Gravitational-Wave Data Analysis: Lecture 1

Peter S. Shawhan

UNIVERSITY OF MARYLAND 1856

Gravitational Wave Astronomy Summer School
May 28, 2012

Outline for Today

- What gravitational wave (GW) data is like
- Characterizing noise
- Frequency-domain representation of data and signals
- Calibration
- Digital filtering basics

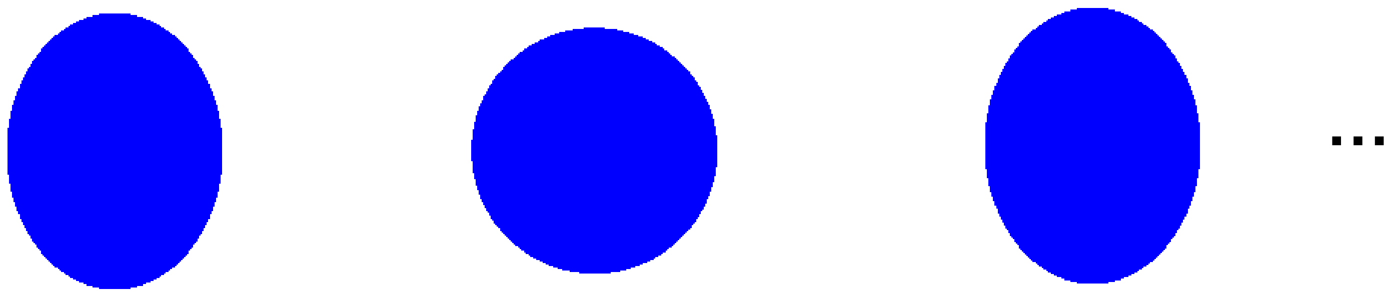
Gravitational Waves: What They Are

GWs are perturbations of the spacetime metric

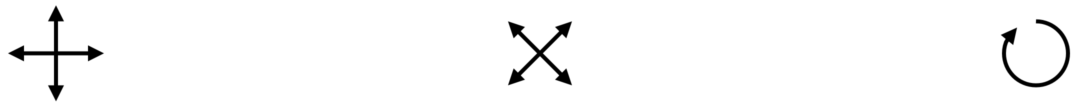
$\Rightarrow$ **Change the effective distance between locally inertial points**

According to GR, there are two polarization components

Wave can be a linear combination of polarization components



"Plus" polarization   "Cross" polarization   Circular polarization



These represent a time-dependent ***strain*** transverse to the direction the wave is traveling – i.e . $h_+(t)$, $h_\times(t)$

GWs can give us a unique view of astrophysical systems & events

GWs are powerful: can drive the dynamics of a system

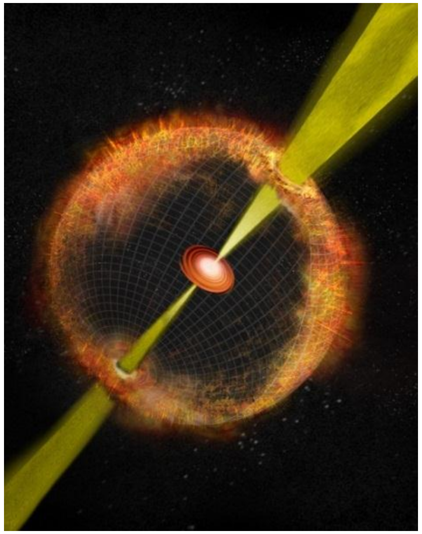
Not scattered by matter

⇒ probe the core engine of the event

Complementary to electromagnetic observations

Reveal “dark” systems such as black hole binaries

Enable tests of GR vs. other theories of gravity



Credit: Bill Saxton, NRAO/AUI/NSF

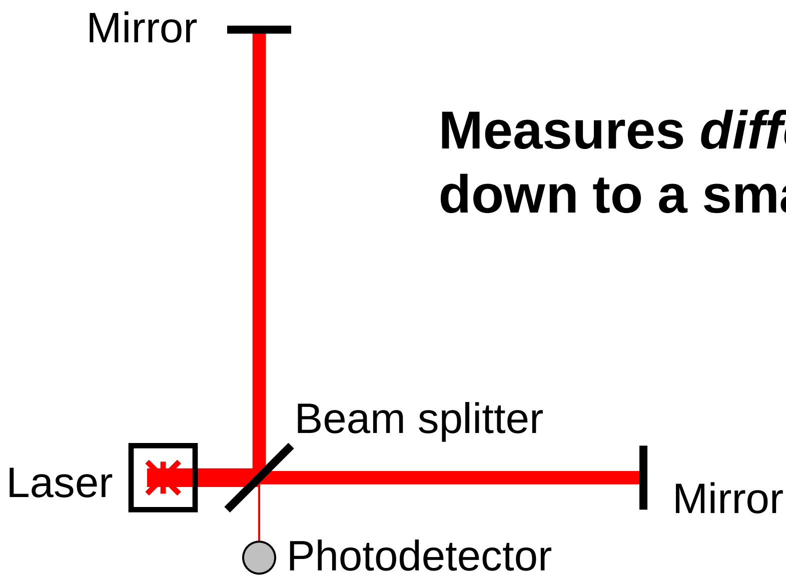
Detecting GWs is a supreme challenge

GWs are incredibly weak by the time they reach Earth – Typical strain ~ $\mathbf{10^{-21}}$ or even smaller !

An unexplored frontier !

Can we really hope to detect them ?!?

Response of a GW Interferometer

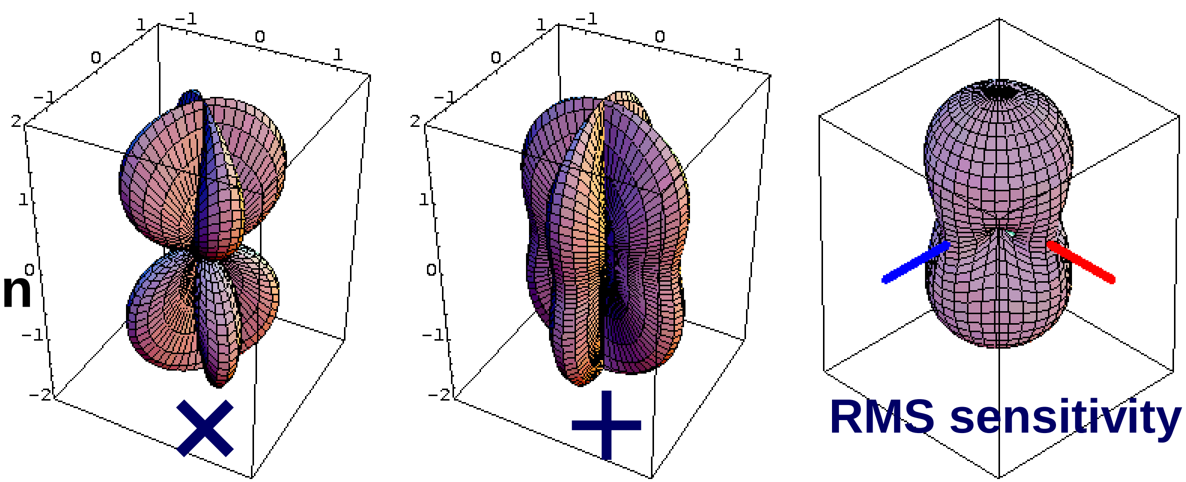


Measures *difference* in effective arm lengths down to a small fraction of a wavelength

In general, a linear combination:

$$h_{\mathrm{det}}(t) = F_+ \, h_+(t) + F_\times \, h_\times(t)$$

Directional sensitivity depends on polarization
in a certain (+,×) basis

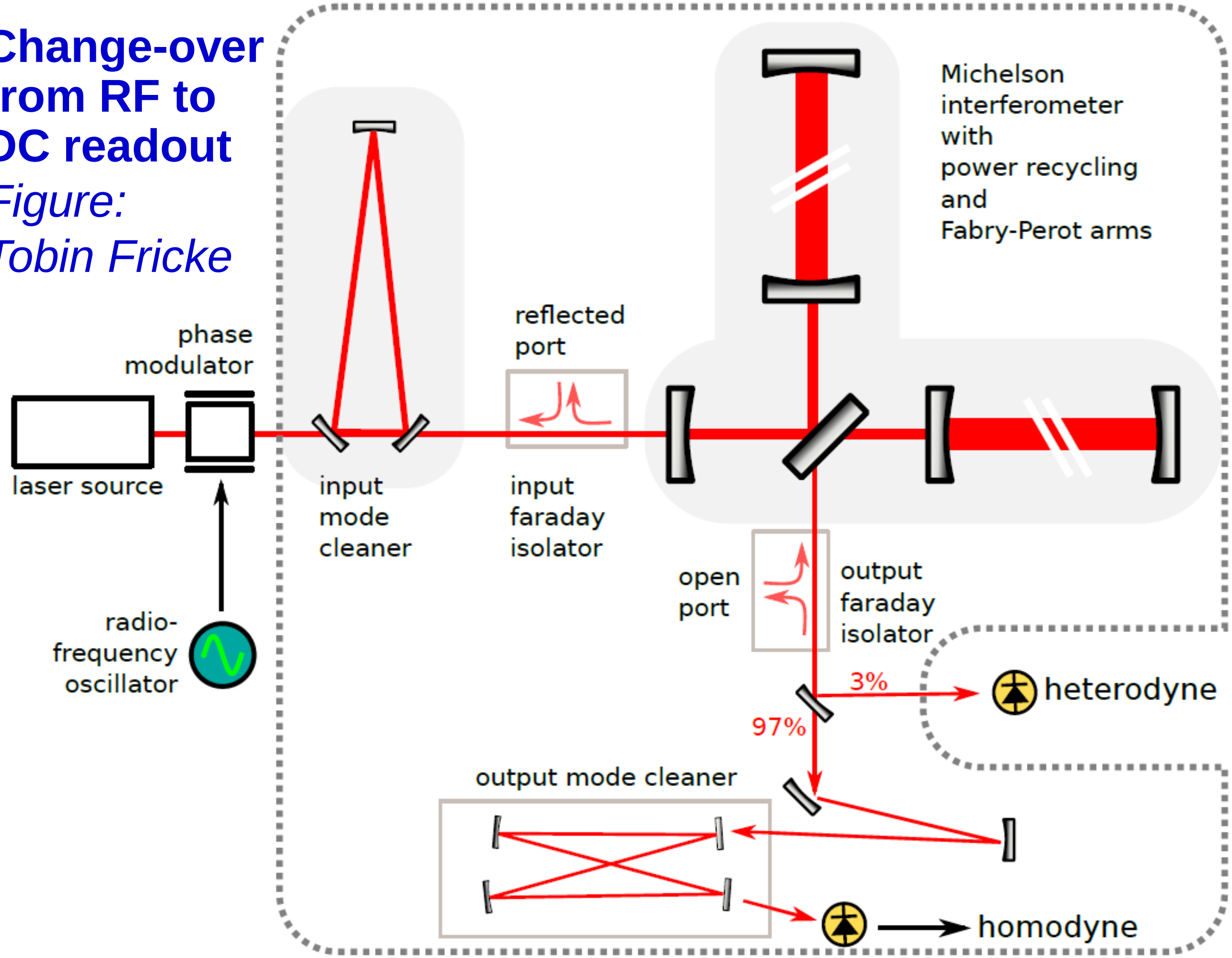


GW Detector Readout – Overview

The eLIGO interferometer

Change-over from RF to DC readout

Figure: Tobin Fricke



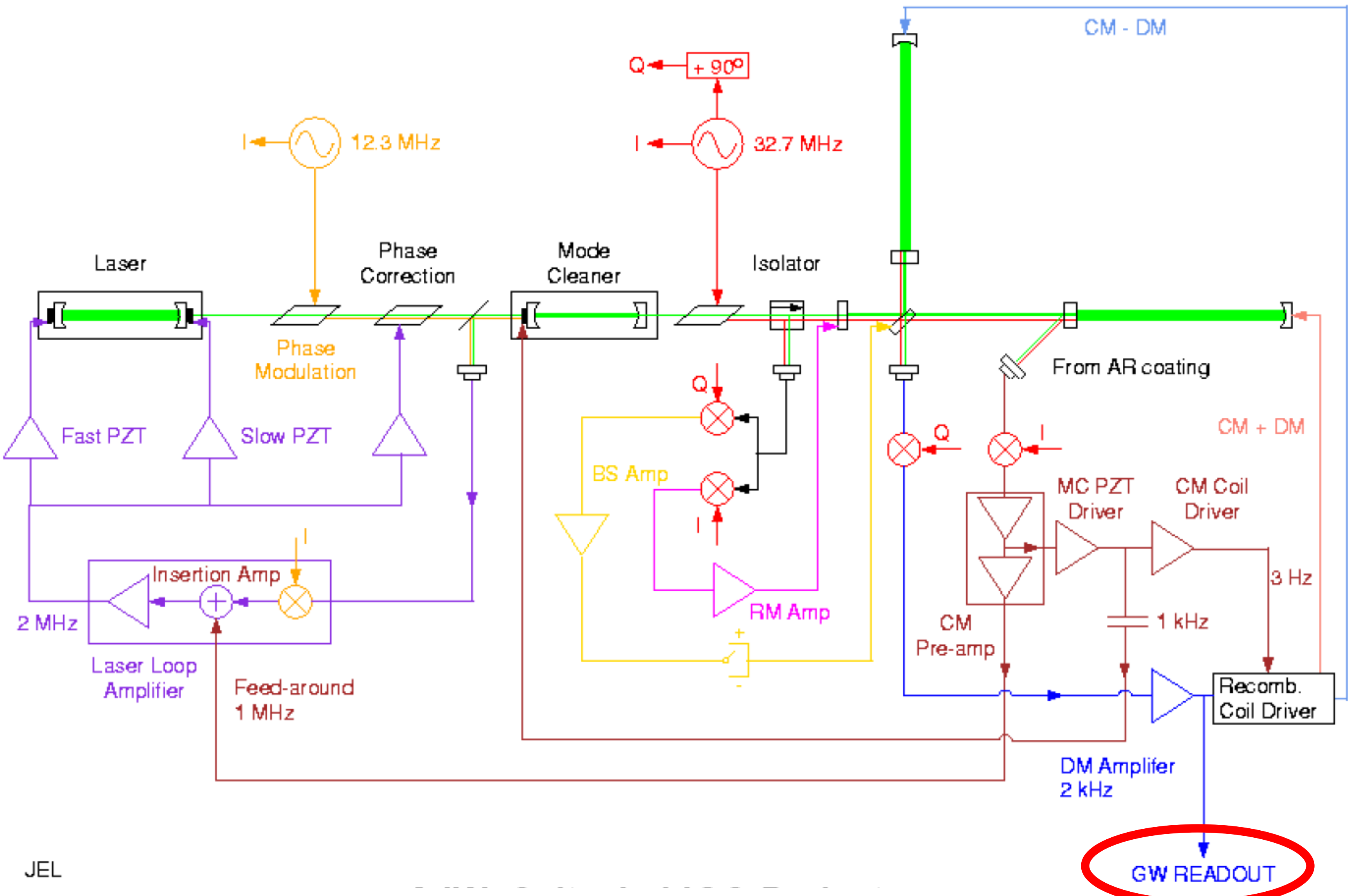
Heterodyne (RF) readout used for initial LIGO/Virgo

- Modulate phase of input light (33 MHz), demodulate signal measured by photodiode
- Perfect destructive interference on avg

Homodyne (DC) readout used for Adv. LIGO/Virgo

- Measure intensity variations
- Arm lengths offset

LIGO Length Sensing and Control
(RF Readout)



JEL

Gravitational-Wave Data

Data = Instantaneous estimate of strain for each moment in time

i.e. demodulated channel sensitive to arm length difference

That's not the whole story – we'll come back to calibration later

Digitized discrete time series recorded in computer files

$$(t_j , x_j)$$

LIGO and GEO sampling rate: 16384 Hz ≡ f_s

VIRGO sampling rate: 20000 Hz

Synchronized with GPS time signal

Common "frame" file format (*.gwf)

***Many* auxiliary channels recorded too**

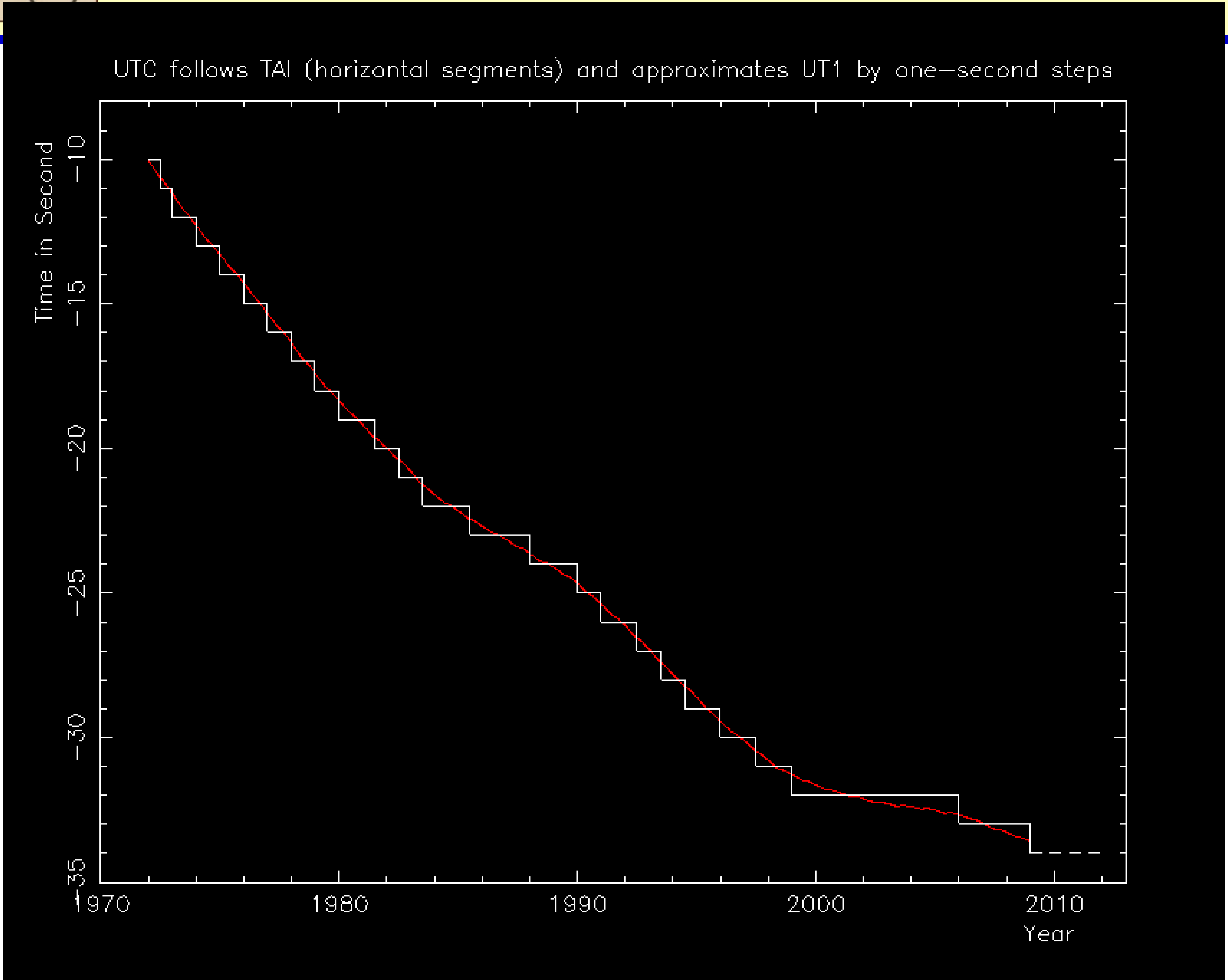
Total data volume: a few megabytes per second per interferometer

Leap Second Coming at End of June

http://hpiers.obspm.fr/iers/bul/bulc/bulletinc.dat

```
Paris, 5 January 2012

Bulletin C 43

To authorities responsible
for the measurement and
distribution of time


                    UTC TIME STEP
                on the 1st of July 2012


A positive leap second will be introduced at the end of June 2012.
The sequence of dates of the UTC second markers will be:

                2012 June 30,     23h 59m 59s
                2012 June 30,     23h 59m 60s
                2012 July  1,      0h  0m  0s

The difference between UTC and the International Atomic Time TAI is:

 from 2009 January 1, 0h UTC, to 2012 July 1  0h UTC  : UTC-TAI = - 34s
 from 2012 July 1,    0h UTC, until further notice    : UTC-TAI = - 35s

Leap seconds can be introduced in UTC at the end of the months of December
or June, depending on the evolution of UT1-TAI. Bulletin C is mailed every
six months, either to announce a time step in UTC or to confirm that there
will be no time step at the next possible date.
```

Leap Seconds – Historical



http://hpiers.obspm.fr/eop-pc/

Relevance of the Sampling Rate

Is 16384 Hz a high enough sampling rate ?

The Sampling Theorem:

Discretely sampled data with sampling rate f_s can completely represent a continuous signal which only has frequency content below the Nyquist frequency, $f_s / 2$

GW signals of interest to ground-based detectors typically stay below a few kHz

e.g. binary neutron star inspiral reaches ISCO at ~1 to 1.5 kHz

Neutron star *f*-modes: ~3 kHz

Black hole quasinormal modes: ~1 kHz for 10 $M_\odot$

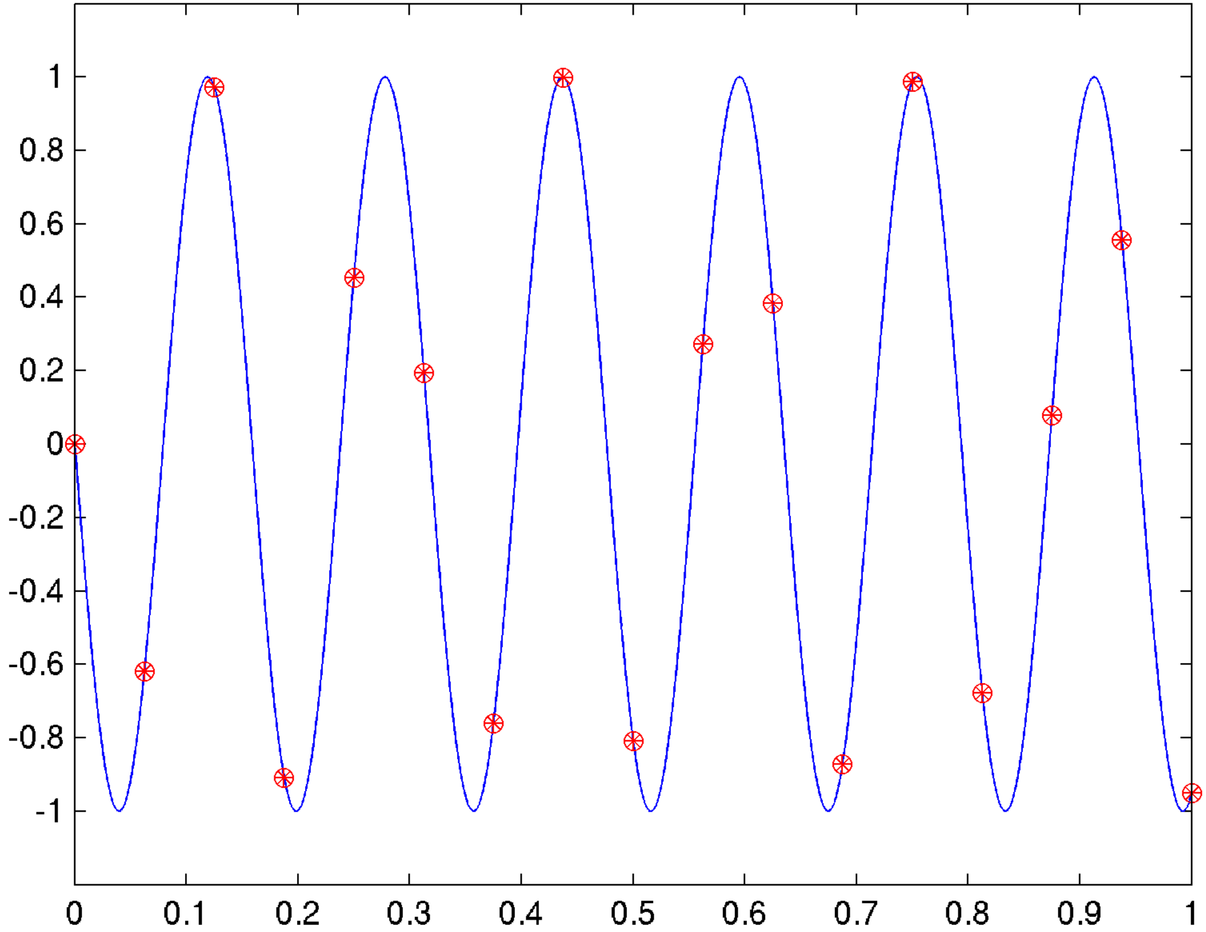
Some core collapse supernova signals could go up to several kHz

What if the signal extends above Nyquist frequency?

Higher frequencies are “aliased” down to lower frequencies

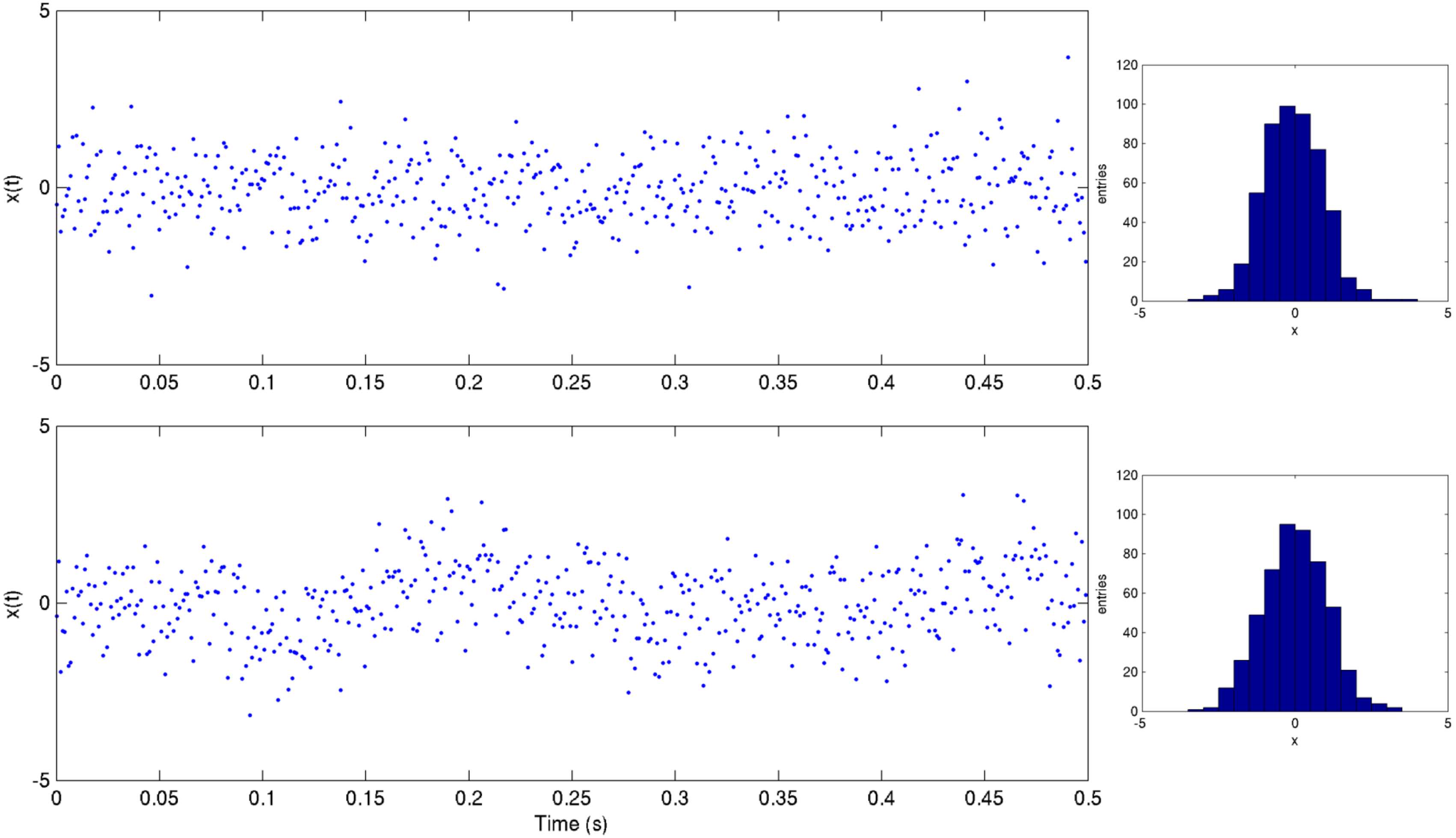
Aliasing

f_s = 16 Hz ; signal frequency = 9.7 Hz



Characterizing Noise

Noise is random, but its *properties* can be characterized



Possible Properties of Noise

Stationary : statistical properties are independent of time

Ergodic process: time averages are equivalent to ensemble averages

Gaussian : A random variable follows Gaussian distribution

For a single random variable, $$p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu_x)^2}{\sigma_x^2}\right]$$

More generally, a *set* of random variables (e.g. a time series) is Gaussian if the joint probability distribution is governed by a covariance matrix

$$C_{xij} := \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$$

such that

$$p(x_1, x_2, \ldots, x_N) = \frac{1}{(2\pi)^{N/2}\sqrt{\det C_x}} \exp\left[-\frac{1}{2}\sum_{i,j=0}^{N-1} C_{xij}^{-1}(x_i - \mu_{xi})(x_j - \mu_{xj})\right]$$

White : Signal power is uniformly distributed over frequency

$\Rightarrow$ Data samples are uncorrelated

Frequency-Domain Representation of a Time Series

Fourier transform

$$\tilde{x}(f) = \int_{-\infty}^{\infty} dt\, x(t) e^{-i2\pi f t}$$

$$\Rightarrow \qquad x(t) = \int_{-\infty}^{\infty} df\, \tilde{x}(f) e^{i2\pi f t}$$

A linear function, complex in general

Defined for all positive *and negative* frequencies

Frequency-Domain Representation of a *Discrete, Finite* Time Series

Time series x_j with N samples at times $t_j = t_0 + j\,\Delta t$

Discrete Fourier transform

$$\tilde{x}_k := \sum_{j=0}^{N-1} x_j \, e^{-i2\pi jk/N}$$

$$\Rightarrow \qquad x_j = \frac{1}{N} \sum_{k=-N/2}^{N/2-1} \tilde{x}_k \, e^{i2\pi jk/N}$$

Frequency spacing is inversely proportional to N

Efficient way to calculate complete discrete Fourier Transform: Fast Fourier Transform (FFT)

Power Spectral Density

Parseval's theorem:

$$\int_{-\infty}^{\infty} dt\, |x(t)|^2 = \int_{-\infty}^{\infty} df\, |\tilde{x}(f)|^2$$

$\Rightarrow$ Total energy in the data can be calculated in either time domain or frequency domain

$|\tilde{x}(f)|^2$ can be interpreted as energy spectral density

When noise (or signal) has infinite extent in time domain, can still define the power spectral density (PSD)

$$\lim_{T\to\infty} \frac{1}{T} \left|\tilde{x}_T(f)\right|^2$$

Watch out for one-sided vs. two-sided PSDs

Estimating the PSD

Generally we need to determine the PSD empirically, using a finite amount of data

Simplest approach: FFT the data, calculate square of magnitude of each frequency component – this is a periodogram

For stationary noise, one can show that the frequency components are statistically independent

This estimate is unbiased (has the correct mean), but has a large variance – so average several periodograms

Alternately, smooth periodogram; give up frequency resolution either way

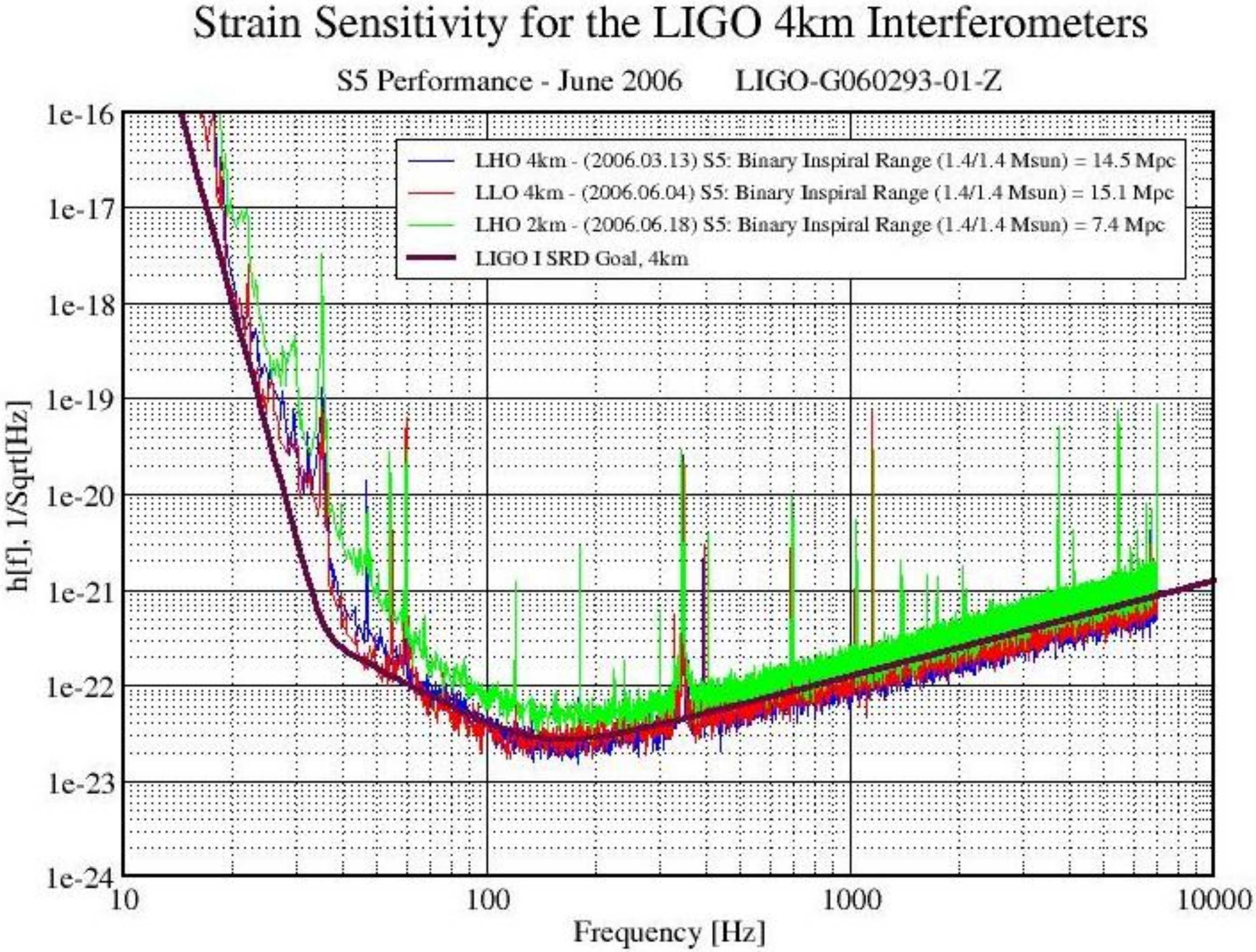
Generally apply a "window" to the data to avoid spectral leakage

Leakage arises from the assumption that the data is periodic!

Tapered window forces data to go to zero at ends of time interval

Welch's method of estimating a PSD averages periodograms calculated from windowed data

Amplitude Spectral Density of LIGO Noise



Interpretation of Time Series Data

Recorded data values are *not* simply proportional to GW strain

A linear system, but that does not guarantee proportionality !

Frequency-dependent amplitude and phase relation (i.e. transfer function)

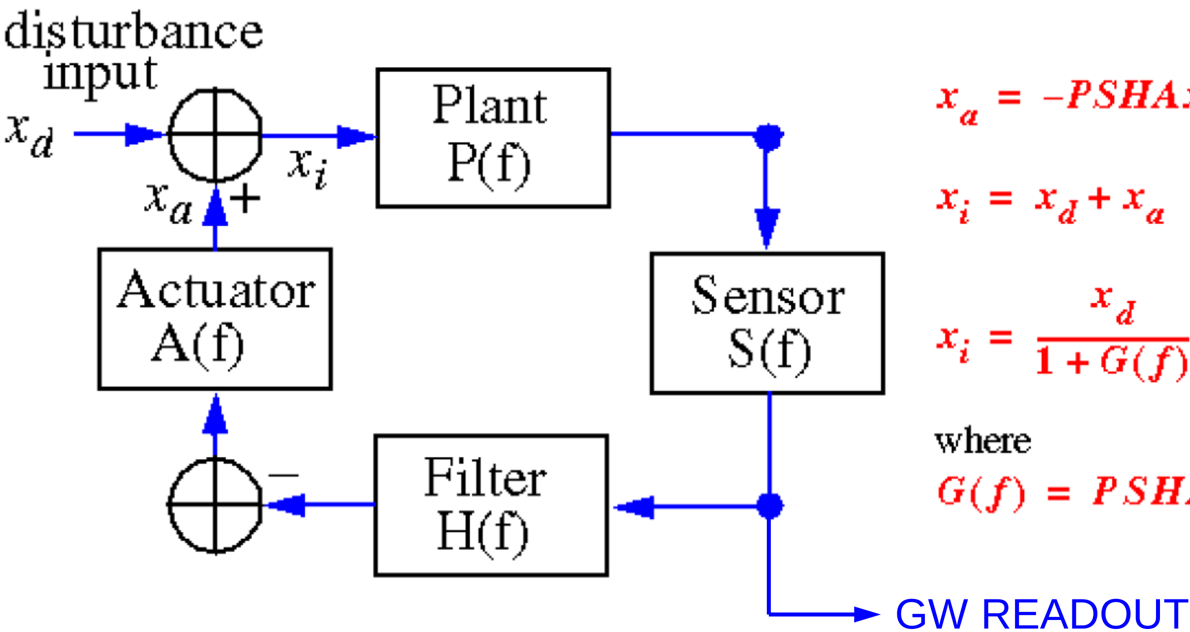
Instrumental and practical reasons

$\Rightarrow$ **Raw time series is a distorted version of GW strain signal**

e.g. a delta-function GW signal produces an output with a characteristic shape and duration ("impulse response")

Want to recover actual GW strain for analysis

Calibration



$$x_a = -PSHAx_i$$

$$x_i = x_d + x_a$$

$$x_i = \frac{x_d}{1 + G(f)}$$

where

$$G(f) = PSHA$$

Monitor P(f) **continuously with "calibration lines"**

Sinusoidal arm length variations with known absolute amplitude

Apply frequency-dependent correction factor to get GW strain

$$h = (\text{GW READOUT}) \times \frac{1 + G(f)}{P(f)\,S(f)}$$

Basics of Digital Filtering

A filter calculates an output time series from a linear combination of the elements of an input time series

Finite Impulse Response (FIR) filter

Calculated *only* from the input time series

Typical form: $y_i = b_0 x_i + b_1 x_{i-1} + b_2 x_{i-2} + \ldots + b_{N-1} x_{i-N}$

Infinite Impulse Response (IIR) filter

Also uses prior elements of the output time series

e.g. $y_i = b_0 x_i + b_1 x_{i-1} + b_2 x_{i-2} + \ldots + b_{N-1} x_{i-N} + a_1 y_{i-1} + a_2 y_{i-2} + \ldots$

Choice of coefficients determines transfer function

Many filter design methods, depending on goals

Causality and phase lag

Linear-phase and zero-phase filters

Watch for transient in filter output at beginning of data stream!

Applications of filtering

High-pass, low-pass, band-pass, band-stop, etc.

Anti-aliasing for down-sampling

Low-pass filter to cut away signal content above new Nyquist frequency

Whitening / Dewhitening

Time for some exercises ...

Based on Matlab – but the UTB laptops have Octave

Work by yourself or with a partner

How to get help:

- Ask me or a neighbor
- Use Matlab's/Octave's built-in help
- Consult a book – I have one here

The items in the handout are intended as a guide

- Feel free to explore !