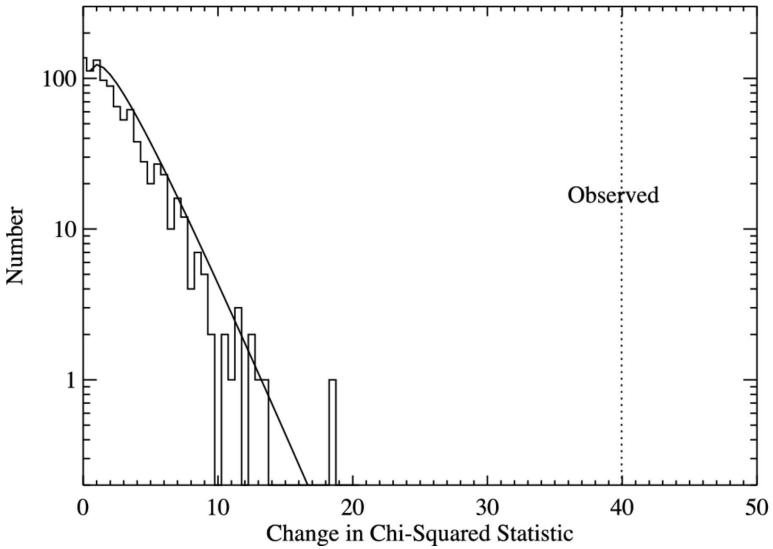
Testing Statistical Hypothesis

1. Analyze the problem: identify the hypothesis and the alternative hypothesis;
2. Choose a test statistics;
3. Compute the test statistics;
4. Determine the frequency distribution of the test statistic under the hypothesis;
5. Make a decision using the distribution as a guide.

Testing the Significance of a Line

The statistical assessment of a line present in a X–ray spectrum is completely different whether it is in *emission* or in *absorption*. The *null hypothesis* is the model without the line, to be tested against the continuum plus line model.

❑ An **emission** line



Because the model is ADDITIVE, we can use as test statistic the $\Delta\chi^2$. By comparing the $\Delta\chi^2$ obtained from the simulations with the one observed we can infer the PCI of the χ^2.

Testing the Significance of a Line

The statistical assessment of a line present in a X–ray spectrum is completely different whether it is in *emission* or in *absorption*.
The *null hypothesis* is the model without the line, to be tested against the continuum plus line model.

❑ An **emission** line

We can form a statistic F_χ which follow the F-distribution with

$$F_\chi = \frac{\Delta\chi^2}{\chi^2_\nu}$$

This is really a test of whether the coefficient of the new term is zero. We can evaluate the PCI from the values of $P(F_\chi, \nu_1, \nu_2)$.
This is the so-called (additive) *F–test*.

Testing the Significance of a Line

The statistical assessment of a line present in a X–ray spectrum is completely different whether it is in *emission* or in *absorption*. The *null hypothesis* is the model without the line, to be tested against the continuum plus line model.

❑ An **emission** line

Protassov *et al.* showed that we can obtain incorrect results if we are testing a hypothesis that is on the boundary of the parameter space. For an emission line we are in this case, and therefore the F–test should not be used, but simulations should.

Testing the Significance of a Line

The statistical assessment of a line present in a X–ray spectrum is completely different whether it is in *emission* or in *absorption*.

The *null hypothesis* is the model without the line, to be tested against the continuum plus line model.

❑ An **absorption** line

In `xspec` the absorption line that mimics a CRSF is modeled as a MULTIPLICATIVE component:

$$\mathrm{CYCLABS}(E) = \exp\left(-\frac{-\tau(WE/E_c)^2}{(E-E_c)^2+W^2}\right)$$

$$\mathrm{GABS}(E) = \exp\left[-\frac{\tau}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(E-E_c)^2}{2\sigma^2}\right)\right]$$

Testing the Significance of a Line

The statistical assessment of a line present in a X–ray spectrum is completely different whether it is in *emission* or in *absorption*.

The *null hypothesis* is the model without the line, to be tested against the continuum plus line model.

❑ An **absorption** line

Because the χ^2 variables are **NOT** independent, and therefore the additive property of the χ^2-distribution is not valid anymore.

Said in other words, the **SUM** of independent χ^2 variables is also χ^2 distributed, but this is not true for the **PRODUCT** of χ^2 variables.

Testing the Significance of a Line

The statistical assessment of a line present in a X–ray spectrum is completely different whether it is in *emission* or in *absorption*.

The *null hypothesis* is the model without the line, to be tested against the continuum plus line model.

❑ An **absorption** line

Therefore we *cannot* use the $\Delta\chi^2$ statistic as test statistic. We can use the F-statistic from the ratio of variances

$$F = \frac{\chi^2_{\nu 1}}{\chi^2_{\nu 2}}$$

This test is critically sensible to non-normality, therefore an Anderson-Darling test should be executed.

By comparing the F obtained from the simulations with the observed one we can determine the PCI of the χ^2.

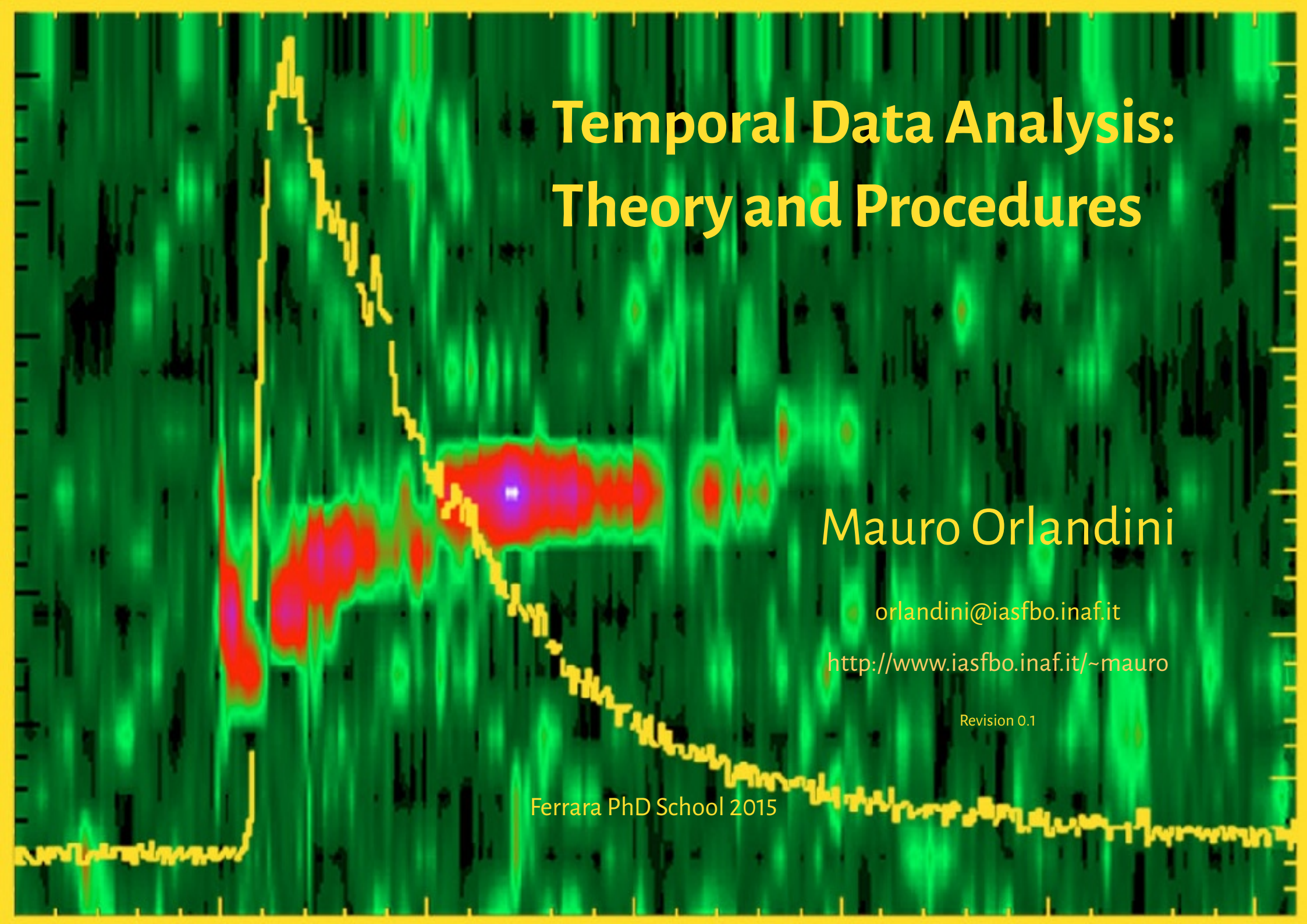
Rule III: Be careful when dealing with lines

- We must choose the right test statistic when assessing the significance of a line: while for an *emission* line is correct to use the $\Delta\chi^2$ statistic, for an *absorption* line (a multiplicative component in `xspec`) the appropriate statistic is the F–statistic.

See you tomorrow!

Tomorrow we will talk about ***Timing Analysis***...

so *estote parati!*



Temporal Data Analysis: Theory and Procedures

Mauro Orlandini

orlandini@iasfbo.inaf.it

http://www.iasfbo.inaf.it/~mauro

Revision 0.1

Ferrara PhD School 2015

Outline

Classification of Physical Data

Any observed data representing a physical phenomenon can be broadly classified as being either *deterministic* or *nondeterministic*. Deterministic data are those that can be described by an explicit mathematical relationship. There are many physical phenomena in practice which produce data that can be represented with reasonable accuracy by explicit mathematical relationships. For example, the motion of a satellite in orbit about the Earth, the potential across a condenser as it discharges through the resistor, the vibration response of an unbalanced rotating machine, or the temperature of water as heat is applied, are all basically deterministic. However, there are many other physical phenomena which produce data that are not deterministic. For example, the height of waves in a confused sea, the acoustic pressure generated by air rushing through a pipe, or the electrical output of a noise generator represent data which cannot be described by explicit mathematical relationships. There is no way to predict an exact value at a future instant of time. These data are random in character and must be described in terms of probability statements and statistical averages rather than explicit equations.

Various special classifications of deterministic and random data will now be discussed.

1.1 Deterministic Data

Data representing deterministic phenomena can be categorized as being either *periodic* or *non periodic*. Periodic data can be further categorized as being either *sinusoidal* or *complex* periodic. Non periodic data can be further categorized as being either *almost-periodic* or *transient*. These various classifications of deterministic data are schematically illustrated in Figure 1.1. Of course, any combination of these forms may also occur. For purposes of review, each of these types of deterministic data, along with physical examples, will be briefly discussed.

1.1.1 Sinusoidal Periodic Data

Sinusoidal data are those types of periodic data which can be defined mathematically by a time-varying function of the form

$$x(t) = X \sin(\omega_0 t + \varphi) \tag{1.1}$$

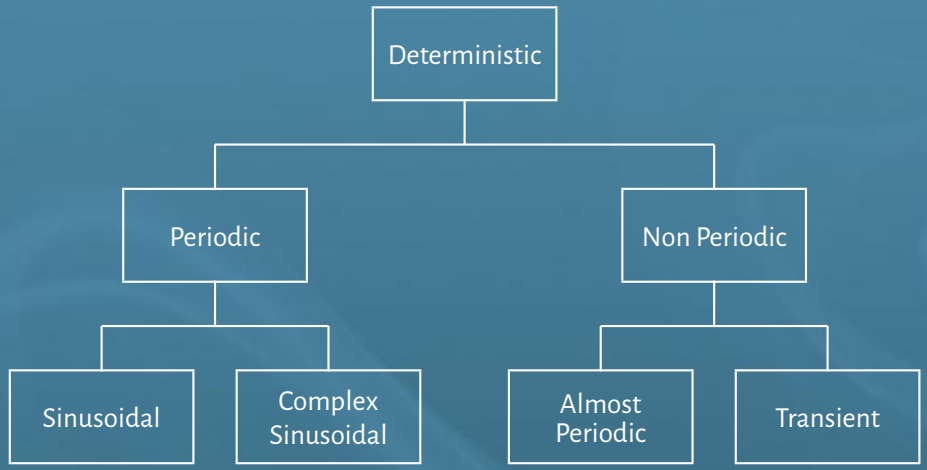


Figure 1.1: Classification of deterministic data

where X is the amplitude, ω_0 is the angular frequency, in units of radians per unit time[1], φ is the initial phase angle (in radians) with respect to the time origin, and $x(t)$ is the instantaneous value at time t. The sinusoidal time history described by (1.1) is usually referred as a sine wave. When analyzing sinusoidal data in practice, the phase angle φ is often ignored.

The time interval required for one full fluctuation or cycle of sinusoidal data is called the period T. The number of cycles per unit time is called the frequency ν.

[1] Not to be confused with the frequency ν, measured in Hz. The two are related by $\omega = 2\pi\nu$.

There are many example of physical phenomena which produce approximately sinusoidal data in practice. The voltage output of an electrical alternator is one example; the vibratory motion of an unbalanced rotating weight is another. Sinusoidal data represent one of the simplest forms of time-varying data from the analysis viewpoint.

1.1.2 Complex Periodic Data

Complex periodic data are those type of periodic data which can be defined mathematically by a time-varying function whose waveform exactly repeats itself at regular intervals such that

$$x(t) = x(t \pm nT) \qquad n = 1, 2, 3, \ldots \tag{1.2}$$

As for sinusoidal data, the time interval required for one full fluctuation is called the *period* T. The angular frequency is called the *fundamental frequency* ω. With few exceptions in practice, complex periodic data may be expanded into a Fourier series according to the following formula (we will return later in

greater detail on that)

$$x(t) = \sum_{k=0}^{\infty} (A_k \cos \omega_k t + B_k \sin \omega_k t) \tag{1.3}$$

with $\omega_k = 2\pi k/T$ and $B_0 = 0$. An alternative way to express the Fourier series is

$$x(t) = X_0 + \sum_{k=1}^{\infty} X_k \cos(\omega_k t + \varphi_k) \tag{1.4}$$

In other words, (1.4) says that complex periodic data consists of a static component and an infinite number of sinusoidal components called *harmonics*, which have amplitudes X_k and phases φ_k. The frequencies of the harmonic components are all integral multiples of ω_1.

Physical phenomena which produce complex periodic data are far more common than those which produce simple sinusoidal data. In fact, the classification of data as being sinusoidal is often only an approximation for data which are actually complex. For example, the voltage output from an electrical alternator may actually display, under careful inspection, some small contributions at higher harmonic

frequencies. In other cases, intense harmonic components may be present in periodic physical data.

1.1.3 Almost-Periodic Data

We have seen that periodic data can be generally reduced to a series of sine waves with commensurately related frequencies. Conversely, the data formed by summing two or more commensurately related sine waves will be periodic. However, the data formed by summing two or more sine waves with arbitrary frequencies will not be periodic.

More specifically, the sum of two or more sine waves will be periodic only when the ratios of all possible pairs of frequencies form rational numbers. This indicates that a fundamental period exists which will satisfy the requirements of (1.2). Hence,

$$x(t) = X_1 \sin(2t + \varphi_1) + X_2 \sin(3t + \varphi_2) + X_3 \sin(7t + \varphi_3)$$

is periodic since $2/3, 2/7$ and $3/7$ are rational numbers (the fundamental period is $T = 1$). On the other

hand,

$$x(t) = X_1 \sin(2t + \varphi_1) + X_2 \sin(3t + \varphi_2) + X_3 \sin(\sqrt{50}t + \varphi_3)$$

is not periodic since $2/\sqrt{50}$ and $3/\sqrt{50}$ are not rational numbers (in this case the fundamental period is infinitely long). The resulting time history in this case will have an "*almost periodic*" character, but the requirement of (1.2) will not be satisfied for any finite value of T.

Based on these discussions, almost periodic data are those types of non periodic data which can be defined mathematically by a time-varying function of the form

$$x(t) = \sum_{k=1}^{\infty} X_k \sin(\omega_k t + \varphi_k) \tag{1.5}$$

with $\omega_j/\omega_k \neq$ rational numbers in all cases. Physical phenomena producing almost periodic data frequently occur in practice when the effects of two or more unrelated periodic phenomena are mixed. A good example is the vibration response in a multiple engine propeller airplane when the engines are out of synchronization.

1.1.4 Transient Non periodic Data

Transient data are defined as all non periodic data other the almost-periodic discussed above. In other words, transient data include all data not previously discusses which can be described by some suitable time-varying function.

Physical phenomena which produce transient data are numerous and diverse. For example, the behavior of the temperature of water in a kettle (relative to room temperature) after the flame is turned off.

1.2 Random Data

Data representing a random physical phenomenon cannot be described by an explicit mathematical relationship because each observation of the phenomenon will be unique. In other words, any given observation will represent only one of the many possible results which might have occurred. For example, assume the output voltage from a thermal noise generator is recordered as a function of time. A

specific voltage time history record will be obtained. However, if a second thermal noise generator of identical construction and assembly is operated simultaneously, a different voltage time history record would result. In fact, every thermal noise generator which might be constructed would produce a different voltage time history record. Hence the voltage time history for any one generator is merely one example of an infinitely large number of time histories which might be occurred.

A single time history representing a random phenomenon is called a *sample function* (or a *sample record* when observed over a finite time interval). The collection of all possible sample functions which the random phenomenon might have produced is called a *random process* or a *stochastic process*. Hence a sample record of data for a random physical phenomenon may be though of as one physical realization of a random process.

Random processes might be categorized as being either *stationary* or *non stationary*. Stationary random processes may be further categorized as being either *ergodic* or *non ergodic*. Non stationary random processes may be further categorized in terms of specific types of non stationary properties. These various

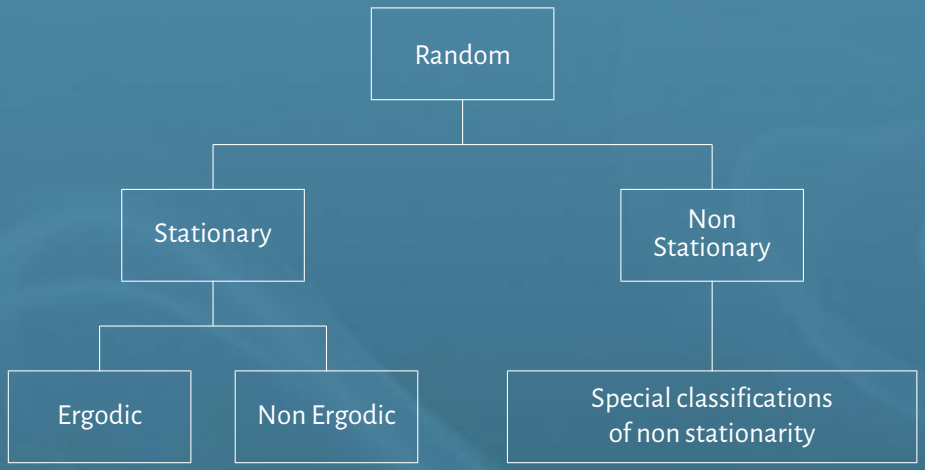


Figure 1.2: Classifications of random data

classifications of random processes are schematically illustrated in Figure 1.2. The meaning and physical significance of these various types of random processes will now be discussed in broad terms.

1.2.1 Stationary Random Processes

When a physical phenomenon is considered in terms of a random process, the properties of the phenomenon can hypothetically be described at any instant of time by computing average values over the collection of sample functions which describe the random process. For example, consider the collection

of sample functions (also called the *ensemble*) which form the random process illustrated in Figure 1.3. The *mean value* (first moment) of the random process at some time t_1 can be computed by taking the instantaneous value of each sample function of the ensemble at time t_1, summing the values, and dividing by the number of sample functions. In a similar manner, a correlation (joint moment) between the values of the random process at two different times (called *autocorrelation function*) can be computed by taking the ensemble average of the product of instantaneous values at two times, t_1 and $t_1 + \tau$. That is, for the random process $\{x(t)\}$, where the symbol $\{\ \}$ is used to denote an ensemble of sample functions, the mean value $\mu_x(t_1)$ and the autocorrelation function $R_x(t_1, t_1 + \tau)$ are given by

$$\mu_x(t_1) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} x_k(t_1) \tag{1.6a}$$

$$R_x(t_1, t_1 + \tau) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} x_k(t_1)\, x_k(t_1 + \tau) \tag{1.6b}$$

where the final summation assumes each sample function is equally likely.

For the general case where $\mu_x(t_1)$ and $R_x(t_1, t_1 + \tau)$ defined in (1.6) vary as time t_1 varies, the random process $\{x(t)\}$ is said to be *non stationary*. For the special case where $\mu_x(t_1)$ and $R_x(t_1, t_1+\tau)$ do not vary as time t_1 varies, the random process $\{x(t)\}$ is said to be *weakly stationary* or stationary in the wide sense. For the weakly stationary processes, the mean value is a constant and the autocorrelation function is dependent only upon the time of displacement τ. That is, $\mu_x(t_1) = \mu_x$ and $R_x(t_1, t_1 + \tau) = R_x(\tau)$.

An infinite collection of higher order moments and joint moments of the random process $\{x(t)\}$ could also be computed to establish a complete family of probability distribution functions describing the process. For the special case where all possible moments and joint moments are time invariant, the random process $\{x(t)\}$ is said to be *strongly stationary* or stationary in the strict sense. For many practical applications, verification of weak stationarity will justify an assumption of strong stationarity.

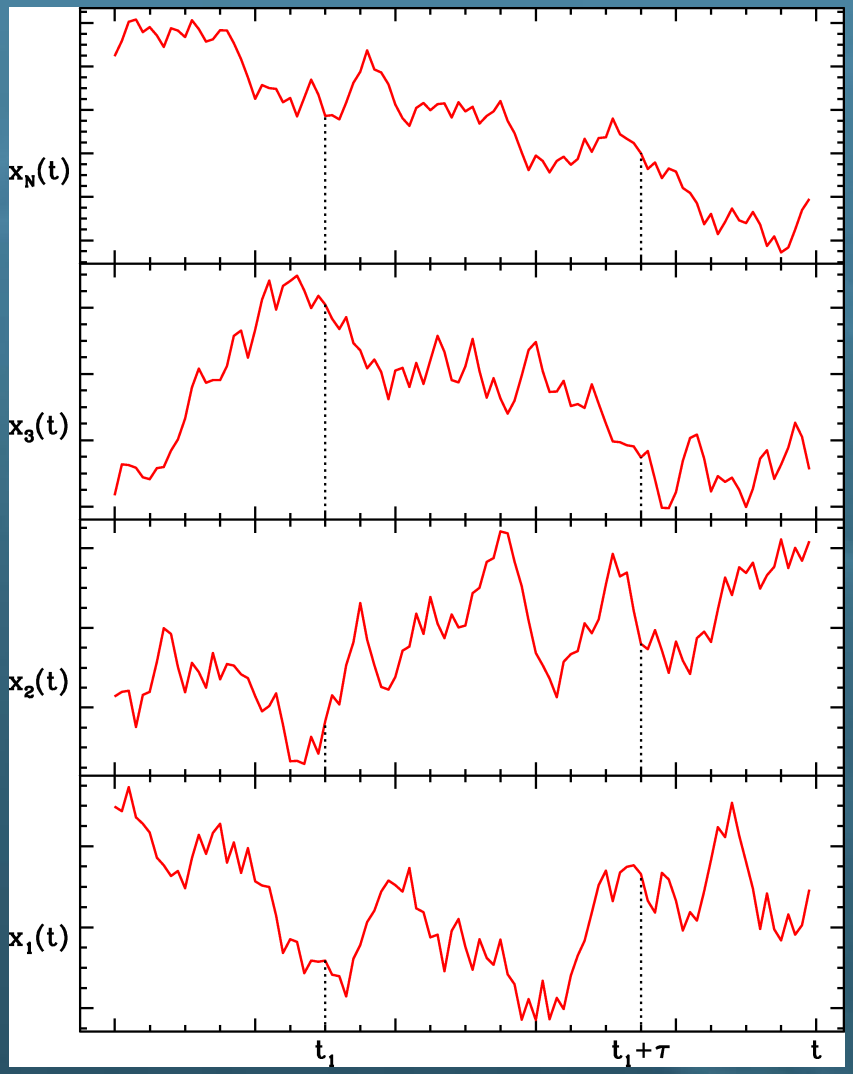


Figure 1.3: Ensemble of sample functions forming a random process

1.2.2 Ergodic Random Processes

The previous section discusses how the properties of a random process can be determined by computing enable averages at specific instants of time. In most cases, however, it is also possible to describe the properties of a stationary random process by computing time averages over specific sample functions in the ensemble. For example, consider the k-th sample function of the random process illustrated in Figure 1.3. The mean value $\mu_x(k)$ and the autocorrelation function $R_x(\tau, k)$ of the k-th sample function are given by

$$\mu_x(k) = \lim_{N\to\infty} \frac{1}{N} \int_0^T x_k(t)dt \tag{1.7a}$$

$$R_x(\tau, k) = \lim_{N\to\infty} \frac{1}{N} \int_0^T x_k(t)x_k(t+\tau)dt \tag{1.7b}$$

If the random process $\{x(t)\}$ is stationary, and $\mu_x(k)$ and $R_x(\tau, k)$ defined in (1.7) do not differ when computed over different sample functions, the random process is said to be *ergodic*. For ergodic ran-

dom processes, the time averaged mean value and autocorrelation function (as well as all other time-averaged properties) are equal to the corresponding ensemble averaged value. That is, $\mu_x(k) = \mu_x$ and $R_x(\tau, k) = R_x(\tau)$. Note that only stationary random process can be ergodic.

Ergodic random processes are clearly an important class of random processes since all processes of ergodic random processes can be determined by performing time averages over a single sample function. Fortunately, in practice, random data representing stationary physical phenomena are generally ergodic. It is for this reason that the properties of stationary random phenomena can be measured properly, in most cases, from a single observed time history record.

1.2.3 Non stationary Random Processes

Non stationary random processes include all random processes which do not meet the requirements for stationarity defined in the previous section. Unless further restrictions are imposed, the properties of non stationary random processes are generally time-varying functions which can be determined only

by performing instantaneous averages over the ensemble of sample functions forming the process. In practice, it is often not feasible to obtain a sufficient number of sample records to permit the accurate measurement of properties by ensemble averaging. This fact has tended to impede the development of practical techniques for measuring and analyzing non stationary random data.

In many cases, the non stationary random data produced by actual physical phenomena can be classified into special categories of non stationarity which simplify the measurement and analysis problem. For example, some type of random data might be described by a non stationary random process $\{y(t)\}$ where each sample function is given by $y(t) = A(t)\,x(t)$. Here $x(t)$ is a sample function from a stationary random process $\{x(t)\}$ and $A(t)$ is a deterministic multiplication factor. In other words, the data might be represented by a non stationary random process consisting of a sample functions with a common deterministic time trend. If non stationary random data fit a specific model of this type, ensemble averaging is not always needed to describe the data. The various desired properties can sometimes be estimated from a single record, as is true for ergodic stationary data.

Harmonic Analysis

Few preliminary remarks are in order: First, we will use the angular frequency ω when we refer to the frequency domain. The unit of the angular frequency is radians/second (or simpler s^{-1}). It is easily converted to the frequency ν (unit in Hz) using the following equation:

$$\omega = 2\pi\nu$$

Second: just let us remember the definition of even and odd functions.

Definition 2.1 (Even and odd functions). A function is said to be even if

$$f(-t) = f(t) \qquad \text{even function}$$

while a function is said to be odd if

$$f(-t) = -f(t) \qquad \text{odd function}$$

Any function can be described in terms of a mixture of even and odd functions, by means of the following

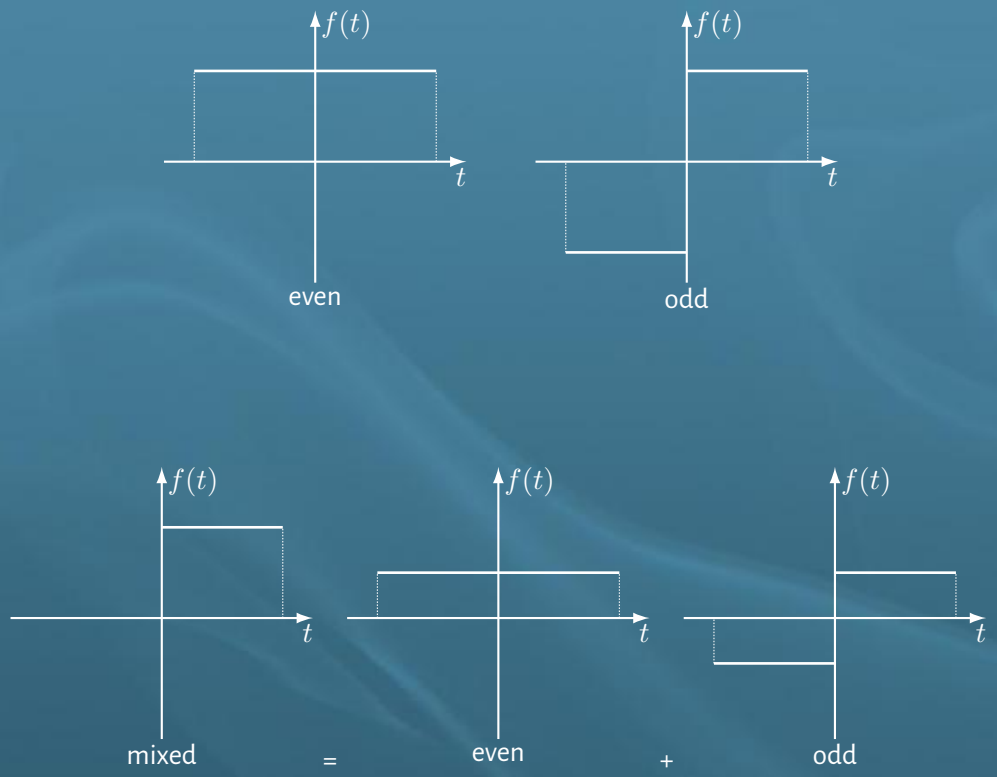


Figure 2.1: Examples of even, odd and mixed functions

decomposition (see Figure 2.1):

$$f_{\text{even}} = \frac{f(t) + f(-t)}{2}$$
$$f_{\text{odd}} = \frac{f(t) - f(-t)}{2}$$

2.1 Fourier Series

This Section will deal with the mapping of *periodic* functions to a series based on the trigonometric functions sine (and odd function) and cosine (even function).

2.1.1 Definition

Any periodic function $f(t)$ can be expanded into a series of trigonometric function, the so called Fourier series, as follows

Definition 2.2 (Fourier series).

$$f(t) = \sum_{k=0}^{\infty} (A_k \cos \omega_k t + B_k \sin \omega_k t) \qquad \text{with } \omega_k = \frac{2\pi k}{T},\ B_0 = 0 \tag{2.1}$$

T is the period of the function $f(t)$. The amplitudes or *Fourier coefficients* A_k and B_k are determined in such a way that the infinite series is identical with the function $f(t)$. Equation (2.1) therefore tells us

that any periodic function can be represented as a superposition of sine-function and cosine-function with appropriate amplitudes – with an infinite number of terms, if need be – yet using only precisely determined frequencies:

$$\omega = 0, \frac{2\pi}{T}, \frac{4\pi}{T}, \frac{6\pi}{T}, \dots$$

2.1.2 Calculation of the Fourier Coefficients

Before we compute the expressions of the Fourier coefficients, we need some tools. In all following integrals we integrate from $-T/2$ to $+T/2$, meaning over an interval with the period T that is symmetrical to $t = 0$. We could also pick any other interval, as long as the integrand is periodic with period T and gets integrated over a whole period. The letters n and m in the formulas below are natural numbers $0, 1, 2, \dots$ Let's have a look at the following

$$\int_{-T/2}^{+T/2} \cos \omega_n t \, dt = \begin{cases} 0 & \text{for } n \neq 0 \\ T & \text{for } n = 0 \end{cases} \tag{2.2}$$

$$\int_{-T/2}^{+T/2} \sin \omega_n t \, dt = 0 \tag{2.3}$$

This results from the fact that the areas on the positive half-plane and the ones on the negative one cancel out each other, provided we integrate over a whole number of periods. Cosine integral for $n = 0$ requires special treatment, as it lacks oscillations and therefore areas can't cancel out each other: there the integrand is 1, and the area under the horizontal line is equal to the width of the interval T. Furthermore, we need the following trigonometric identities:

$$\begin{aligned}
\cos\alpha\cos\beta &= \frac{1}{2}\left[\cos(\alpha-\beta)+\cos(\alpha+\beta)\right] \\
\sin\alpha\sin\beta &= \frac{1}{2}\left[\cos(\alpha-\beta)-\cos(\alpha+\beta)\right] \\
\sin\alpha\cos\beta &= \frac{1}{2}\left[\sin(\alpha-\beta)+\sin(\alpha+\beta)\right]
\end{aligned} \tag{2.4}$$

Using these identities we can demonstrate that the system of basis functions consisting of $(\sin\omega_k t,\ \cos\omega_k t)$ with $k = 0, 1, 2, \ldots$ is an *orthogonal system*. This means that

$$\int_{-T/2}^{+T/2} \cos\omega_n t \cos\omega_m t \, dt = \begin{cases} 0 & \text{for } n \neq m \\ T/2 & \text{for } n = m \neq 0 \\ T & \text{for } n = m = 0 \end{cases} \tag{2.5}$$

$$\int_{-T/2}^{+T/2} \sin\omega_n t \sin\omega_m t \, dt = \begin{cases} 0 & \text{for } n \neq m,\ n = 0,\ m = 0 \\ T/2 & \text{for } n = m \neq 0 \end{cases} \tag{2.6}$$

$$\int_{-T/2}^{+T/2} \sin\omega_n t \cos\omega_m t \, dt = 0 \tag{2.7}$$

Please note that our basis system is not an *orthonormal system*, i.e. the integrals for $n = m$ are not normalized to 1. What's even worse, the special case of $n = m = 0$ in (2.5) is a nuisance, and will keep bugging us again and again.

Using the above orthogonality relations, we are able to calculate the Fourier coefficients straight away. We need to multiply both sides of (2.1) by $\cos\omega_k t$ and integrate from $-T/2$ to $+T/2$. Due to the or-

thogonality, only terms with $k = k'$ will remain; the second integral will always disappear. This gives us:

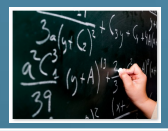
$$A_k = \frac{2}{T}\int_{-T/2}^{+T/2} f(t)\cos\omega_k t\,dt \qquad \text{for } k \neq 0 \tag{2.8}$$

$$A_0 = \frac{1}{T}\int_{-T/2}^{+T/2} f(t)\,dt \tag{2.9}$$

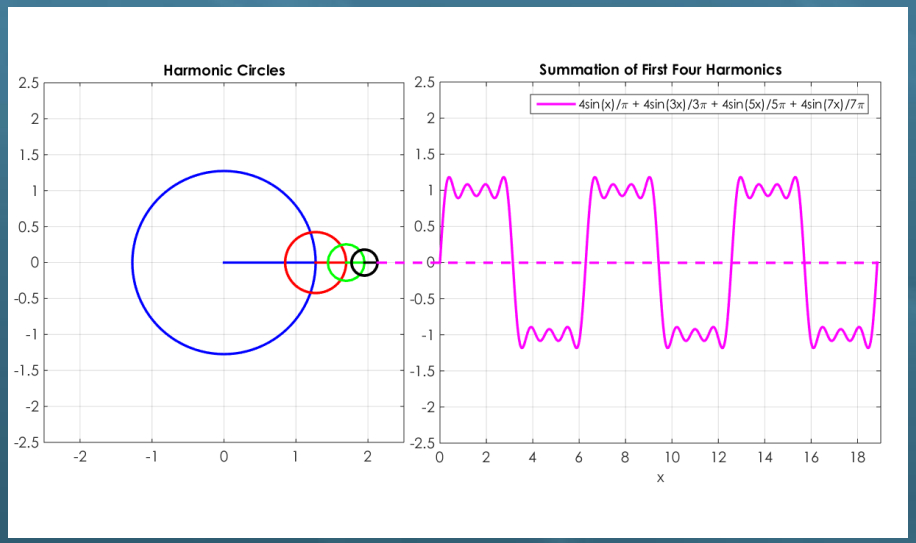
Please note the prefactors $2/T$ or $1/T$, respectively, in (2.8) and (2.9). Equation (2.9) simply is the average of the function $f(t)$. Now let's multiply both sides of (2.1) by $\sin\omega_k t$ and integrate from $-T/2$ to $+T/2$. We now have:

$$B_k = \frac{2}{T}\int_{-T/2}^{+T/2} f(t)\sin\omega_k t\,dt \qquad \text{for all } k \tag{2.10}$$

Equations (2.8) and (2.10) may also be interpreted like: by weighting the function $f(t)$ with $\cos\omega_k t$ or $\sin\omega_k t$, respectively, we “pick” the spectral components from $f(t)$, when integrating, corresponding to the even or odd components, respectively, of the frequency ω_k.



Ex. 2.1 *Calculation of Fourier coefficients: Constant and triangular functions*



Run the Fourier applet

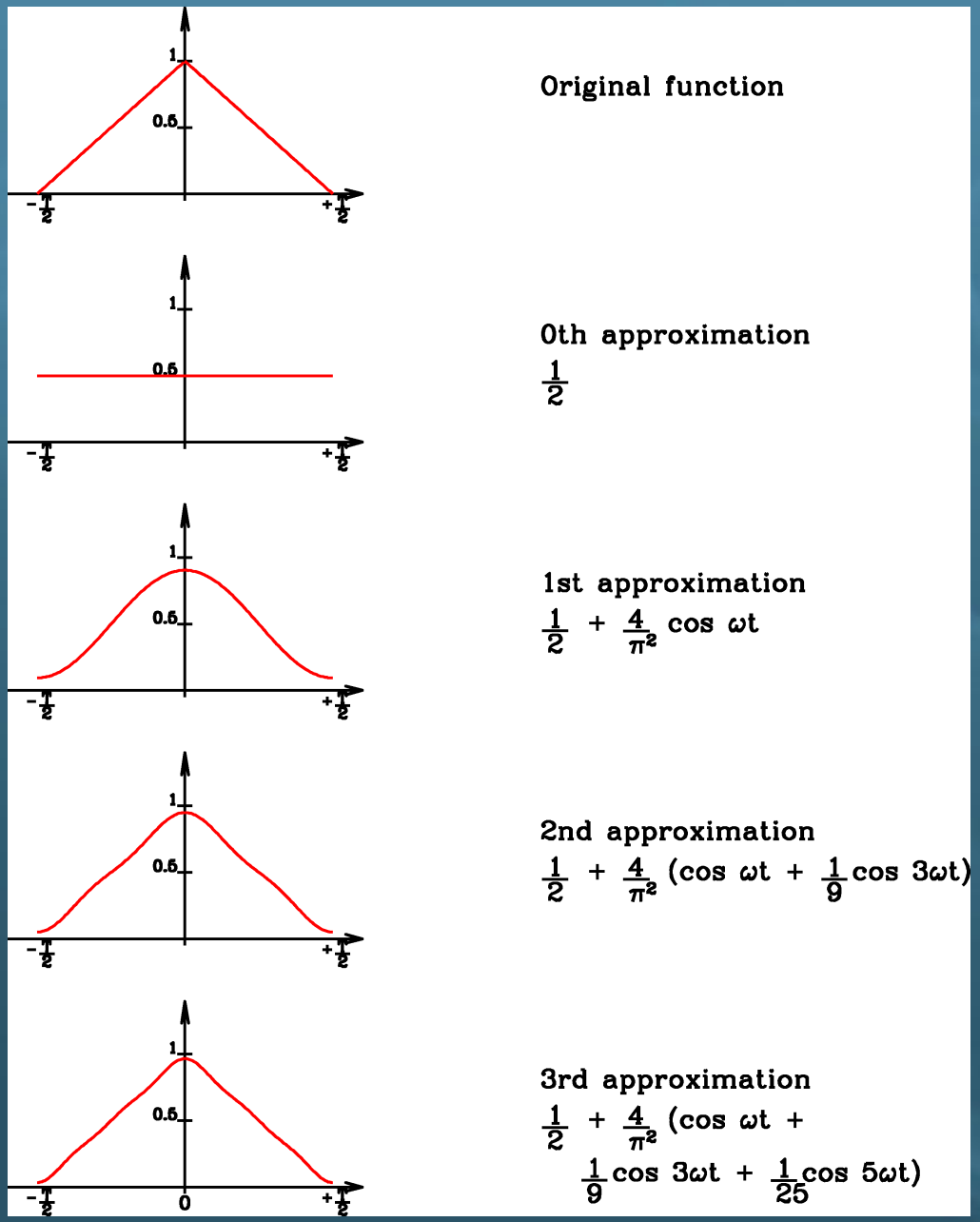


Figure 2.2: The triangular function and consecutive approximations by a Fourier series with more and more terms

2.1.3 Fourier Series and Music

While listening to music, we are able to clearly distinguish the sound produced by different instruments. The sound coming from a flute is quite different from the sound coming from a violin, even if they play the same note[2].

In musical terms, this difference is called ***timbre*** and it was von Helmholtz, in the second half of the XIX century, who understood that (von Helmholtz H. 1885. *"On the sensations of tone as a physiological basis for the theory of music"*)

> *"Each vibratory motion of the air in the ear canal, corresponding to a musical sound, can always be uniquely regarded as the sum of a number of vibratory movements."*

or, in mathematical terms, the timbre can be easily explained in terms of Fourier decomposition of the signal.

[2] In general terms, we call ***pitch*** of a sound the "perceived" frequency of a musical note, and it is related to amount of Fourier frequencies we (that is, our ears) are able to distinguish.

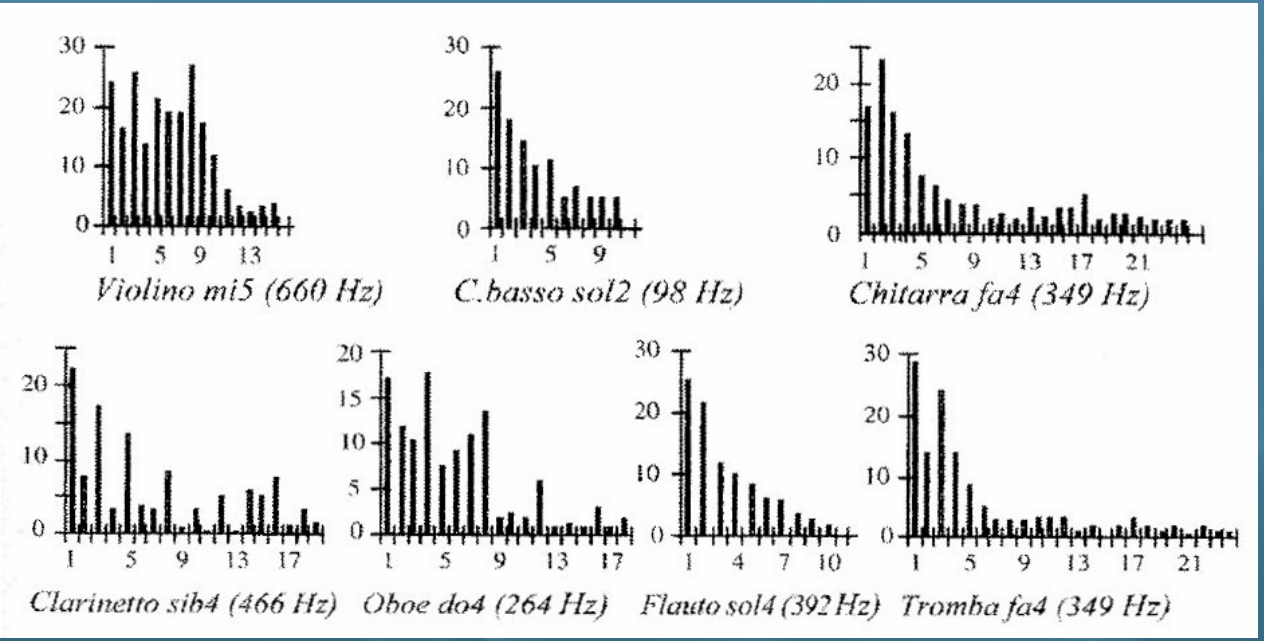


Figure 2.3: Comparison of a Fourier decomposition of a musical signal played by different instruments. On the abscissa we list the Fourier frequency index k, while the Y axis shows the power, in dB, emitted in the single harmonics (from Olson H.F., *"Music, Physics and Engineering"*).

Indeed, if we apply Eq. (2.1) and extract the Fourier coefficients for various instruments we will obtain something like shown in Figure 2.3: it is evident that the harmonic content of different instruments is quite different.

For example, for the violin we have that the first Fourier frequencies are quite intense (and the brilliance of the violin sound is due the fact that these harmonics peak in the region where our ear is more sensi-

tive). On the other hand, for the clarinet (here in the *chalumeau* registry) the even harmonics are quite faint, giving raise to its characteristic "hollow" sound. The typical metallic sound of the trumpet is due to the presence of very high harmonics, beyond the 21st.

It is interesting to observe that the harmonic content is different not only for different instruments playing the same note, but also for the same note played by the same instrument (the *La* played by a violin on the *La* string (not fingered) and the *La* played on the *Re* string (fingered)), or the same note played in different octaves. As an example, in Figure 2.4 we show the harmonic content of all the *Do*'s in the piano. Note how the lack of harmonics in the higher *Do*'s, like *Do6* and *Do7*, makes them an almost "pure sound". Furthermore, for the *Do1* we can notice that the fundamental frequency and the lower harmonics are fainter than the harmonics between the 10th and the 15th. Despite this, our ears recognize the sound as a *Do1*. This phenomenon is called "*virtual pitch*", and it is the demonstration that our brain is a "Fourier analyzer".

Our brain is therefore able to decompose an acoustic signal in its Fourier components, and is able to

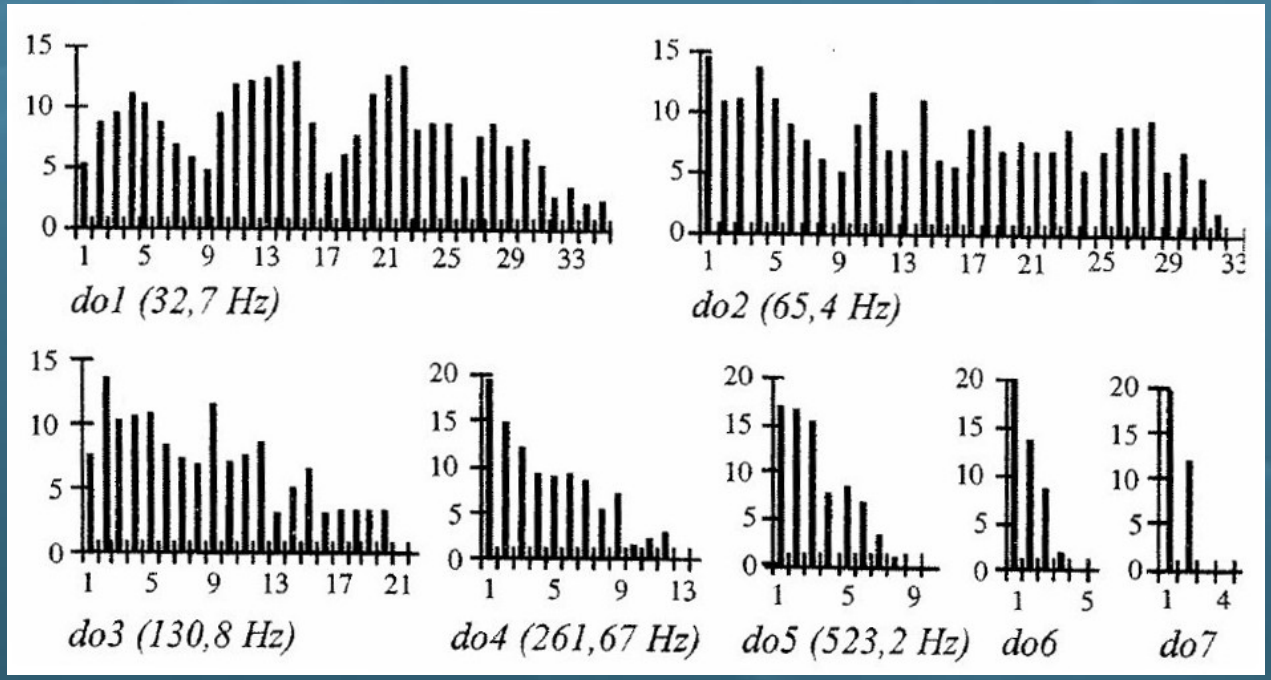


Figure 2.4: Comparison of Fourier decomposition of different *Do*'s played by a piano, from the left to the right of the keyboard. Note that for the *Do1* the fundamental and the lower harmonics are fainter than the higher ones, therefore the pitch is somewhat "virtual", while the lack of harmonics in *Do6* and *Do7* makes them an almost pure sound.

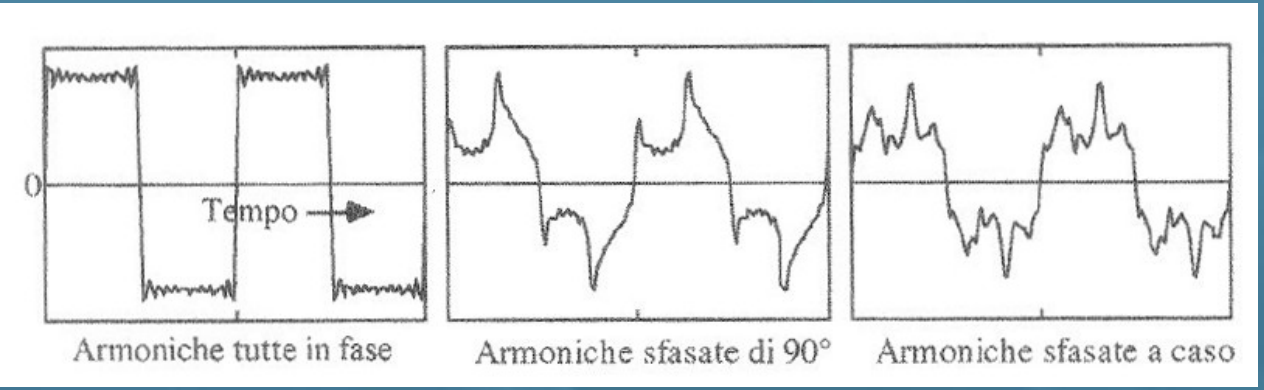


Figure 2.5: Effects of phase shifts among harmonics in a complex signal: *Left:* a square wave obtained by summing the first 21 harmonics all in phase among each other. *Central:* harmonics shifted by $\pi/2$. *Right:* random shift.

perceive each of them, independently of their phase relationships. In this perspective, the sentence that Leibniz' wrote in a letter to Christian Goldbach on April 17, 1712 was prophetic: "*Musica est exercitium arithmeticæ occultum nescientis se numerare animi*"[3].

The fact that the human brain is not able to perceive phase differences among harmonics is very important: indeed, the timbre of an instrument would change during the emission of a sound because of the different velocities of the harmonics along a string. To illustrate this phenomenon, in Figure 2.5 we show a square wave obtained by summing the first 21 harmonics, all in phase among each other. In the

[3] *Music is a hidden arithmetic exercise of the soul, which does not know that it is counting.*

central panel we show the shape of the wave obtained by introducing a phase shift of $\pi/2$, while in the right panel the shift is random. While the wave shapes are completely different, if the signals are sent to a loudspeaker they are indistinguishable to the human ear.

2.1.4 Fourier Series in Complex Notation

In (2.1) the index k starts from 0, meaning that we will rule out *negative* frequencies in our Fourier series. The cosine terms didn't have a problem with negative frequencies. The sign of the cosine argument doesn't matter anyway, so we would be able to go halves as far as the spectral intensity at the positive frequency $k\omega$ was concerned: $-k\omega$ and $k\omega$ would get equal parts, as shown in Figure 2.6. As frequency $\omega = 0$ (a frequency as good as any other frequency $\omega \neq 0$) has no "brother", it will not have to go halves. A change of sign for the sine-terms arguments would result in a change of sign for the corresponding series term. The splitting of spectral intensity like "between brothers" (equal parts of $-\omega_k$ and $+\omega_k$ now will have to be like "between sisters": the sister for $-\omega_k$ also gets 50%, but hers is *minus* 50%!

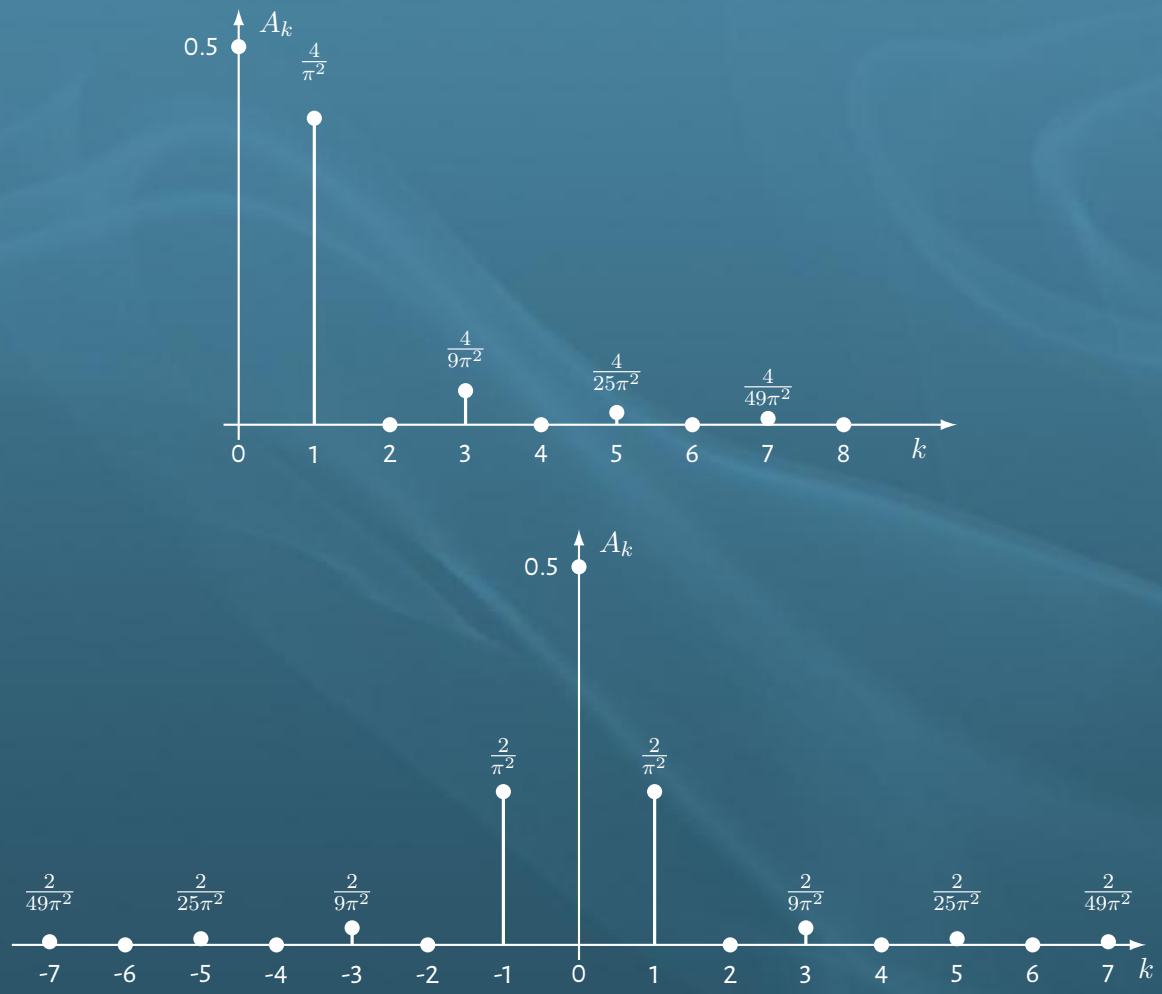


Figure 2.6: Plot of the "triangular function" Fourier frequencies: *Top:* Only positive frequencies; *Bottom:* Positive and negative frequencies.

Instead of using (2.1) we might as well use:

$$f(t) = \sum_{k=-\infty}^{+\infty} (A'_k \cos \omega_k t + B'_k \sin \omega_k t) \tag{2.11}$$

where, of course, the following is true: $A'_{-k} = A'_k$, $B'_{-k} = -B'_k$. The formulas for the computation of A'_k and B'_k for $k > 0$ are identical to (2.8) and (2.10), though the lack the extra factor 2. Equation (2.9) for A_0 stays unaffected by this. This helps us avoid to provide a special treatment for the constant term. Now we're set and ready for the introduction of complex notation. In the following we'll always assume that $f(t)$ is a real function. Generalizing this for complex $f(t)$ is no problem. Our most important tool is Euler identity:

$$e^{i\alpha t} = \cos \alpha t + i \sin \alpha t \tag{2.12}$$

where i is the imaginary unit ($i^2 = -1$). This allows us to rewrite the trigonometric functions as

$$\cos \alpha t = \frac{1}{2}(e^{i\alpha t} + e^{-i\alpha t})$$
$$\sin \alpha t = \frac{1}{2i}(e^{i\alpha t} - e^{-i\alpha t}) \tag{2.13}$$

Inserting these relations into (2.1) we obtain

$$f(t) = A_0 + \sum_{k=1}^{\infty} \left(\frac{A_k - iB_k}{2} e^{i\omega_k t} + \frac{A_k + iB_k}{2} e^{-i\omega_k t} \right) \tag{2.14}$$

If we define

$$\begin{aligned} C_0 &= A_0 \\ C_k &= \frac{A_k - iB_k}{2} \\ C_{-k} &= \frac{A_k + iB_k}{2}, \qquad k = 1, 2, 3, \ldots \end{aligned} \tag{2.15}$$

we finally get

$$f(t) = \sum_{k=-\infty}^{\infty} C_k e^{i\omega_k t} \qquad \omega_k = \frac{2\pi k}{T} \tag{2.16}$$

Now C_k can be formulated in general terms as

$$C_k = \frac{1}{T} \int_{-T/2}^{+T/2} f(t)\, e^{-i\omega_k t}\, dt \qquad \text{for } k = 0, \pm 1, \pm 2, \ldots \tag{2.17}$$

Please note that there is a negative sign in the exponent. Please also note that the index k runs from $-\infty$ to $+\infty$ for C_k, whereas it runs from 0 to $+\infty$ for A_k and B_k.

2.1.5 Partial Sums, Parseval Equation

For practical work, infinite Fourier series have to get terminated at some stage, regardless. Therefore, we only use a partial sum, say until we reach $k_{\text{max}} = N$. This Nth partial sum then is:

$$S_N = \sum_{k=0}^{N} (A_k \cos \omega_k t + B_k \sin \omega_k t) \tag{2.18}$$

Terminating the series results in the following squared error:

$$\delta_N^2 = \frac{1}{T}\int_T \left[f(t) - S_N(t)\right]^2 \, dt \tag{2.19}$$

The T below the integral symbol means integration over a full period. This definition will become plausible in a second if we look at the discrete version:

$$\delta_N^2 = \frac{1}{T}\sum_{j=0}^{N}(f_j - s_j)^2 \tag{2.20}$$

Please note that we divide by the length of the interval, to compensate for integrating over the interval T. Now we know that the following is correct for the infinite series:

$$\lim_{N\to\infty} S_N = \sum_{k=0}^{\infty}(A_k \cos\omega_k t + B_k \sin\omega_k t) \tag{2.21}$$

provided the A_k and B_k happen to be the Fourier coefficients. Does this also have to be true for the Nth partial sum? Isn't there a chance the mean squared error would get smaller, if we used other coefficients

instead of Fourier coefficients? That's not the case! To prove it, we'll now insert (2.18) and (2.19) in (2.21), leave out $\lim_{N\to\infty}$ and get:

$$
\begin{aligned}
\delta_N^2 &= \frac{1}{T}\left\{\int_T f^2(t)\,dt - 2\int_T f(t)S_N(t)\,dt + \int_T S_N^2(t)\,dt\right\} \\
&= \frac{1}{T}\left\{\int_T f^2(t)\,dt\right. \\
&\quad -2\int_T \sum_{k=0}^{\infty}(A_k\cos\omega_k t + B_k\sin\omega_k t)\sum_{k=0}^{N}(A_k\cos\omega_k t + B_k\sin\omega_k t)\,dt \\
&\quad \left.+\int_T \sum_{k=0}^{N}(A_k\cos\omega_k t + B_k\sin\omega_k t)\sum_{j=0}^{N}(A_j\cos\omega_j t + B_j\sin\omega_j t)\,dt\right\} \\
&= \frac{1}{T}\left\{\int_T f^2(t)\,dt - 2TA_0^2 - 2\frac{T}{2}\sum_{k=1}^{N}(A_k^2+B_k^2) + TA_0^2 + \frac{T}{2}\sum_{k=1}^{N}(A_k^2+B_k^2)\right\} \\
&= \frac{1}{T}\int_T f^2(t)\,dt - A_0^2 - \frac{1}{2}\sum_{k=1}^{N}(A_k^2+B_k^2)
\end{aligned}
\tag{2.22}
$$

Here, we made use of the somewhat cumbersome orthogonality properties (2.5), (2.6) and (2.7). As the A_k^2 and B_k^2 always are positive, the mean squared error will drop monotonically while N increases.

Ex. 2.2 *Approximating the triangular function*

As δ_N^2 is always positive, we finally arrive from (2.22) at the Bessel inequality

$$\frac{1}{T}\int_T f^2(t)\,dt \geq A_0^2 + \frac{1}{2}\sum_{k=1}^{N}(A_k^2 + B_k^2) \tag{2.23}$$

For the border-line case of $N \to \infty$ we get the Parseval equation:

$$\frac{1}{T}\int_T f^2(t)\,dt = A_0^2 + \frac{1}{2}\sum_{k=1}^{\infty}(A_k^2 + B_k^2) \tag{2.24}$$

Parseval equation may be interpreted as follows: $1/T \int f^2(t)\,dt$ is the mean squared "signal" within the time domain, or – more colloquially – the *information content*. Fourier series don't lose this information content: it's in the squared Fourier coefficients.

2.2 Continuous Fourier Transformation

Contrary to Section 2.1, here we won't limit things to periodic $f(t)$. The integration interval is the entire real axis $(-\infty, +\infty)$. For this purpose we'll look at what happens at the transition from a series-representation to an integral-representation:

$$\text{Series:} \quad C_k = \frac{1}{T} \int_{-T/2}^{+T/2} f(t)\, e^{-i\omega_k t}\, dt$$

$$\text{Continuous:} \quad \lim_{T\to\infty} (TC_k) = \int_{-\infty}^{+\infty} f(t)\, e^{-i\omega t}\, dt$$

2.2.1 Definition

Let us define the *Forward Fourier Transformation* and the *Inverse Fourier Transformation* as follows:

Definition 2.3 (Forward Fourier transformation).

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t}\, dt \tag{2.25}$$

Definition 2.4 (Inverse Fourier transformation).

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) e^{+i\omega t} \, d\omega \tag{2.26}$$

Please note that in the case of the forward transformation, there is a minus sign in the exponent (cf. (2.17)), in the case of the inverse transformation, this is a plus sign. In the case of the inverse transformation, $1/2\pi$ is in front of the integral, contrary to the forward transformation.

The asymmetric aspect of the formulas has tempted many scientists to introduce other definitions, for example to write a factor $1/\sqrt{2\pi}$ for forward as well as inverse transformation. That's no good, as the definition of the average $F(0) = \int_{-\infty}^{+\infty} f(t) \, dt$ would be affected.

Now let us demonstrate that the inverse transformation returns us to the original function. For the forward transformation, we often will use $\mathrm{FT}(f(t))$, and for the inverse transformation we will use $\mathrm{FT}^{-1}(F(\omega))$.

We will begin with the inverse transformation and insert:

Proof.

$$\begin{aligned} f(t) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(\omega)\, e^{i\omega t}\, d\omega = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\omega \int_{-\infty}^{+\infty} f(t')\, e^{-i\omega t'}\, e^{i\omega t}\, dt' \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} f(t')\, dt' \int_{-\infty}^{+\infty} e^{i(t-t')\omega}\, d\omega \\ &= \int_{-\infty}^{+\infty} f(t')\, \delta(t-t')\, dt' = f(t) \end{aligned}$$

where $\delta(t)$ is the Dirac δ-function[4].

Note that for $f(t) = 1$ we have

$$\begin{aligned} \mathrm{FT}(\delta(t)) &= 1 \\ \mathrm{FT}^{-1}(1) &= 2\pi\delta(\omega) \end{aligned} \tag{2.27}$$

We realize the dual character of the forward and inverse transformations: a very slowly varying function $f(t)$ will have a very high spectral density for very small frequencies; the spectral density will go down

[4] The δ-function is actually a distribution. Its value is zero anywhere except when its argument is equal to zero. In this case it is ∞.

quickly and rapidly approaches 0. Conversely, a quickly varying function $f(t)$ will show spectral density over a very wide frequency range.

Now let us discuss an important example: the Fourier transform of the "rectangular" function (see Section 2.3.1 for a detailed discussion)

$$f(t) = \begin{cases} 1 \quad \text{for } -T/2 \leq t \leq +T/2 \\ 0 \quad \text{else} \end{cases}$$

Its Fourier transform is

$$F(\omega) = 2\int_0^{+T/2} \cos \omega t \, dt = T\frac{\sin \omega T/2}{\omega T/2} \tag{2.28}$$

The imaginary part is 0, as $f(t)$ is even. The Fourier transformation of a rectangular function, therefore, is of the type $\sin x/x$. Some authors use the expression $\text{sinc}(x)$ for this case. The "c" stands for *cardinal*. The functions $f(t)$ and $F(\omega)$ are shown in Figure 2.7.

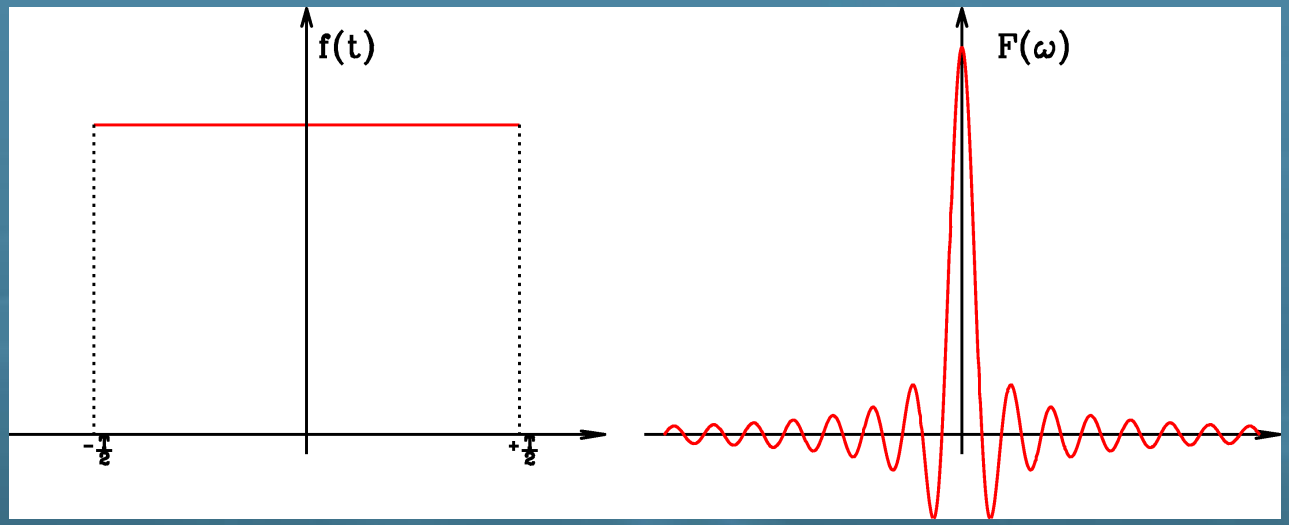


Figure 2.7: The rectangular function *(left)* and its Fourier transform sinc(x) *(right)*

Ex. 2.3 *Fourier transformation of relevant functions: Gaussian, bilateral exponential, unilateral exponential*

2.2.2 Convolution, Parseval Theorem

2.2.2.1 Convolution

The convolution of a function $f(t)$ with another function $g(t)$ is defined as:

Definition 2.5 (Convolution).

$$h(t) = \int_{-\infty}^{+\infty} f(\xi)\, g(t-\xi)\, d\xi \equiv f(t) \otimes g(t) \tag{2.29}$$

Please note that there is a minus sign in the argument of $g(t)$. The convolution is commutative, distributive, and associative. This means

commutative: $$f(t) \otimes g(t) = g(t) \otimes f(t)$$

distributive: $$f(t) \otimes (g(t) + h(t)) = f(t) \otimes g(t) + f(t) \otimes h(t)$$

associative: $$f(t) \otimes (g(t) \otimes h(t)) = (f(t) \otimes g(t)) \otimes h(t)$$

As an example of convolution, let us take a pulse that looks like an unilateral exponential function

$$f(t) = \begin{cases} e^{-t/\tau} & \text{for } t \geq 0 \\ 0 & \text{else} \end{cases} \tag{2.30}$$

Any device that delivers pulses as a function of time, has a finite rise-time/decay-time, which for simplicity's sake we'll assume to be a Gaussian

$$g(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{t^2}{\sigma^2}\right) \tag{2.31}$$

That is how our device would represent a δfunction – we can't get sharper than that. The function $g(t)$, therefore, is the device's resolution function, which we'll have to use for the convolution of all signals we want to record. An example would be the bandwidth of an oscilloscope. We then need:

$$S(t) = f(t) \otimes g(t) \tag{2.32}$$

where $S(t)$ is the experimental, *smeared* signal. It's obvious that the rise at $t = 0$ will not be as steep,

and the peak of the exponential function will get "ironed out". We'll have to take a closer look:

$$
\begin{aligned}
S(t) &= \frac{1}{\sigma\sqrt{2\pi}} \int_0^{+\infty} e^{-\xi/\tau} \exp\left(-\frac{1}{2}\frac{(t-\xi)^2}{\sigma^2}\right) d\xi \\
&= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{t^2}{\sigma^2}\right) \int_0^{+\infty} \exp\left[-\frac{\xi}{\tau} + \frac{t\xi}{\sigma^2} - \frac{1}{2}\frac{\xi^2}{\sigma^2}\right] d\xi \\
&= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{t^2}{\sigma^2}\right) \exp\left(\frac{1}{2}\frac{t^2}{\sigma^2}\right) \exp\left(-\frac{t}{\tau}\right) \exp\left(\frac{\sigma^2}{2\tau^2}\right) \times \\
&\qquad \int_0^{+\infty} \exp\left\{-\frac{1}{2\sigma^2}\left[\xi - \left(t - \frac{\sigma^2}{\tau}\right)\right]^2\right\} d\xi \qquad (2.33) \\
&= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{t}{\tau}\right) \exp\left(\frac{\sigma^2}{2\tau^2}\right) \times \\
&\qquad \int_{-(t-\sigma^2/\tau)}^{+\infty} \exp\left(-\frac{1}{2}\frac{\xi'^2}{\sigma^2}\right) d\xi' \qquad \text{with } \xi' = \xi - \left(t - \frac{\sigma^2}{\tau}\right) \\
&= \frac{1}{2} \exp\left(-\frac{t}{\tau}\right) \exp\left(\frac{\sigma^2}{2\tau^2}\right) \operatorname{erfc}\left(\frac{\sigma}{\tau\sqrt{2}} - \frac{t}{\sigma\sqrt{2}}\right)
\end{aligned}
$$

Here, $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$ is the complementary error function, where

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt \tag{2.34}$$

Figure 2.8 shows the result of the convolution of the exponential function with the Gaussian. The following properties immediately stand out: (i) The finite time resolution ensures that there is also a signal at negative times, whereas it was 0 before convolution. (ii) The maximum is not at $t = 0$ any more. (iii) What can't be seen straight away, yet is easy to grasp, is the following: the center of gravity of the exponential function, which was at $t = \tau$, doesn't get shifted at all upon convolution.

Now we prove the extremely important *Convolution Theorem*:

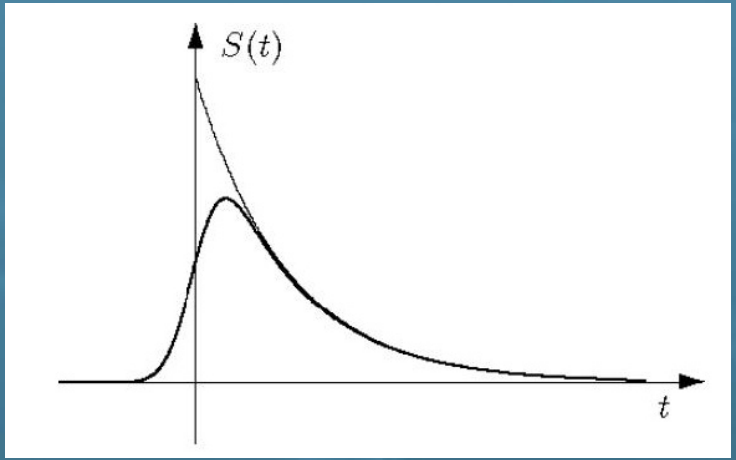


Figure 2.8: Result of the convolution of an unilateral exponential function with a Gaussian. The exponential function without convolution is indicated with the thin line

Theorem 2.1 (Convolution theorem). *Let be*

$$f(t) \leftrightarrow F(\omega)$$

$$g(t) \leftrightarrow G(\omega)$$

Then

$$h(t) = f(t) \otimes g(t) \leftrightarrow H(\omega) = F(\omega) \times G(\omega) \tag{2.35}$$

The *convolution integral* becomes, through Fourier transformation, a product of the Fourier-transformed ones.

Proof.

$$\begin{aligned} H(\omega) &= \int\int f(\xi)g(t-\xi)\,d\xi\; e^{-i\omega t}\,dt \\ &= \int f(\xi)\,e^{-i\omega\xi}\left[\int g(t-\xi)\,e^{-i\omega(t-\xi)}\,dt\right]\,d\xi \\ &\overset{t'=t-\xi}{=} \int f(\xi)\,e^{-i\omega\xi}\,d\xi\,G(\omega) \\ &= F(\omega)\,G(\omega) \end{aligned}$$

The integration boundaries $\pm\infty$ did not change by doing that, and $G(\omega)$ does not depend on ξ. The *inverse* Convolution theorem is:

Theorem 2.2 (Inverse convolution theorem). *Let be*

$$f(t) \leftrightarrow F(\omega)$$

$$g(t) \leftrightarrow G(\omega)$$

Then

$$h(t) = f(t) \times g(t) \leftrightarrow H(\omega) = \frac{1}{2\pi} F(\omega) \otimes G(\omega) \tag{2.36}$$

Proof.

$$
\begin{aligned}
H(\omega) &= \int f(t)\, g(t)\, e^{-i\omega t}\, dt \\
&= \int \left(\frac{1}{2\pi} \int F(\omega')\, e^{+i\omega' t}\, d\omega' \times \frac{1}{2\pi} \int G(\omega'')\, e^{+i\omega'' t}\, d\omega'' \right) e^{-i\omega t}\, dt \\
&= \frac{1}{(2\pi)^2} \int F(\omega') \int G(\omega'') \underbrace{\int e^{i(\omega'+\omega''-\omega)t}\, dt}_{=2\pi\delta(\omega'+\omega''-\omega)}\, d\omega' d\omega'' \\
&= \frac{1}{2\pi} \int F(\omega')\, G(\omega-\omega')\, d\omega' \\
&= \frac{1}{2\pi} F(\omega) \otimes G(\omega)
\end{aligned}
$$

Contrary to the Convolution Theorem (2.35) in (2.36) there is a factor $1/2\pi$ in front of the convolution of the Fourier transforms.

A widely popular exercise is the *unfolding* of data: the instruments' resolution function "smears out" the quickly varying functions, but we naturally want to reconstruct the data to what they would look like if

the resolution function was infinitely good – provided we precisely knew the resolution function. In principle, that's a good idea – and thanks to the Convolution Theorem, not a problem: you Fourier-transform the data, divide by the Fourier-transformed resolution function and transform it back. For practical applications it doesn't quite work that way. As in real life, we can't transform from $-\infty$ to $+\infty$, we need low-pass filters, in order not to get "swamped" with oscillations resulting from cut-off errors. Therefore, the advantages of unfolding are just as quickly lost as gained. Actually, the following is obvious: whatever got "smeared" by finite resolution, can't be reconstructed unambiguously. Imagine that a very pointed peak got eroded over millions of years, so there's only gravel left at its bottom. Try reconstructing the original peak from the debris around it! The result might be impressive from an artist's point of view, an artifact, but it hasn't got much to do with the original reality.

Ex. 2.4 *Convolution: Gaussian frequency distribution. Lorentzian frequency distribution*

2.2.2.2 Cross Correlation

Sometimes, we want to know if a measured function $f(t)$ has anything in common with another measured function $g(t)$. Cross correlation is ideally suited to that.

Definition 2.6 (Cross correlation).

$$h(t) = \int_{-\infty}^{+\infty} f(\xi)\, g^*(t+\xi)\, d\xi \equiv f(t) \star g(t) \tag{2.37}$$

Important note: Here, there is a plus sign in the argument of g, therefore we don't mirror $g(t)$. For even functions $g(t)$ this doesn't matter. The asterisk $*$ means complex conjugated. We may disregard it for real functions. The symbol $\star$ means cross correlation, and is not to be confounded with $\otimes$ for folding. Cross correlation is associative and distributive, yet not commutative. That's not only because of the complex-conjugated symbol, but mainly because of the plus sign in the argument of $g(t)$. Of course, we want to convert the integral in the cross correlation to a product by using Fourier transformation.

Theorem 2.3 (Cross correlation). *Let be*

$$f(t) \leftrightarrow F(\omega)$$

$$g(t) \leftrightarrow G(\omega)$$

Then

$$h(t) = f(t) \star g(t) \leftrightarrow H(\omega) = F(\omega) \times G^*(\omega) \tag{2.38}$$

Proof.

$$\begin{aligned} H(\omega) &= \int \int f(\xi)\, g^*(t+\xi)\, d\xi\, e^{-i\omega t}\, dt \\ &= \int f(\xi) \left[\int g^*(t+\xi)\, e^{-i\omega t}\, dt \right] d\xi \\ &= \int f(\xi)\, G^*(+\omega)\, e^{+i\omega t}\, d\xi \\ &= F(\omega) \times G^*(\omega) \end{aligned}$$

In the third passage we used the first shifting rule with $\xi = -a$. In the last passage we use the following identity:

$$\begin{aligned} G(\omega) &= \int g(t)\, e^{-i\omega t}\, dt \\ G^*(\omega) &= \int g^*(t)\, e^{+i\omega t}\, dt \\ G^*(-\omega) &= \int g^*(t)\, e^{-i\omega t}\, dt \end{aligned}$$

The interpretation of (2.38) is simple: if the spectral densities of $f(t)$ and $g(t)$ are a good match, i.e. have much in common, then $H(\omega)$ will become large on average, and the cross correlation $h(t)$ will also be large, on average. Otherwise, if $F(\omega)$ would be small e.g. where $G^*(\omega)$ is large and vice versa, so that there is never much left for the product $H(\omega)$. Then also $h(t)$ would be small, i.e. there is not much in

common between $f(t)$ and $g(t)$.

2.2.2.3 Autocorrelation

The autocorrelation function is the cross correlation of a function $f(t)$ with itself. You may ask, for what purpose we'd want to check for what $f(t)$ has in common with $f(t)$. Autocorrelation, however, seems to attract many people in a magical manner. We often hear the view, that a signal full of noise can be turned into something really good by using the autocorrelation function, i.e. the signal-to-noise ratio would improve a lot. Don't you believe a word of it! We'll see why shortly.

Definition 2.7 (Autocorrelation).

$$h(t) = \int f(t)\, f^*(\xi + t)\, d\xi \tag{2.39}$$

From its definition and the cross-correlation theorem (2.38) we have the so called

Theorem 2.4 (Wiener-Khinchin).

$$\begin{aligned} f(t) &\leftrightarrow F(\omega) \\ h(t) = f(t) \star f(t) &\leftrightarrow H(\omega) = F(\omega) \times F^*(\omega) = |F(\omega)|^2 \end{aligned} \tag{2.40}$$

We may either use the Fourier transform $F(\omega)$ of a noisy function $f(t)$ and get angry about the noise in $F(\omega)$, or we first form the autocorrelation function $h(t)$ from the function $f(t)$ and are then happy about the Fourier transform $H(\omega)$ of function $h(t)$. Normally, $H(\omega)$ does look a lot less noisy, indeed. Instead of doing it the roundabout way by using the autocorrelation function, we could have used the square of the magnitude of $F(\omega)$ in the first place. We all know, that a squared representation in the ordinate always pleases the eye, if we want to do cosmetics to a noisy spectrum. Big spectral components will grow when squared, small ones will get even smaller. But isn't it rather obvious that squaring doesn't change anything to the signal-to-noise ratio? In order to make it “look good”, we pay the price of losing linearity.

2.2.2.4 The Parseval Theorem

The autocorrelation function also comes in handy for something else, namely for deriving Parseval theorem. We start out with (2.39), insert especially $t = 0$, and get Parseval theorem:

Theorem 2.5 (Parseval theorem).

$$h(0) = \int |f(t)|^2 \, dt = \frac{1}{2\pi} \int |F(\omega)|^2 \, d\omega \tag{2.41}$$

The second equal sign is obtained by inverse transformation of $|F(\omega)|^2$ where, for $t = 0$, $e^{i\omega t}$ becomes unity.

Equation (2.41) states that the *information content* of the function $f(x)$ – defined as integral over the square of the magnitude – is just as large as the *information content* of its Fourier transform $F(\omega)$ (same definition, but with $1/(2\pi)$).

2.3 Window Functions

By necessity, every observed signal we process must be of finite extent. The extent may be adjustable and selectable, but it must be finite. If we observe the signal for T units of time, in order to apply our *continuous* Fourier transformations we perform a so-called *period extension* of our data, as shown in Figure 2.3. It is evident that if the periodic extension of a signal does not commensurate with its natural period, then discontinuities at the boundaries will be present. These discontinuities will introduce spurious frequencies, responsible for spectral contributions over the entire set of frequencies. This effect is called **spectral leakage**.

Windows are weighting functions applied to data to reduce spectral leakage associated with finite observations intervals. From one viewpoint, the window is applied to data (as a multiplicative weighting) to reduce the order of the discontinuity of the periodic extension. This is accomplished by matching as many orders of derivatives (of the weighted data) as possible at the boundary. The easiest way to achieve

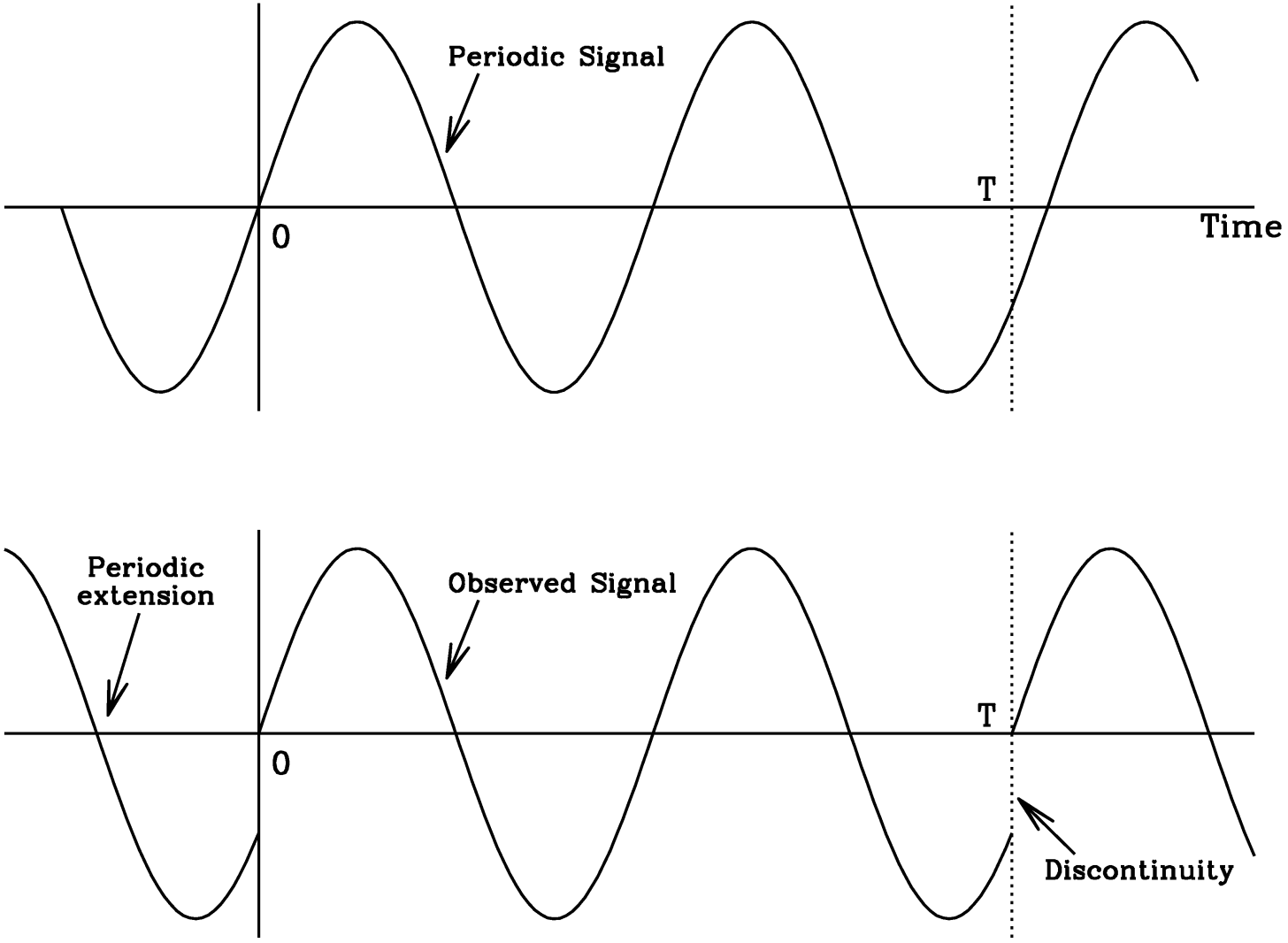


Figure 2.9: Periodic extension of a sinusoidal signal not periodic in the observation interval

this matching is by setting the value of these derivatives to zero or near zero. Thus windowed data are smoothly brought to zero at the boundaries so that the periodic extension of the data is continuous in many orders of derivatives.

From another viewpoint, the window is multiplicatively applied to the Fourier frequencies so that a signal of arbitrary frequency will exhibit a significant component for frequencies close to the Fourier frequencies. Of course both viewpoints lead to identical results.

All window functions are, of course, even functions. The Fourier transforms of the window function therefore don't have an imaginary part. We require a large dynamic range in order to better compare window qualities. That's why we'll use logarithmic representations covering equal ranges. And that's also the reason why we can't have negative function values. To make sure they don't occur, we'll use the power representation, i.e. $|F(\omega)|^2$.

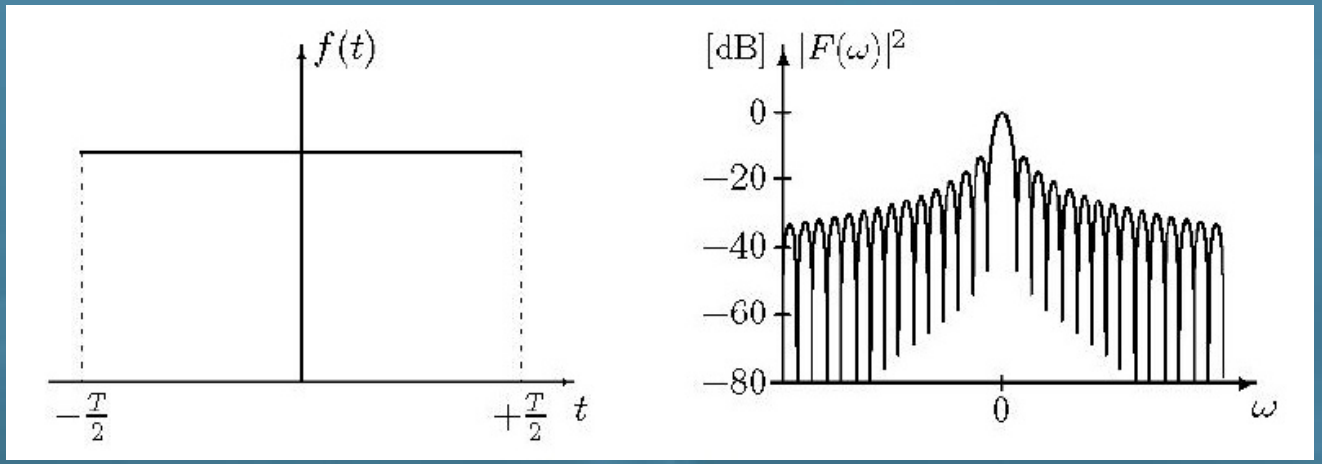


Figure 2.10: Rectangular window function and its Fourier transform in power representation.

2.3.1 The Rectangular Window

$$f(t) = \begin{cases} 1 \quad \text{for } -T/2 \leq t \leq T/2 \\ 0 \quad \text{else} \end{cases} \tag{2.42}$$

has the power representation of the Fourier transform (see (2.28)):

$$|F(\omega)|^2 = T^2 \left(\frac{\sin(\omega T/2)}{\omega T/2} \right)^2 \tag{2.43}$$

The rectangular window and this function are shown in Figure 2.10.

2.3.1.1 Zeroes

Where are the zeros of this function? We'll find them at $\omega T/2 = l\pi$ with $l = 1, 2, 3, \ldots$ and without the zero! The zeros are equidistant, the zero at $l = 0$ in the numerator gets plugged by a zero in the denominator.

2.3.1.2 Intensity at the Central Peak

Now we want to find out how much intensity is at the central peak, and how much gets lost in the side-bands (sidelobes). To get there, we need the first zero at $\omega T/2 = \pi$ or $\omega = \pm 2\pi/T$ and:

$$\int_{-2\pi/T}^{+2\pi/T} T^2 \left(\frac{\sin(\omega T/2)}{\omega T/2}\right)^2 d\omega \overset{\omega T/2 = x}{=} T^2 \frac{2}{T^2} 2 \int_0^{+2\pi} \frac{\sin^2 x}{x} dx = 4T\,\mathrm{Si}(2\pi) \tag{2.44}$$

where $\mathrm{Si}(x)$ is the sine integral, defined as

$$\mathrm{Si}(x) \equiv \int_0^x \frac{\sin y}{y} dy \tag{2.45}$$

The last passage in (2.44) may be proved as follows. We start out with

$$\int_0^\pi \frac{\sin^2 x}{x}\, dx$$

and integrate per parts with $u = \sin^2 x$ and $v = -1/x$:

$$\begin{aligned}\int_0^\pi \frac{\sin^2 x}{x}\, dx &= \left.\frac{\sin^2 x}{x}\right|_0^\pi + \int_0^\pi \frac{2\sin x \cos x}{x}\, dx \\ &= 2\int_0^\pi \frac{\sin 2x}{2x}\, dx \\ &\overset{2x=y}{=} \mathrm{Si}(2\pi)\end{aligned} \tag{2.46}$$

By means of the Parseval theorem (2.41) we get the total intensity

$$\int_{-\infty}^{+\infty} T^2 \left(\frac{\sin(\omega T/2)}{\omega T/2}\right)^2 d\omega = 2\pi \int_{-T/2}^{+T/2} 1^2\, dt = 2\pi T \tag{2.47}$$

The ratio of the intensity at the central peak to the total intensity is therefore

$$\frac{4T\,\mathrm{Si}(2\pi)}{2\pi T} = \frac{2}{\pi}\,\mathrm{Si}(2\pi) = 0.903$$

This means that $\approx 90\%$ of the intensity is in the central peak, whereas some 10% are "wasted" in the sidelobes.

2.3.1.3 Sidelobe Suppression

Now let's determine the height of the first sidelobe. To get there, we need:

$$\frac{d|F(\omega)|^2}{d\omega} = 0 \qquad \text{or also} \qquad \frac{dF(\omega)}{d\omega} = 0 \tag{2.48}$$

and this occurs when

$$\frac{d}{dx}\frac{sinx}{x} = 0 \overset{x=\omega T/2}{=} \frac{x\cos x - \sin x}{x^2}$$

Solving this transcendental equation gives us the smallest possible solution $x = 4.4934$ or $\omega = 8.9868/T$.

Inserting this value in $|F(\omega)|^2$ results in:

$$\left|F\left(\frac{8.9868}{T}\right)\right|^2 = T^2 \times 0.04719 \tag{2.49}$$

For $\omega = 0$ we get $|F(0)|^2 = T^2$, the ratio of the first sidelobe height to the central peak height is therefore 0.04719. It is customary to express ratios between two values spanning several order of magnitude in decibel (short dB). The definition of decibel is

$$\text{dB} \equiv 10 \log_{10} x \tag{2.50}$$

Quite regularly people forget to mention what the ratio's based on, which can cause confusion. Here we're talking about intensity-ratios. If we're referring to amplitude-ratios, (that is, $F(\omega)$), this would make precisely a factor of two in logarithmic representation! Here we have a sidelobe suppression (first sidelobe) of:

$$10 \log_{10} 0.04719 = -13.2 \text{ dB}$$

2.3.1.4 3 dB Bandwidth

As the $10 \log_{10}(1/2) = -3.0103 \approx -3$, the 3 dB bandwidth tells us where the central peak has dropped to half its height. This is easily calculated as follows

$$T^2 \left(\frac{\sin(\omega T/2)}{\omega T/2} \right)^2 = \frac{1}{2} T^2$$

Using $x = \omega T/2$ we have

$$\sin^2 x = \frac{x}{2} \qquad \text{or } \sin x = \frac{x}{\sqrt{2}}$$

This transcendental equation has the following solution:

$$x = 1.3915, \qquad \text{thus } \omega_{3\,\mathrm{dB}} = \frac{2.783}{T}$$

This gives the total width $(\pm\omega_{3\,\mathrm{dB}})$:

$$\Delta\omega = \frac{5.566}{T} \tag{2.51}$$

This is the slimmest central peak we can get using Fourier transformation. Any other window function will lead to larger 3 dB-bandwidths. Admittedly, it's more than nasty to stick more than 10% of the information into the sidelobes. If we have, apart from the prominent spectral component, another spectral component, with – say – an approx. 10 dB smaller intensity, this component will be completely smothered by the main component's sidelobes. If we're lucky, it will sit on the first sidelobe and will be visible; if we're out of luck, it will fall into the gap (the zero) between central peak and first sidelobe and will get swallowed. So it pays to get rid of these sidelobes.

Warning! This 3 dB-bandwidth is valid for $|F(\omega)|^2$ and not for $F(\omega)$! Since one often uses $|F(\omega)|$ or

the cosine-/sine-transformation one wants the 3 dB-bandwidth thereof, which corresponds to the 6 dB-bandwidth of $|F(\omega)|^2$. Unfortunately, you cannot simply multiply the 3 dB-bandwidth of $|F(\omega)|^2$ by $\sqrt{2}$, you have to solve a new transcendental equation. However, it's still good as a first guess because you merely interpolate linearly between the point of 3 dB-bandwidth and the point of the 6 dB-bandwidth. You'd overestimate the width by less than 5%.

2.3.2 Windowing or Convolution?

In principle, we have two possibilities to use window functions:

- Either you weight, i.e. you multiply, the input by the window function and subsequently Fourier-transform, or
- You Fourier-transform the input and convolute the result with the Fourier transform of the window function.

According to the Convolution Theorem (2.35) we get the same result. What are the pros and cons of both procedures? There is no easy answer to this question. What helps in arguing is thinking in discrete data. Take, e.g., a weighting window. Let's start with a reasonable value for the its parameter, based on considerations of the trade-off between 3 dB-bandwidth (i.e. resolution) and sidelobe suppression. In the case of windowing we have to multiply our input data, say N real or complex numbers, by the window function which we have to calculate at N points. After that we Fourier-transform. Should it turn out that we actually should require a better sidelobe suppression and could tolerate a worse resolution – or vice versa – we would have to go back to the original data, window them again and Fourier-transform again.

The situation is different for the case of convolution: we Fourier-transform without any bias concerning the eventually required sidelobe suppression and subsequently convolute the Fourier data (again N numbers, however in general complex!) with the Fourier-transformed window function, which we have to calculate for a sufficient number of points. What is a sufficient number? Of course, we drop

the sidelobes for the convolution and only take the central peak! This should be calculated at least for five points, better more. The convolution then actually consists of five (or more) multiplications and a summation for each Fourier coefficient. This appears to be more work; however, it has the advantage that a further convolution with another, say broader Fourier-transformed window function, would not require to carry out a new Fourier transformation. Of course, this procedure is but an approximation because of the truncation of the sidelobes. If we included all data of the Fourier-transformed window function including the sidelobes, we had to carry out N (complex) multiplications and a summation per point, already quite a lot of computational effort, yet still less than a new Fourier transformation. This could be relevant for large arrays, especially in two or three dimensions like in image processing and tomography.

Temporal Analysis in X–ray Astronomy

Now we will apply all the mathematical tools developed in Part I to real data. In particular we will explore the techniques that are commonly used in timing studies of X–ray sources. The regime we will be referring to is that of equidistantly binned timing data, the background noise of which is dominated by counting statistics. If there are gaps in the data, they are far apart and the data are not "sparse" in the sense that nearly all time bins are empty. This kind of data are eminently suited to analysis with FFT techniques, and the discussed methods will be based on these techniques.

3.1 Power Spectra in X–ray Astronomy

If we indicate with x_k, $k = 0, 1, 2, \ldots, N-1$, the number of photons detected in bin k by our instrument, then the discrete Fourier transform a_j, with $j = -N/2, \ldots, N/2-1$, decomposes this signal into N sine waves. The following expressions describe the signal transform pair:

Definition 3.8 (Discrete Fourier transform in X–ray astronomy).

$$a_j = \sum_{k-0}^{N-1} x_k \, e^{2\pi i j k/N} \qquad j = -\frac{N}{2}, \ldots, \frac{N}{2} - 1 \tag{3.1a}$$

$$x_k = \frac{1}{N} \sum_{j=-N/2}^{N/2-1} a_j \, e^{-2\pi i j k/N} \qquad k = 0, 1, \ldots, N-1 \tag{3.1b}$$

Important Note: In X–ray astronomy it is customary the use of the convention as in Press *et al.* Accordingly, the prefactor $1/N$ is present in the *inverse* discrete Fourier transform and not in the *direct* one. The consequence is that a_0 will not be anymore the average, but the total number of counts $N_{\text{ph}} = \sum_k x_k$. As we said before, it is only a question of convention.

If the signal is an equidistant time series of length T, so that x_k refers to a time $t_k = k(T/N)$, then the transform is an equidistant "frequency series", and a_j refers to a frequency $\omega_j = 2\pi\nu_j = 2\pi j/T$. The time step is $\delta t = T/N$; the frequency step is $\delta\nu = 1/T$.

Note that $a_{-N/2} = \sum_k x_k \, e^{-\pi i k} = \sum_k x_k \, (-1)^k = a_{N/2}$, and that a_0 is nothing else that the total

number of detected photons $a_0 = \sum_k x_k \equiv N_{\text{ph}}$.

We have already seen that the Parseval theorem relates the a_j and x_k:

$$\sum_{k=0}^{N-1} |x_k|^2 = \frac{1}{N} \sum_{j=-N/2}^{N/2-1} |a_j|^2 \tag{3.2}$$

This implies that there is a relation between the summed squared modulus of the Fourier amplitudes and the total variance of the data:

$$\begin{aligned}
\text{Var}(x_k) &= \sum_k (x_k - \bar{x})^2 = \sum_k x_k^2 - 2\bar{x} \underbrace{\sum_k x_k}_{N\bar{x}} + N(\bar{x})^2 \\
&= \sum_k x_k^2 - N(\bar{x})^2 = \sum_k x_k^2 - \frac{1}{N}\left(\sum_k x_k\right)^2 \\
&\overset{(3.2)}{=} \frac{1}{N}\sum_j |a_j| - \frac{1}{N}|a_0|^2
\end{aligned}$$

Therefore we have

$$\mathrm{Var}(x_k) = \frac{1}{N} \sum_{\substack{j=-N/2 \\ j\neq 0}}^{N/2-1} |a_j|^2 \tag{3.3}$$

Adopting the normalization used by Leahy *et al.* (1983), we will define

Definition 3.9 (Power spectrum).

$$P_j \equiv \frac{2}{N_{\mathrm{ph}}} |a_j|^2 \qquad j = 0, 1, 2, \ldots, \frac{N}{2} \tag{3.4}$$

where N_{ph} is the total number of photons.

Taking into account that for real data $|a_j| = |a_{-j}|$ and that the term at the Nyquist frequency occurs only once in (3.3), we find the expression for the total variance in terms of P_j:

$$\mathrm{Var}(x_k) = \frac{N_{\mathrm{ph}}}{N} \left(\sum_{j=1}^{N/2-1} P_j + \frac{1}{2} P_{N/2} \right) \tag{3.5}$$

Note the difference in the indexing of a_j and P_j. Often the variance is expressed in terms of the fractional

root-mean-square (rms) variation in the x_k:

$$\text{rms} = \frac{\sqrt{\frac{1}{N}\,\text{Var}(x_k)}}{\bar{x}} = \sqrt{\frac{1}{N_{\text{ph}}}\left(\sum_{j=1}^{N/2-1} P_j + \frac{1}{2}\,P_{N/2}\right)} \tag{3.6}$$

Sometimes rms is expressed in terms of percentage, and is then called the "percentage rms variation".

A sinusoidal signal $x_k = A\sin(2\pi\nu_j t_k)$ at the Fourier frequency ν_j will cause a spike at ν_j in the power spectrum with

$$P_{j,\text{sine}} = \frac{1}{2}\,\frac{N^2}{N_{\text{ph}}}\,A^2 \tag{3.7}$$

The reason for choosing this apparently rather awkward normalization for the powers lies in the statistical properties of the noise power spectrum, to be described later.

Finally, let us discuss on relation between the *sampled* x_k (with Fourier transform a_j) and the *continuous* function $x(t)$ (with Fourier transform $a(\nu)$). It is easy to understand that x_k is given by a **double multiplication** by two functions:

Window Function

$$w(t) = \begin{cases} 1 & 0 \leq t < T \\ 0 & \text{else} \end{cases} \tag{3.8}$$

Sampling Function

$$i(t) = \sum_{k=-\infty}^{+\infty} \delta\left(t - k\frac{T}{N}\right) \tag{3.9}$$

Therefore in order to obtain the power spectrum of x_k we must perform a double convolution with both the window and the sampling functions. The power spectrum of the shifted window function is (see Section 2.3.1):

$$|W(\nu)|^2 = \left|\frac{\sin \pi\nu T}{\pi\nu}\right|^2 \tag{3.10}$$

The Fourier transform on an infinitely extended periodic series of δ-functions is

$$I(\nu) = \frac{N}{T} \sum_{k=-\infty}^{+\infty} \delta\left(\nu - k\frac{N}{T}\right) \tag{3.11}$$

The functions $w(t)$ and $i(t)$, together with the corresponding power spectra $W(\nu)$ and $I(\nu)$, are shown

in Figure 3.1.

The convolution of $a(\nu)$ with $W(\nu)$ causes all features in the power spectrum to become wider. We have already seen that the convolution with a δ-function at ν_0 causes a shift of the function by ν_0: $f(\nu) * \delta(\nu - \nu_0) = f(\nu - \nu_0)$ Therefore the convolution of $a(\nu)$ with $I(\nu)$, which is a series of δ-functions with spacing N/T results in a convolved function $a(\nu) * I(\nu)$ that repeats every N/T frequency units.

For a real signal $x(t)$ we have, as before, $a(-\nu) = a^*(\nu)$, so that $|a(\nu)|^2 = |a(-\nu)|^2$: the power spectrum is symmetric with respect to $v = 0$. The final result is that the power spectrum of the convolved function $|a(\nu) * I(\nu)|^2$ is reflected around the Nyquist frequency $\nu_{N/2} = \frac{1}{2}N/T$. This causes features with a frequency exceeding the Nyquist frequency by ν_x to appear also at a frequency $\nu_{N/2} - \nu_x$, a phenomenon we have already seen and known as *aliasing*.

From their definitions, is is straightforward to show that the discrete Fourier amplitudes a_j are the values at the Fourier frequencies $\nu_j = j/T$ of the windowed and aliased continuous Fourier transform $a_{\mathrm{WI}}(\nu)$

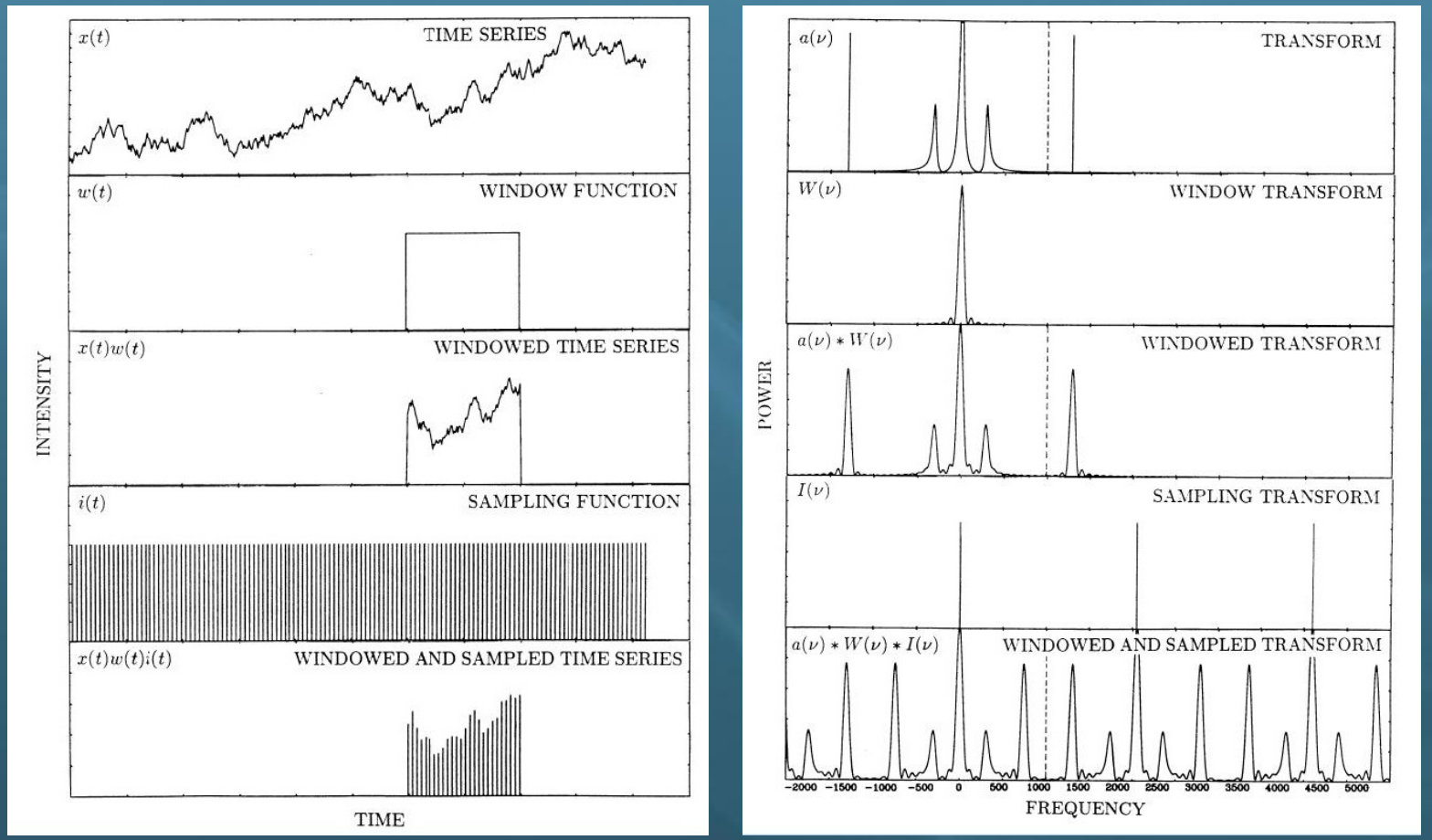


Figure 3.1: *Left:* Obtaining the discrete time series x_k involves the application of the two function $w(t)$ (window function) and $i(t)$ (sampling function). The bottom panel show the final results. *Right:* The discrete Fourier transform a_j of x_k is obtained out of the continuous Fourier transform by a double convolution. These are the power spectra corresponding to the various Fourier transforms. Vertical dashed lines indicate the Nyquist frequency

$$a_{\mathrm{WI}}(\nu) = a(\nu) * W(\nu) * I(\nu) = \int_{-\infty}^{+\infty} x(t)\, w(t)\, i(t)\, e^{2\pi i \nu t}\, dt = \int_{-\infty}^{+\infty} x(t) \sum_{k=0}^{N-1} \delta\left(t - k\frac{T}{N}\right) e^{2\pi i \nu t}\, dt = \sum_{k=0}^{N-1} x\left(k\frac{T}{N}\right) e^{2\pi i \nu k T/N} \tag{3.12}$$

so that $a_{\mathrm{WI}}(j/T) = a_j$. Explicitly performing the convolution of $a(\nu)$ with $I(\nu)$ we finally have

$$a_j = a_{\mathrm{WI}}(j/T) = a_{\mathrm{W}}(j/T) * a_I(j/T) = \frac{N}{T} \sum_{k=-\infty}^{+\infty} a_{\mathrm{W}}\left(v_k - k\frac{N}{T}\right) \tag{3.13}$$

were we used (3.11) and where $\nu_j = j/T$ and $a_{\mathrm{W}} = a(\nu) * W(\nu)$.

To summarize: the transition from the continuous Fourier transform to the discrete Fourier transform involves two operations: windowing, a convolution with the function $W(\nu)$ which is essentially a peak with a width $\delta\nu = 1/T$ plus sidelobes, and aliasing, a reflection of features above the Nyquist frequency back into the range $(0, \nu_{N/2})$. Windowing is caused by the finite extent, aliasing by the discrete sampling of the data.

In practice, aliasing is not so much of a problem as one might fear, as the data are not really discretely sampled at intervals $\delta t = T/N$, but rather binned into time bins with a width δt. This is equivalent of convolving with the "binning window"

$$b(t) = \begin{cases} N/T & -\frac{T}{2N} < t < \frac{T}{2N} \\ \\ 0 & \text{else} \end{cases} \tag{3.14}$$

before the discrete sampling. Applying the inverse convolution theorem, we can see that the effect of this on the Fourier transform will be that $a(\nu)$ is multiplied with the transform of $b(t)$:

$$B(\nu) = \frac{\sin \pi \nu T/N}{\pi \nu T/N} \tag{3.15}$$

This function drops from a value of 1 at $\nu = 0$ to 0 at $\nu = N/T$; halfway, at the Nyquist frequency it has the value $2/\pi$. The effect of this multiplication is a considerable repression of the high-frequency features that could be aliased back into the frequency range $(0, \nu_{N/2})$. This is understandable: the effect

of the binning is nothing else than averaging the time series over the bin width T/N so that variations with a frequency close to N/T are largely averaged out.

The problem caused by the windowing can be more serious: the "leakage" cause by the finite width of the central peak of $W(\nu)$ and its sidelobes can strongly distort steep power spectra (they becomes less steeper) and it can spread out δ-functions over the entire power spectrum.

3.2 Power Spectral Statistics

In general, signal processing is devoted to *detection* and *estimation*. Detection is the task of determining if a specific signal set is present in the observation, while estimation is the task of obtaining the values of the parameters describing the signal. The process of detecting something in a power spectrum against a background of noise has several steps. To quantify the power of the source signal, that is to determine what the power signal $P_{j,\text{signal}}$ would have been in the absence of noise, we must consider the *interaction*

between the noise and the signal.

As our starting point we will make the assumption that our signal will be due to the *sum* of two *independent* processes: signal and noise. This corresponds to assume $x_k = x_{k,\text{signal}} + x_{k,\text{noise}}$. For the linearity of the Fourier transform if b_j and c_j are the Fourier transforms of $x_{k,\text{signal}}$ and $x_{k,\text{noise}}$, then $a_j = b_j + c_j$. This means that a similar properties **does not apply** to power spectra:

$$|a_j|^2 = |b_j + c_j|^2 = |b_j|^2 + |c_j|^2 + \text{cross terms} \tag{3.16}$$

If the noise is random and uncorrelated, and if many powers are averaged, then the cross terms will tend to average out to zero, and we can write down

$$P_j = P_{j,\text{signal}} + P_{j,\text{noise}} \tag{3.17}$$

3.2.1 The Probability Distribution of the Noise Powers

For a wide range of type of noise, the noise powers $P_{j,\text{noise}}$ follow a χ^2 distribution with 2 degrees of freedom (dof). Indeed, if A_j and B_j are the Fourier coefficient of the noise signal, then the Parseval theorem says that $P_{j,\text{noise}} = A_j^2 + B_j^2$. But A_j and B_j are linear combinations of the x_k, therefore if x_k are normally distributed, then the A_j and B_j do as well, so that $P_{j,\text{noise}}$, by definition, is distributed according to the χ^2 distribution with 2 dof.

If the x_k follow some other probability distribution, for example the Poisson distribution, then it follows from the central limit theorem that for "certain" conditions on this other distribution (i.e. for large N), the A_j and B_j will be approximately normally distributed.

In practice, one finds out that noise powers are nearly always χ^2 distributed, not only for Poisson noise, but also for many other type of noise.

The power spectrum normalization defined in (3.4) is chosen in such a way that if the noise in the photon

counting data x_k is pure Poissonian counting noise, then the distribution of the P_j, noise is exactly given by a χ^2 distribution with 2 dof. Therefore the probability to exceed a certain threshold power level P_{thr} is given by

$$\text{Prob}(P_{j,\text{noise}} > P_{\text{thr}}) = Q(P_{\text{thr}}|2) \qquad j = 1, 2, \ldots, N/2 - 1 \tag{3.18}$$

where the integral probability of the χ^2 is defined as

$$Q(\chi^2|n) = \frac{1}{2^{n/2}\,\Gamma\left(\frac{n}{2}\right)} \int_{\chi^2}^{\infty} t^{\frac{n}{2}-1}\, e^{-\frac{t}{2}}\, dt \tag{3.19}$$

where n is the number of dof.

Because the $P_{j,\text{noise}}$ follow this distribution, the power spectrum is very noisy; the standard deviation of the noise powers is equal to their mean value:

$$\sigma_{P_j} = \langle P_j \rangle = 2 \tag{3.20}$$

Two more or less equivalent methods are often used to decrease the large variance of the $P_{j,\text{noise}}$.

- Rebin the power spectrum, averaging b consecutive frequency bins;
- Divide the data up into M equal segments, transform these segments each individually and then average the resulting M power spectra, each normalized according to (3.4). The N_{ph} is now the number of photons in *each* transform.

These two methods, of course, degrade the frequency resolution.

Because the time required to computed the Fourier transform of N data point using an FFT algorithm is proportional to $N \log N$, there is a computational advantage in the second method; the time saving factor is about $1 + \log M / \log N$.

For a variable source, a further advantage of the second method is the possibility to follow the variations of the power spectra as a function of time and/or intensity (see Figure 3.2).

The first method, on the other hand, has the advantage of producing a power spectrum that extends to

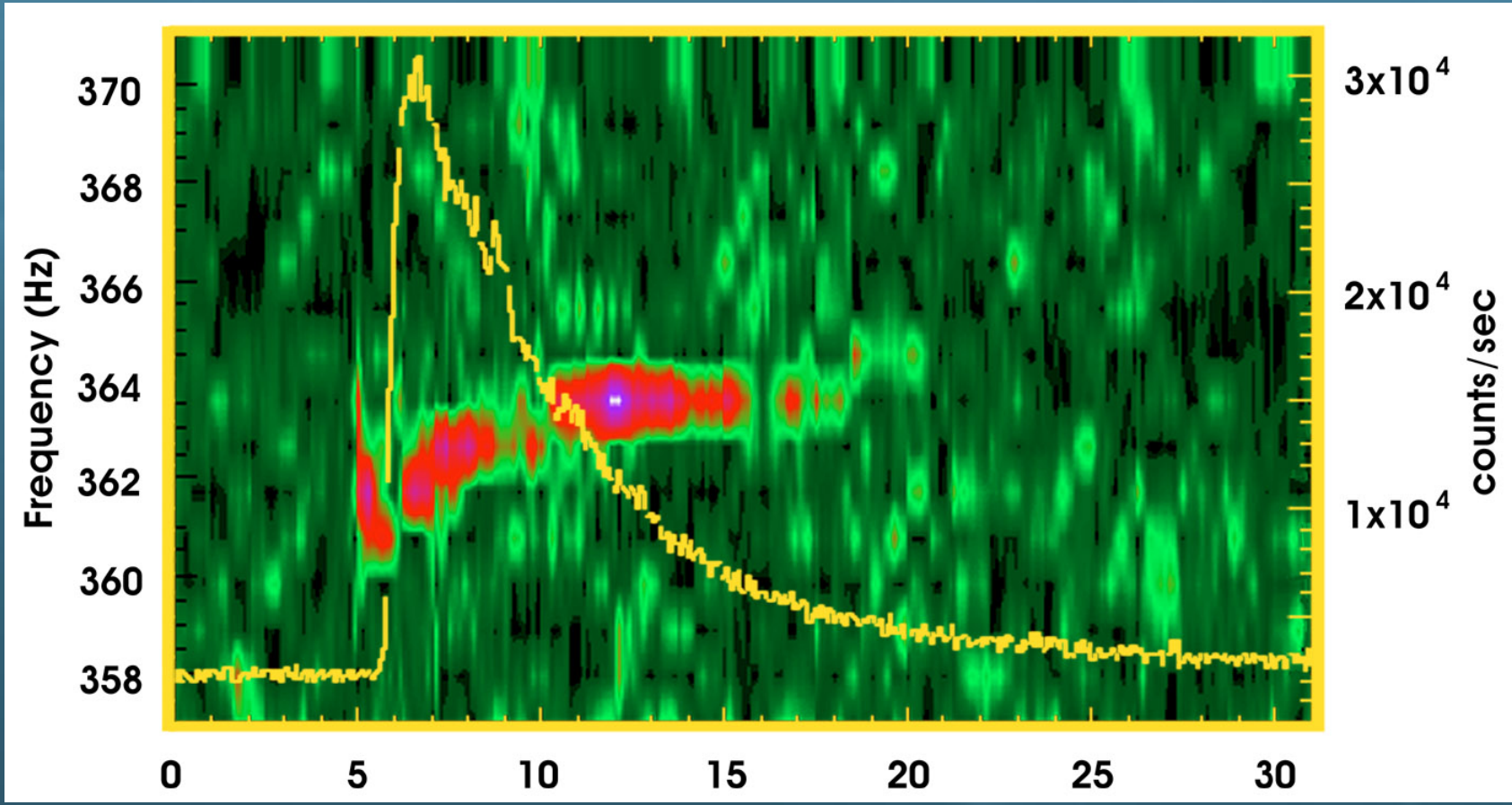


Figure 3.2: Dynamic power spectrum of the low mass X–ray binary 4U 1728–34. The color map shows increasing power in the order green, red, blue, and white. The time increases along the horizontal axis with a resolution of 1 sec. The total time shown is 32 sec. During the burst (the yellow line) the source exhibits pulsation at ~ 363 Hz (Strohmayer *et al*. 1996. *ApJ* 469, L9)

lower frequencies. It is of course possible to combine both methods: each power in the final spectrum will be the average of Mb original powers.

Because of the additive properties of the χ^2 distribution, the sum of Mb powers is distributed according to the χ^2 distribution with $2\,Mb$ dof, so that the probability for a given power $P_{j,\text{noise}}$ in the average spectrum to exceed a P_{thr} will be

$$\text{Prob}(P_{j,\text{noise}} > P_{\text{thr}}) = Q(Mb\,P_{\text{thr}}|2\,Mb) \tag{3.21}$$

For large Mb this distribution tends asymptotically to a normal distribution with a mean of 2 and a standard deviation of $2/\sqrt{Mb}$:

$$\lim_{Mb\to\infty} \text{Prob}(P_{j,\text{noise}} > P_{\text{thr}}) = Q_{\text{Gauss}}\left(\frac{P_{\text{thr}}-2}{2/\sqrt{Mb}}\right) \tag{3.22}$$

where the integral probability of the normal distribution is

$$Q_{\text{Gauss}}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt \tag{3.23}$$

3.2.2 The rms Variation in the Source Signal

Assuming that the signal power spectrum has been properly separated from the total power spectrum, we can convert the signal power into the rms variation of the source signal x_k using the expression

$$\text{rms} = \sqrt{\frac{b \sum_j P_{j,\text{signal}}}{N_{\text{ph}}}} \tag{3.24}$$

where P_j is an Mb times averaged power and where N_{ph} is the number of photons *per transform*.

3.3 Power Spectral Searches Made Easy

In this section we collect all previous results into a “how-to” recipe for testing the power spectrum for a weak signal using equal statistically independent trials

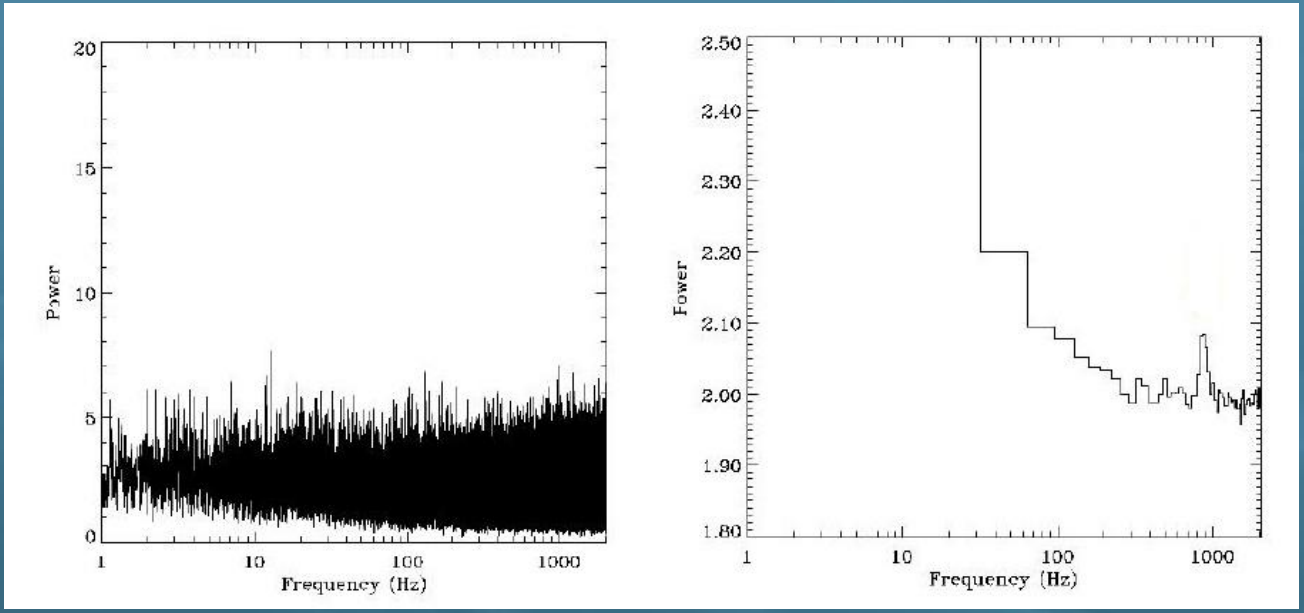


Figure 3.3: Effect of choosing the binning size in detecting weak features: the case of the kHz QPO in 4U 1728–34. The same data are shown in both the two panels, but the right bin size reveals the QPO at ~ 800 Hz

① Determine the M and b. The optimal choice for Mb is that which approximately matches the expected width of the power spectral feature one desires to detect, $\Delta\nu \geq Mb/T_{\text{obs}}$ (see Figure 3.3 for the effects of choosing the right b). Note that gaps in the data or the desire to observe the time evolution of the power spectrum may dictate M.

② Calculate the M power spectra normalized according to (3.4). Note that x_k is the number of pho-

tons in bin k and N_{ph} is the number of photons in *one* power spectrum.

③ Average the M power spectra.

④ Observe the noise power distribution. Is the noise power spectrum flat? Is its mean level equal to 2? If so, the noise is probably dominated by Poissonian counting statistics. If not, it is necessary to find out why.

⑤ Determine the detection level.

⑥ Check the average spectrum for powers exceeding the detection level.

⑦ Quantify the signal power in terms of a detection or an upper limit.

⑧ Convert the signal power into the relative rms variation of the source signal, defined as

$$\text{rms} = \sqrt{\frac{1}{N}\sum_{k}(\text{RATE}_k - \langle\text{RATE}\rangle)^2} \tag{3.25}$$

and compute the *excess* variance

$$\text{Excess Variance} = \sqrt{\text{rms}^2 - \frac{1}{N}\sum_k \text{ERROR}_k^2} \tag{3.26}$$

⑨ To say more about the signal, you need to model its power spectrum.

3.4 Type of Variability

In the previous Section we were left with the last point in our "how-to" with the problem of modeling a power spectrum. In this section we will deal with the problem of linking the *shape* of a power spectrum with the statistical processes that originated the timing variability. In Figure 3.4 we show a schematic power spectrum of an X–ray source displaying characteristic features: a continuum described in terms of $1/f$ noise, a Quasi-Periodic Oscillation (QPO) and a sharp peak due to a coherent signal (in this case the rotation period of the object). We have already discussed on the Poissonian level; now we will now

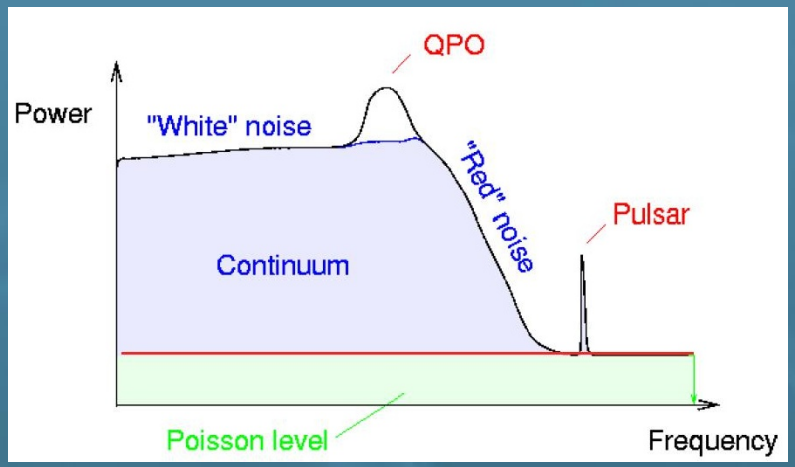


Figure 3.4: Noise classification in astronomical power spectra

analyze in details the other components.

3.4.1 $1/f$ Noise

Definition 3.10 ($1/f$ noise). $1/f$ refers to the phenomenon of the spectral density, $S(f)$, having the form

$$S(f) = K\, f^{-\alpha} \tag{3.27}$$

where f is the frequency.

$1/f$ noise is an intermediate between the well understood white noise with no correlation in time and random walk (Brownian motion) noise with no correlation between increments (see Figure 3.5). Brownian motion is the integral of white noise, and integration of a signal increases the exponent α by 2 whereas the inverse operation of differentiation decreases it by 2. Therefore, $1/f$ noise can not be obtained by the simple procedure of integration or of differentiation of such convenient signals. Moreover, there are no simple, even linear stochastic differential equations generating signals with $1/f$ noise. The widespread occurrence of signals exhibiting such behavior suggests that a generic mathematical explanation might exist. Except for some formal mathematical descriptions like fractional Brownian motion (half-integral of a white noise signal), however, no generally recognized physical explanation of $1/f$ noise has been proposed. Consequently, the ubiquity of $1/f$ noise is one of the oldest puzzles of contemporary physics and science in general.

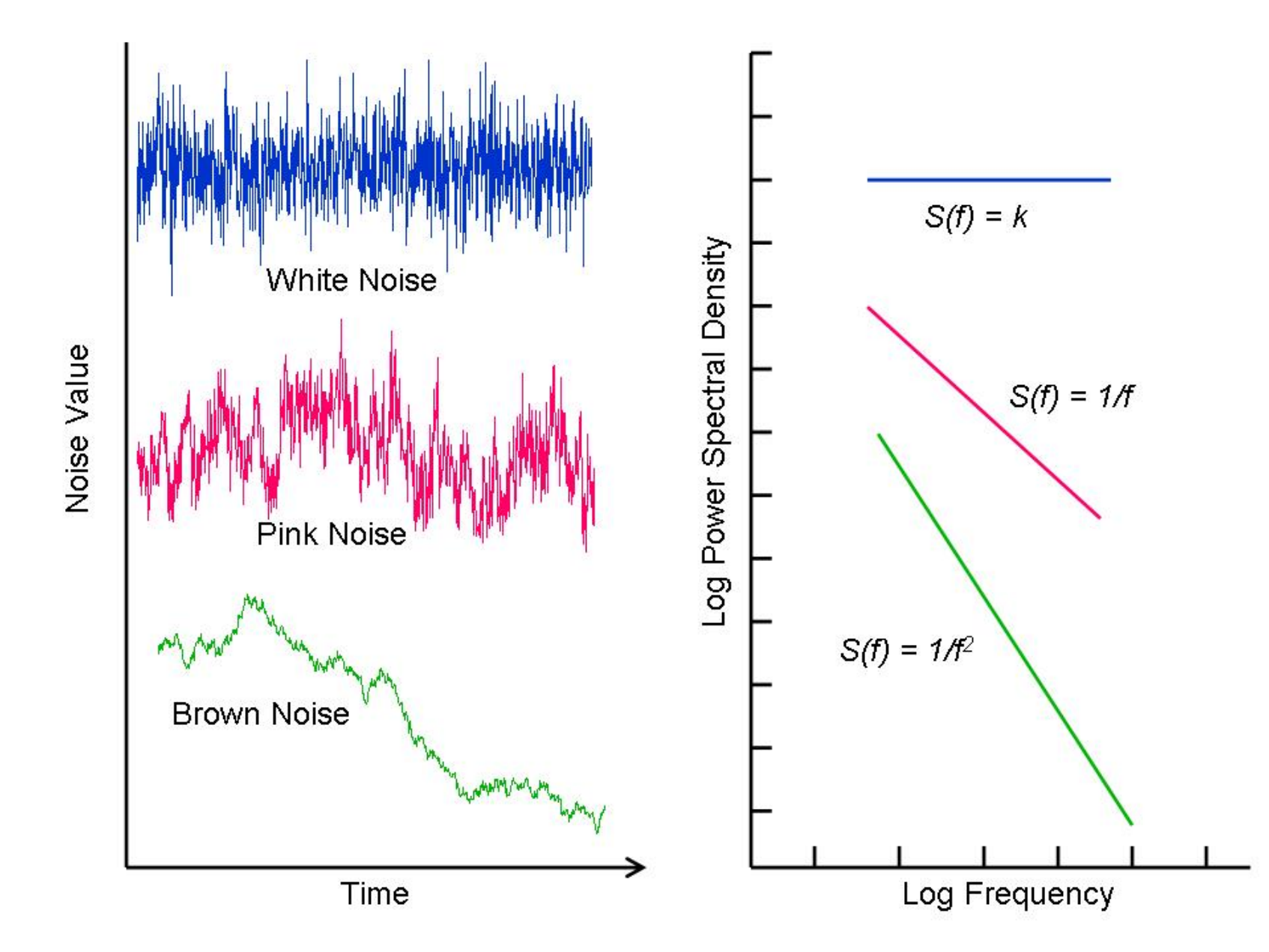


Figure 3.5: Examples of $1/f$ noise: on the left the time series and on the right the corresponding power spectrum

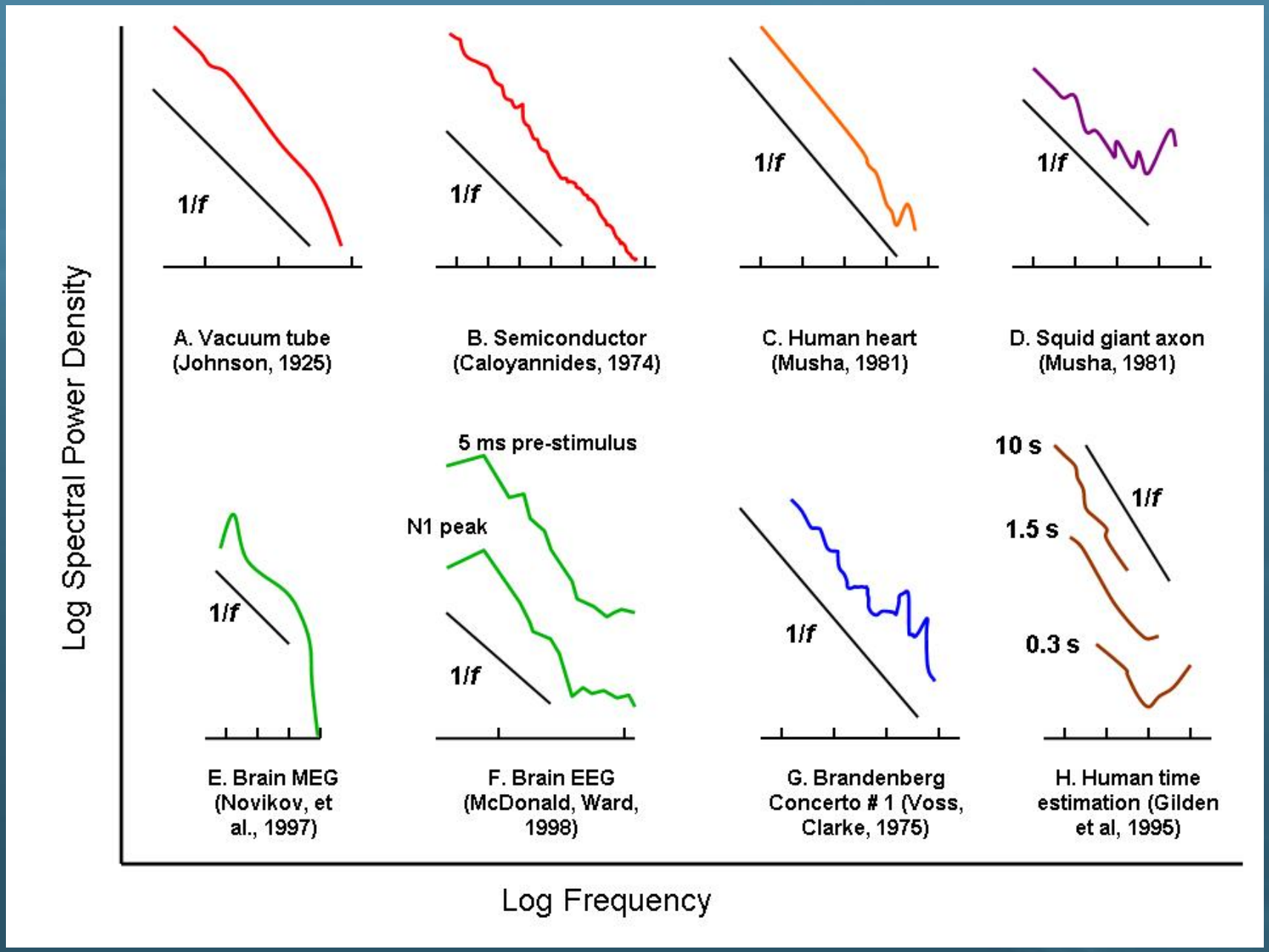


Figure 3.6: Examples of $1/f$ noise observed in both in the natural world and in man-made processes from physics, biology, neuroscience, and psychology

The case of $\alpha = 1$, or *pink noise,* is both the canonical case, and the one of most interest, but the more general form, where $0 < \alpha \leq 3$, is sometimes referred to simply as $1/f$. $1/f^{\alpha}$ noise is of interest because it occurs in many different systems, both in the natural world and in man-made processes (see Figure 3.6) from physics, biology, neuroscience, and psychology.

Although 1/f noise appears in many natural systems and has been intensively studied for decades with many attempts to describe the phenomenon mathematically, researchers have not yet been able to agree on a unified explanation. Thus, there exist at present several formulations of systems that give rise to $S(f) = K/f^{\alpha}$.

3.4.2 Shot Noise Process

First, let t_k be a Poisson point process. A shot noise process is obtained by attaching to each t_k a relaxation function (unilateral exponential function)

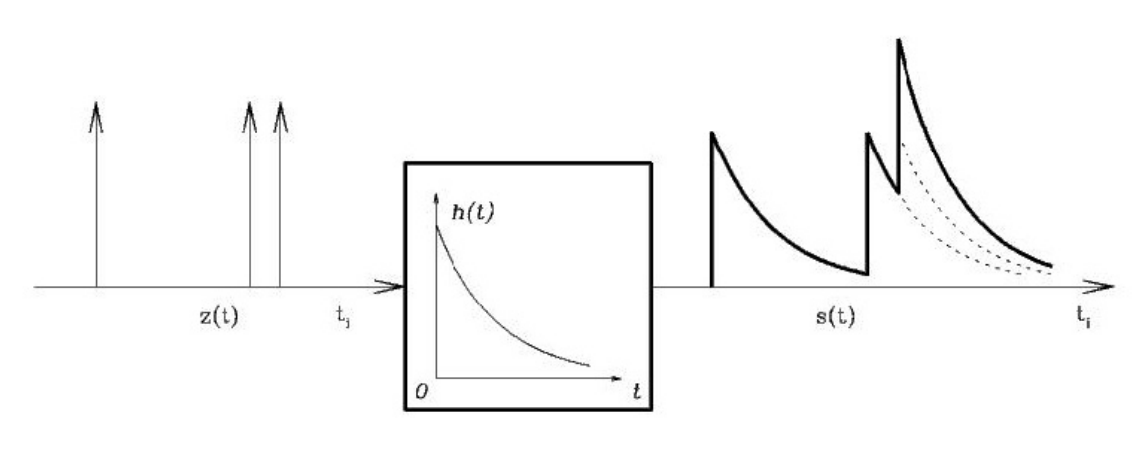


Figure 3.7: The shot noise process

$$N(t) = N_0 e^{-\lambda t}, \qquad t \geq 0$$

and summing on k (see Figure 3.7). The Fourier transform of the shot noise process is (see Page 46)

$$S(f) = \lim_{T \to \infty} \frac{1}{T} \langle \mid F(f) \mid^2 \rangle = \frac{N_0^2 n}{\lambda^2 + f^2}$$

where n is the average rate at which t_k occur, and T is the interval over which the process is observed. As

we have already seen, the power spectrum of an unilateral exponential function is a Lorentzian function. For an aggregation of shot noise processes with λ uniformly distributed on $[\lambda_1, \lambda_2]$, the power spectrum is

$$S(f) = \begin{cases} N_0^2 n & \text{if } 0 \ll f \ll \lambda_1 \ll \lambda_2 \\ \dfrac{N_0^2 n \pi}{2f(\lambda_2 - \lambda_1)} \cdot \dfrac{1}{f} & \text{if } \lambda_1 \ll f \ll \lambda_2 \\ N_0^2 n \cdot \dfrac{1}{f^2} & \text{if } 0 \ll \lambda_1 \ll \lambda_2 \ll f \end{cases}$$

If the impulse response function is a power law, $N_0 x^{-\beta}$, the process is called *fractal* shot noise, and the power spectrum is of the form

$$S(f) \approx \frac{k}{f^{2(1-\beta)}}$$

When $\beta = 1/2$, we obtain $S(f) \approx 1/f$.

3.5 Fitting Power Spectra Continuum with Lorentzians

Instead of describing the observed power spectrum continua in terms of $1/f$ noise, recently it has become quite popular a different approach. The power spectrum from X–ray sources like low-mass X–ray binaries (LMXB) can be described in terms of a flat-top continuum at low frequencies that becomes steeper at high frequencies, with bumps and wiggles. This continuum can be fit *without* the need of power-law components, but as a sum of Lorentzians, some of which are broad (Belloni *et al.* 2002).

The power spectra are described as the sum of Lorentzian components $L(\nu)$ of the form

$$L(\nu) = \frac{r^2 \Delta}{\pi} \frac{1}{\Delta^2 + (\nu - \nu_0)^2} \tag{3.28}$$

where r is the integrated fractional rms (over $-\infty$ to $+\infty$) of each Lorentzian and Δ is its Half-Width at Half Maximum (HWHM=FWHM/2). The power spectra are then displayed in a $\nu\, P_\nu$ plot. The frequency

$\nu_{\rm max}$ at which the $\nu\, P_\nu$ attains its maximum is

$$\nu_{\rm max} = \sqrt{\nu_0^2 + \Delta^2} = \nu_0 \sqrt{1 + \frac{1}{4Q^2}} \tag{3.29}$$

where $Q \equiv \nu_0/2\Delta$ is called quality factor. Note that $\nu_{\rm max} \geq \nu_0$: the difference is small for narrow features but becomes large in the case of broad ones. In Figure 3.8 we show an example of such a fit for two LMXB: XTE 1118+480 and 1E 1724–3045.

With this phenomenological modelization it is possible to use a limited number of fit components and compare the power spectra of different sources. But what is the *physical mechanism* responsible for the observed shape of the power spectra is still an open issue.

3.6 Quasi-Periodic Oscillations (QPO)

Quasi-Periodic Oscillations (QPO) are broad features observed in the power spectra of many X–ray sources. They are described in terms both of a Gaussian or a Lorentzian shape. As discussed above, the Lorentzian

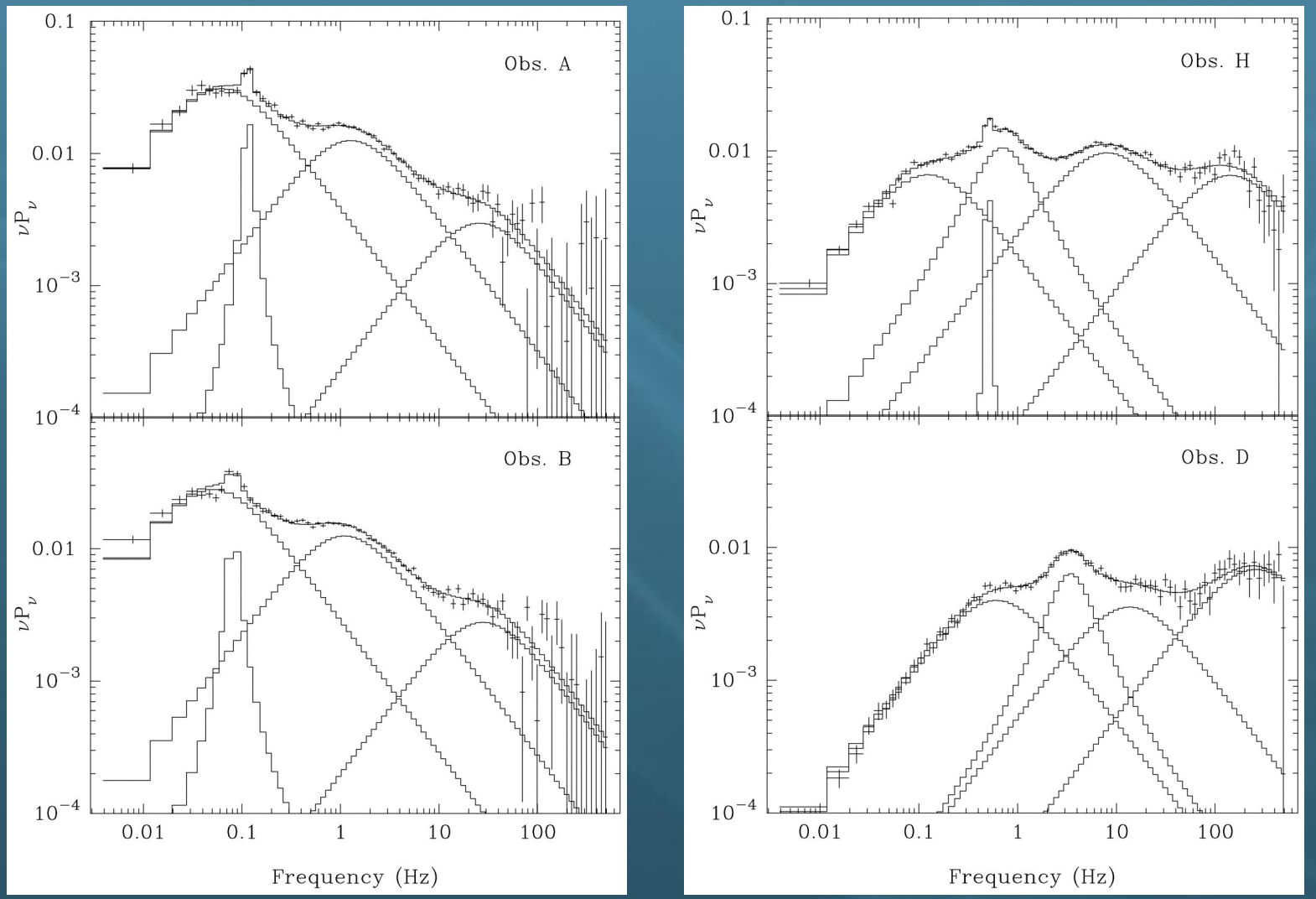


Figure 3.8: Power spectra in $\nu\ P_\nu$ form for two LMXB XTE 1118+480 (left) and 1E 1724–3045 (right). The presence of more power at high frequency for 1E 1724–3045 is interpreted as due to the presence of a neutron star in the system (while XTE 1118+480 should hosts a black hole)

shape has a physical basis as due to a shot noise process. The QPO can be therefore characterized by its centroid frequency LC, its width LW, and its normalization LN. Instead of LN t is customary to give the QPO percentage rms, defined as

$$\text{percentage rms} = 100 \sqrt{\frac{I}{\langle \text{RATE} \rangle}} \tag{3.30}$$

where I is the Lorentzian integral, defined as

$$I = \frac{\pi}{2} \text{LC} \times \text{LW}$$

and $\langle \text{RATE} \rangle$ is the source average count rate. Sometimes, for a QPO is given the quality factor Q, defined as

$$Q\text{-factor} = \frac{\text{LC}}{\text{LW}} \tag{3.31}$$

In Figure 3.9 we show a Lorentzian fit to a QPO observed in the low-mass X-ray binary and atoll source

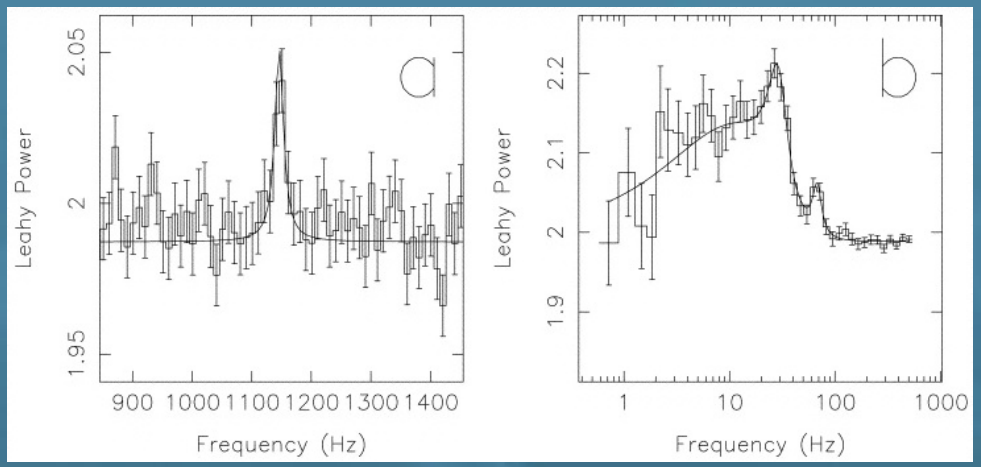


Figure 3.9: Typical Leahy normalized power spectra in the energy range of 2–18 keV. (a) The kHz QPO at 1150 Hz; (b) the complex high-frequency noise and the 67 Hz QPO. From Wijnands *et al.* 1998. *ApJ* 495, L39

4U 1735-44.

3.7 Analysis of a Coherent Signal

In X–ray astronomy (and in astronomy in general) the detection of coherent signal is quite common: for example, we detect periodic signal as due to star pulsations, pulse periods in pulsars, orbital modulations and eclipses, precession.

Two methods of analysis are used to examine data for evidence for periodic signals: FFT and epoch folding. In general, both techniques have certain advantages and disadvantages in their application. There latter are worsened both by the presence of gaps in the data and the large number of statistically independent frequencies which could, in principle, be examined.

Epoch folding is more sensitive to non sinusoidal pulse shapes encountered in X–ray astronomy. Furthermore, the technique is relatively insensitive to randomly occurring gaps in the data so long as the net pulse phase coverage is reasonably uniform. Epoch folding is, however, extremely computer time-consuming (even if now the increased CPU power of the current computers makes this issue less important).

The FFT, on the other hand, is extremely efficient. However, the FFT is difficult to interpret in the presence of gaps in the data (and in this case is better to use the Lomb-Scargle periodogram technique).

The epoch folding consists in folding the data modulo a trial period and then grouping the observations according to phase, in order to obtain a high signal-to-noise profile. The χ^2 statistics is then used to test

the high signal-to-noise profile for uniformity. This statistic is χ^2_{n-1} distributed, where n is the number of phase bins. By varying the trial period we can build a χ^2 vs period diagram and find out the one that gives the maximum χ^2 (that is, the rejection of the uniformity hypothesis). Because the χ^2 distribution resembles a triangular distribution (see Figure 3.10), it can be often well be fit by a Gaussian function, the mean of which may be considered the "*best*" coherent period P present in the data. The FWHM of the χ^2 distribution should be of the order of $\sim P^2/T$, where T is the total elapsed time of the observation. Of course this method works if there are not intrinsic period variations, like the one due to orbital motions. In this case it is necessary to perform a time transformation that makes the signal coherent. This transformation is called a *timing model*. The timing model predicts a model profile, or *template*, that is correlated to the *average* profile so that a phase offset can be determined. When multiplied by the instantaneous pulse period, that phase yields a time offset that can be added to a high-precision reference point on the profile (for example, the edge of the profile) to create the time-of-arrival or TOA, as shown in Figure 3.11. The general procedure to derive information on the source from the measured TOAs is

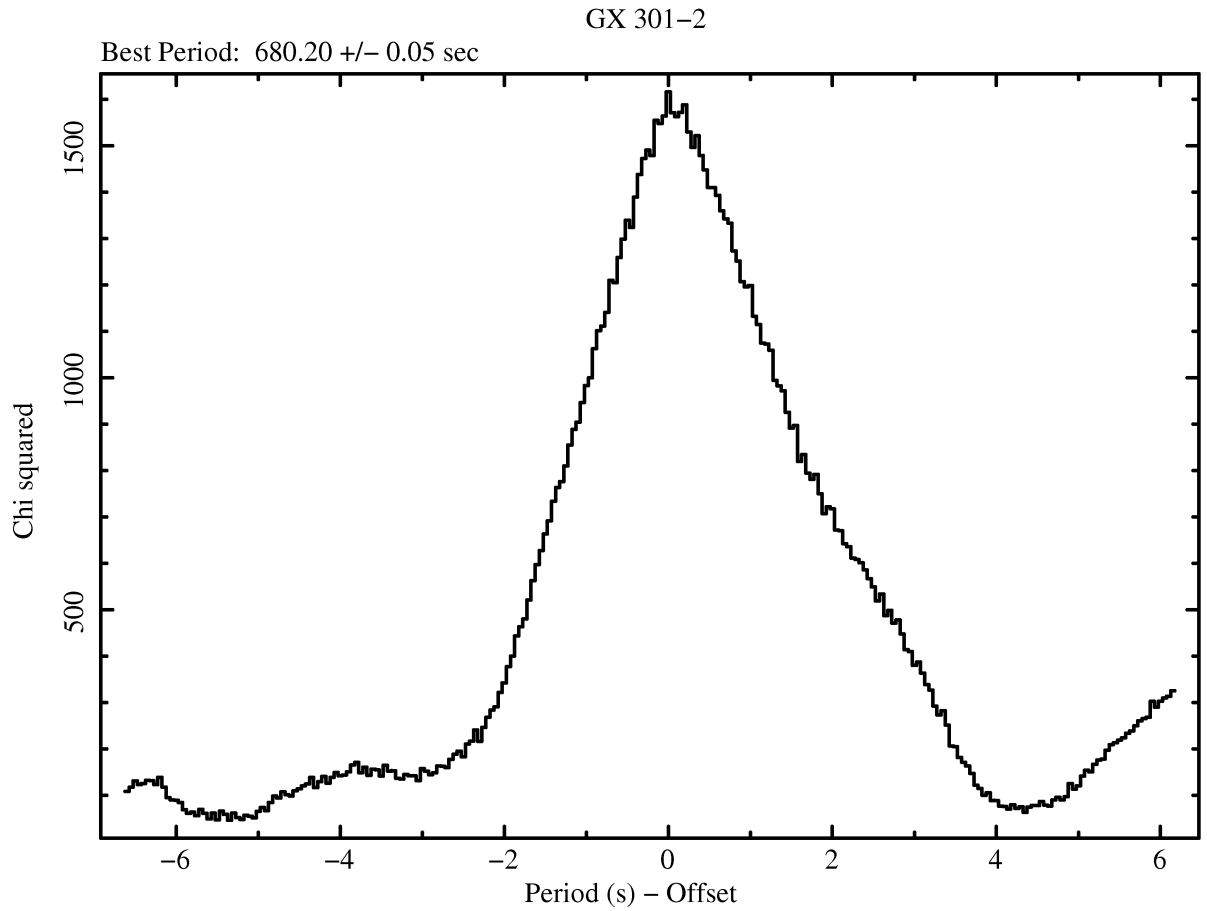


Figure 3.10: Pulse period of the X–ray binary pulsar GX 301–2 obtained by means of the epoch folding technique

depicted in Figure 3.12.

The TOA of the pulse number n is, by definition,

$$t_n = t_0 + n\,P \tag{3.32}$$

where t_0 is a reference time (usually the start of the observation). When including intrinsic period variations we can perform a Taylor expansion in P and write down

$$t_n = t_0 + n\,P + \frac{1}{2}n^2\,P\dot{P} + \frac{1}{6}n^3\,P^2\ddot{P} + \cdots \tag{3.33}$$

Equation (3.33) can be inverted and expressed in terms of the pulse phase φ at time $(t - t_0)$

$$\varphi = \varphi_0 + f_0(t - t_0) + \frac{1}{2}\dot{f}(t - t_0)^2 + \frac{1}{6}\ddot{f}(t - t_0)^3 + \cdots \tag{3.34}$$

where $f_0 = 1/P$ and f is the frequency. The precision with which a TOA can be determined is approximately equal to the duration of a sharp pulse feature (e.g., the leading edge) divided by the signal-to-

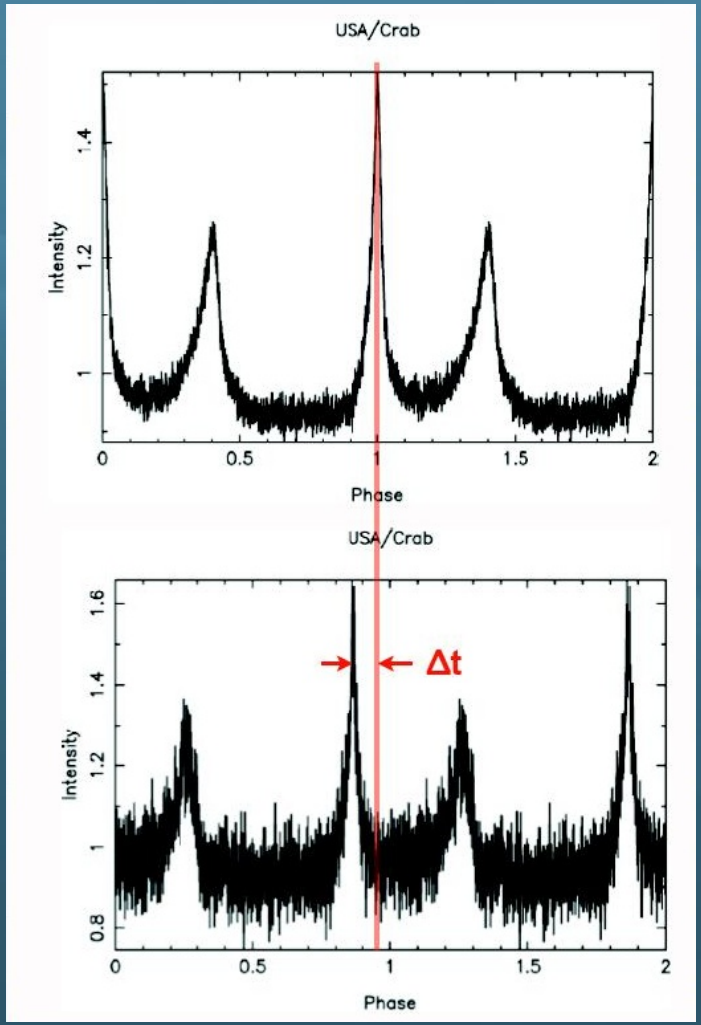


Figure 3.11: Measuring the phase shift with respect to a template profile

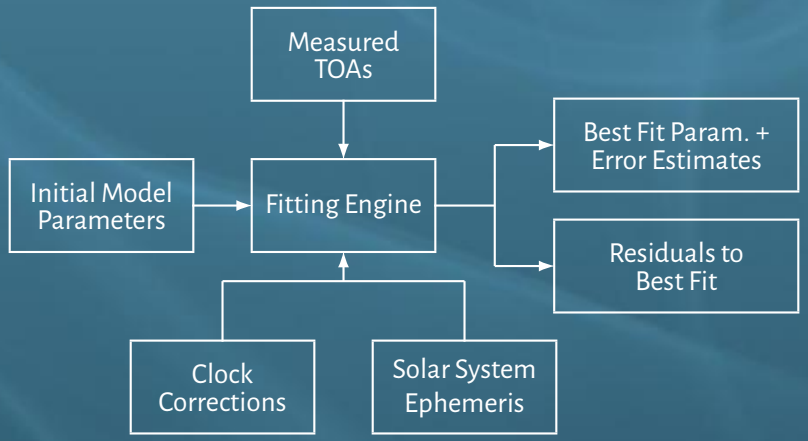


Figure 3.12: General procedure to derive information on the source from the measured TOAs

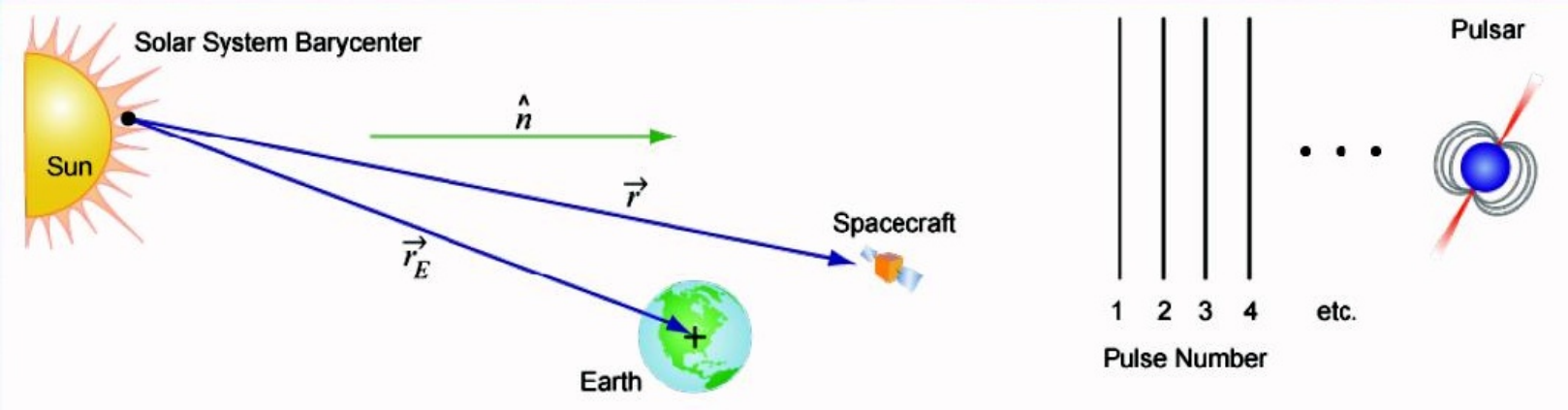


Figure 3.13: The time of arrivals of pulses from celestial objects are referenced to the nearly inertial reference frame of the Solar System barycenter

noise ratio of the average profile. It is usually expressed in terms of the width of the pulse features W_{f} in units of the period P, the pulse period P, and the signal-to-noise ratio SNR such that $\sigma_{\mathrm{TOA}} \propto W_{\mathrm{f}} P/\mathrm{SNR}$. Therefore strong, fast pulsars with narrow pulse profiles provide the best arrival times.

Before proceeding, it is better to eliminate the very first cause of period variations: the motion of the Earth and/or of the spacecraft. This is done by referencing the TOAs to a nearly inertial reference frame: the Solar System barycenter (see Figure 3.13).

The variation of the TOAs due to the orbital motion of the pulsar in the binary system can be written as

an addition term in Equation (3.33) of the form

$$\cdots + \frac{a_x \sin i}{c} F(e, \omega, \tau, \theta) \tag{3.35}$$

where $a_x \sin i$ is the projected semi-major axis of the pulsating star orbiting with an inclination angle i in the binary system. The function $F(e, \omega, \tau, \theta)$ represents the eccentric orbit of the pulsar around the center of mass of the binary system, where e is the eccentricity, ω the longitude of the periastron, τ the time of periastron passage, and $\theta \equiv 2\pi(t - \tau)/P_{\text{orb}}$ is the mean anomaly (see Figure 3.14 for the definition of the orbital parameters). In particular, we have that

$$F(e, \omega, \tau, \theta) = (1 - e^2) \frac{\sin(\upsilon + \omega)}{1 + e \cos \upsilon} \tag{3.36}$$

where the true anomaly υ can be calculated from the observed mean anomaly θ using the relations

$$\tan \frac{\upsilon}{2} = \sqrt{\frac{1+e}{1-e}} \tan \frac{E}{2}; \qquad E - e \sin E = \theta \tag{3.37}$$

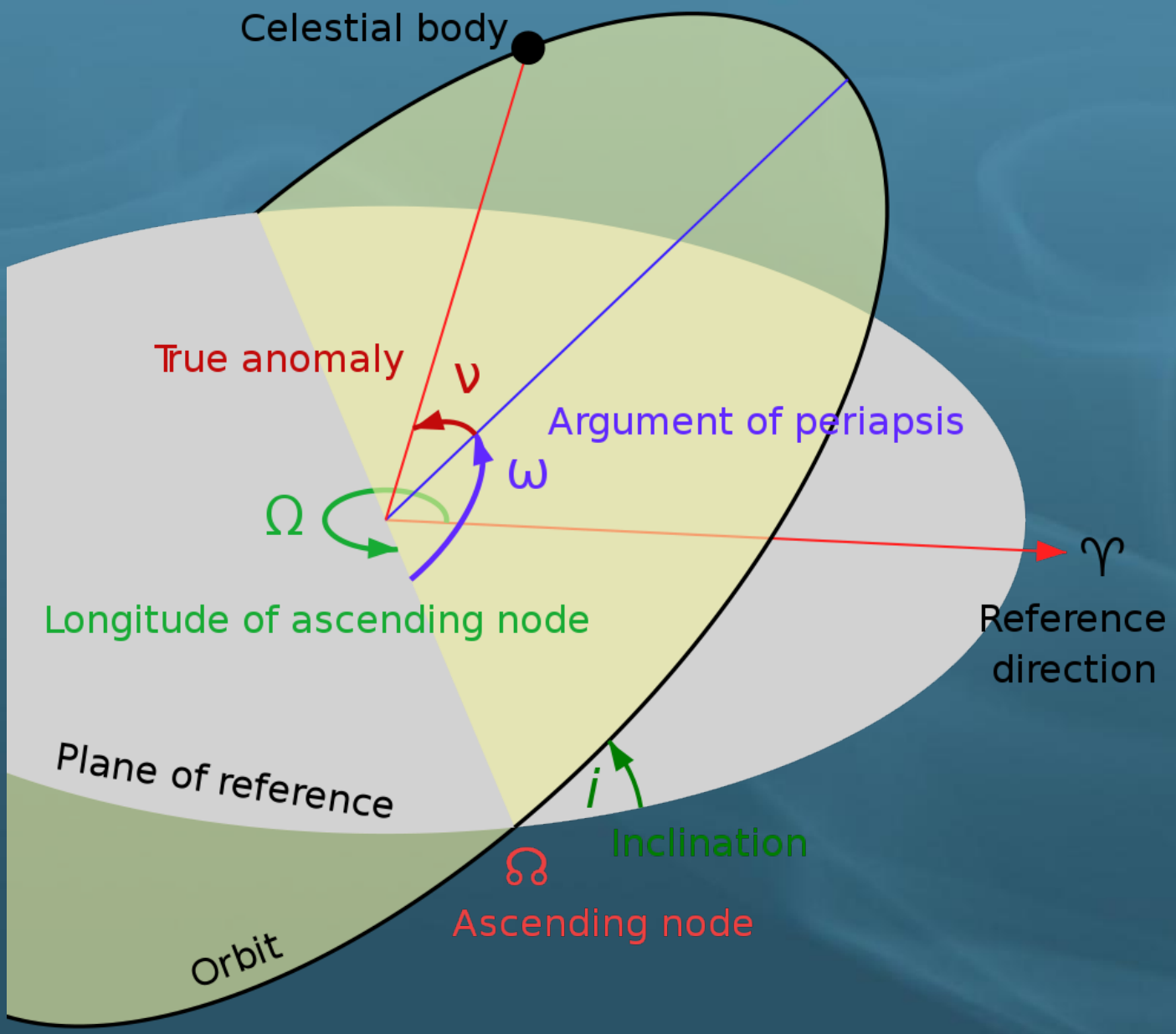


Figure 3.14: In this diagram, the orbital plane (yellow) intersects a reference plane (gray). The intersection is called the line of nodes, as it connects the center of mass with the ascending and descending nodes

The last relation is the so-called *Kepler Equation*. By fitting the series of TOAs with Equation (3.33) plus (3.35), we are able to obtain the pulse period P at time t_0, its time derivatives $\dot{P}$ and $\ddot{P}$, together with the orbital parameters of the pulsar $a \sin i, e, \omega, \tau$ and P_{orb}. An example of the Doppler curve, representing the pulse delay times due to the light transit across the binary orbit, is shown in Figure 3.15 for the X–ray binary pulsar Hercules X–1.

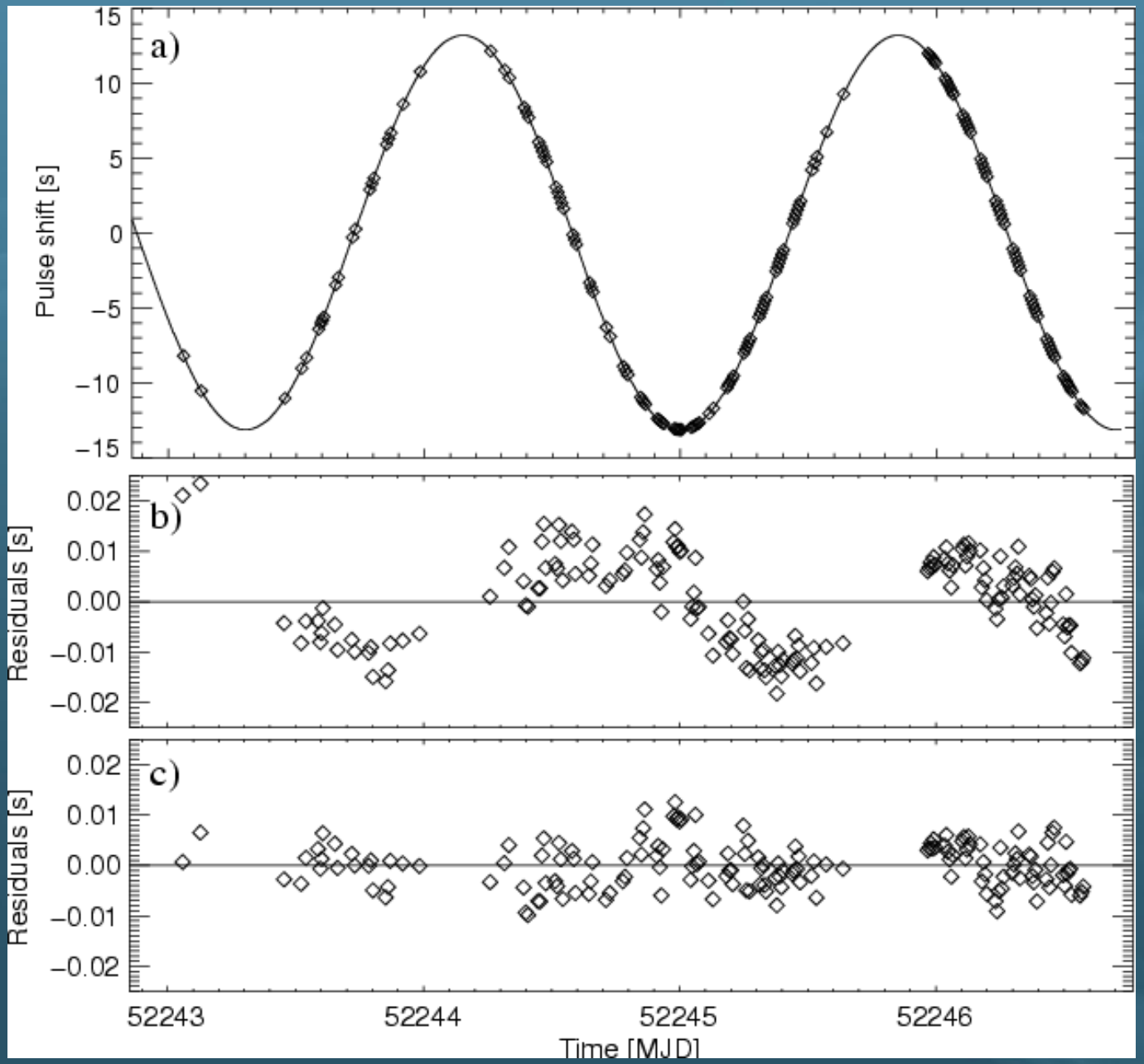


Figure 3.15: Delays of the TOA in Her X–1 due to its orbital motion and theoretical sine curve for the 35 day orbital period. The residuals refer to different orbital parameters solutions. From Staubert *et al.* 2009

Bibliography

Reference Textbooks

- **Random Data: Analysis and Measurement Procedures**, Bendat J.S. and Piersol A.G., Wiley-Interscience (2010)
- **Fourier Transformation for Pedestrian**, Butz T., Springer-Verlag (2006)
- **Handbook of Pulsar Astronomy**, Lorimer D.R. and Kramer M., Cambridge University Press (2004)
- **The Fourier Transform and its Applications**, Osgood Brad G., 30 video lectures (about 50 min each) available online from http://see.stanford.edu/SEE/courseinfo.aspx?coll=84d174c2-d74f-493d-92ae-c3f45c0ee091
- **Numerical Recipes: The Art of Scientific Computing**, Press W.H., Flannery B.P., Teukolsky S.A. and Vetterling W.T. Cambridge University Press (2007) available online from http://www.nr.com
- **Fourier Techniques in X-ray Timing**, van der Klis M., in *Timing neutron Stars*, H. Ögelman and E.P.J. van den Heuvel (eds), Kluwer Academic Pub., p.27–69 (1989)

In-depth Readings

Belloni T, Psaltis D., van der Klis M. 2002. *A unified description of the timing features of accreting X–ray binaries.* ApJ 572, 392

Davies S.R. 1990. *An improved test for periodicity.* MNRAS 244, 93

Harris F.J. 1978. *On the use of windows for harmonic analysis with the discrete Fourier transform*. Proc. IEEE 66, 51

Jerri A.J. 1977. *The Shannon sampling theorem – Its various extensions and applications: A tutorial review.* Proc. IEEE 65, 1565

Leahy D.A., Darbro W., Elsner R.F., Weisskopf M.C., Sutherland P.G., Kahn S., Grindlay J.E. 1983. *On searches for pulsed emission with application to four globular clusters X–ray sources: NGC 1851, 6441, 6624 and 6712.* ApJ 266, 160

Lomb N.R. 1976. *Least-squares frequency analysis of unequally spaced data*. Astrophys. Space Sci. 39, 447

Nowak M.A., Vaughan B.A., Wilms J., Dove J.B., Begelman M.C. 1999. *RXTE observation of Cygnus X–1: II. Timing analysis.* ApJ 510, 874

Scargle J.D. 1982. *Studies in astronomical time series analysis. II. Statistical aspect of spectral analysis of unevenly spaced data*. ApJ 263, 835

Shannon C.E. 1998. *Communication in the presence of noise.* Proc. IEEE 86, 447

Staubert R., Klochkov D., Wilms J. 2009. *Updating the orbital ephemeris of Hercules X-1: Rate of decay and eccentricity of the orbit.* A&A 500, 883

Vaughan B.A., van der Klis M., Wood K.S., Norris J.P., Hertz P., Michelson P.F., van Paradijs J., Lewin W.H.G., Mitsuda K., Penninx W. 1994. *Searches for millisecond pulsations in low-mass X-ray binaries. II.* ApJ 435, 362

Where to find these class notes?

The class notes of this Course, both in slide format and booklet format, can be downloaded from the following web page

☞ http://www.iasfbo.inaf.it/~mauro/Didattica/Timing

Thank you for your attention…

…and PATIENCE!

We are all in the gutter,
but some of us are looking at the stars.

(Oscar Wilde)