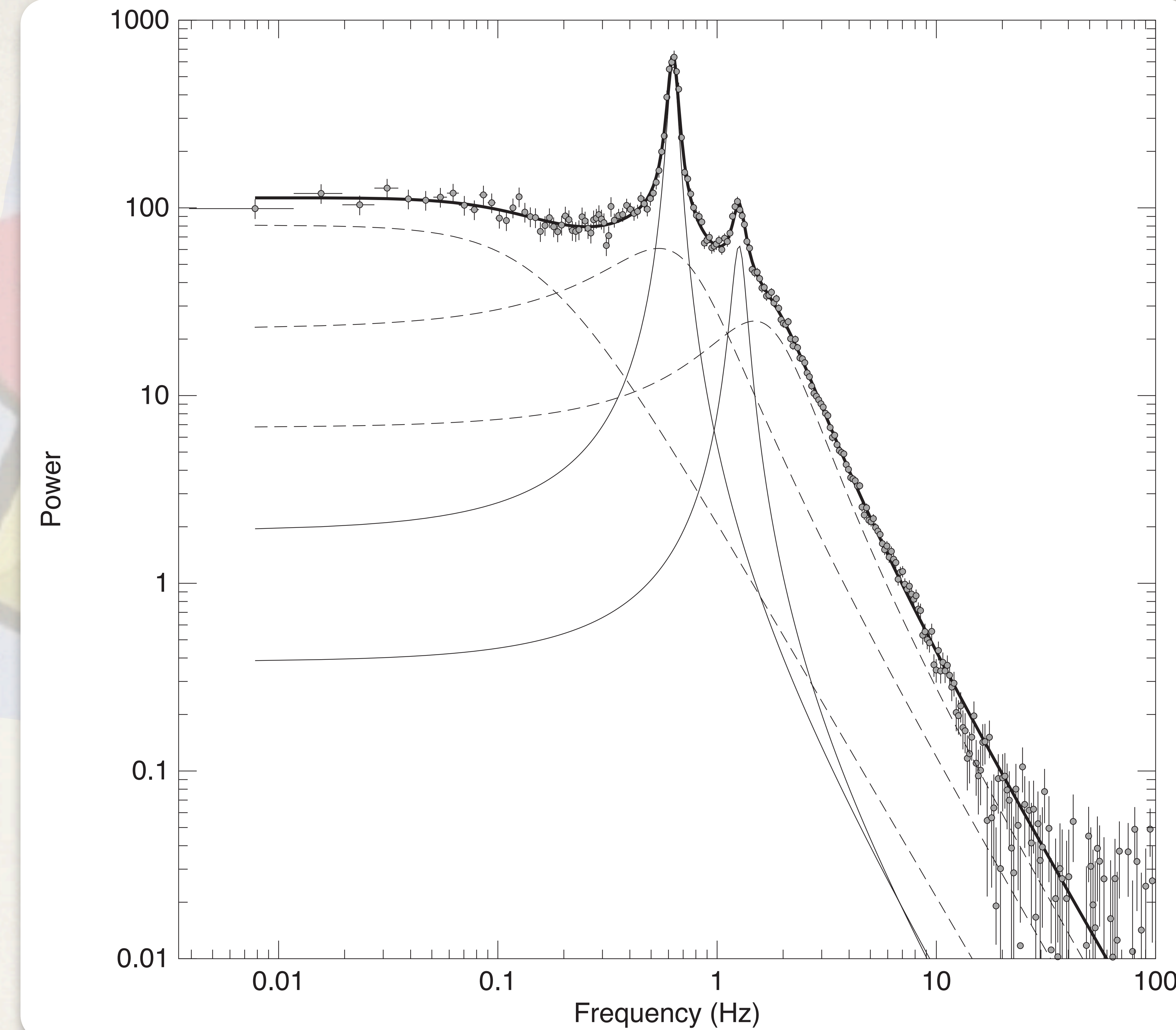
Time series analysis

Timing methods in X-ray Astronomy

Time series analysis

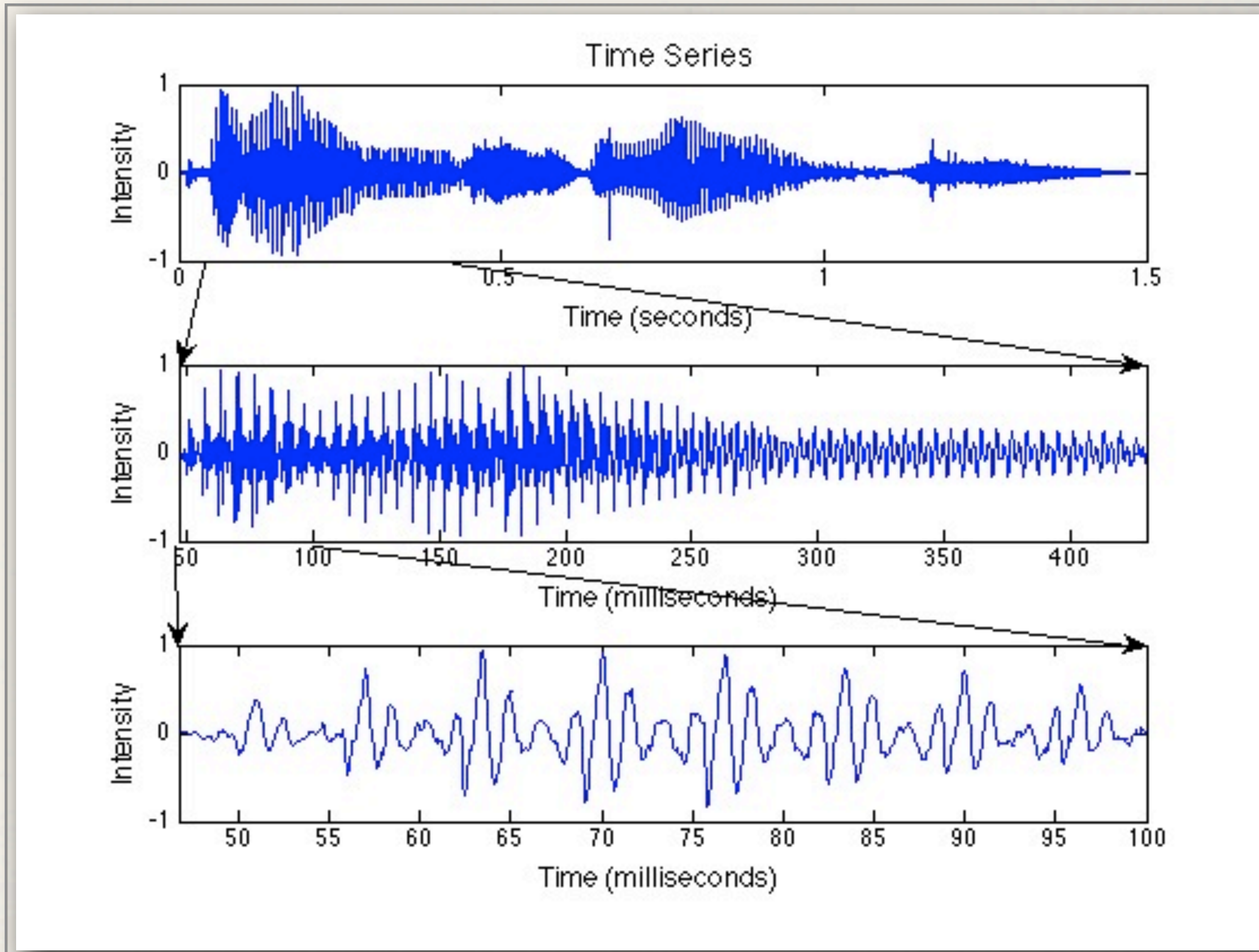
Tomaso Belloni (INAF - Osservatorio Astronomico di Brera)


1000
100
10
1
0.1
0.01
Power
0.01
0.1
1
10
100
Frequency (Hz)

Time series

$h(t)$ or $z(t)$

- 1-d sequence
- Many obvious examples
- Large literature on many fields



Time series and frequency

- Time is important
- Different representation
- Frequency domain
- Fourier analysis
- Heat conduction → discontinuity

Joseph Fourier (1768-1830)

Fourier transform

- Fourier transform equations
- h(t) and H(f): two representations of the same equation
- Linear transformation
- Decomposition on sine waves
- $\sin(2\pi f_0 t) \Leftrightarrow \delta(f-f_0)$
- Invariant to time shift

$$h(t) = \int_{-\infty}^{\infty} H(f) e^{2\pi i f t} df$$

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-2\pi i f t} dt$$

h(t)	H(f)
Real	H(-f) = [H(f)]*
Even	H(-f) = H(f) [even]
Odd	H(-f) = -H(f) [odd]
Real & Even	H(f) is real and even
Real & Odd	H(f) is imaginary and odd

Other basic properties

- Correlation

$$Corr(g,h) = \int_{-\infty}^{\infty} g(t+\tau)h(\tau)d\tau \iff G(f)H^*(f)$$

- Autocorrelation

Autocorrelation is the fourier transform of the power spectrum

$$Corr(g,g) = \int_{-\infty}^{\infty} g(t+\tau)g(\tau)d\tau \iff |G(f)|^2$$

- Parseval's theorem $$\int_{-\infty}^{\infty} |h(t)|^2 dt = \int_{-\infty}^{\infty} |H(f)|^2 df$$ Total power in the signal

One-sided vs. two-sided

- Power spectral density (PSD)

$$P_h(f) \equiv |H(f)|^2 \qquad -\infty < f < \infty$$

- One-sided

$$P_h(f) \equiv |H(f)|^2 + |H(-f)|^2 \qquad 0 \leq f < \infty$$

- If h(t) is real

$$P_h(f) \equiv 2|H(f)|^2$$

Recap

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-2\pi i f t} dt$$

- Fourier transform: decomposition on a base of sinusoids
- Sum of correlation with sinusoids
- h(t) extends from -∞ to +∞
- PSD over frequency gives signal power
- We have real signals...
- ... but we don't have either continuous or infinite signals

Discrete Fourier transform

- Sampled function: x_k (k=1,...,N), total length T [N numbers]

Discrete FT

$$a_j = \sum_{k=0}^{N-1} x_k e^{2\pi ijk/N}$$ (j=-N/2,...,N/2-1)

- Here times are t_k = kT/N, frequencies are j/T
- Time step: ΔT = T/N
- Frequency step: Δν = 1/T

$$x_k = \frac{1}{N} \sum_{j=-N/2}^{N/2-1} a_j e^{-2\pi ijk/N}$$

Inverse FT

Uncertainty principle (I)

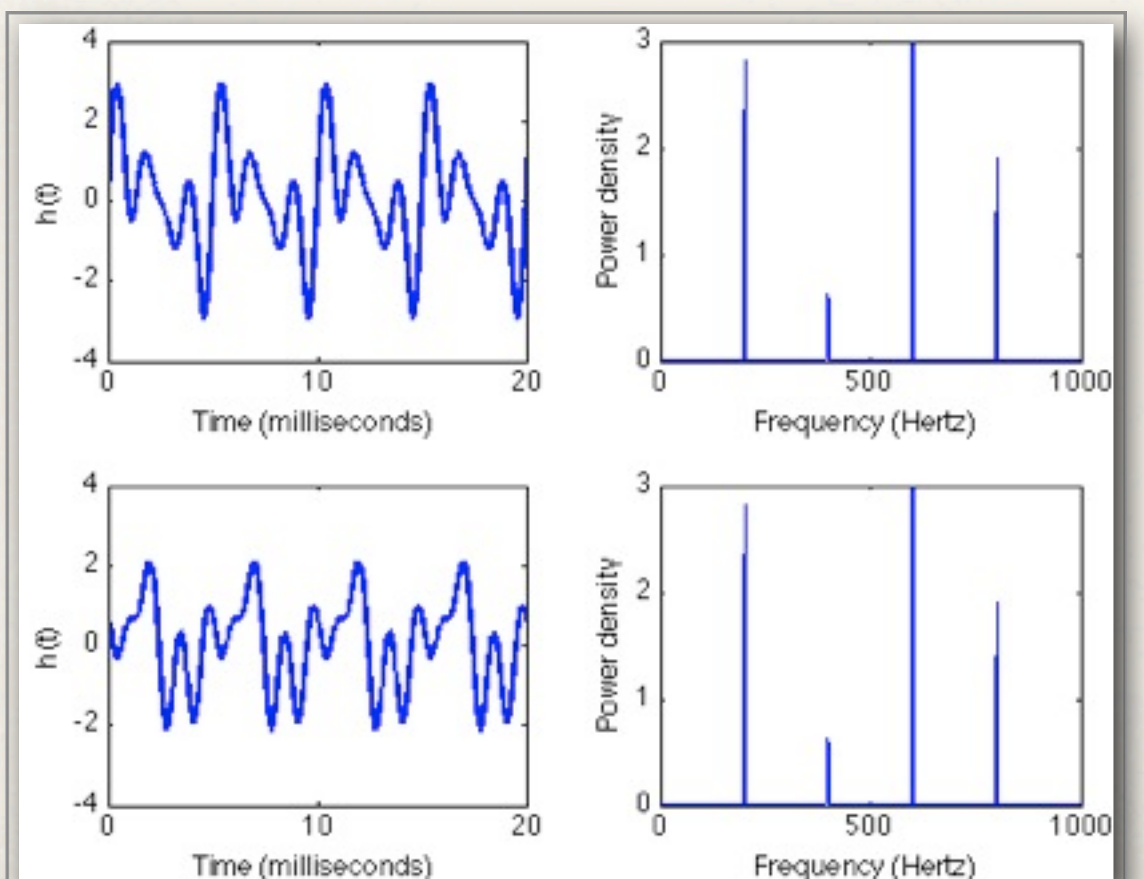
- Frequency resolution: $\Delta\nu = 1/T$
- Time resolution: T (length of the sample of N measurements)
- The longer your measurement, the higher your frequency resolution
- This is important in time-frequency analysis (non-stationary signals)
- Formal version of UP much more complex

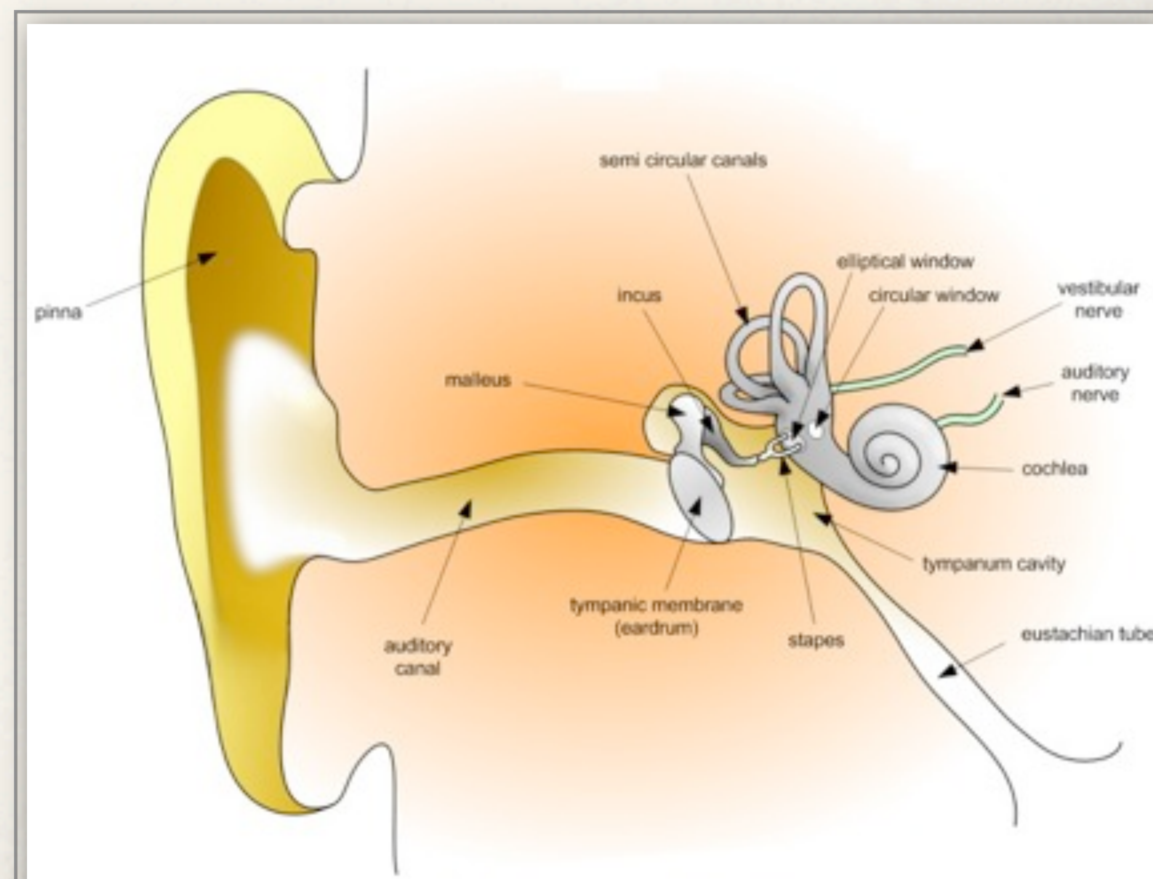
No loss of information

- N numbers in input - N numbers in output (for real signals H(-f) = [H(f)]*, but values are complex)
- Highest frequency: $\nu_{N/2} = \frac{1}{2}\frac{N}{T}$ Nyquist frequency
- Critical sampling of a sine wave is two sample points per cycle
- If you sample less, you get the wrong period (wait..)
- Notice that H(f) is complex for real input
- Also: $a_0 = \sum_{k=0}^{N-1} x_k e^{2\pi i 0 k/N} = \sum_k x_k \equiv N_{counts}$

Power density spectrum

- If we ignore the phases of the a_j's: $$P = \frac{2}{N_{phot}}|a|^2 \quad \text{(j=0,...,N/2)}$$
- Again, analogous to hearing system



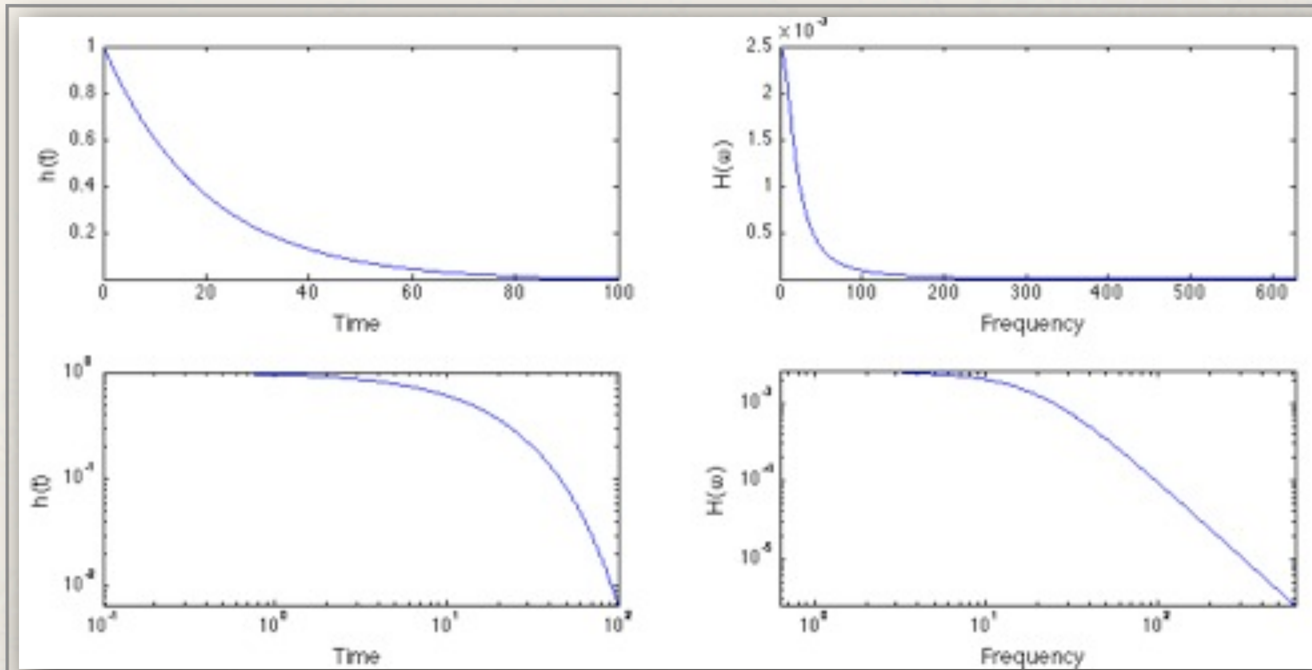


Power density spectrum

- An example: (continuous) transform of a one-sided exponential

$$h(t) = e^{-\lambda t} \qquad H(f) = \frac{1}{2\pi i f + \lambda} \equiv \frac{1}{i\omega + \lambda}$$

$$P(f) = |H(f)|^2 = \frac{1}{\omega^2 + \lambda^2}$$



Power density spectrum

- Non-linear transformation

$$x_k = y_k + z_k \Rightarrow \begin{aligned} a_j &= b_j + c_j \\ |a_j|^2 &= |b_j|^2 + |c_j|^2 + crossterms \end{aligned}$$

- If independent (random noise added), cross terms average out to zero

Finite duration and sampling

- How can one connect continuous and discrete FT?

$$a(\nu) = \int_{-\infty}^{\infty} h(t) e^{-2\pi i \nu t} dt \qquad\qquad a_j = \sum_{k=0}^{N-1} h_k e^{-2\pi i j k / N}$$

- Continuous time series: *h(t) [-∞,+∞]*

- Discrete time series: h_k *[k=0, ... , N-1]*

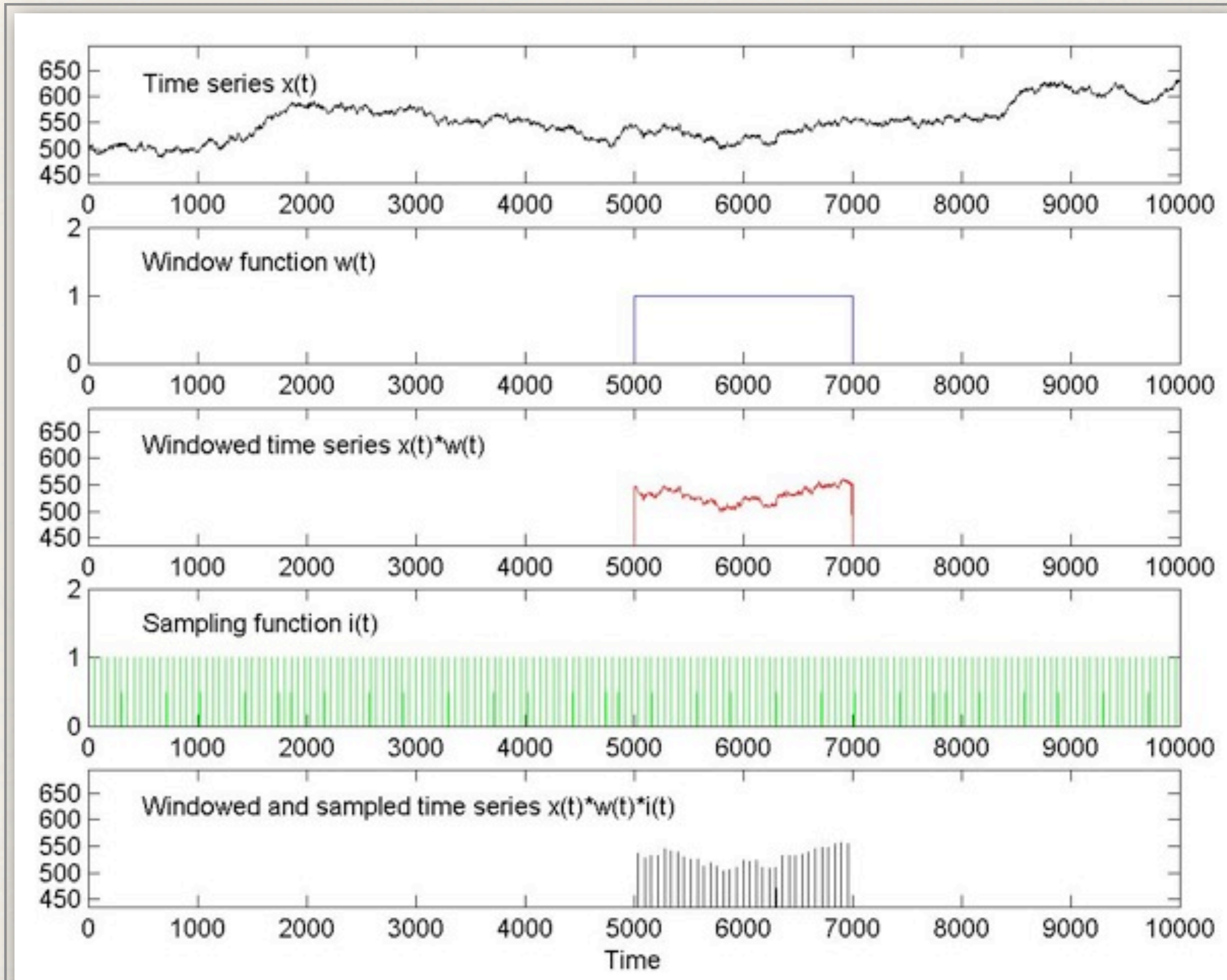
Finite duration and sampling

- We multiply $$h_k = h(t)w(t)i(t)$$

- *w(t)*: window function $$w(t) = \begin{cases} 1 & 0 \le t < T \\ 0 & otherwise \end{cases}$$

- *i(t)*: sampling function $$i(t) = \sum_{k=-\infty}^{\infty} \delta(t - \frac{kT}{N})$$

Finite duration and sampling



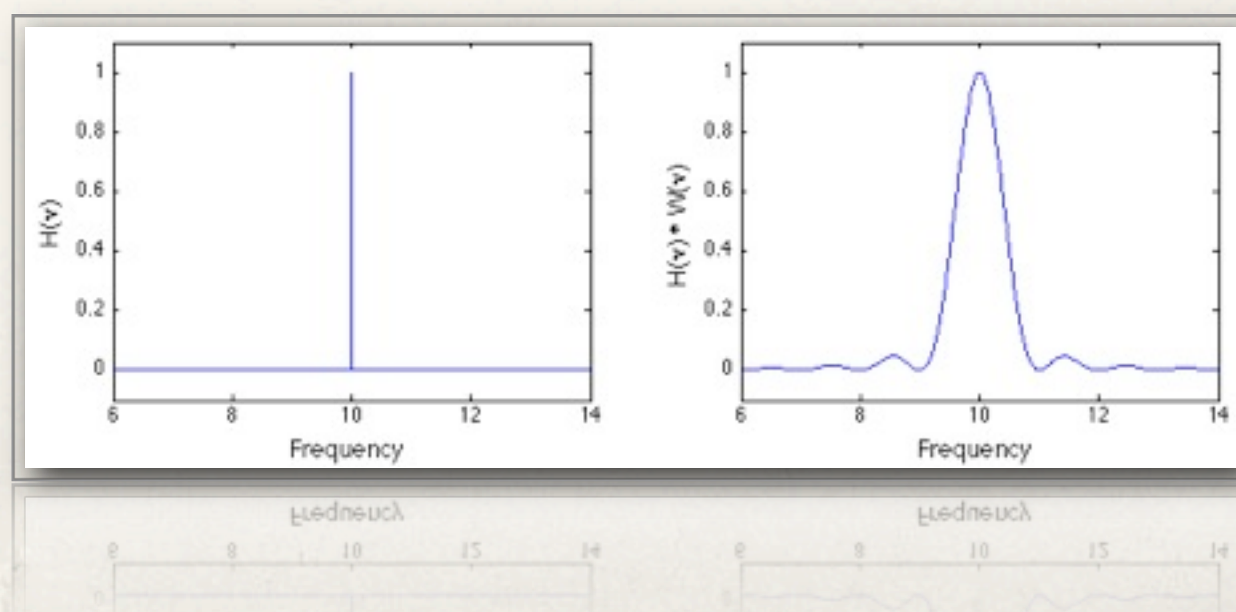
Convolution theorem: windows

- The transform of the product of two functions is the convolution of the transforms

$$x(t)y(t) \Longleftrightarrow a(\nu) * b(\nu) \equiv \int_{-\infty}^{\infty} a(\mu)b(\nu - \mu)d\mu$$

$$|W(\nu)|^2 \equiv \left| \int_{-\infty}^{\infty} w(t)e^{-2\pi\nu T} \right|^2 = \left| \frac{\sin \pi\nu T}{\pi\nu} \right|^2$$

Broadening of peaks



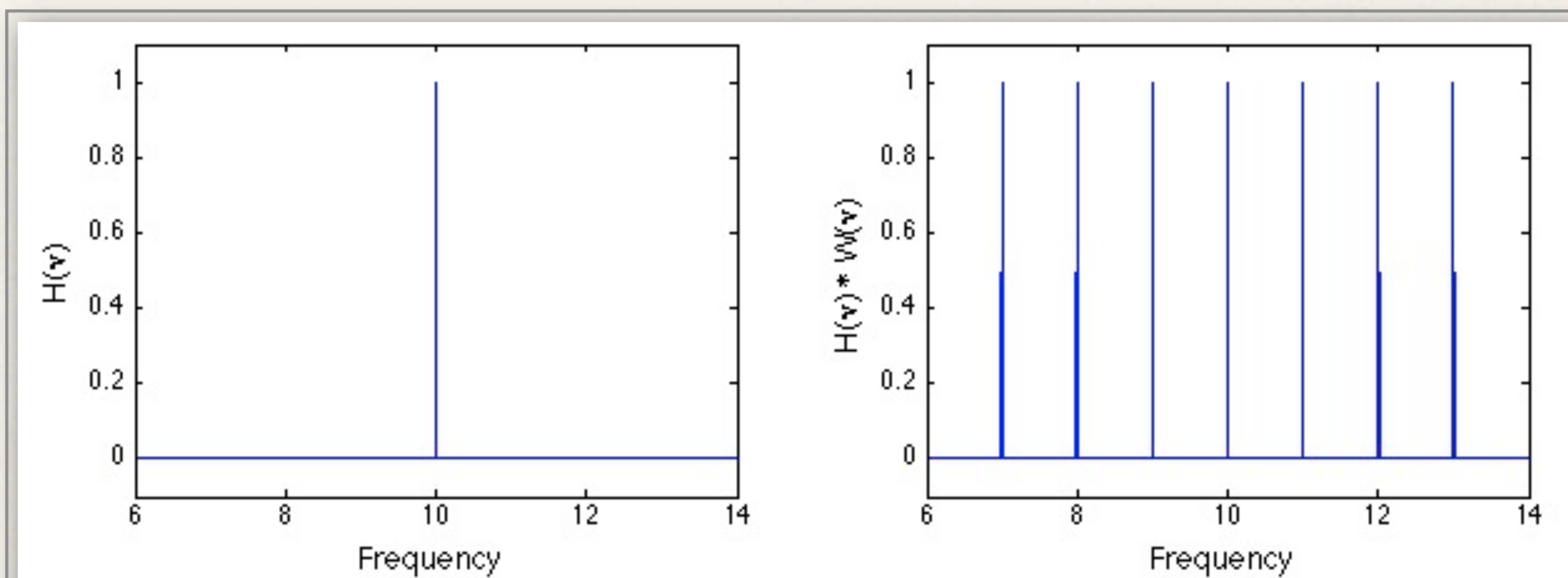
Convolution theorem: sampling

- The transform of the product of two functions is the convolution of the transforms

$$I(\nu) \equiv \int_{-\infty}^{\infty} i(t) e^{-2\pi\nu i t} dt = \frac{N}{T} \sum_{\ell=-\infty}^{\infty} \delta\left(\nu - \ell\frac{N}{T}\right)$$

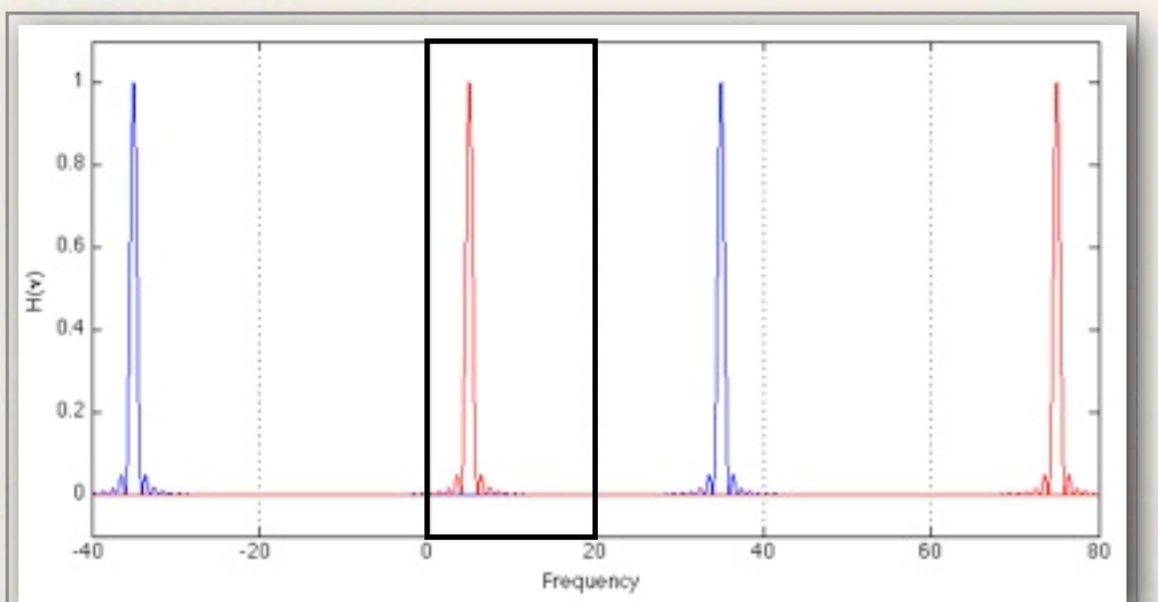
- Infinite series of δ functions, with spacing $N/T = 2\,\nu_{Nyq}$

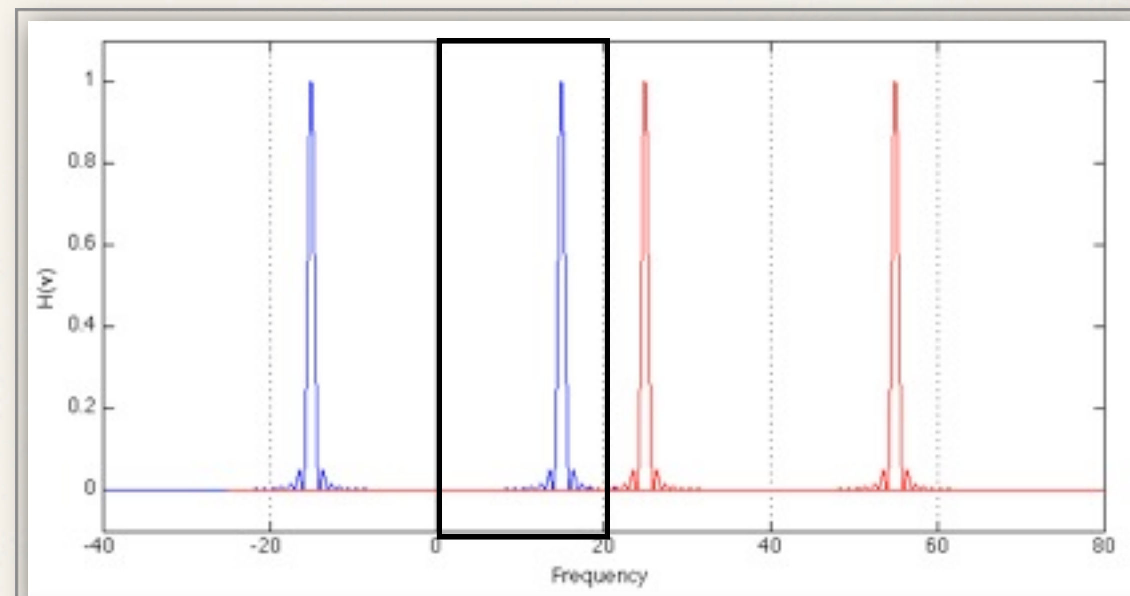
$>\nu_{Nyq}!$

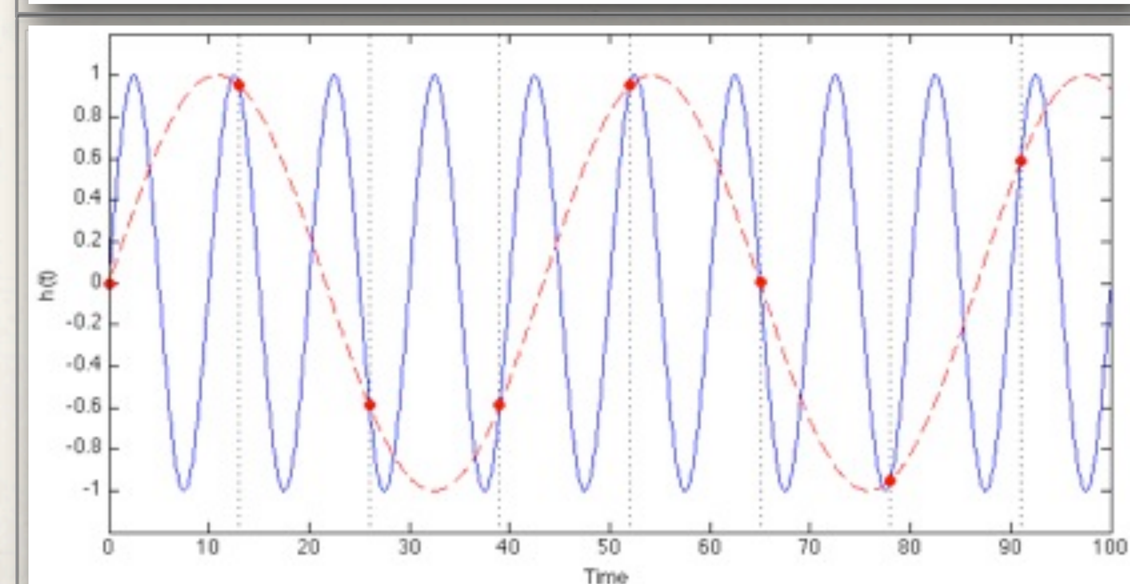


Aliasing

- FT is symmetric in frequency for a real signal
- Alias repeats it every $2\nu_{Nyq}$
- Problem is signal above ν_{Nyq}







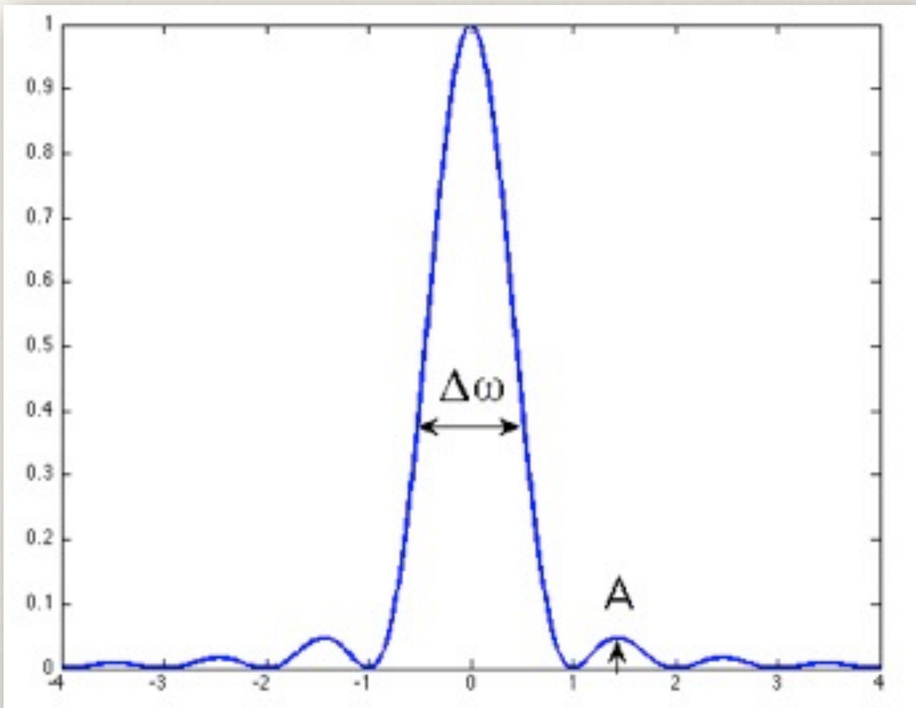
Summary of discrete FT effects

- WINDOW: broadening & sidebands
- SAMPLING: aliasing

- Aliasing not such a big problem for high-energy astronomy
- Binning, not sampling
- Suppression of high frequencies

Window effects

- Window effect is a problem:
  - It broadens delta peaks
  - It flattens the slopes of noise components (sidelobes)
- The longer the observation, the better

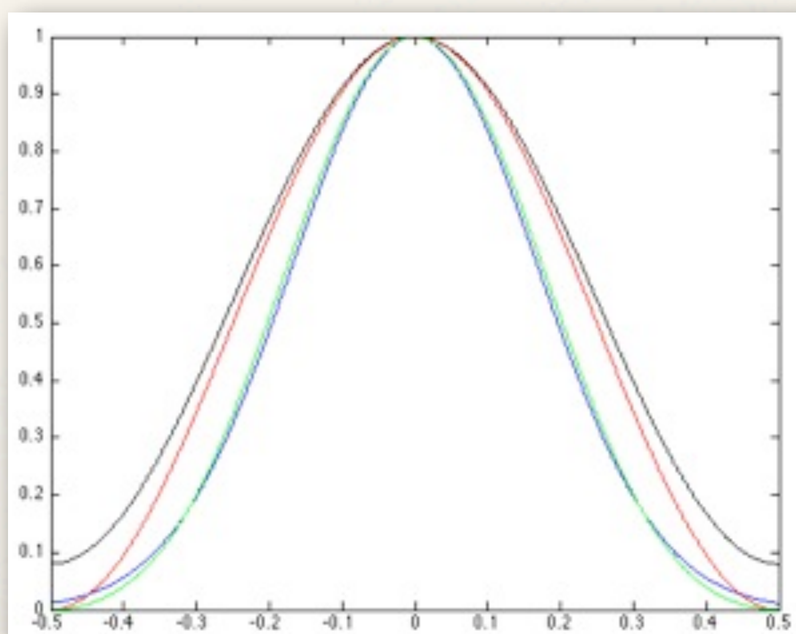


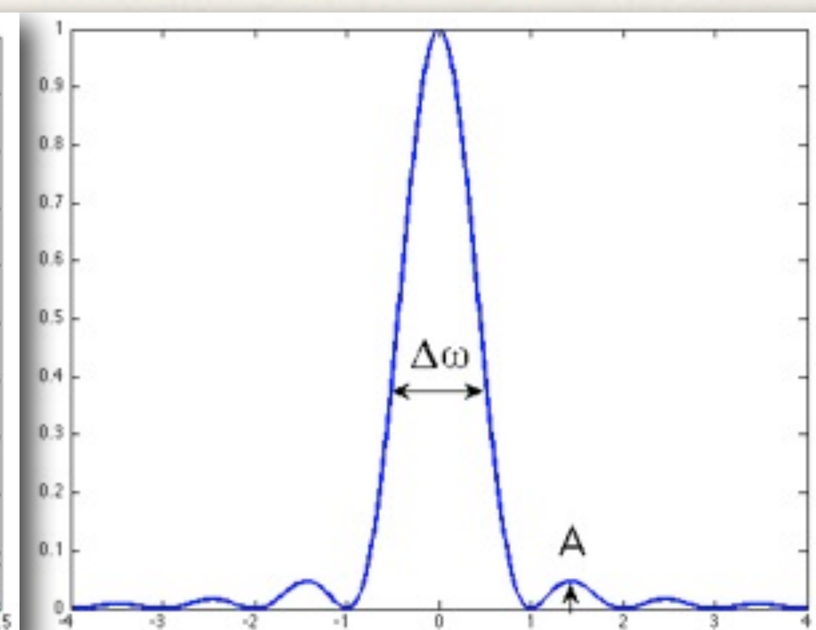
Window carpentry

- We can use different windows

Window	Δω	A	p	Function
Boxcar	0.89	-13db	2	1
Hamming	1.36	-43db	2	$0.54+0.46\cos(2\pi t)$
Gaussian	1.55	-55db	2	$\exp(-18t^2)$
Hanning	1.44	-32db	5	$\cos^2(\pi t)$
Blackman	1.68	-58db	5	$0.42+0.5\cos(2\pi t)+0.08\cos(4\pi t)$

- We lose some signal



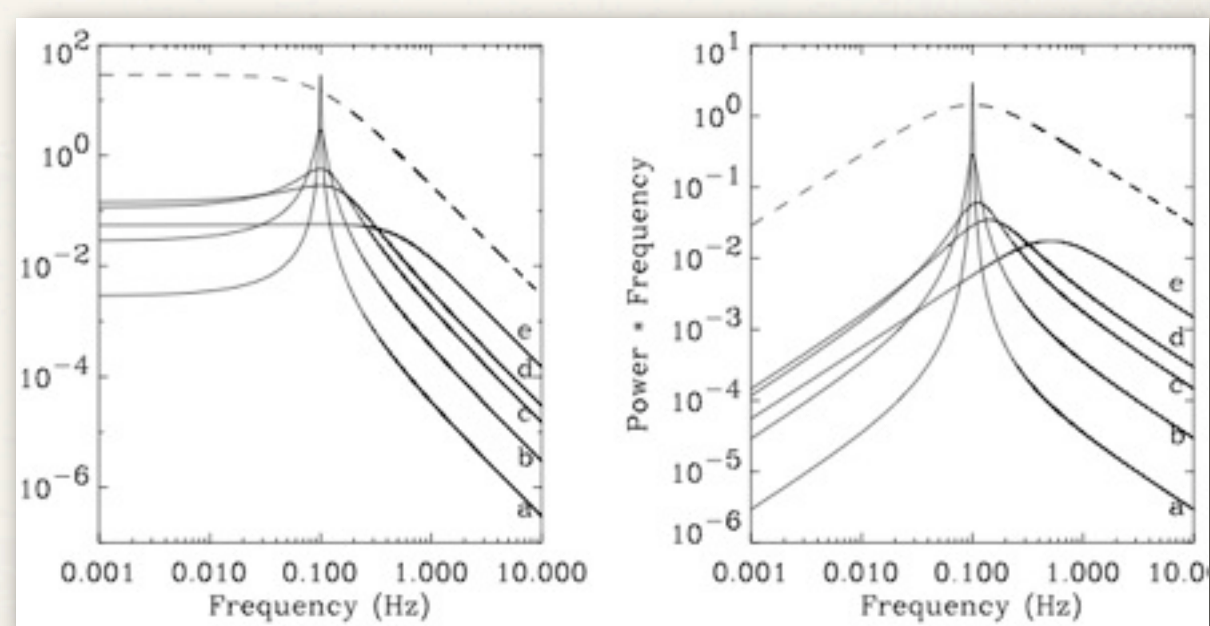


Power spectra units

- A power spectrum is in units of Hz^{-1}
- It scales with the square of the intensity: variance
- If we divide by the square of the intensity, we get the fractional variance (squared rms)
- The square root of its integral is the total fractional rms
- Useful to compare amount of variability

Power spectrum plots

- Multiply the power spectrum by the frequency
- Obtain a νP_ν representation
- Useful to see where the power per decade peaks
- Characteristic frequencies are peaks in νP_ν (later)



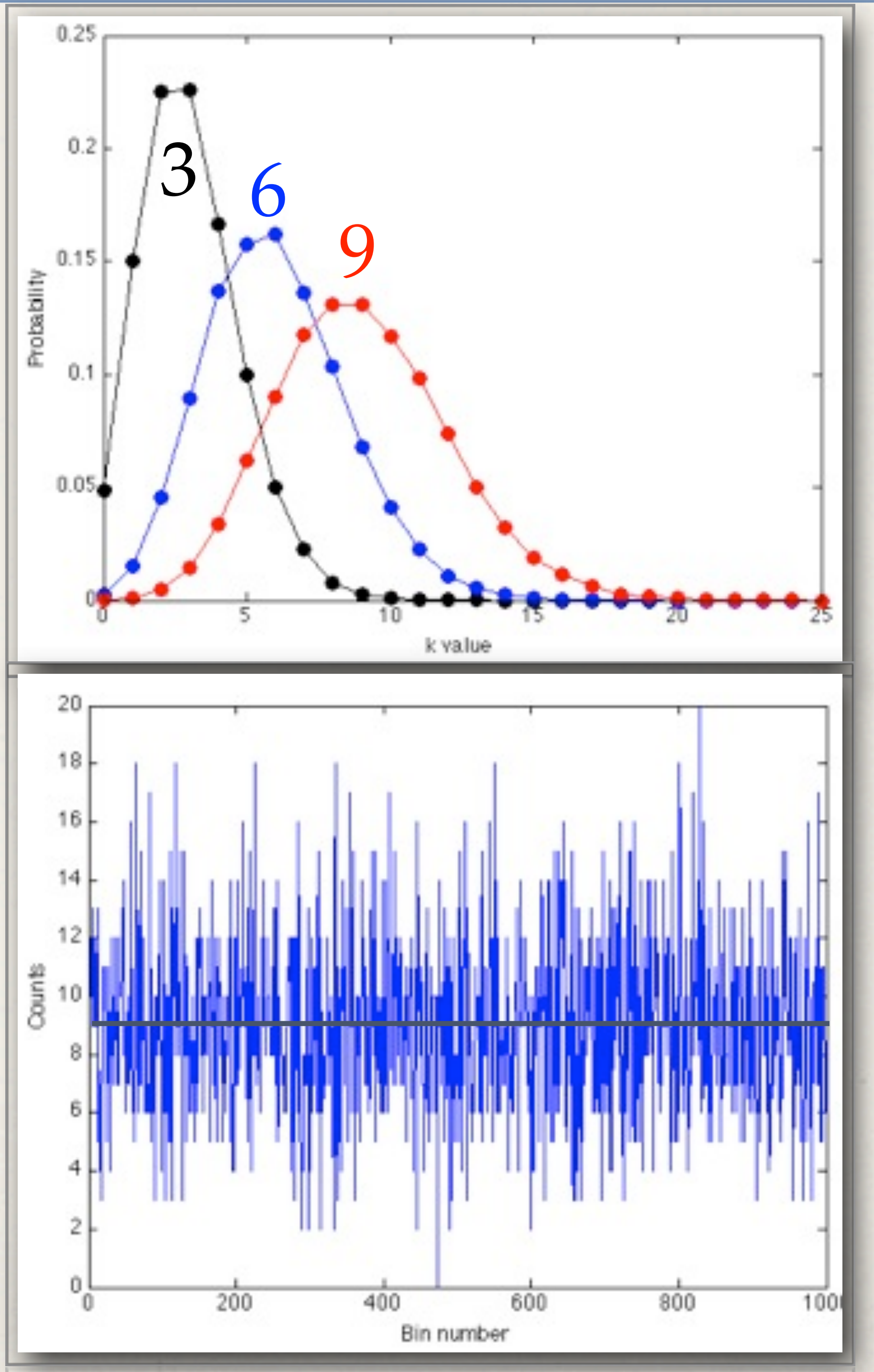
Timing methods in X-ray Astronomy

Tomaso Belloni (INAF - Osservatorio Astronomico di Brera)

Poisson noise effects

$$p(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

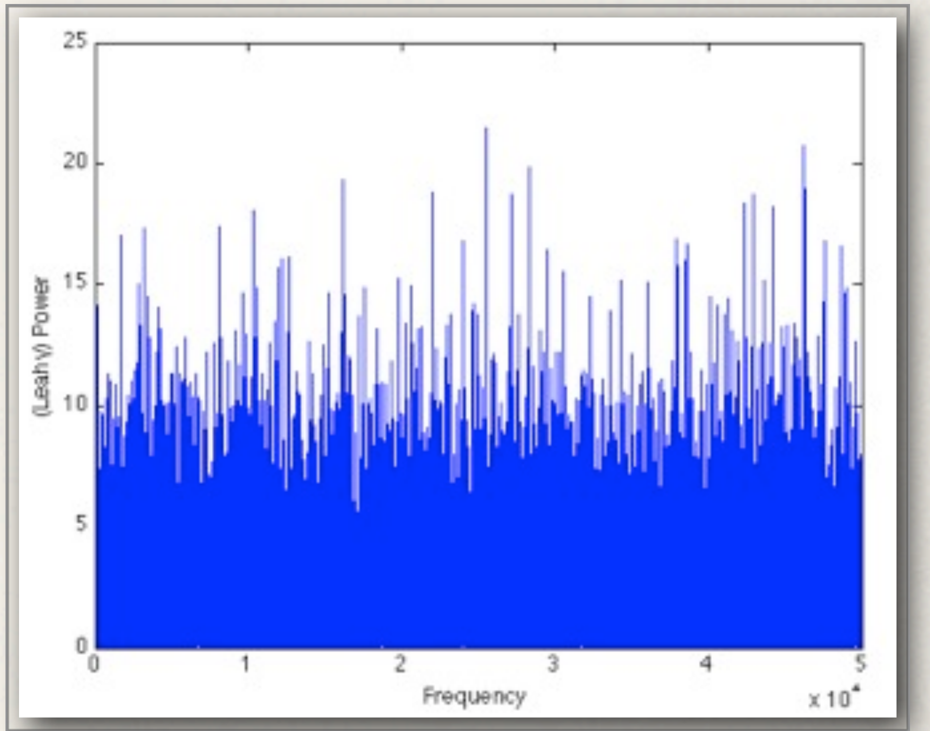
- Counting detector
- Counting noise
- Background negligible
- Independent arrival times
- Exponential waiting time between photons



Power spectrum normalization

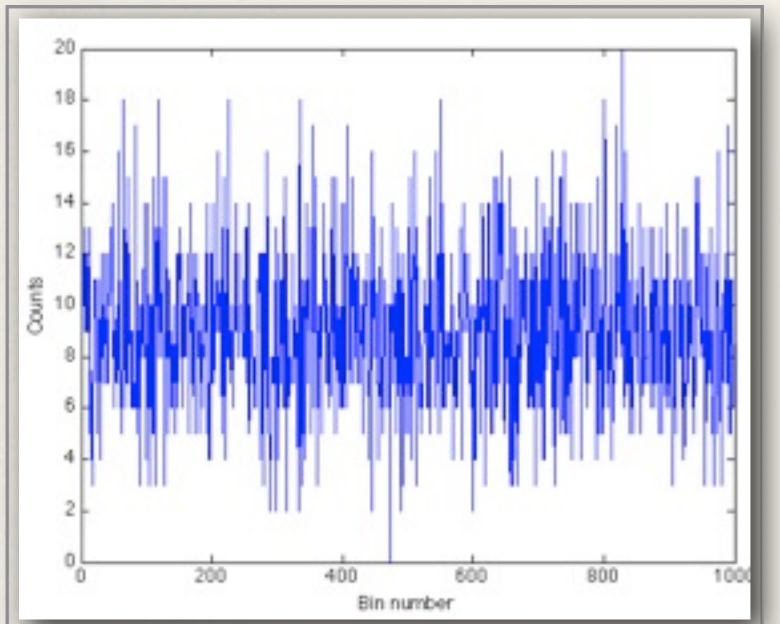
Leahy Norm.

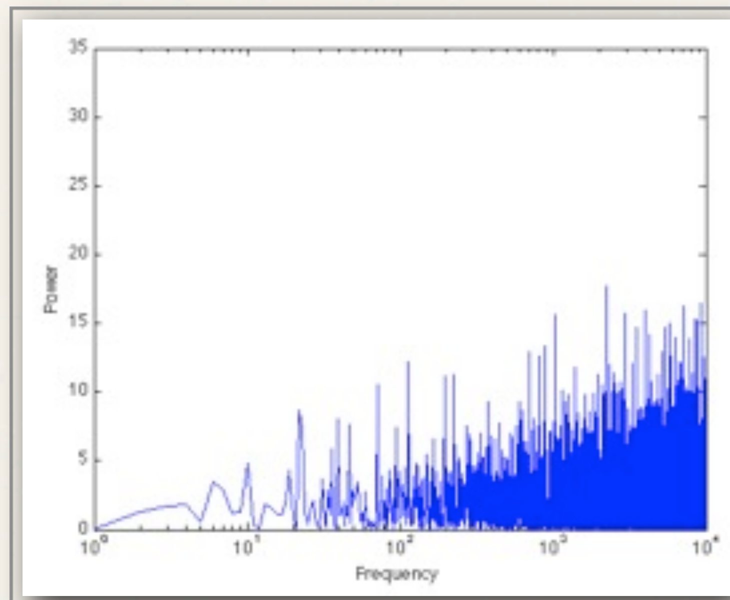
$$P = \frac{2}{N_{phot}} |a|^2$$

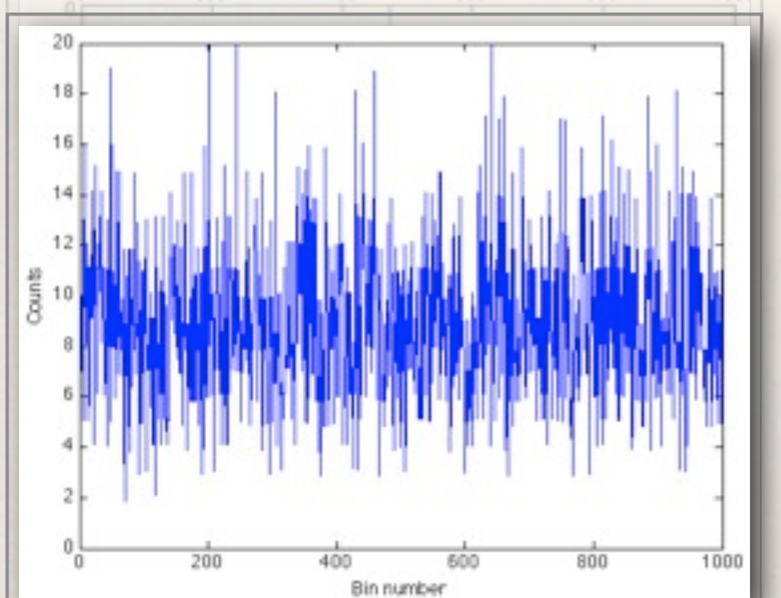


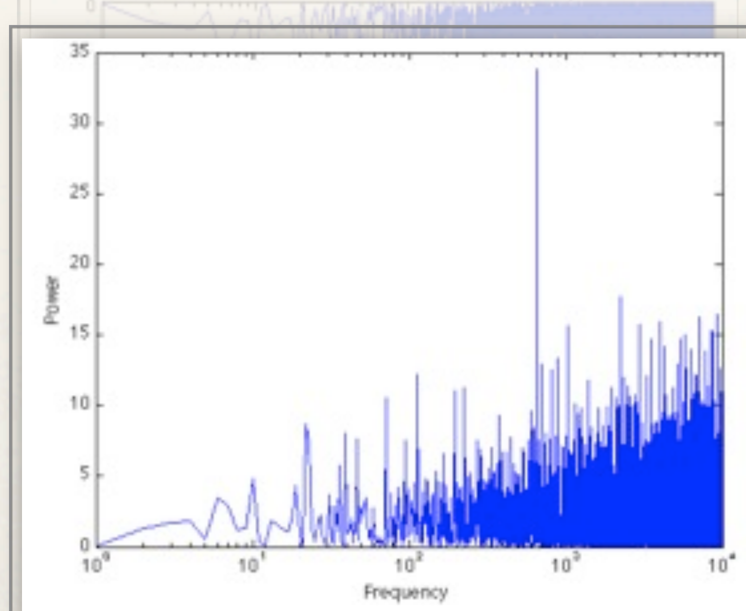
- With this choice, noise power a χ^2 with 2 d.o.f.
- Most noises do
- Average power is 2. I can calculate statistics
- Noise & signal independent: $P_{total} = P_{signal} + P_{noise}$
- Not always so... (count rate!)
- More complex: deadtime

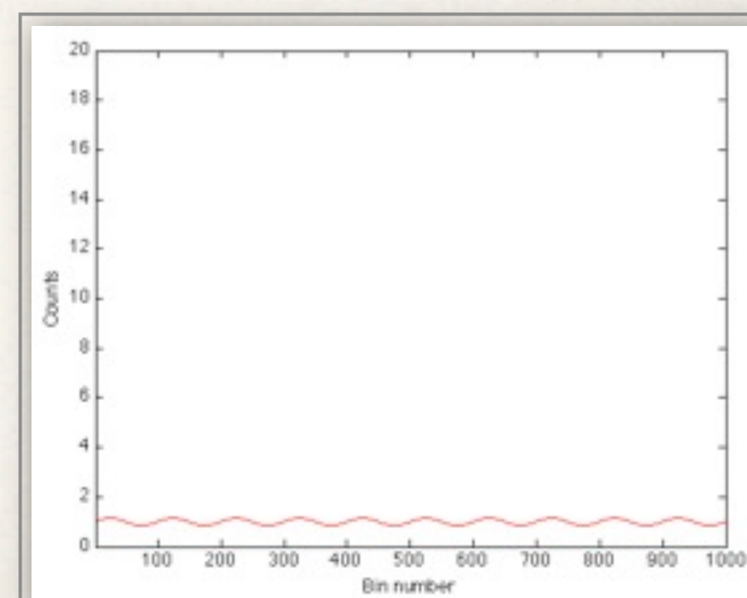
Power of Power spectrum





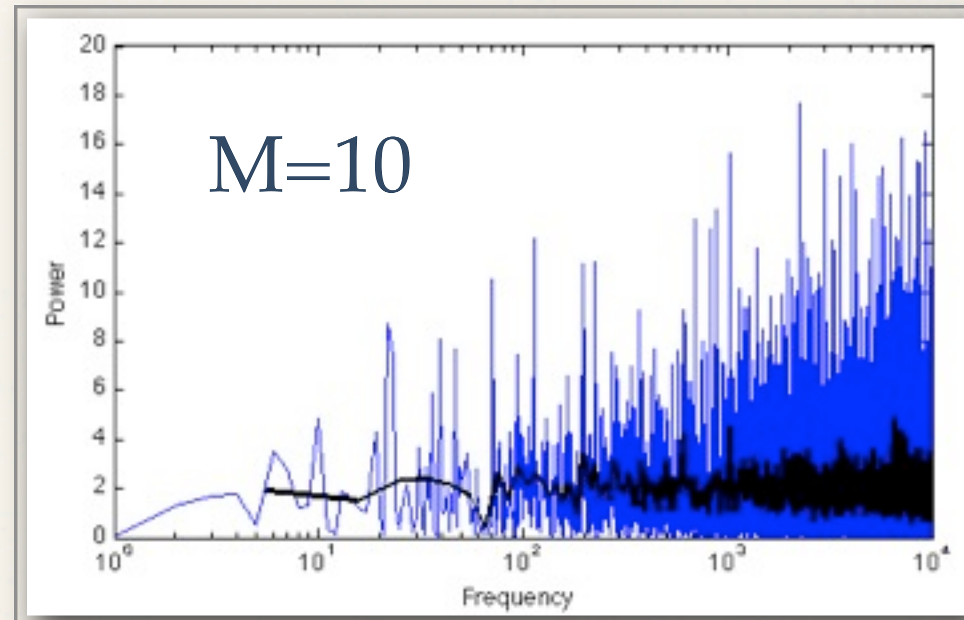






Noisy noise

- Power spectrum of noise is very noisy! $\sigma_{P_j} =< P_j >= 2$
- Increasing length or Δt not useful
- Two ways out:
  - a) Frequency rebinning by M
  - b) Time slicing by W and averaging powers



2 with 2MW dof

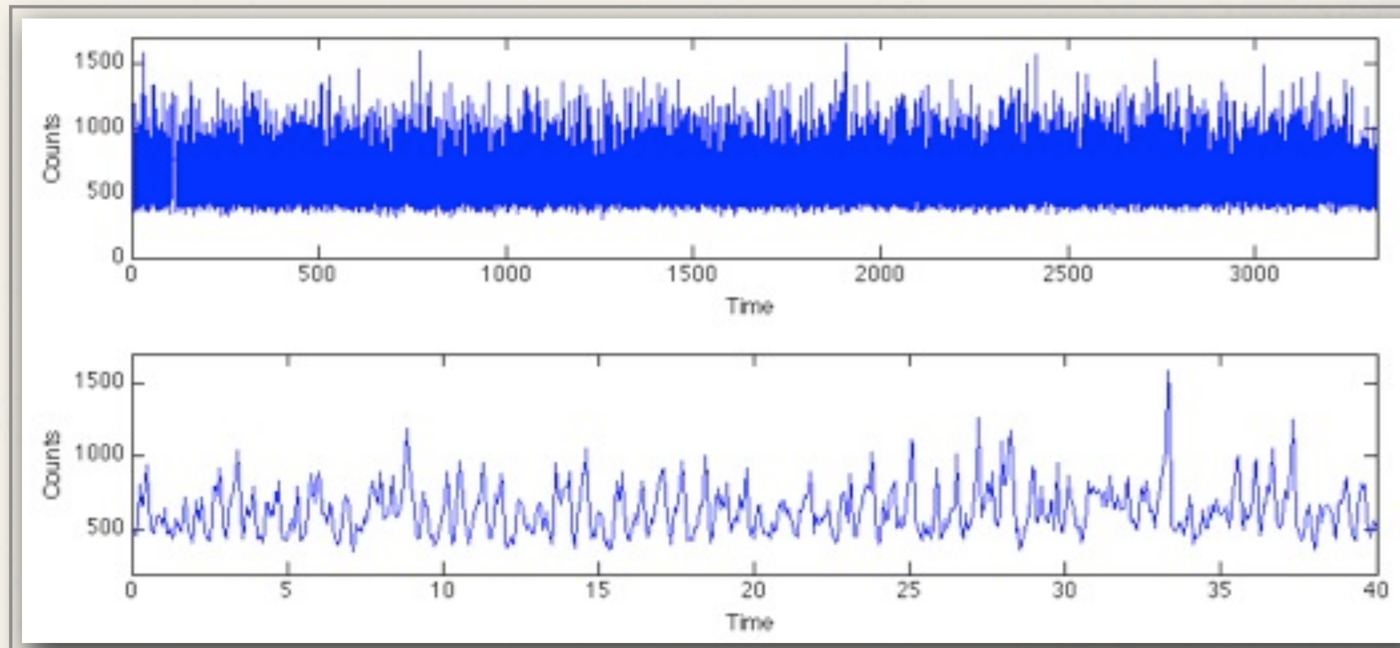
2MW distribution scaled by MW

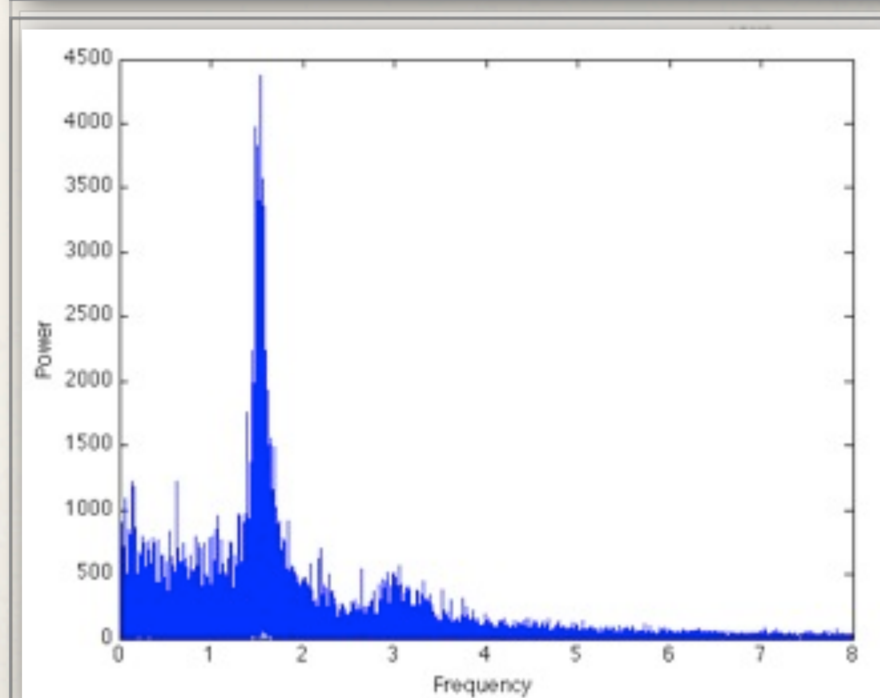
Mean: 2

Standard dev: $\frac{2}{\sqrt{MW}}$

Full power spectrum

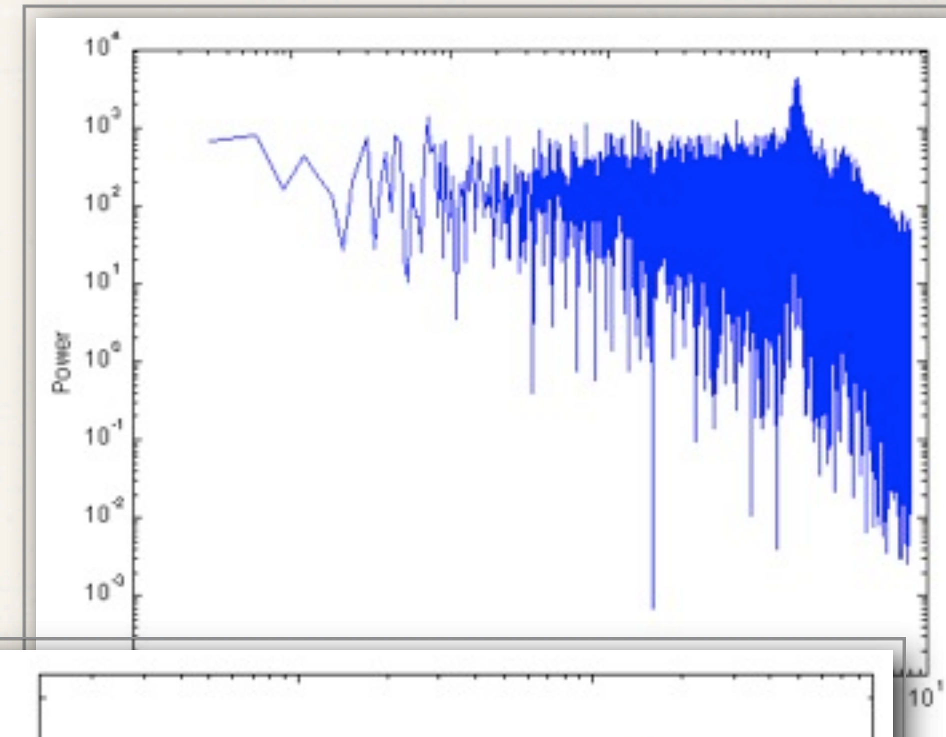
- RXTE light curve
- t = 1/16 seconds
- T = 3325 seconds
- Something can be seen by eye in the light curve
- Full power spectrum
- High-power signal, no coherent peak

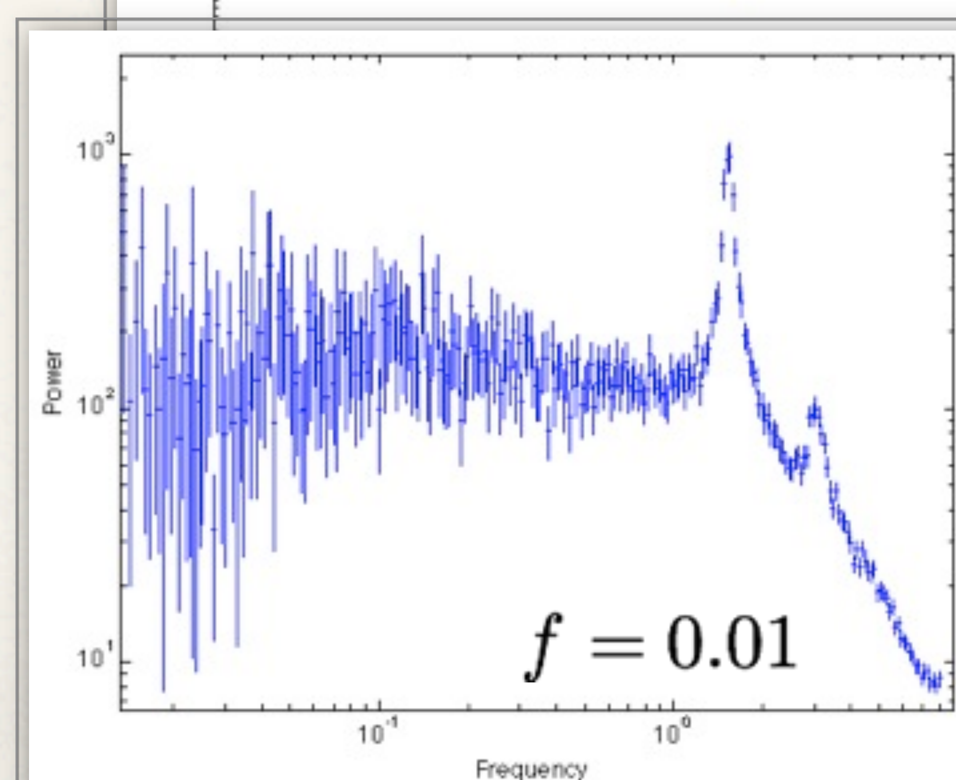




Log space and rebinning

- Log-log plot more appropriate for all frequencies
- Errors are 100%
- Frequency rebinning (M)
- Log-rebinning: $\Delta\nu_j = \Delta\nu_{j-1} * (1 + f)$
- Error bars, better shape
- Poisson level below scale



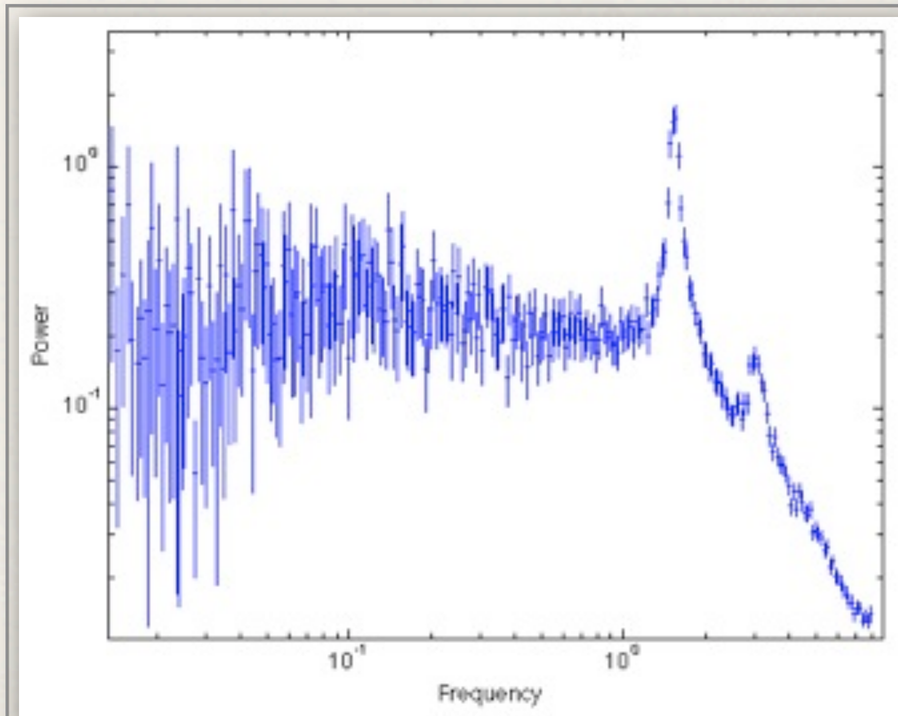


Normalization of power spectra

- Leahy normalization very useful for statistics
- Power ∝ square intensity
- Remove it by dividing by square intensity: rms (Belloni) normalization
- Caveat: from Leahy to rms^2
- Meaning: squared rms per decade
- Root of integral gives fractional rms

$$P_{rms} = \frac{P}{C^2}$$

$$P_{rms} = \frac{P_{Leahy}}{C}$$

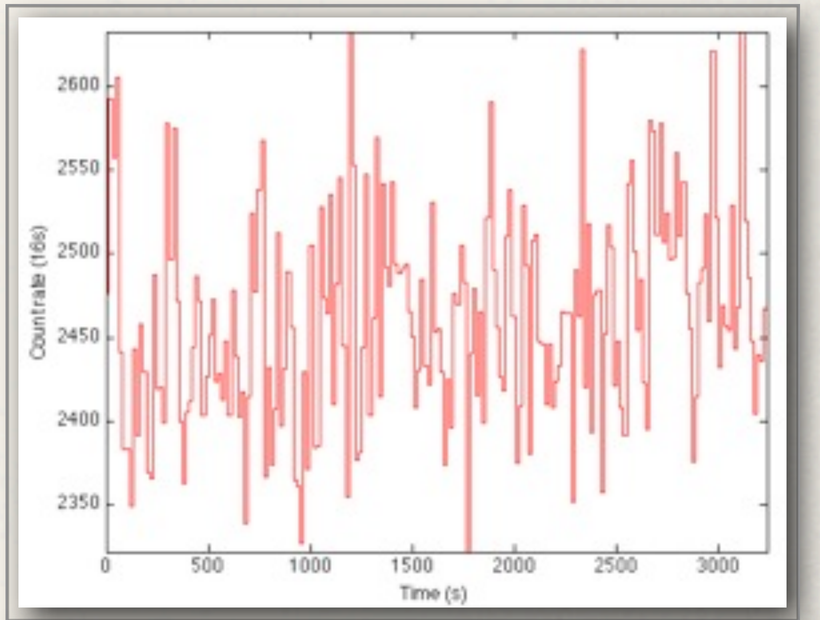


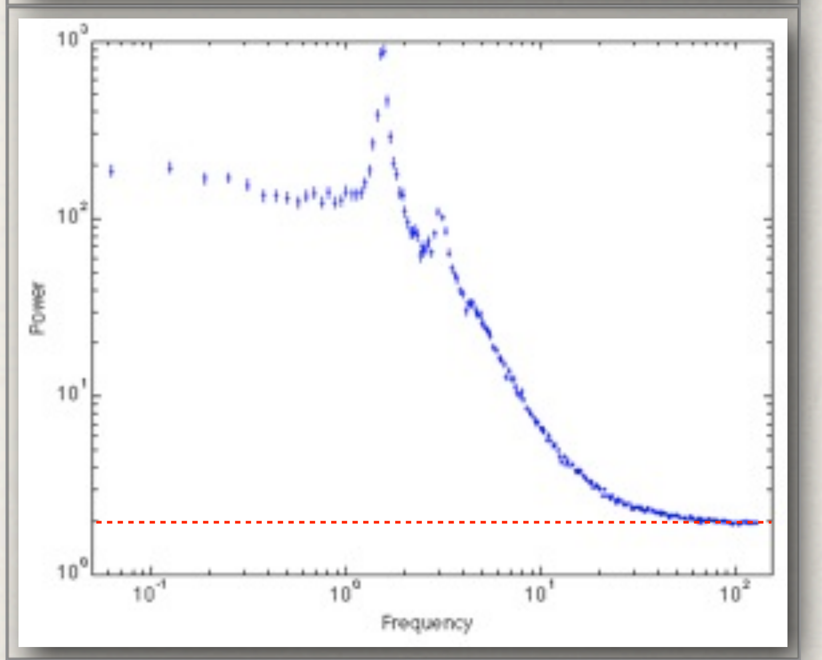
A note about rebinning

- Coherent peak: narrow power distribution - least rebinning - the longer the observation span, the better
- Broad peak: broad power distribution - rebinning helps - length of observation not crucial
- Very important for maximizing sensitivity

W: Welch Power Spectrum

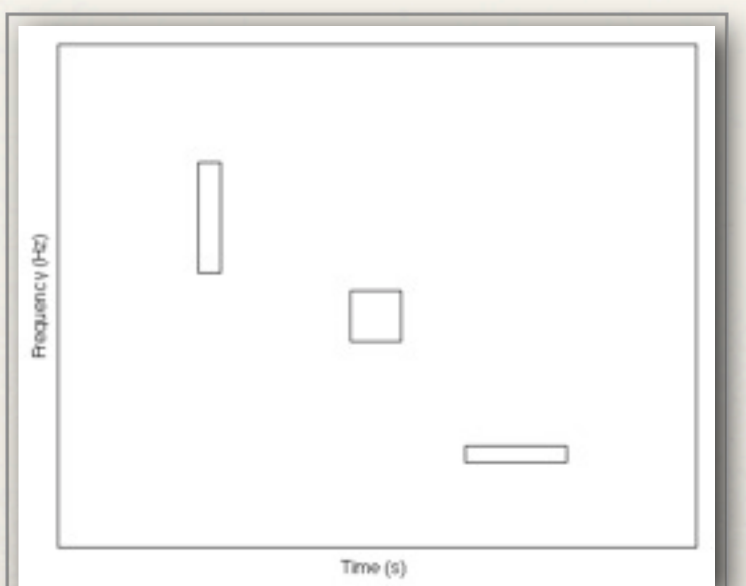
- If signal stationary
- Slice the signal
- Power spectrum of slices
- Add the W slices
- Sliding slices are also possible (statistics?)
- Windowing is also possible

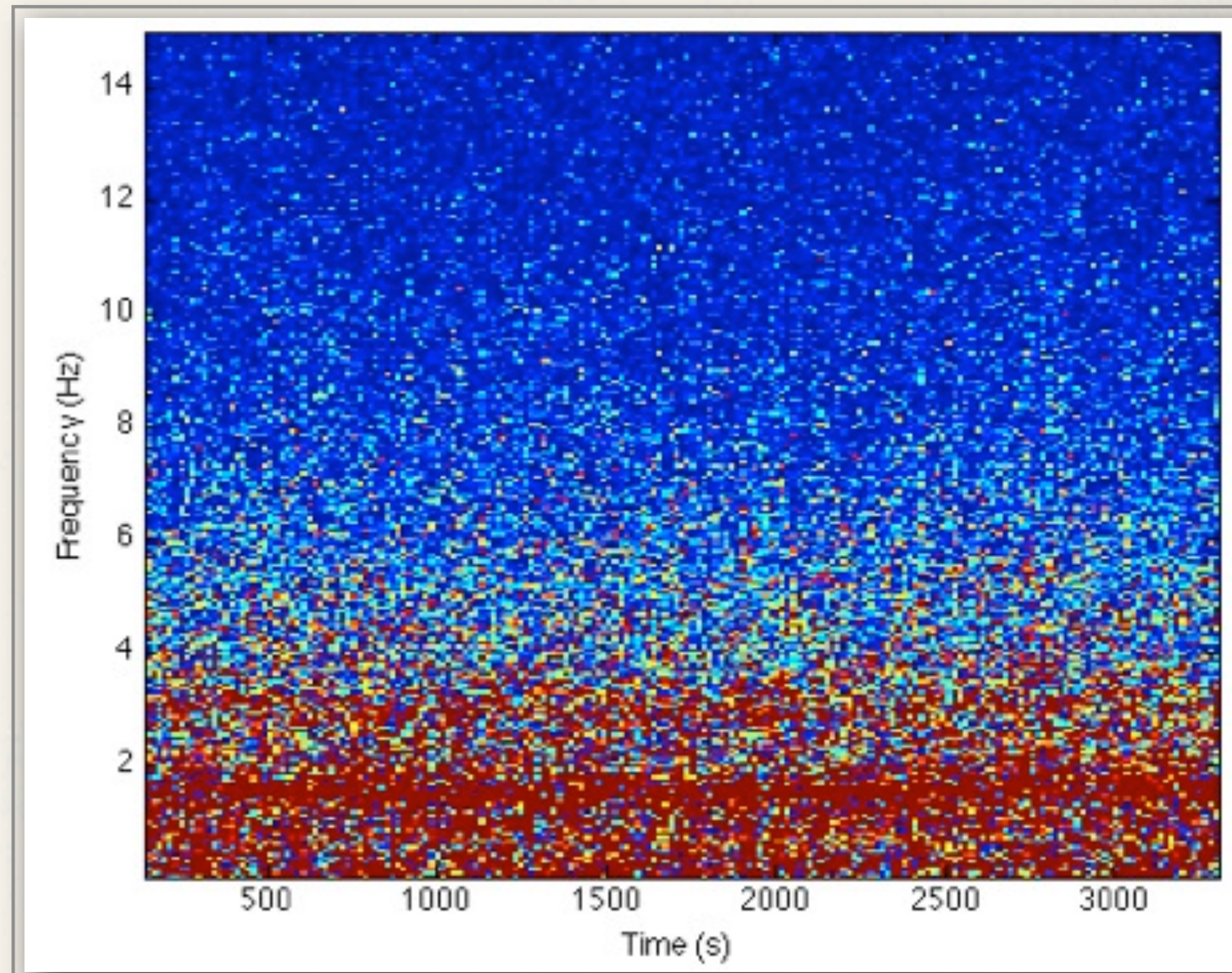




Time-frequency analysis

- If signal is not stationary
- No average of power spectra
- Image: time-frequency-power
- Uncertainty principle





Shift ‘n’ add technique

- Used for twin high-frequency peaks
- You see one, not the other
- The one you see moves
- Correct for the movement, align the spectra in an additive way
- More complex: multiplicative technique (tricky to implement)

Instrumental dead time

- After a photon, dead time
- Introduces correlations between photons (no Poisson!)
- It must be as small as possible and well-known and modeled
- Two types of dead time:
    - Paralyzable — Every incident event causes a dead time t_d even if it's not detected
    - Non-paralyzable — Only a detected event causes a dead time t_d

Paralyzable dead time

- If incident rate r_{in} is very high, no detected counts at all!
- Detected rate: $r_0 = r_{in} e^{-r_{in} t_d}$ $\qquad \lim_{r_{in} \to +\infty} r_0 = 0$
- In RXTE/PCA, for binning time $t_b \geq t_d$

$$< P_j >= 2 \times \left[1 - 2r_0 t_d \left(1 - \frac{t_d}{2t_b}\right)\right] - 2\frac{N-1}{N} r_0 t_d \left(\frac{t_d}{t_b}\right) \cos\left(\frac{2\pi j}{N}\right)$$

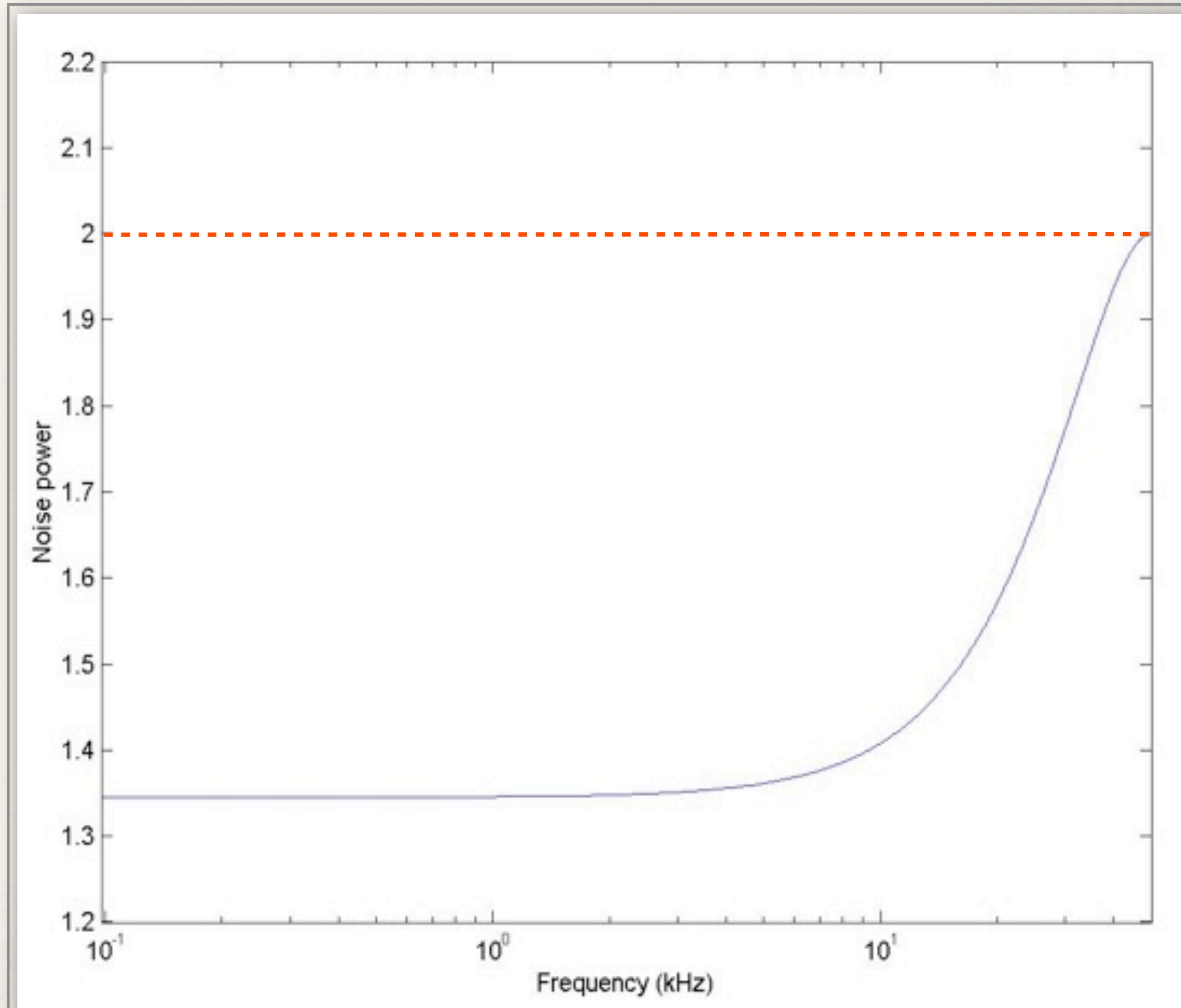
Paralyzable dead time

$r_{in} = 20$ kcts/s

$r_0 = 16.385$ kcts/s

$t_d = t_b = 10\ \mu s$

$N = 1024$



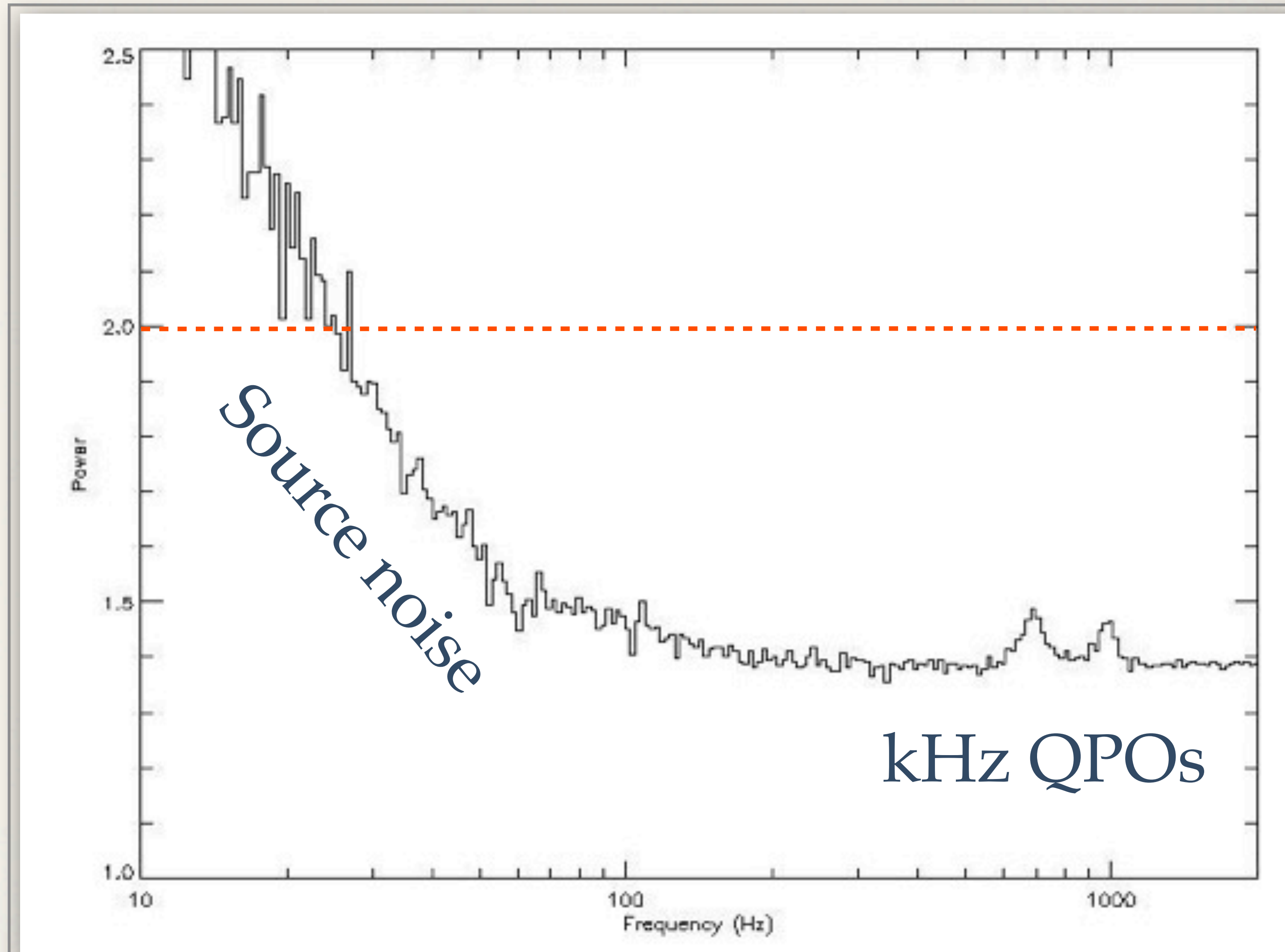
Non-paralyzable dead time

- If incident rate r_{in} is very high, one count every t_d
- Detected rate: $r_0 = \dfrac{r_{in}}{1 + r_{in} t_d}$ $\qquad \lim_{r_{in} \to +\infty} r_0 = t_d^{-1}$
- Formula is even more complicated, result is similar
    - Depression of noise level @ low frequencies (correlation)
    - Peak @ t_d (quasi-periodicity)

Paralyzable dead time: Sco X-1

$r_0 = 10^5$ cts/s

$t_d = 10\ \mu s$



Fitting power spectra

- Fit with typical minimization (χ^2)
- Rebinning is important for χ^2
- Error estimation vs. significance
- Limit in power an NOT rms

- Coherent peaks: distribution of powers and number of trials

Number of trials

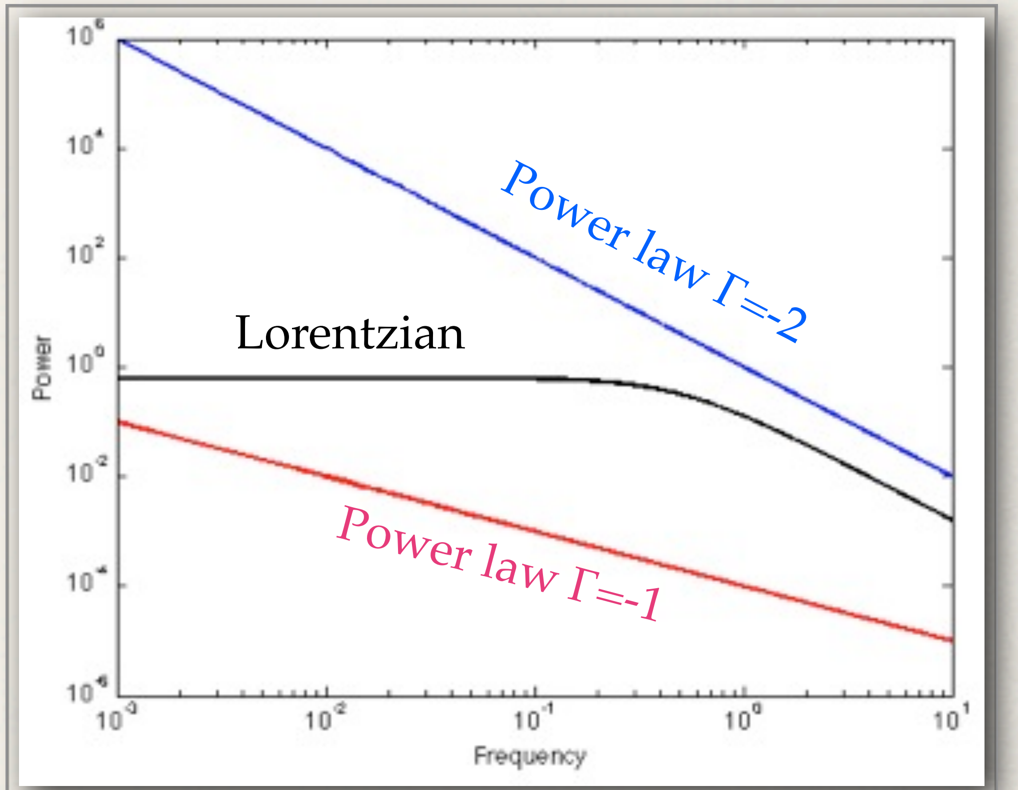
- Important statistical concept
- Should be done correctly, but if P is small can be approximated

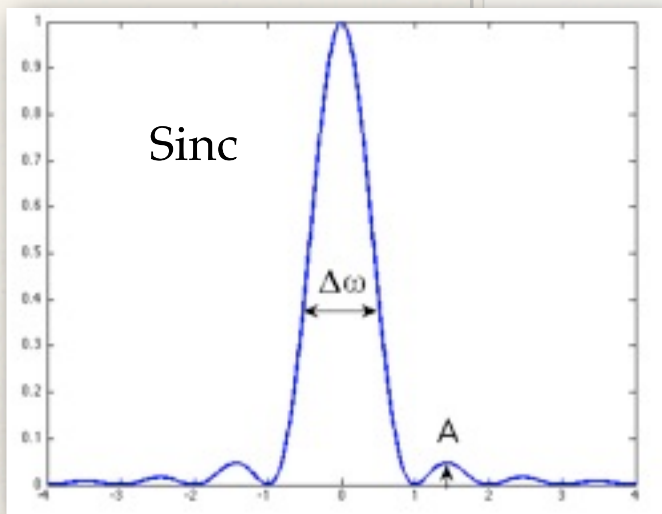
$$\tilde{P}_{chance} = P_{chance} \times N_{trials}$$

- IMPORTANT: how to estimate N_{trials}
- For Power Spectra: number of *independent* frequencies

Continuum components

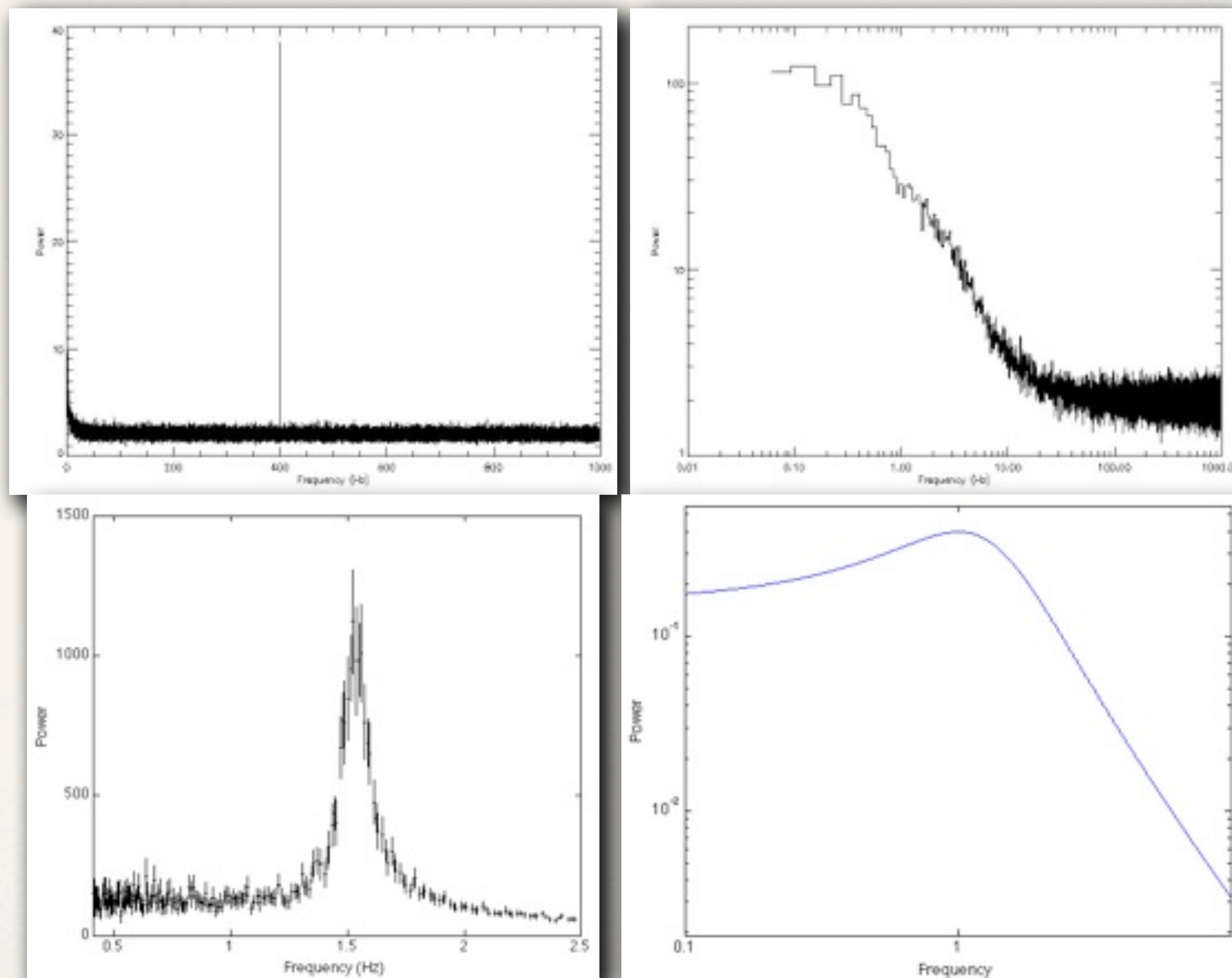
- Very important for accreting sources
- Slope is limited by the window
- Window overflow
- Γ=-2 is the steepest value
- If an issue (pulsar noise): exotic methods





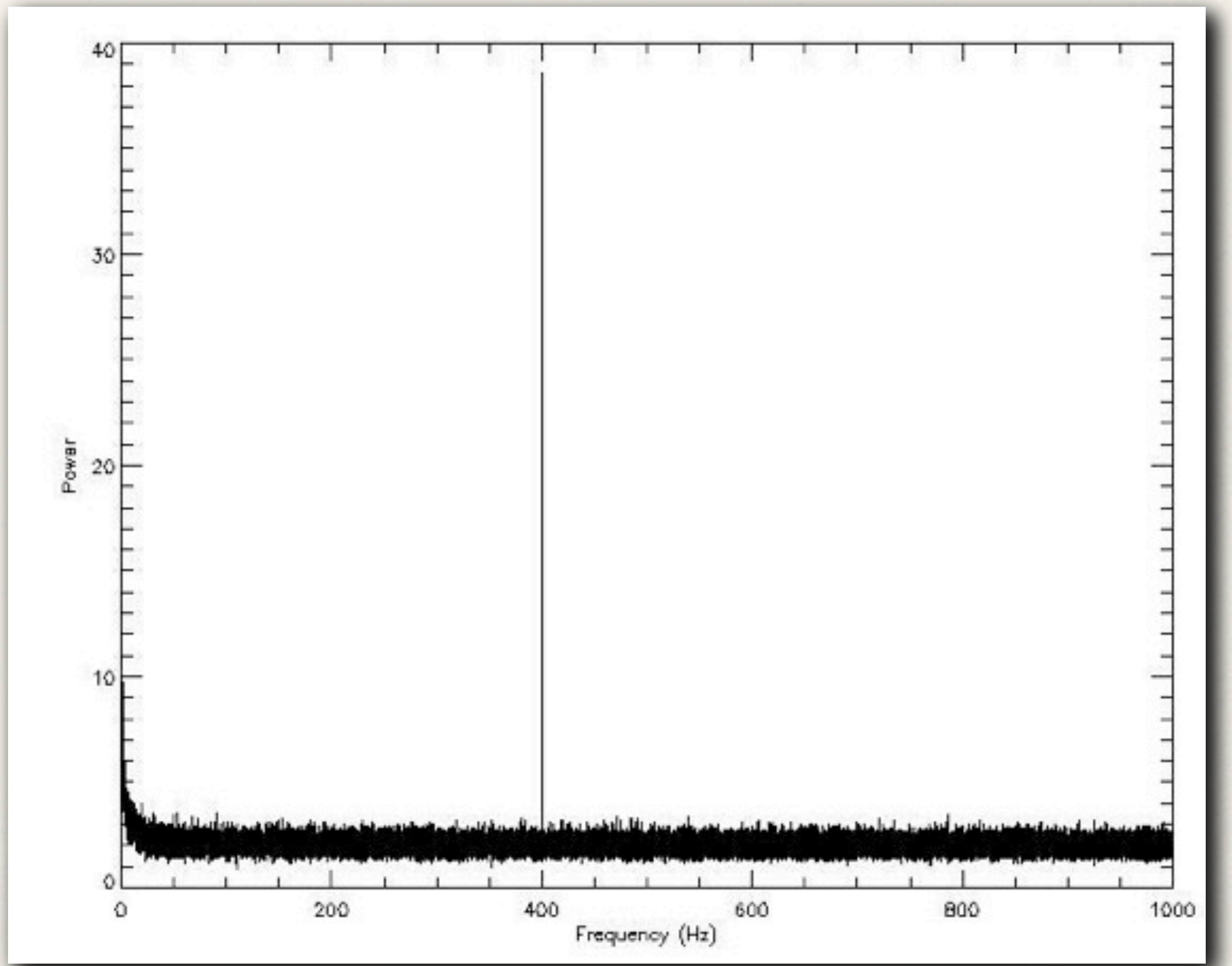
Main types of signals

- Coherent pulsation
- Broad-band noise
- Broad peak (QPO)
- “Peaked-noise”



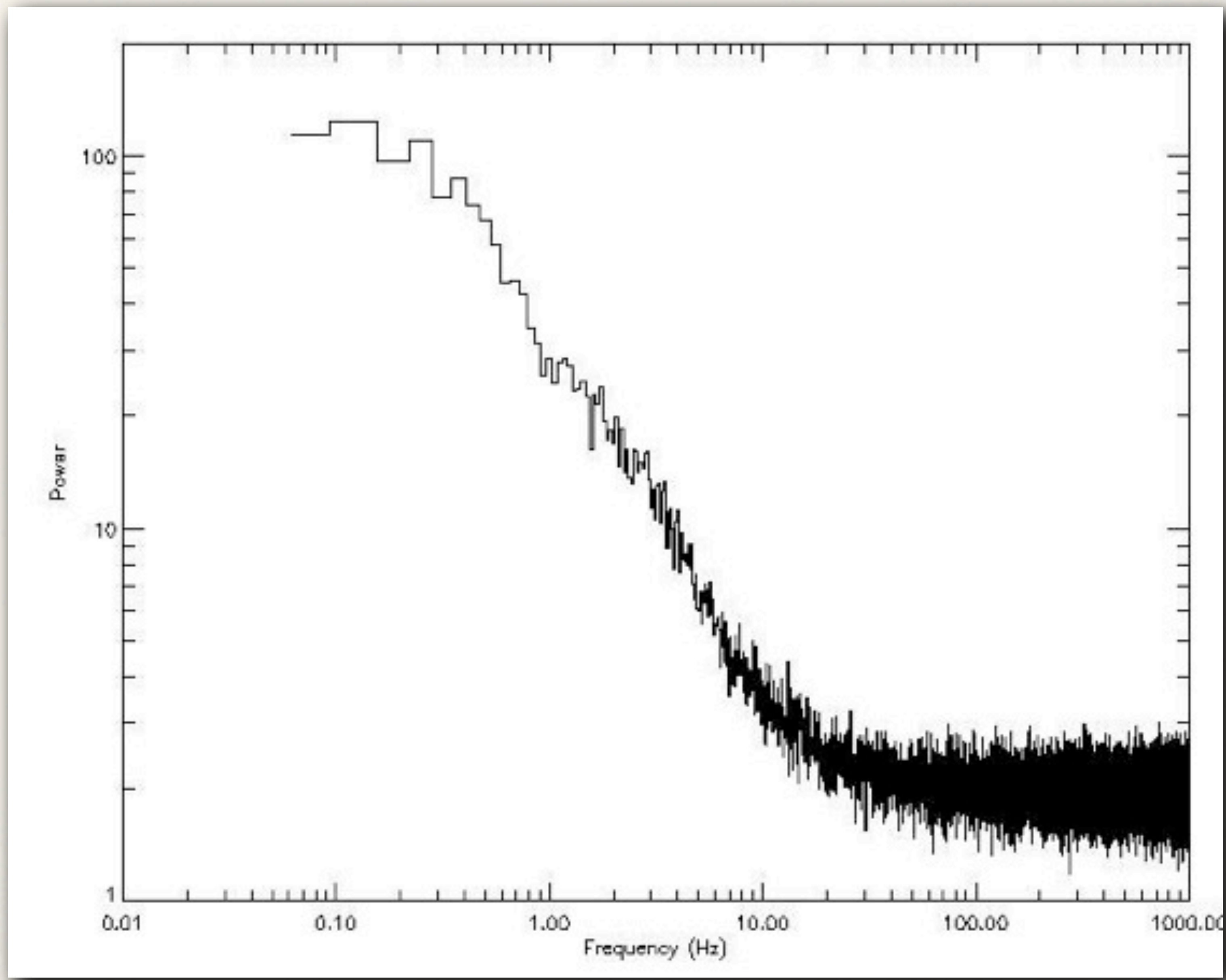
Main types of signals

- Coherent pulsation
- Broad-band noise
- Broad peak (QPO)
- "Peaked-noise"



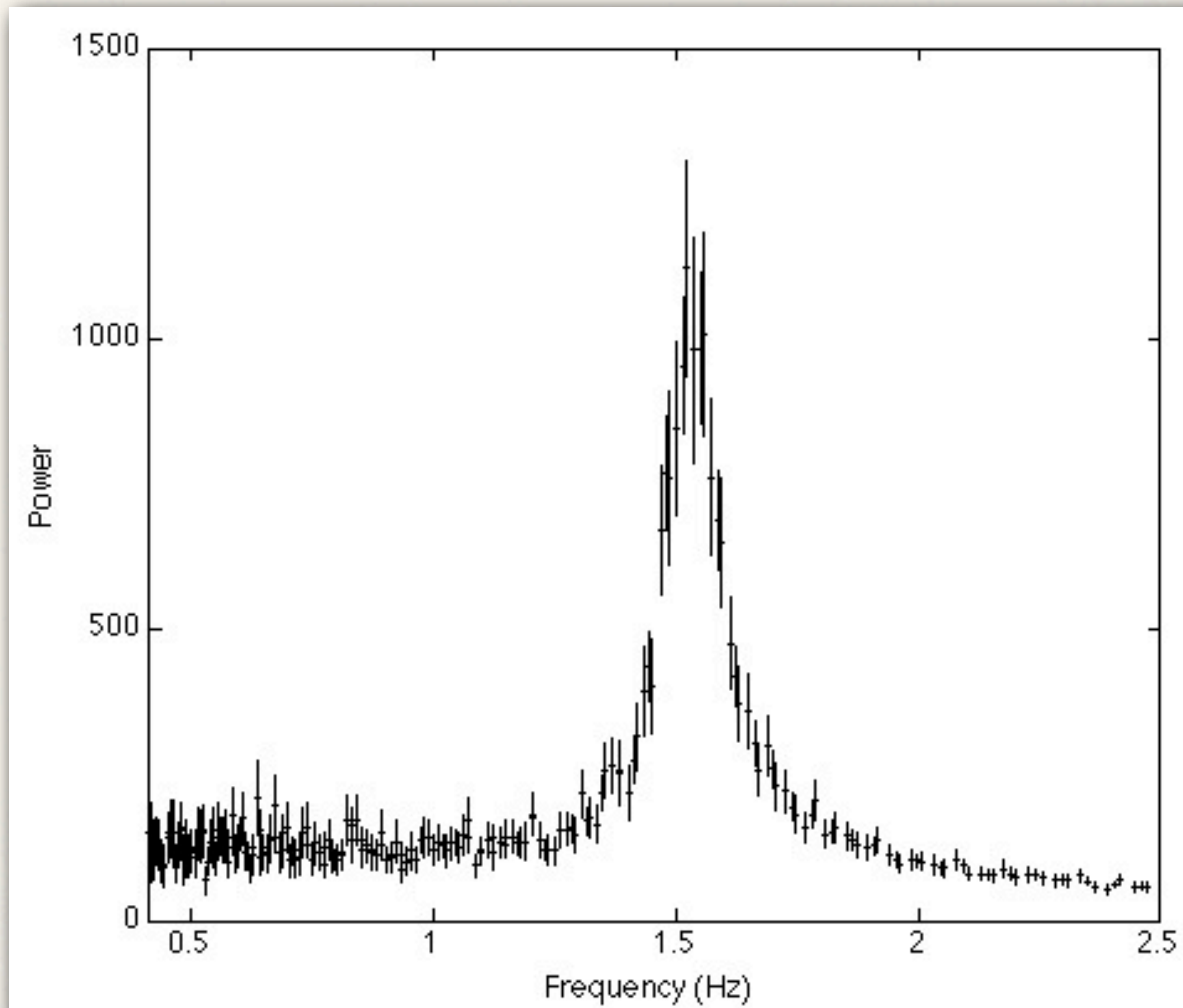
Main types of signals

- Coherent pulsation
- Broad-band noise
- Broad peak (QPO)
- “Peaked-noise”



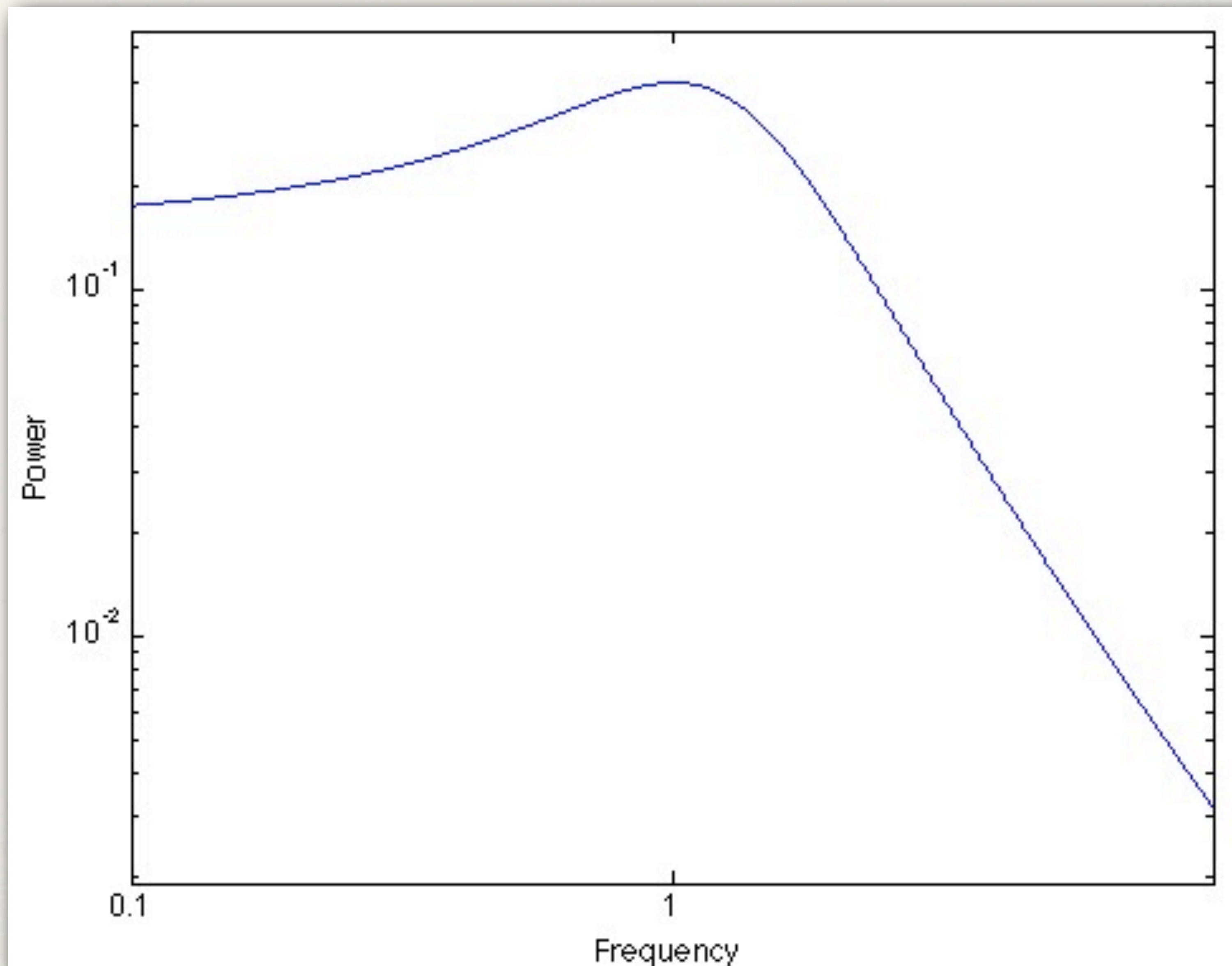
Main types of signals

- Coherent pulsation
- Broad-band noise
- Broad peak (QPO)
- "Peaked-noise"



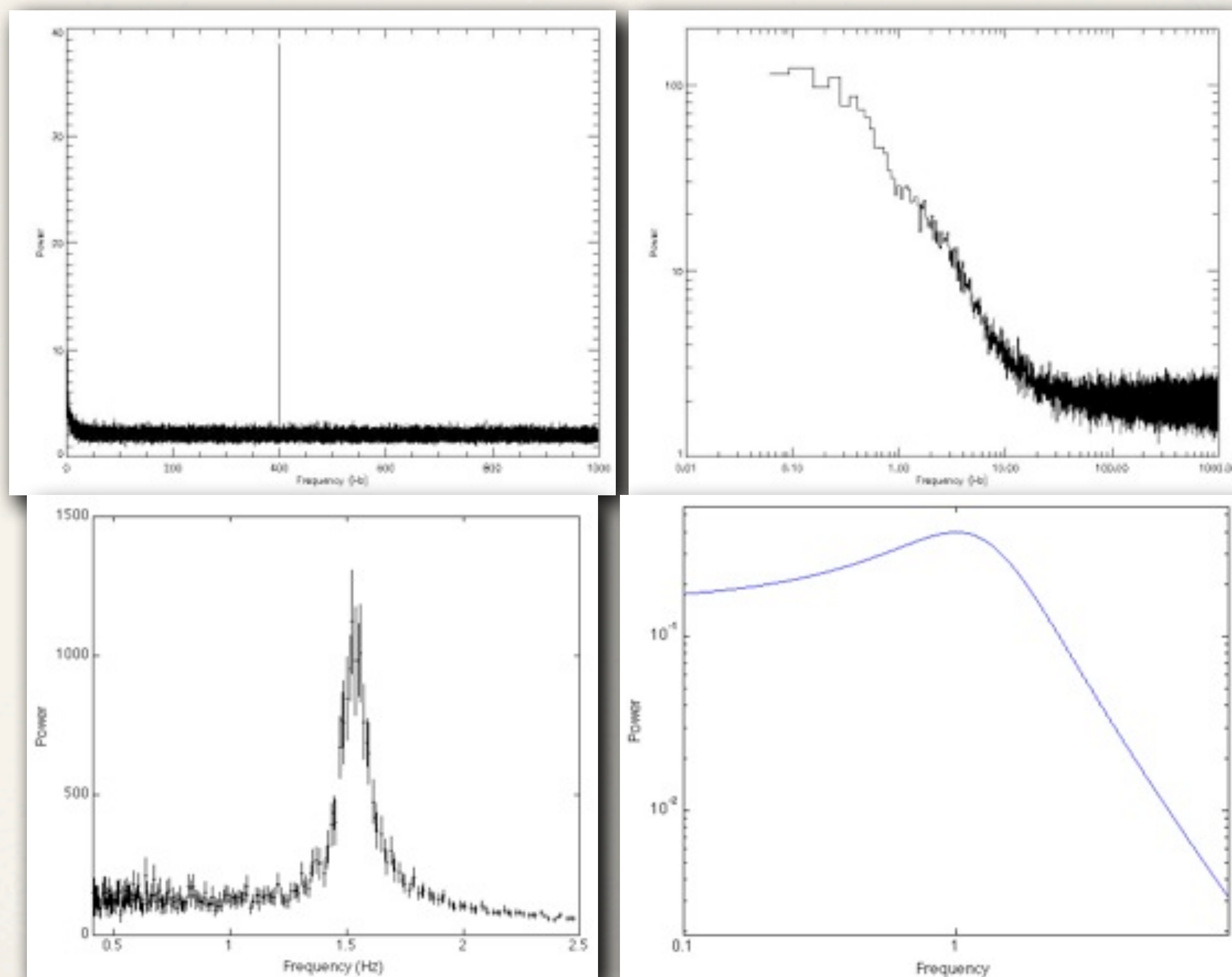
Main types of signals

- Coherent pulsation
- Broad-band noise
- Broad peak (QPO)
- “Peaked-noise”

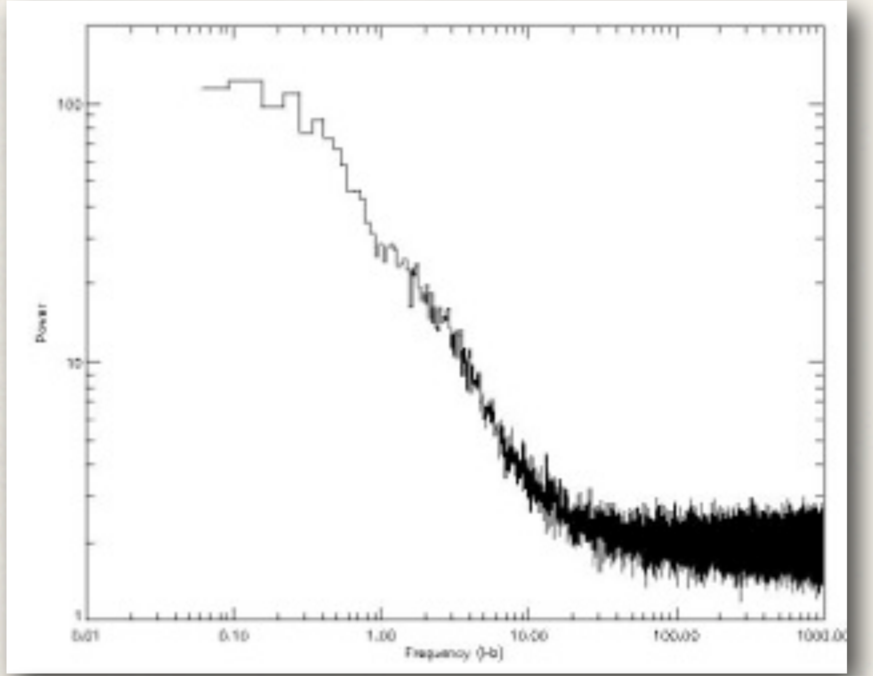


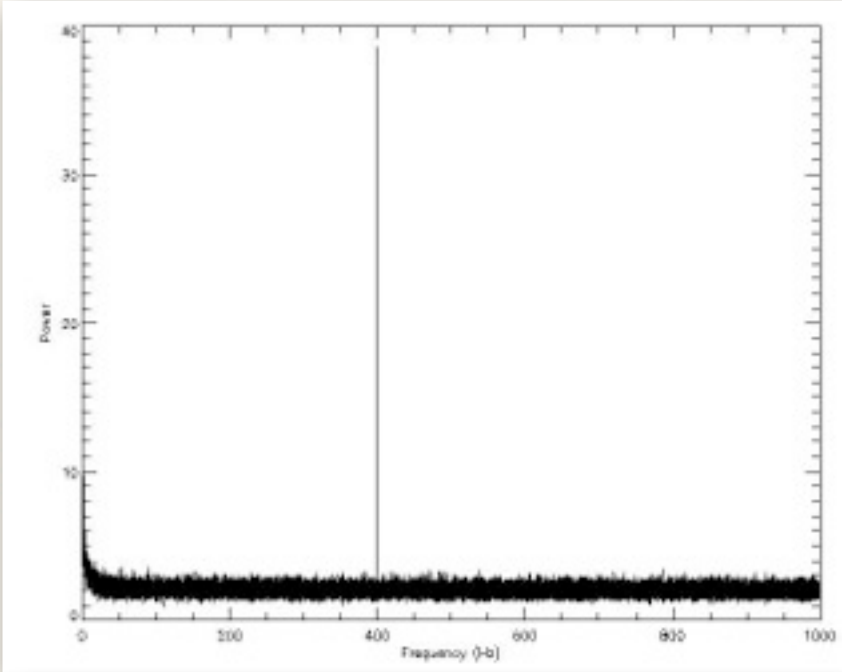
Main types of signals

- Coherent pulsation
- Broad-band noise
- Broad peak (QPO)
- "Peaked-noise"

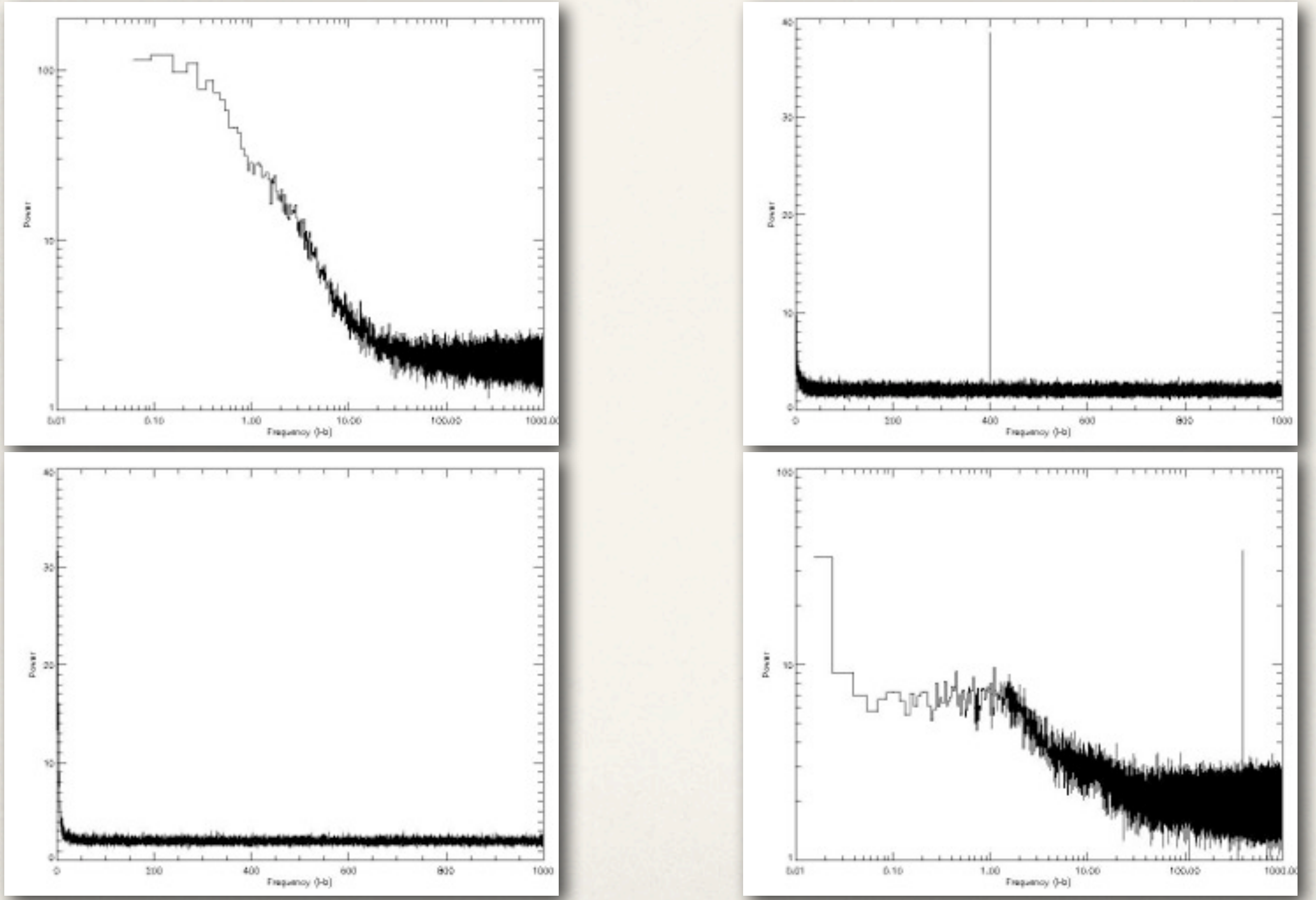


A word on representation





A word on representation

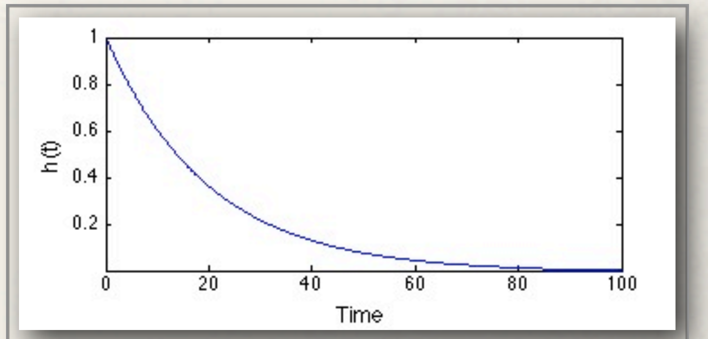


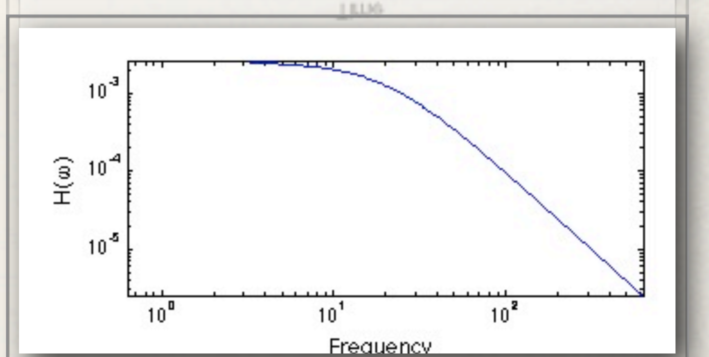
The Lorentzian (zero-centered)

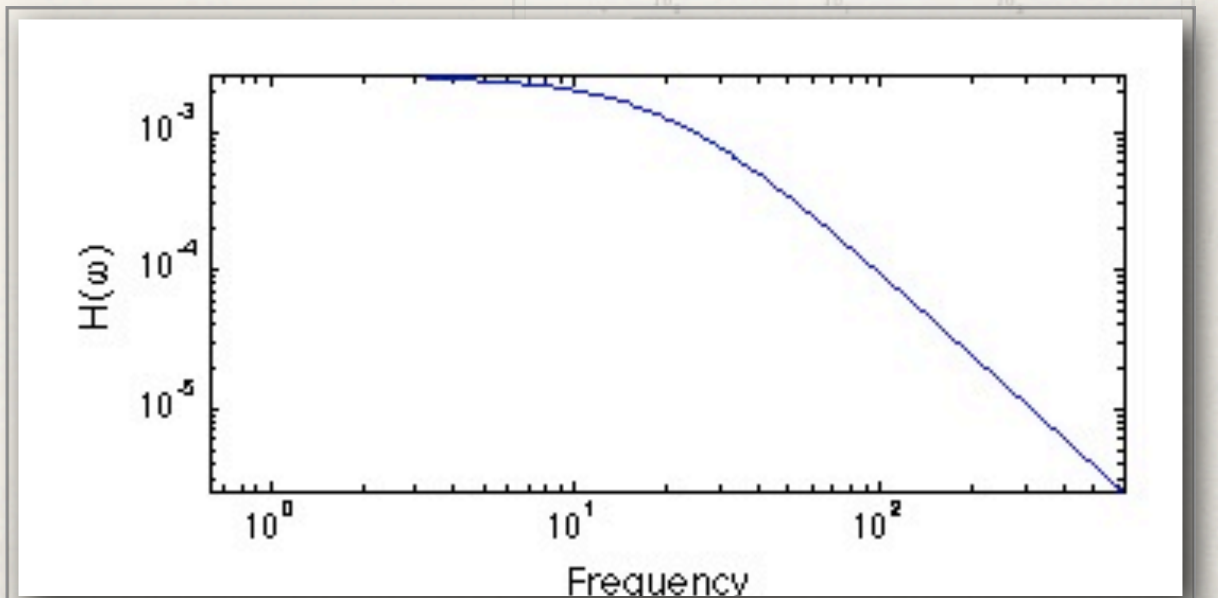
- Power spectrum of a one-sided exponential

$$L(\nu; N, \Delta) = \frac{\Delta}{2\pi} \frac{1}{\nu^2 + (\frac{\Delta}{2})^2}$$

- Good for modeling broad-band noise components (flat-top)





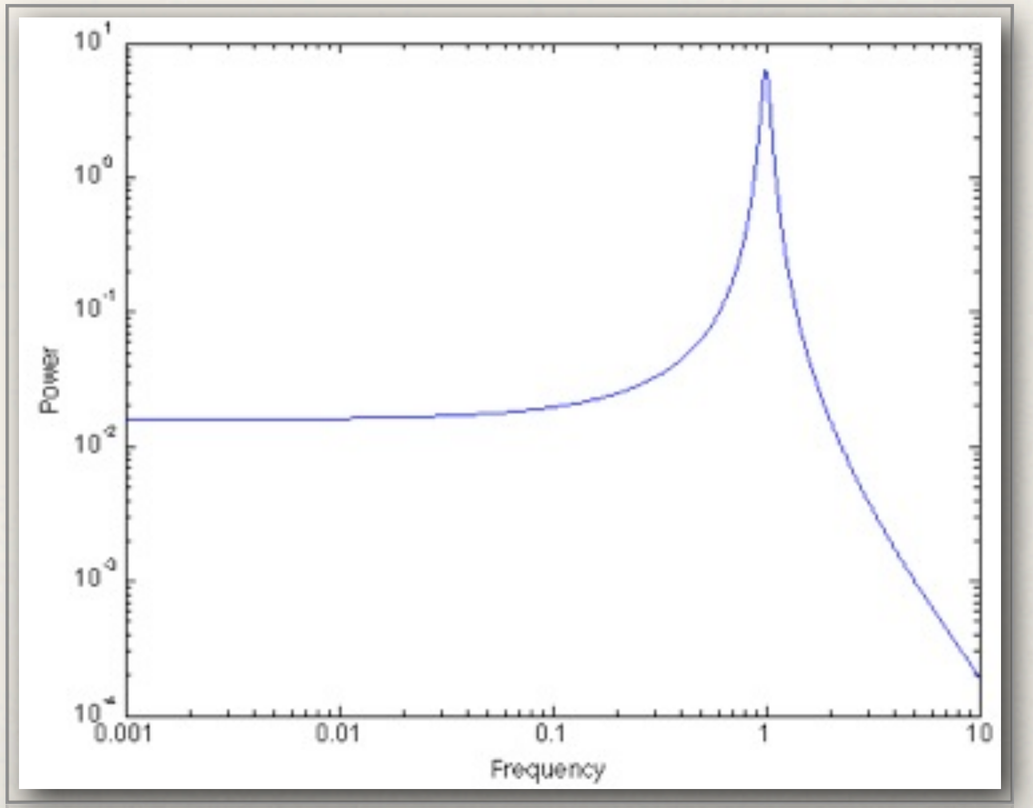


The Lorentzian

- Centroid of Lorentzian not at zero

$$L(\nu; N, \nu_0, \Delta) = \frac{\Delta}{2\pi}\frac{1}{(\nu-\nu_0)^2 + (\frac{\Delta}{2})^2}$$

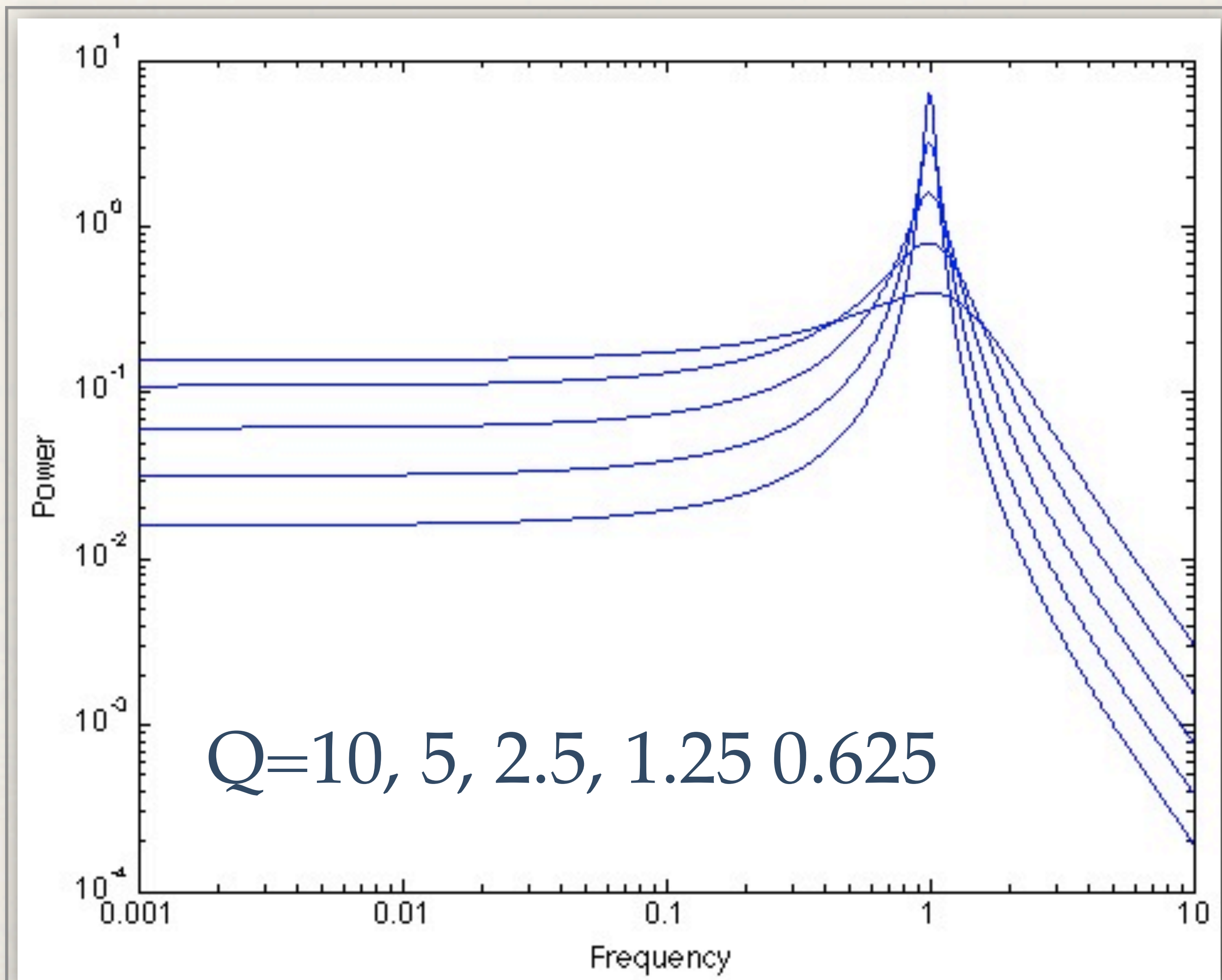
- Good for modeling Quasi-Periodic Peaks



The Quality factor Q

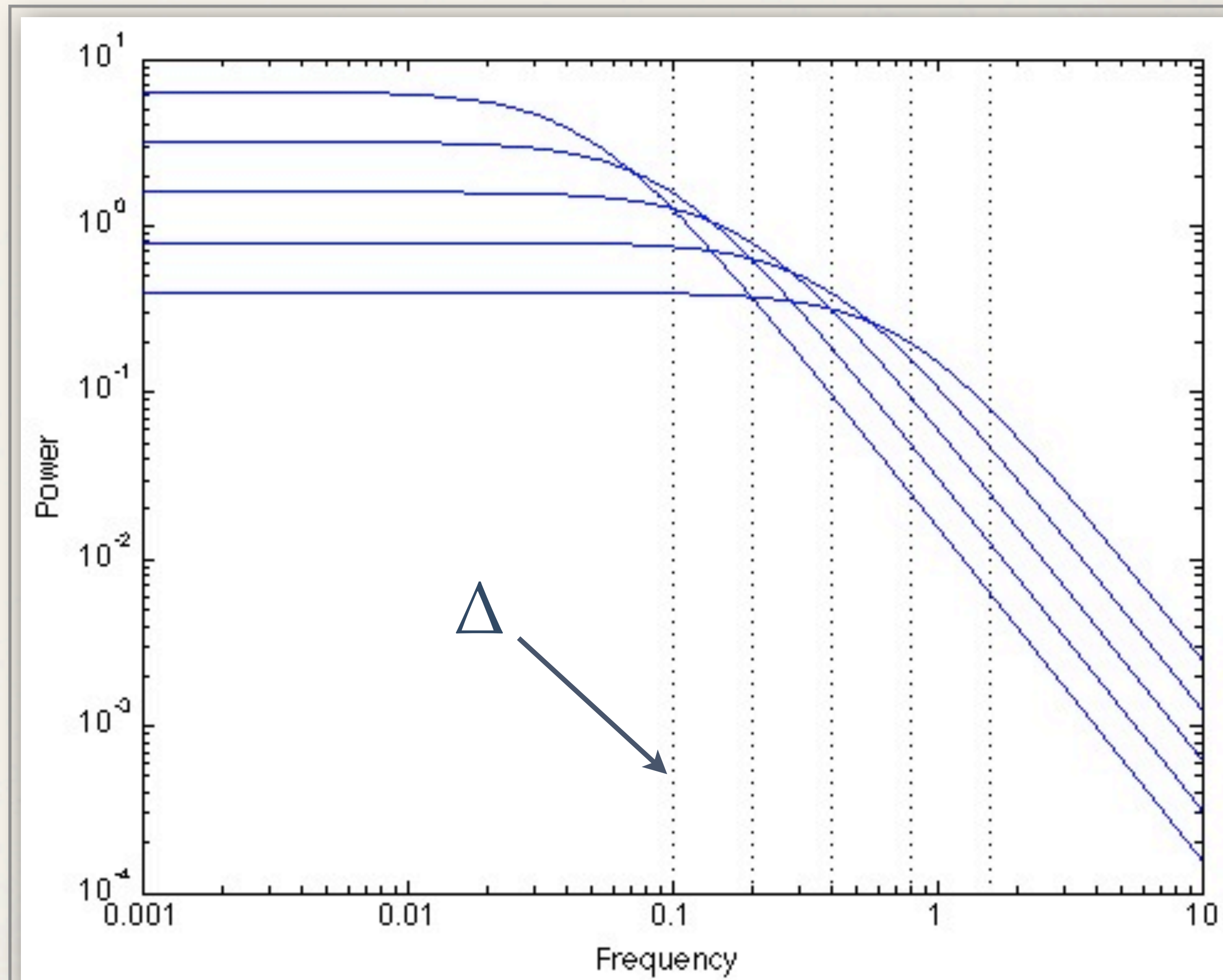
- To quantify the coherence of a component

$$Q = \frac{\nu_0}{\Delta}$$



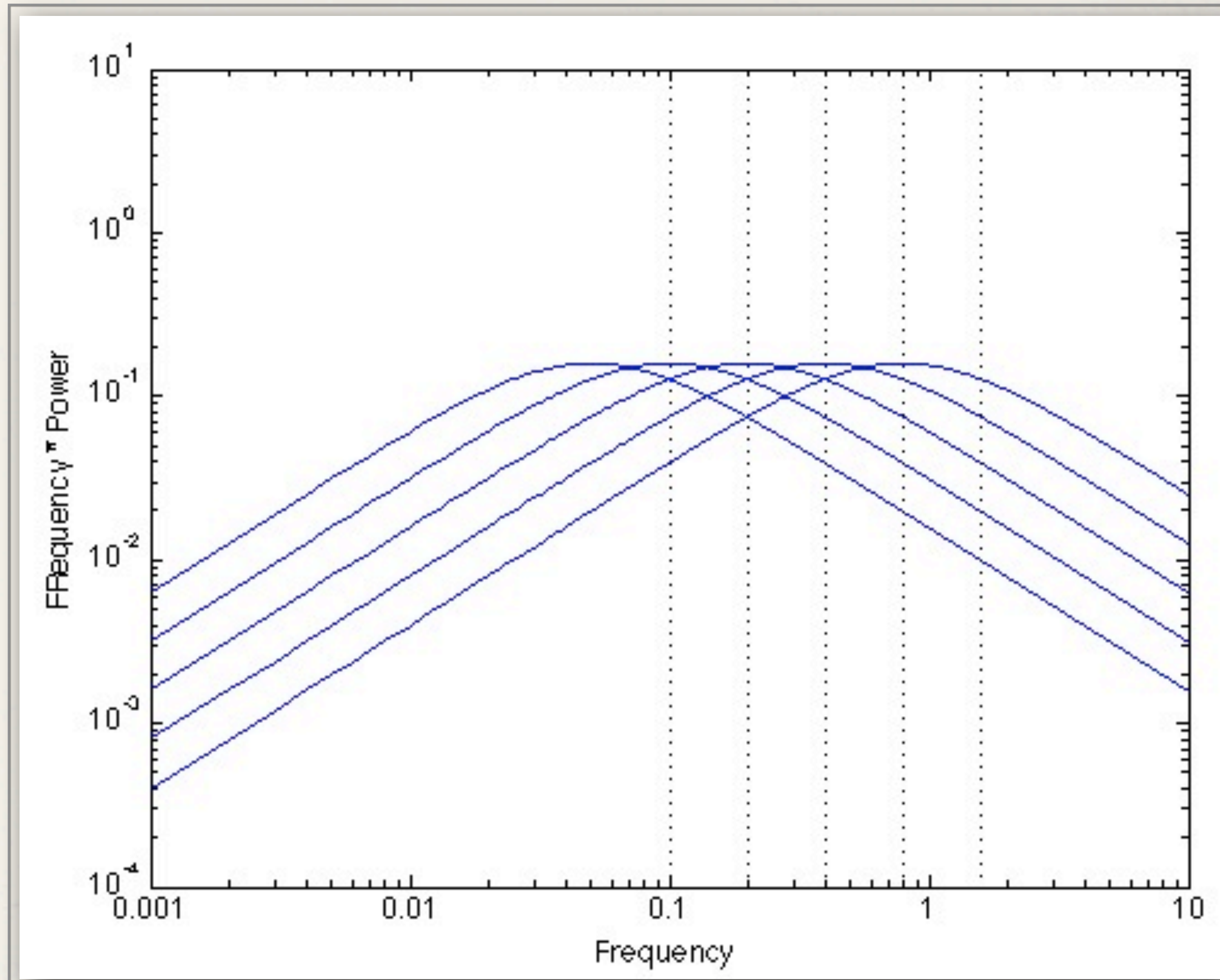
Q=0: the peak without quality

- Here $\nu_0 = 0$, equal N
- Notice position of the break
- Factor of two higher



Better representation

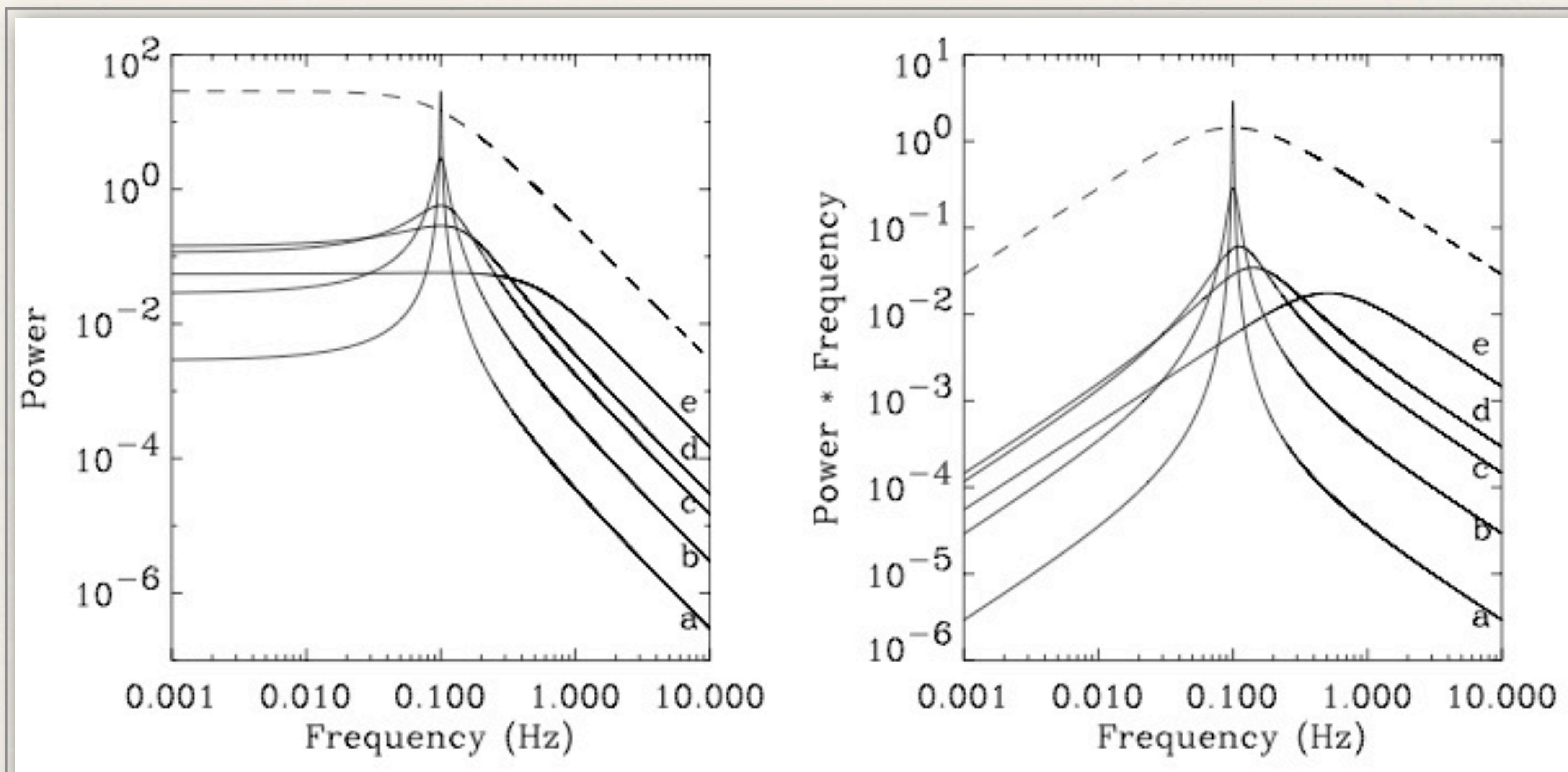
- In νP_ν the effect is the same
- Better value is $\Delta/2$
- But... how do I treat things homogeneously and how do I treat peaked noise?



Characteristic frequency

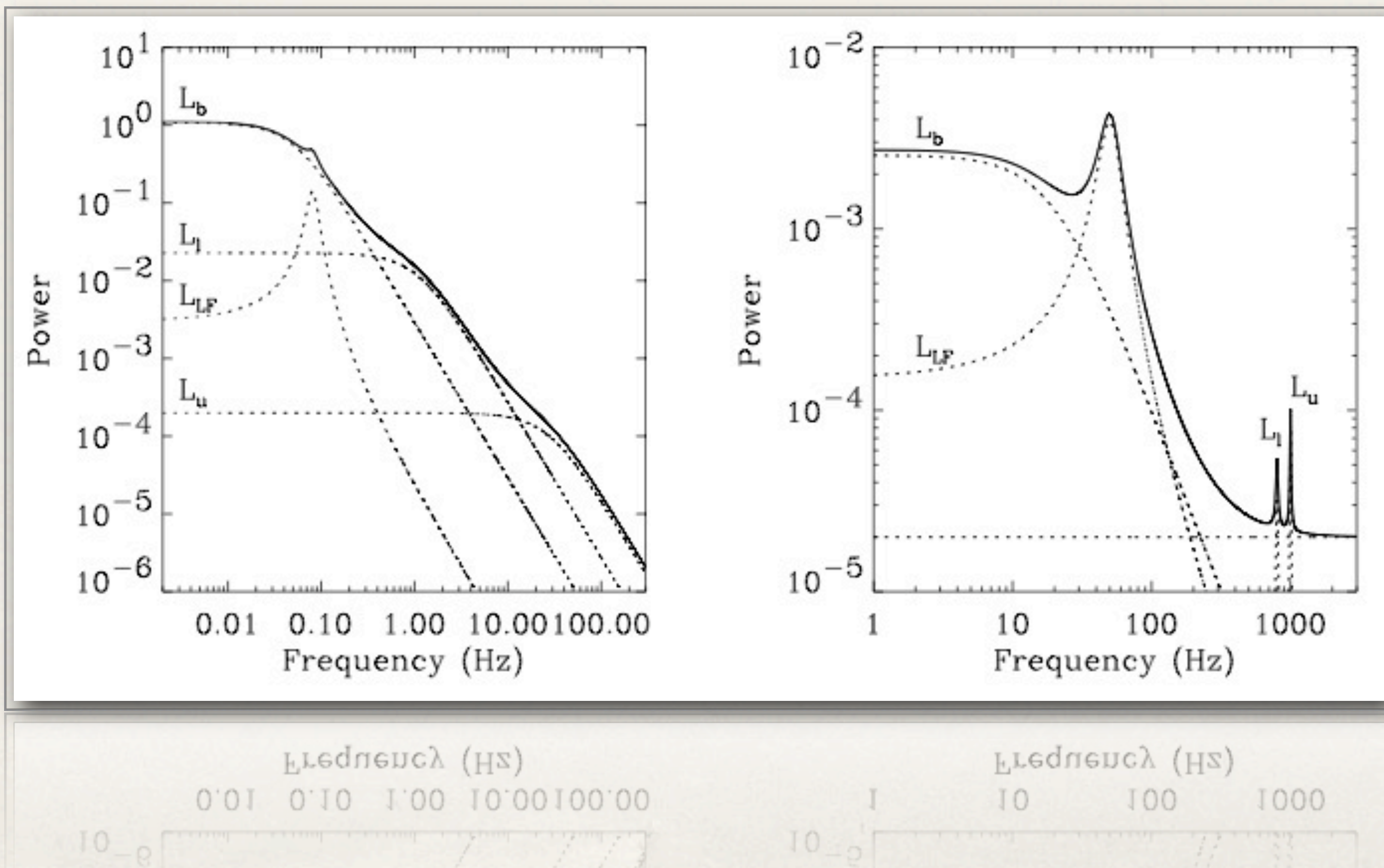
- We can use the peak in νP_ν

$$\nu_{max} = \sqrt{\nu_0^2 + \frac{\Delta}{2}^2}$$



Lorentzian decomposition

- With these tools we can fit power spectra



No physical backing (yet)

- Power spectrum of a damped oscillator
- Also called Cauchy distribution
- Even if it looks like a Lorentzian, it might not be a Lorentzian

Dealing with gaps

- Some solutions are obvious:
    - Welch method (skip gaps)
    - Zero padding (or local average)
- Other methods are available: Lomb-Scargle
    - Good for general uneven sampling
    - Equivalent to linear least-square fit to sin+cos
    - Statistically robust

Lomb-Scargle periodogram

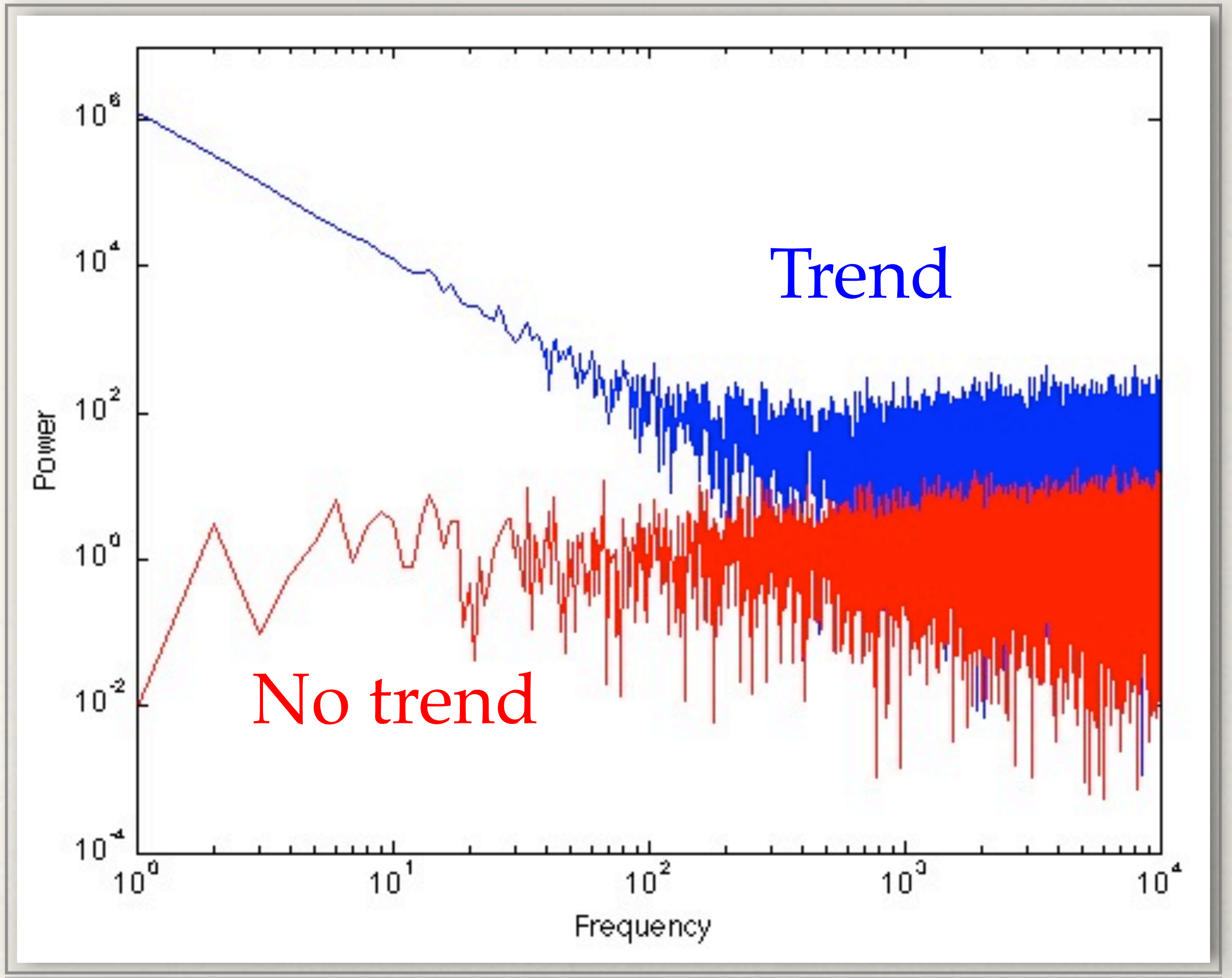
- h_j sampled at t_j

$$P_N(\omega) = \frac{1}{2\sigma^2}\left\{\frac{[\sum_j (h_j - \bar{h})\cos\omega(t_j-\tau)]^2}{\sum_j \cos^2\omega(t_j-\tau)} + \frac{[\sum_j (h_j - \bar{h})\sin\omega(t_j-\tau)]^2}{\sum_j \sin^2\omega(t_j-\tau)}\right\}$$

- where: $\tan(2\omega\tau) = \frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j}$ ensures shift independence
- Powerful method: it can go beyond "Nyquist"

Beware of trends!

- A trend is a modification to the window
- Must be de-trended
- Same about possible drop outs

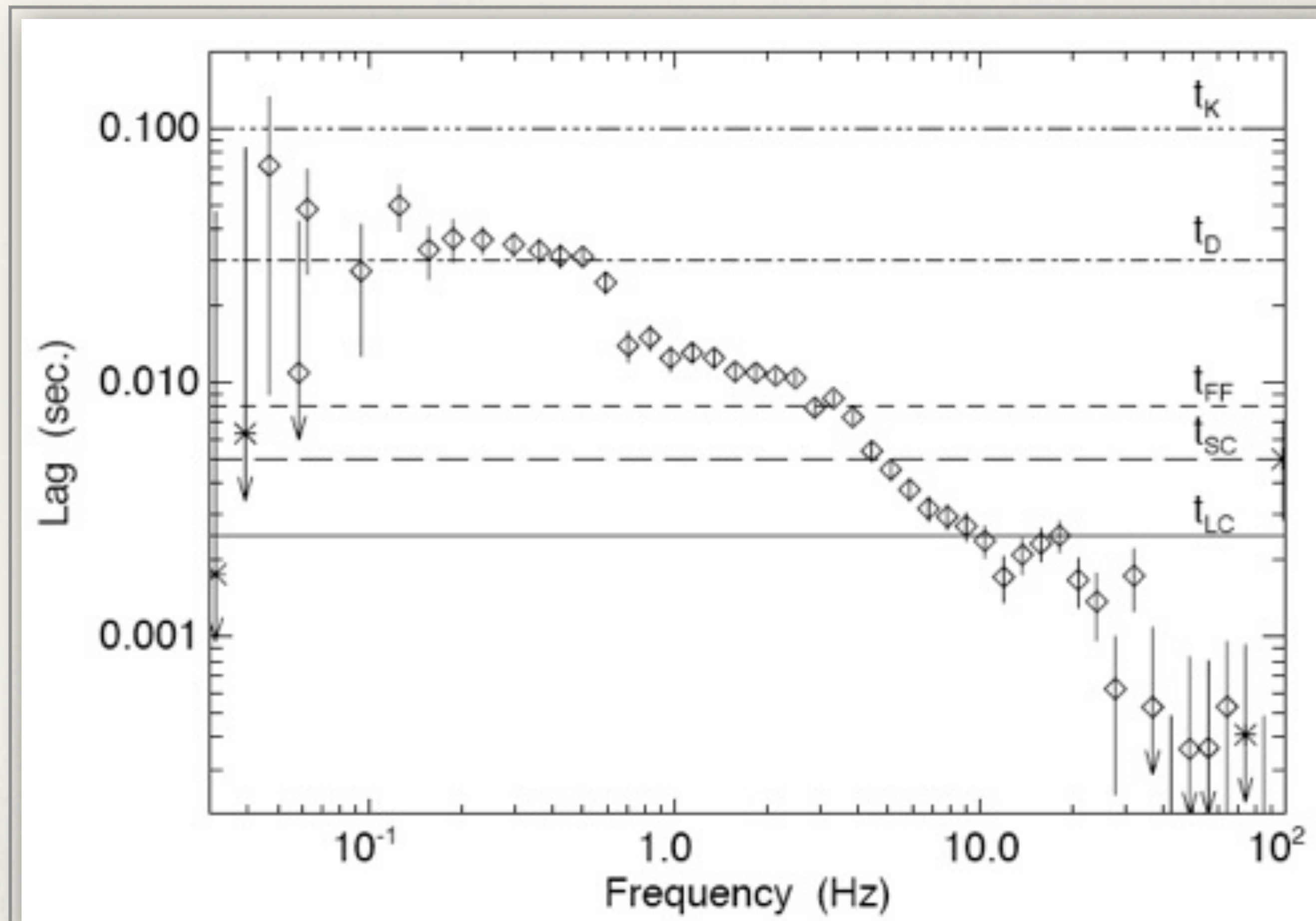


Cross-spectrum

- Power spectrum: amplitudes of the FFT
- We throw away the phases
- If we take *two* time series *f(t)* & *g(t)*, the phases make more sense
- Cross-spectrum: $C_{f,g}(\nu) = F_f^*(\nu) \times F_g(\nu)$
- If *f*=*g*, it becomes the power spectrum
- What is it useful for?

Phase/time lags

- The phases give us the phase delay between the two time series
- Not easy to interpret, can be linked to physical models
- Time lags: phase ϕ/ν
- Additional technical details (not shown)



Auto/cross-correlation

- The power spectrum is the FT of the autocorrelation

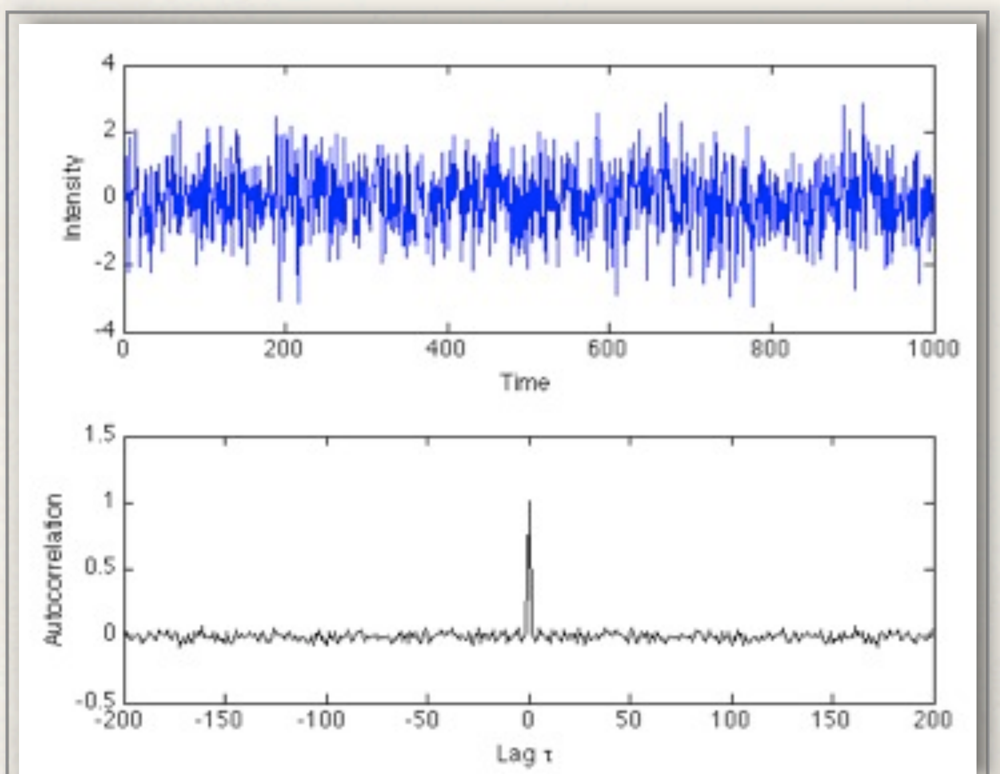
$$Corr(g,g) = \int_{-\infty}^{\infty} g(t+\tau)g(\tau)d\tau \Longleftrightarrow |G(f)|^2$$

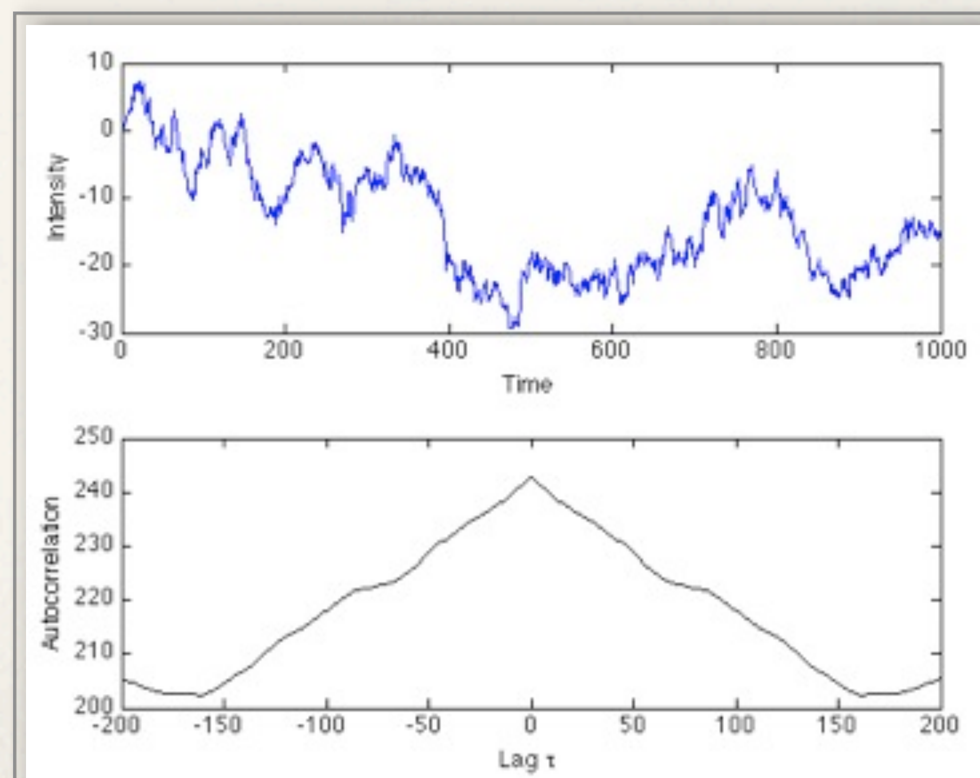
- Autocorrelation is real and even, power spectrum is real and even
- The cross spectrum is the FT of the crosscorrelation
- Power- an cross-spectrum contain more information (if you can afford them because of statistics)

Autocorrelation

$$Corr(g,g) = \int_{-\infty}^{\infty} g(t+\tau)g(\tau)d\tau$$

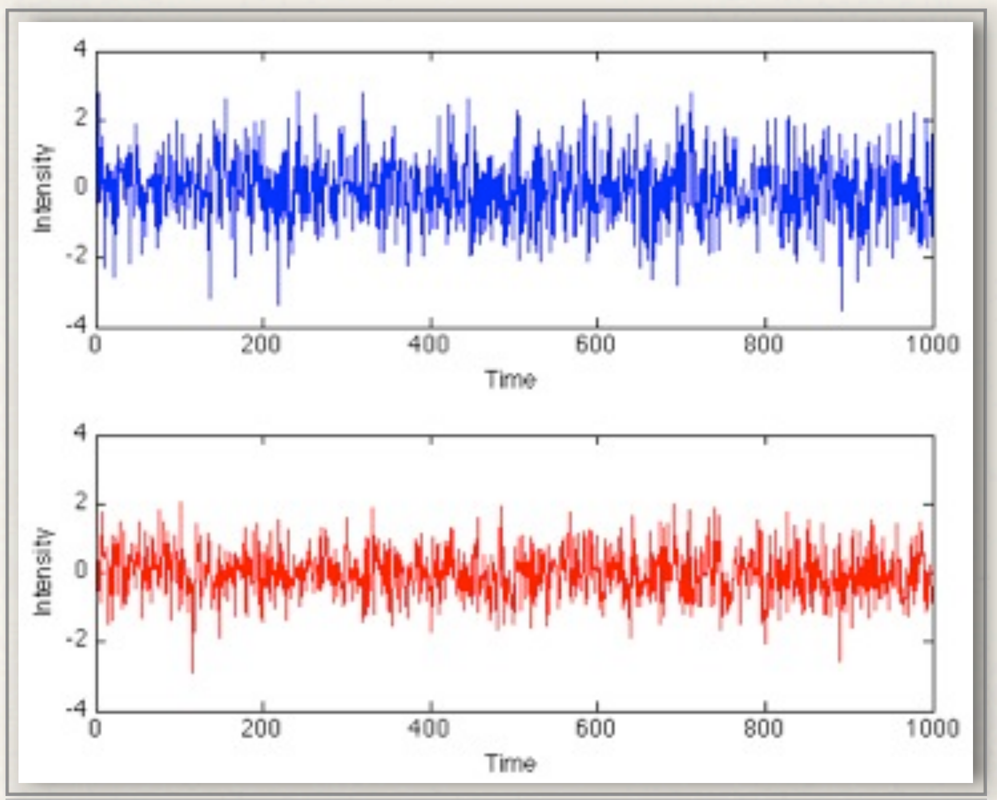
- Uncorrelated noise: ACF is zero everywhere but at $\tau=0$ [variance]
- Biased ACF: dividing by N
- Unbiased ACF: dividing by $N-|m|$

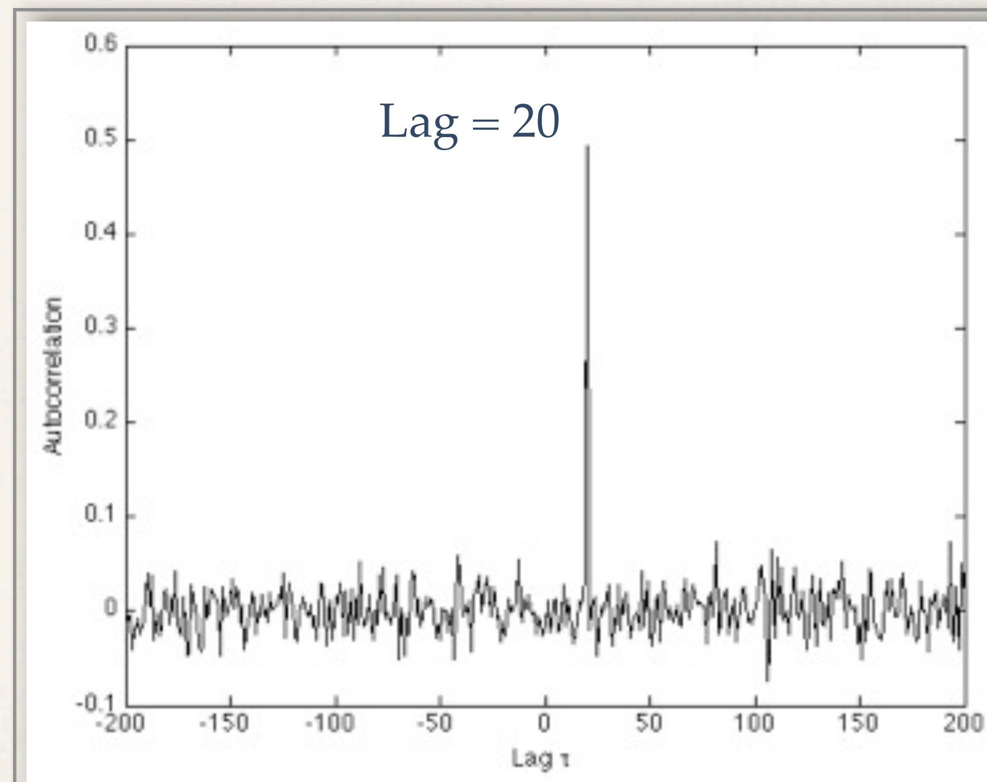




Crosscorrelation

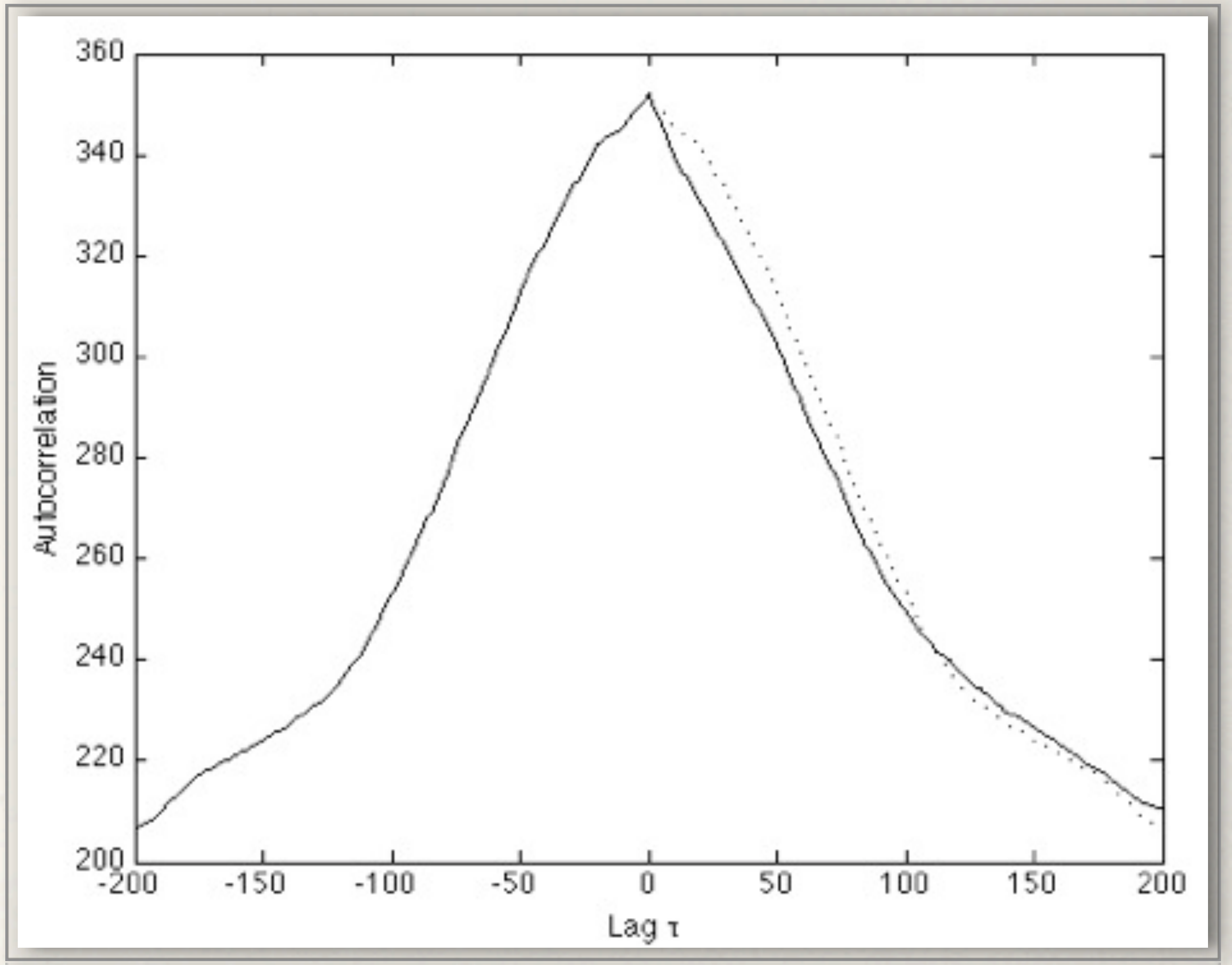
- Uncorrelated series: CCF is zero everywhere
- Simple shift: peak somewhere





When do I have a lag?

- *"The CCF peaks at 0, therefore there is no measurable lag"*
- NO!
- CCF is a superposition of sinusoids of different periods
- Any asymmetry implies a lag



Coherent signals: barycentric corr.

- The Earth moves and rotates, the satellite also moves
- This has an effect on the period (doppler modulation)..
- .. and on the absolute phase

- Times are corrected to the barycenter of the solar system
- Standard routines and ephemeris
- Not relevant for aperiodic signals

Period folding I: χ^2 test

- Photon arrival times t_j
- For trial period produce phases $\phi_j = \mathrm{Frac}\left(\frac{t_j}{P}\right)$
- Put photon in appropriate phase bin
- Test vs. constancy (χ^2)
- If time bins and not times, easy to generalize
- Problem: binning and statistics (few photons?)

Period folding II: Z^2 test

- Photon arrival times t_j
- For trial period produce phases $\phi_j = \mathrm{Frac}\left(\frac{t_j}{P}\right)$
- Compute $$Z_n^2 = \frac{2}{N}\sum_{k=1}^{n}\left[\left(\sum_{j=1}^{N}\cos k\phi_j\right)^2 + \left(\sum_{j=1}^{N}\sin k\phi_j\right)^2\right]$$

  where n is the desired number of harmonics
- Z is distributed as a χ^2 with $2n$ d.o.f.
- Good for small number of photons [Rayleigh test]

Additional complications

- There can be a significant period derivative
- If your pulsar is in a binary system, there is Doppler effect
- Easy to lose a pulsation
- Power spectrum smeared, folding as well
- Must factorize possible orbit in the solution
- Many free parameters