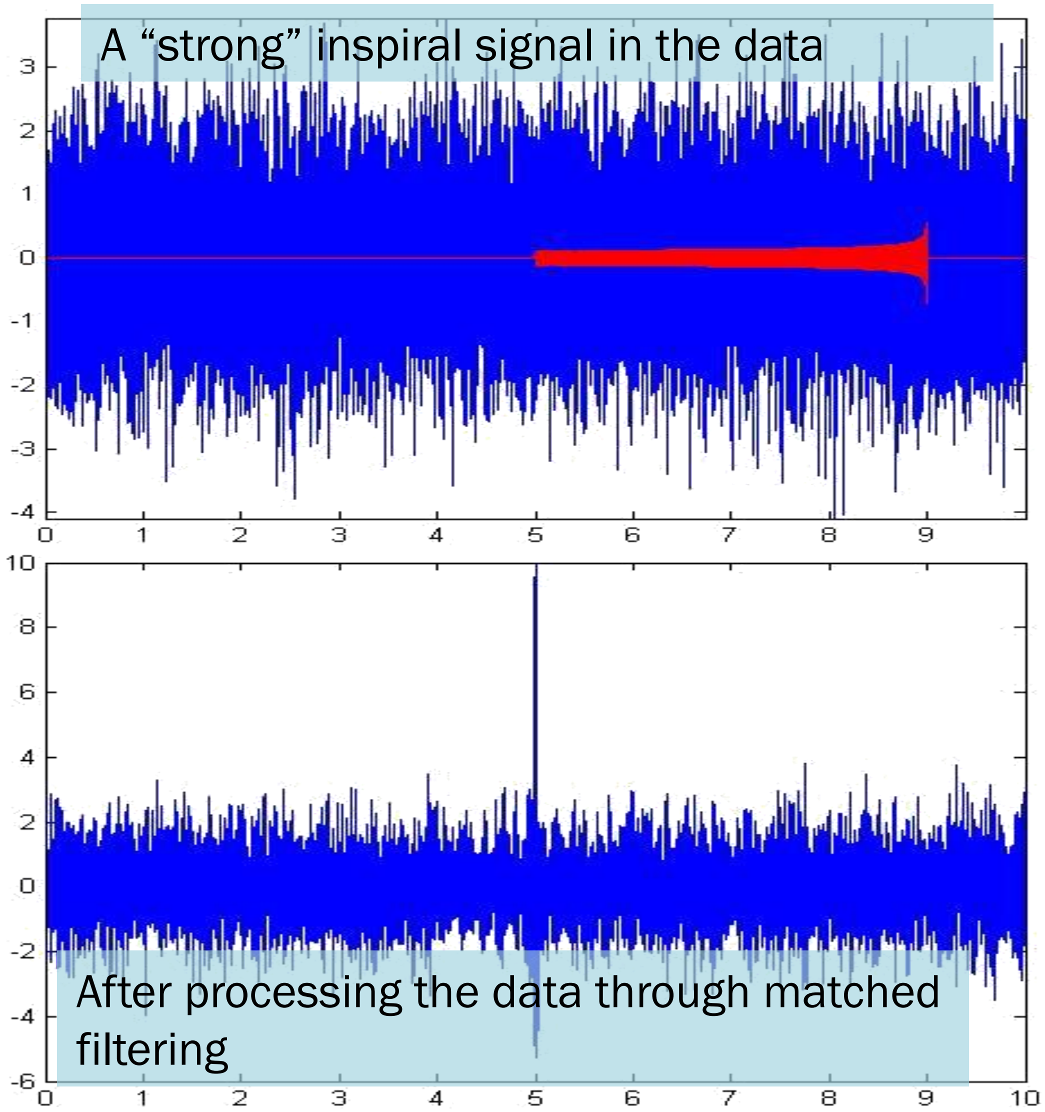
Noise

Noise in GW data analysis

- GW data in current detectors is noise dominated and will remain so for future detectors
- The main objective of data analysis in GW astronomy is to extract signals from noise dominated data



Probability theory refresher

Random variable and sample space

- Appendix A of "Swarm intelligence methods for statistical regression"

Random variable	Trial	Trial outcome (or value)	Sample space
• A quantity whose value cannot be predicted	• The process of obtaining the value of a random variable	• The value of a random variable in a trial	• The set of all possible trial values

Probability density function

- Continuous or discrete sample space $\Rightarrow$**Continuous** or **Discrete** random variable
- Continuous random variable X: **Probability density function** $p_X(x)$

$$\Pr(X \in [x, x + dx]\,) = p_X(x)dx$$

$$\Pr(X \in \mathbb{A} \subset \mathbb{R}^1\,) = \int_{\mathbb{A}} dx\, p_X(x)$$

$$\Rightarrow \int_{-\infty}^{\infty} p_X(x)dx \;= 1$$

Frequentist and Bayesian

- Frequentist: Obtain probability of an event from its frequency of occurrence in large (approaching infinity) number of trials

$$\Pr(X \in [x, x + dx]) = p_X(x)dx = \text{Frequency of getting a value in } [x, x + dx]$$

  - Example: Throwing a coin multiple times and estimating the frequency of getting heads

- Bayesian: Probability of an event is our degree of belief in it

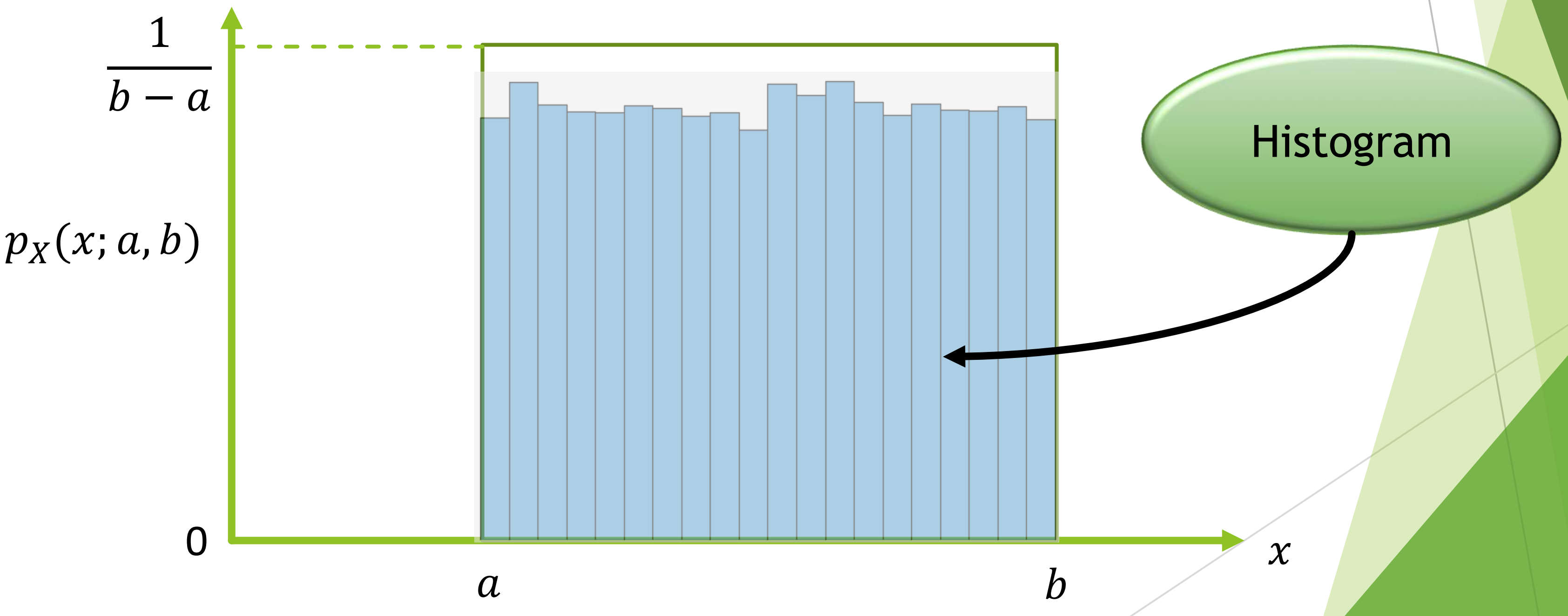
$$\Pr(X \in [x, x + dx]) = p_X(x)dx = My\ belief\ that\ X\ is\ in\ [x, x + dx]$$

  - Example: I believe that there is a 70% chance of rain in the next 1 hour

Uniform pdf

$$p_X(x;a,b) = U(x;a,b) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

Matlab: `rand`

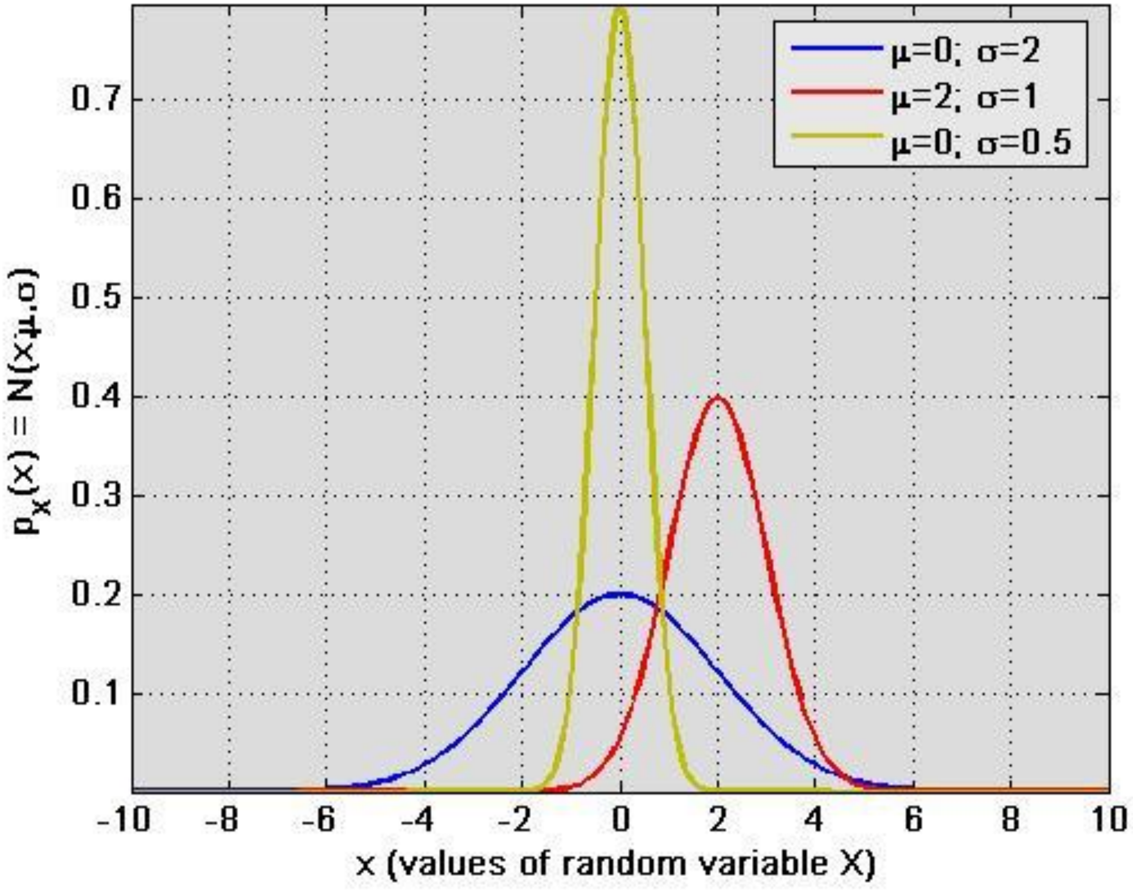


Normal pdf

$$p_X(x;\mu,\sigma) = N(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Matlab: `randn`

- Also called “Gaussian” pdf in the physics literature
- Two parameters:
  - Mean (μ)
  - Standard deviation (σ)



Expectation

Consider any function $f(X)$ of a random variable X

Expectation of $f(X)$: $E[f(X)] = \int_{-\infty}^{\infty} f(x) p_X(x) dx$

$f(x) = x$

mean of $p_X(x)$

$f(x) = x^n$

$\mathbf{n^{th}}$ order moment of $p_X(x)$

$f(x) = \ln p_X(x)$

Information Entropy of $p_X(x)$

$f(x) = (x - E[X])^n$: n^{th} moment around the mean ("**central moment**")

$n = 2 \rightarrow$ "variance" σ^2 ("standard deviation" : σ)

Joint probability density function

- Joint pdf of two continuous random variables X and Y
$$\Pr(X \in [x, x+dx] \text{ AND } Y \in [y, y+dy] \text{ in a trial}) = p_{XY}(x,y)dx\,dy$$
$$\Pr\big((X,Y) \in \mathbb{A} \subset \mathbb{R}^2\big) = \iint_{\mathbb{A}} p_{XY}(x,y)dxdy$$
- Similarly, joint pdf of N random variables $(X_1, X_2, \dots, X_N)$
$$\Pr\big((X_1, X_2, \dots, X_N) \in \mathbb{V} \subset \mathbb{R}^N\big) = \int_{\mathbb{V}} p_{X_1 X_2 \dots X_N (x_1, x_2, \dots, x_N)}\, d^N x$$
$$\Rightarrow \int_{\mathbb{R}^N} p_{X_1 X_2 \dots X_N (x_1, x_2, \dots, x_N)}\, d^N x = 1$$
- Random variables X and Y are said to be **statistically independent** if
$$p_{XY}(x,y) = p_X(x)p_Y(y)$$

Joint Expectations

$$E[f(X_1,\ldots,X_N)] = \int_{-\infty}^{\infty} f(x_1,\ldots,x_N)\,p_{X_1,\ldots,X_N}(x_1,\ldots,x_N)\,dx_1,\ldots,dx_N$$

$E[(X_1 - E[X_1])^{m_1}(X_2 - E[X_2])^{m_2}\ldots(X_N - E[X_N])^{m_N}]$ is called the $m_1 + m_2 + \ldots + m_N$ order joint (central) moment

- Of course, there are many possible joint moments of a given order
- Of special importance is the joint **central moment of second order**

$C_{ij} = E[(X_i - E[X_i])(X_j - E[X_j])]$

C_{ij} is called the **Covariance** of X_i and X_j

Covariance Matrix *C*

Symmetric positive definite:

$\bar{x}\boldsymbol{C}\bar{x}^T > 0$ for $\bar{x} \in \mathbb{R}^N \neq 0$

Bivariate Normal pdf

$$p_{XY}(x,y;\bar{\mu},\mathbf{C}) = \frac{1}{2\pi|\mathbf{C}|^{1/2}}\exp\left(-\frac{1}{2}(\bar{x}-\bar{\mu})^T\mathbf{C}^{-1}(\bar{x}-\bar{\mu})\right)$$

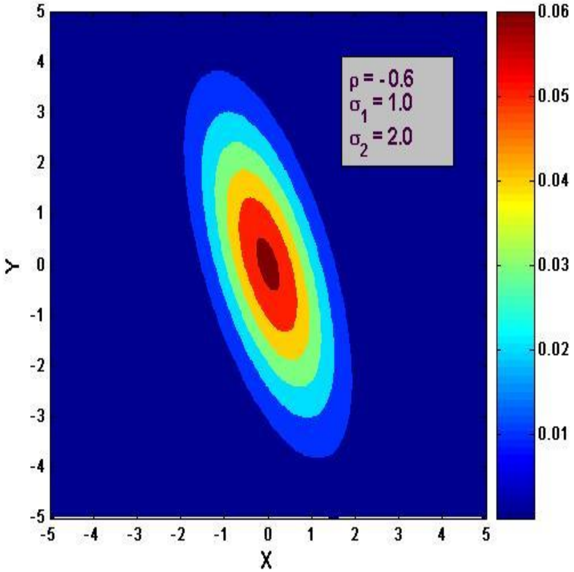
$$\bar{x} = \begin{pmatrix} x \\ y \end{pmatrix};\ \ \bar{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$$

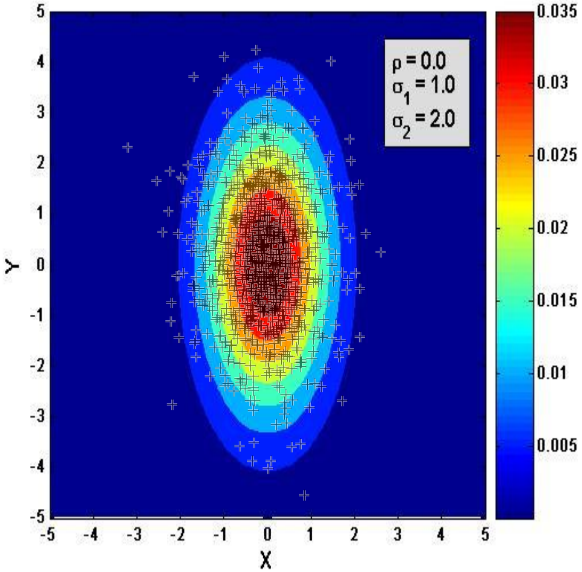
$$\mathbf{C} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix};\ -1 < \rho < 1$$

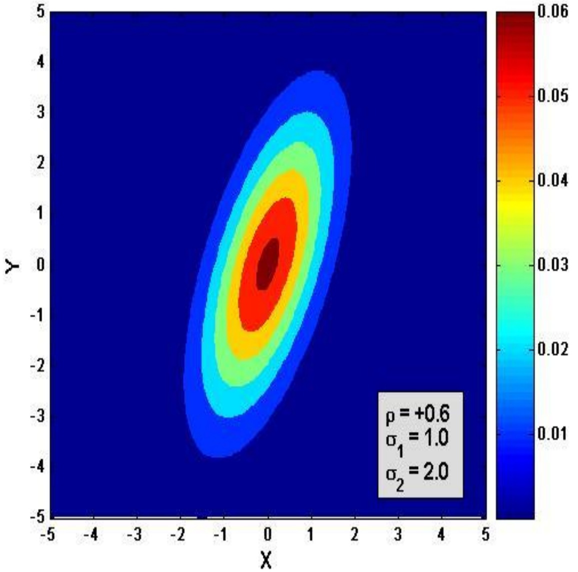
$|\boldsymbol{C}|$ is the determinant of $\boldsymbol{C}$

$\rho = 0 \Rightarrow p_{XY}(x,y;\bar{\mu},C) = p_X(x;\mu_x,\sigma_x)p_Y(y;\mu_y,\sigma_y)$: Statistical independence

$\mu_x = \mu_y = 0$
$\sigma_x = 1.0, \sigma_y = 2.0$
$\rho \in \{-0.6, 0, 0.6\}$







Multivariate Normal pdf

$$p_{\bar{X}}(\bar{x};\bar{\mu},\boldsymbol{C}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{C}|^{1/2}} \exp(-\frac{1}{2}(\bar{x}-\bar{\mu})^T\boldsymbol{C}^{-1}(\bar{x}-\bar{\mu}))$$

- $\bar{\mu} \in \mathbb{R}^N$: Mean vector, $\mu_i = E[X_i]$
- $\boldsymbol{C}$ is the covariance matrix $\rightarrow$ $\mathrm{C}_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$
- $\boldsymbol{C}$ is **positive definite** and **symmetric** $\Rightarrow$ Ellipsoid in $\mathbb{R}^N$
- Orientation of ellipsoid given by eigenvectors of **C**
- Dimensions of the ellipsoid given by eigenvalues of **C**
- If $\boldsymbol{C}$ is a diagonal matrix $\Rightarrow$ All elements of $\bar{X}$ are statistically independent

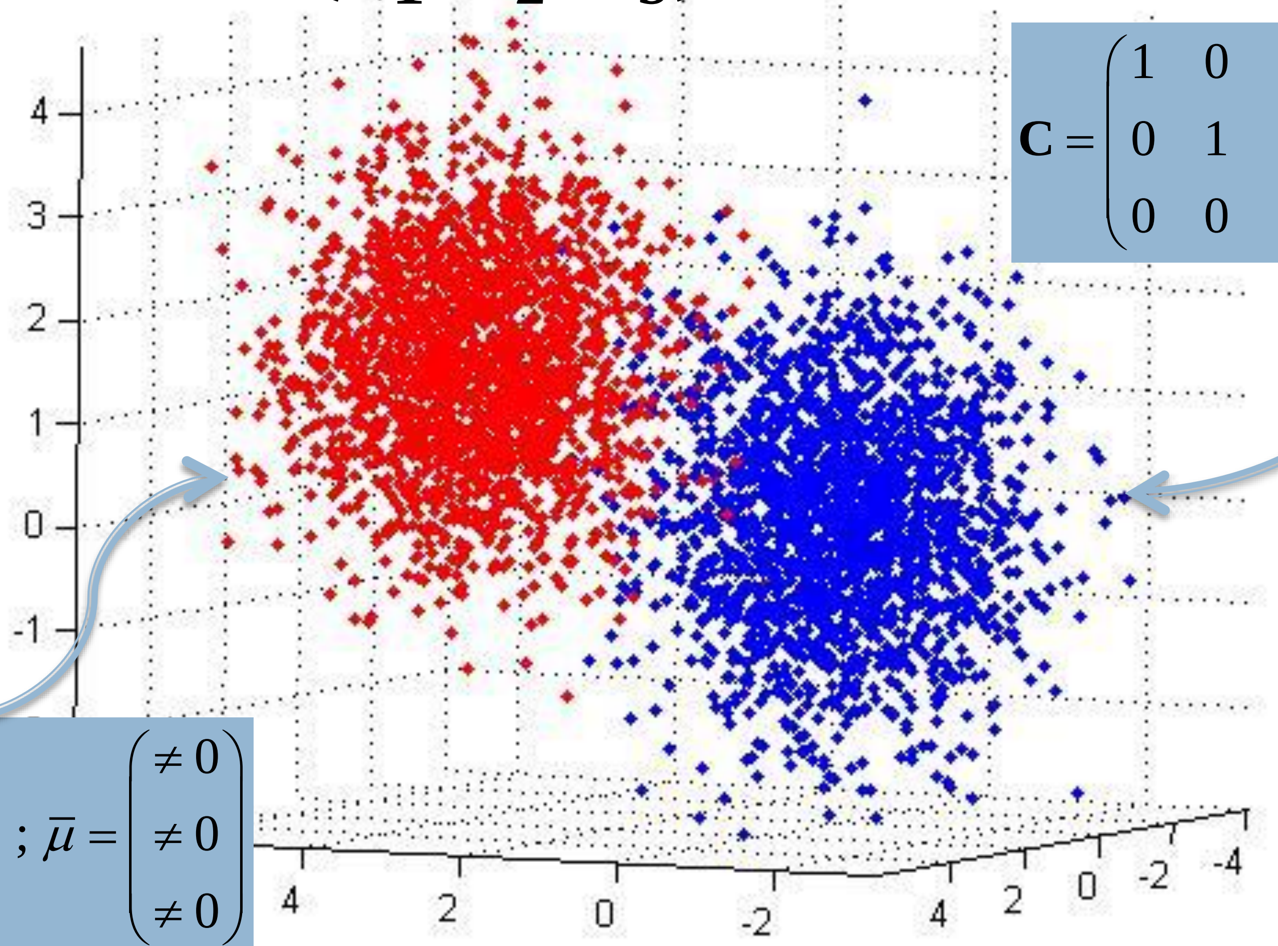
Geometrical View

$$(X_1, X_2, X_3) \in \mathbb{R}^3$$

Trials outcome $(x_1, x_2, \ldots, x_N) \in \mathbb{R}^N$ can be visualized as a point in N-dimensional space: Joint pdf tells us how the points are going to be distributed

Example: Trial outcomes from Tri-variate normal pdf

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} ; \bar{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$



$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} ; \bar{\mu} = \begin{pmatrix} \neq 0 \\ \neq 0 \\ \neq 0 \end{pmatrix}$$

Marginal and conditional pdf

- Marginalization:

$$p_X(x) = \int_{-\infty}^{\infty} p_{XY}(x, y)dy$$

- Conditional pdf gives the probability of $X \in [x, x + dx]$ given that Y has taken value y in a trial

$$p_{X|Y}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)}$$

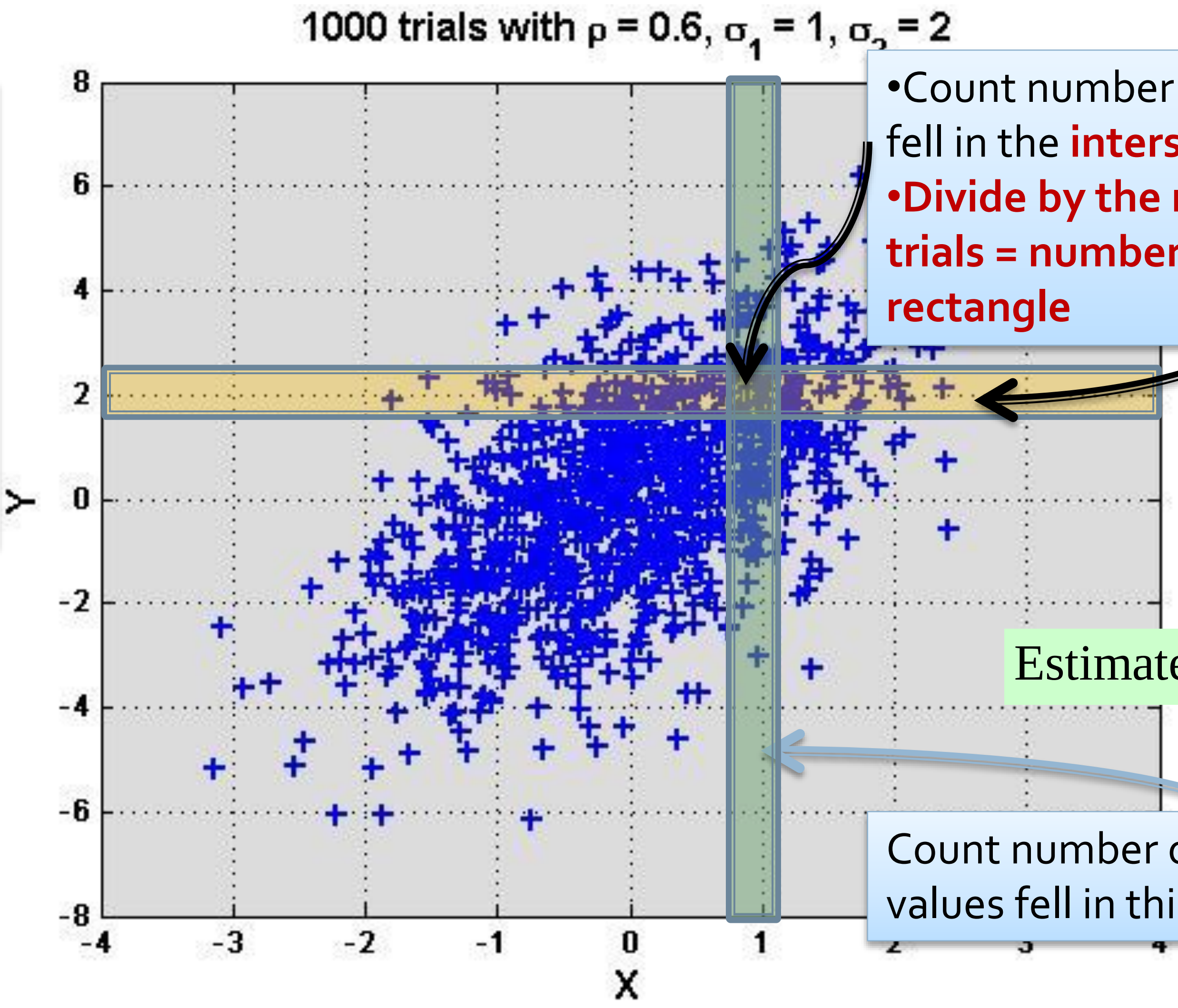
Marginal and conditional pdf

Marginalization:

$$p_X(x) = \int_{-\infty}^{\infty} p_{XY}(x,y)dy$$

Conditional PDF:

$$p_{X|Y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)}$$



•Count number of times the trial values fell in the **intersection (square)**

•**Divide by the number of effective trials = number of trials in this rectangle**

Estimate $p_{X|Y}(x \mid y)$

Estimate marginal pdf $p_X(x)$

Count number of times the trial values fell in this rectangle

Stochastic process (NOISE)

Stochastic process: definition

- An **ordered** sequence of random variables is called a **Stochastic Process** (or Noise)

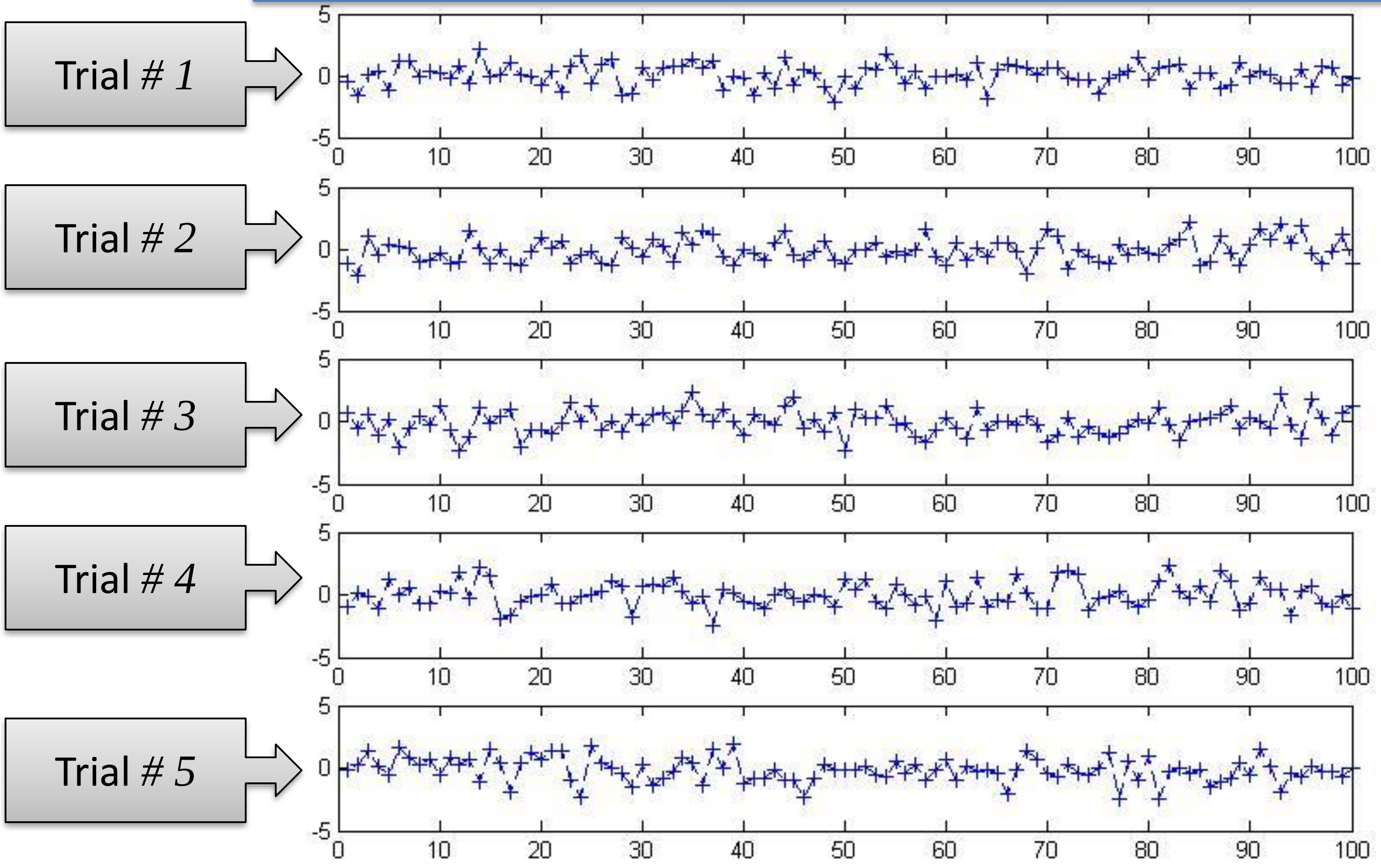
$$\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots$$

- Changing the ordering of the random variables changes the stochastic process
- Stochastic processes are countably infinite sequences in general
- In practice, we always deal with some finite **sub-sequence**

$$\bar{X} = \left(X_p, X_{p+1}, \dots, X_{p+N-1}\right) \equiv (X_1, X_2, \dots, X_N)$$

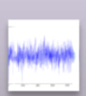
Realizations of a stochastic process

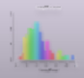
Different **realizations** of a stochastic process subsequence $\bar{X} = (X_1, X_2, \dots, X_{100})$



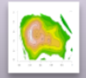
Data realization

- Download any segment of data from GWOSC

-  The noise in that segment is a realization of some underlying stochastic process $\Rightarrow$ the data segment is also a realization of some stochastic process

-  Any analysis results obtained from the given data will be affected by the noise realization in it

- Since noise values are random and cannot be predicted, we cannot say how much of the data is noise and how much is not $\Rightarrow$ We cannot say how much the results have been affected by noise

-  We can only make probabilistic statements about the effect of noise on results

Gaussian noise

- The joint pdf of **any** sub-sequence is a multivariate normal pdf

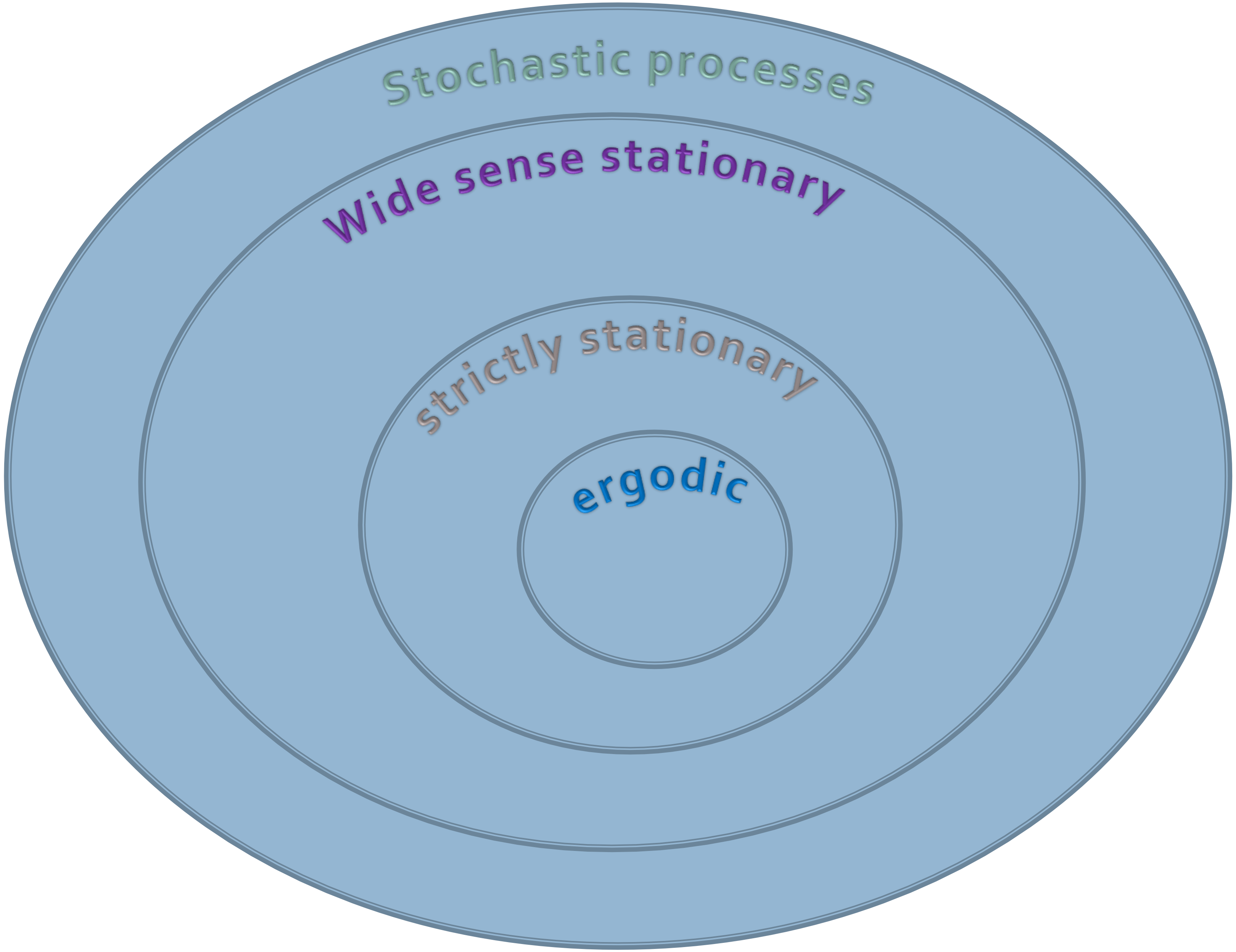
$$p_{\bar{X}}(\bar{x}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{C}|^{1/2}} \exp(-\frac{1}{2}||\bar{x} - \bar{\mu}||^2)$$

- $\bar{x} = (x_0, x_1, \dots, x_{N-1}) \in R^N$ (row vector)
- $E[X_i] = \mu_i$
- $C_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$: Covariance matrix
- $|\boldsymbol{C}|$: Determinant of $\boldsymbol{C}$
- $||\bar{x}||^2 = \langle \bar{x}, \bar{x} \rangle$ where $\langle \bar{x}, \ \bar{y} \rangle = \ \bar{x}\boldsymbol{C}^{-1}\bar{y}^T$

Classification of stochastic processes

Stationary stochastic process

- **Wide-sense stationary** process: All first and second order moments are time-translation independent
  - $E[\ X_k\ ]$ is constant, independent of k
  - $E[\ X_k X_{k+m}\ ]$ is independent of k and dependent only on m
  - $\Rightarrow E[(X_k - E[X_k])^2] = \sigma^2 = const.$
    - Makes sense to talk about the mean and variance of the stochastic process; The mean and variance can be different for different elements of a general stochastic process
- **Strictly stationary** stochastic process: moments of **all order** are time-translation independent


Stochastic processes
Wide sense stationary
strictly stationary
ergodic

Autocovariance

- Autocovariance sequence of a wide-sense stationary stochastic process (with mean $\mu = E[X_j], \forall j$)

$$\phi(m) = E[(X_k - \mu)(X_{k+m} - \mu)]$$

(No dependence on k because of wide-sense stationarity)

- It can be proven that:

$$\phi(0) = \sigma^2 \text{ : variance}$$

$$\phi(-m) = \phi(m) \text{ : even sequence}$$

$$\phi(0) > \phi(m \neq 0)$$

The definitions are more convenient if one switches to continuous index (“time”): $\phi(\tau) = E[(X(t) - \mu)(X(t + \tau) - \mu)]$

Power spectral density (PSD)

Power Spectral Density

Wide-sense stationary process with autocovariance $\phi(\tau)$

$$S_n(f) = \int_{-\infty}^{\infty} d\,\tau \phi(\tau) e^{-2\pi i f \tau}$$

$S_n(f)$ is real and symmetric as $\phi(\tau)$ is an even function
(Sign of phase in the Fourier transform does not matter)

Physical interpretation

The variance of the noise process

$$\sigma^2 = \phi(0) = \int_{-\infty}^{\infty} d\,f S_n(f)$$

Hence, $S_n(f)df$ can be interpreted as the noise variance contributed by the band $[f, f + df] \Rightarrow S_n(f)$ is the density of the variance ("power") in Fourier frequency ("spectrum") $\Rightarrow$

Power Spectral Density

White noise

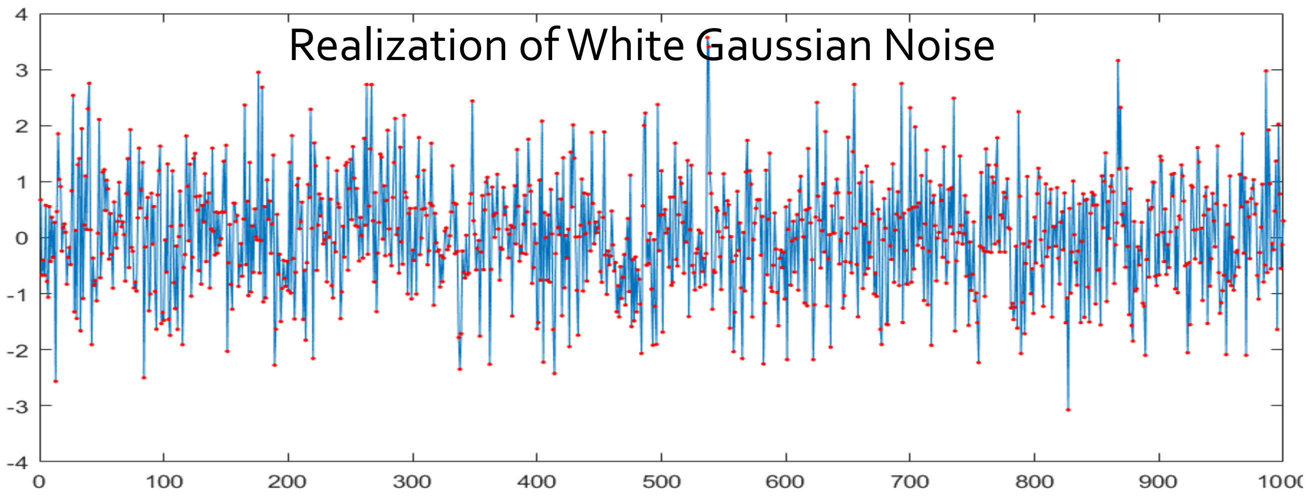
- Wide-sense stationary noise with autocovariance:

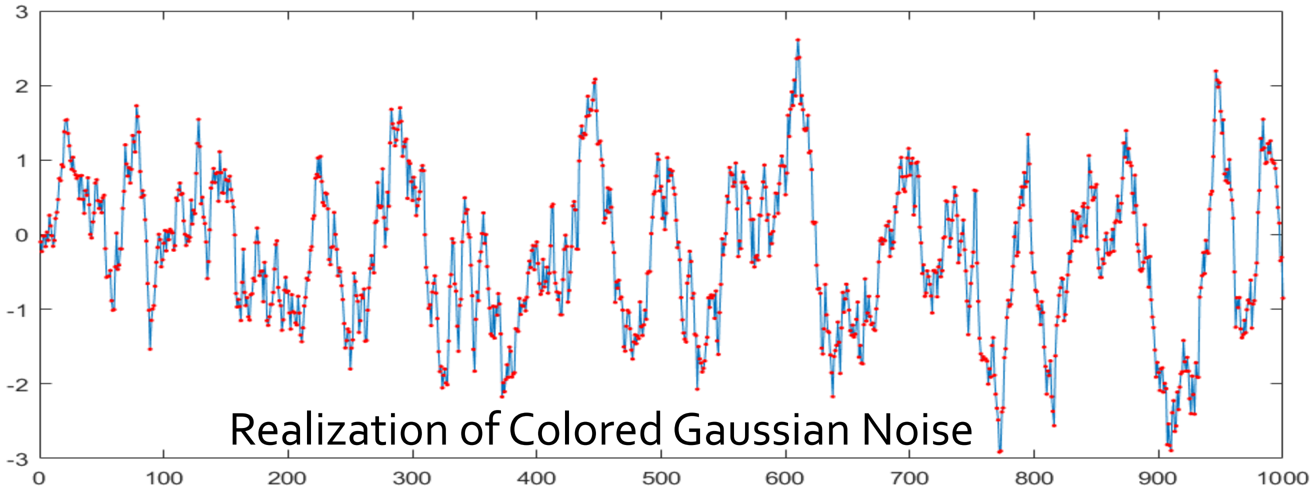
$$\phi(\tau) = \sigma^2 \delta(\tau) \Rightarrow S_n(f) = const.$$

is called **White Noise** because the power contributed by a given band is the same at all frequencies

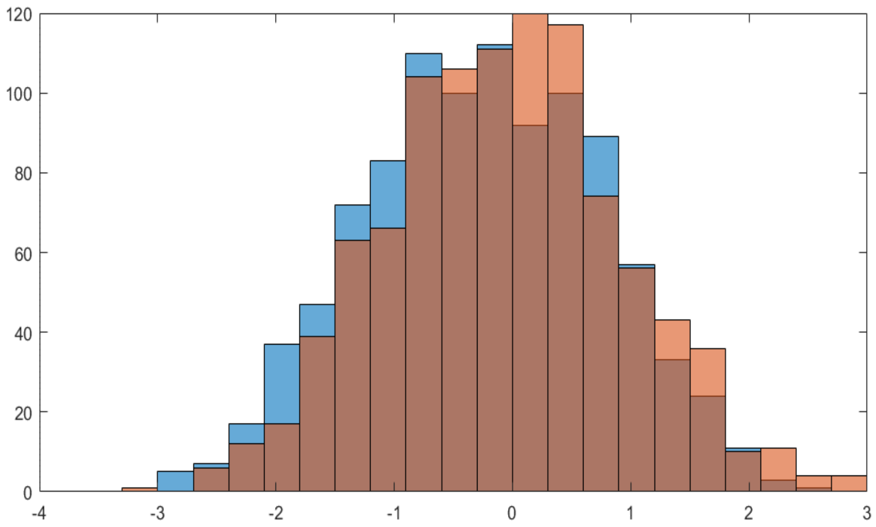
  - Analogous to white light: Equal intensity of light at all visible frequencies
- $\phi(\tau) = \sigma^2 \delta(\tau) \Rightarrow$ No covariance between samples $\Rightarrow C_{ij} = 0$ for $i \neq j$
- Simple example of white noise: iid noise
  - Independent and Identically distributed noise
  - iid noise realization: Draw independent trial values from the same pdf

White and colored noise





Here, the two noise processes have the same marginal pdf $X_i \sim N(0,1) \Rightarrow$ marginal pdf is not enough to describe noise

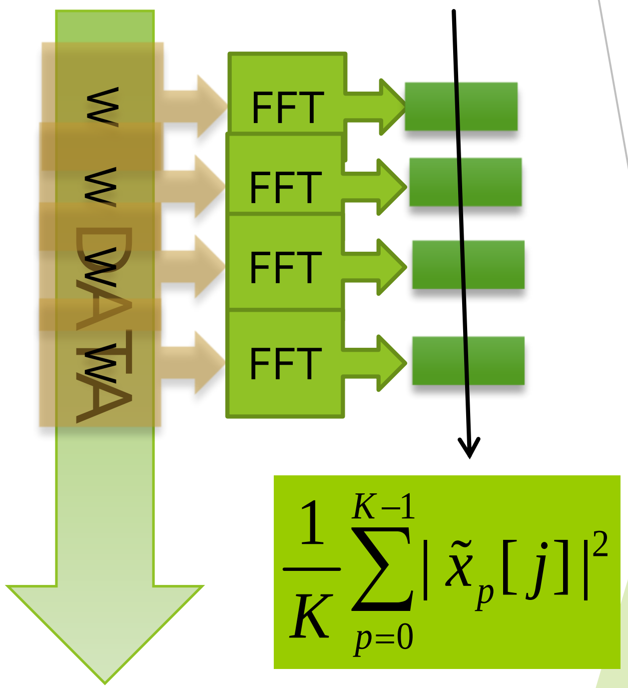


Welch’s method

- Welch’s method of overlapping windows
  - Already a built-in function in Matlab: `psd` (old) `pwelch` (new)
- Compute FFTs of K short overlapping windowed segments
- Calculate modulus squared of the FFT

  PSD estimate: $S_n[j] = \frac{1}{K}\sum_{p=0}^{K-1} |\tilde{x}_p[j]|^2$

- Matlab: `pwelch(x,nWin,[],[],fs)`
  - `x` : data vector
  - `nWin`: number of samples in each short segment
  - `fs` : sampling frequency of the data"



Simulating noise

White Gaussian Noise

- WGN is iid Normal noise
- Simply draw independent trial values from the same normal pdf

$$p_X(x) = N(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- Joint pdf of N samples of zero mean WGN:

$$p_{\bar{X}}(\bar{x}) = \prod_{i=1}^{N} p_X(x_i) = \frac{1}{\left(\sqrt{2\pi}\right)^N \sigma^N}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{N-1} x_i^2\right)$$

- Non-zero mean WGN realization can be generated by adding μ to every value of WGN realization
- Matlab: `randn(1,n)` generates n sample realization of zero mean WGN

Wiener-Khinchin theorem

Input noise process $x(t)$ with PSD $S_{in}(f)$

Linear time invariant system

$$y(t) = \int_{-\infty}^{\infty} x(t')\, h(t-t')dt'$$

$h(t)$: system impulse response

$T(f) = \int_{-\infty}^{\infty} dt\, h(t) e^{-2\pi i\, f t}$: System transfer function

Output noise process $y(t)$ with PSD $S_{out}(f)$

$$S_{out}(f) = S_{in}(f)|T(f)|^2$$

Generating stationary colored Gaussian noise

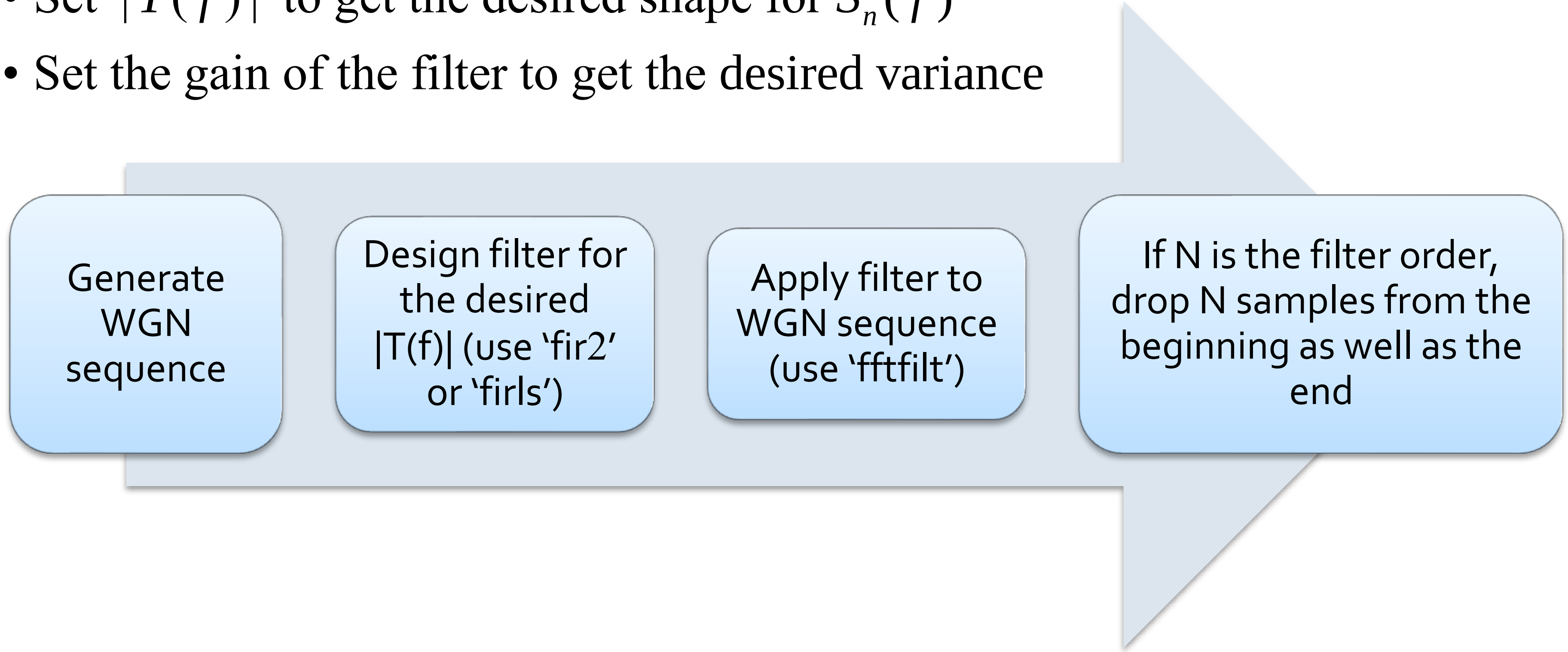
Pass WGN through a filter

$T(f)$: Transfer function of the filter

$S_n(f)$: PSD of the output noise

$S_n(f) = |T(f)|^2$

- Set $|T(f)|$ to get the desired shape for $S_n(f)$
- Set the gain of the filter to get the desired variance



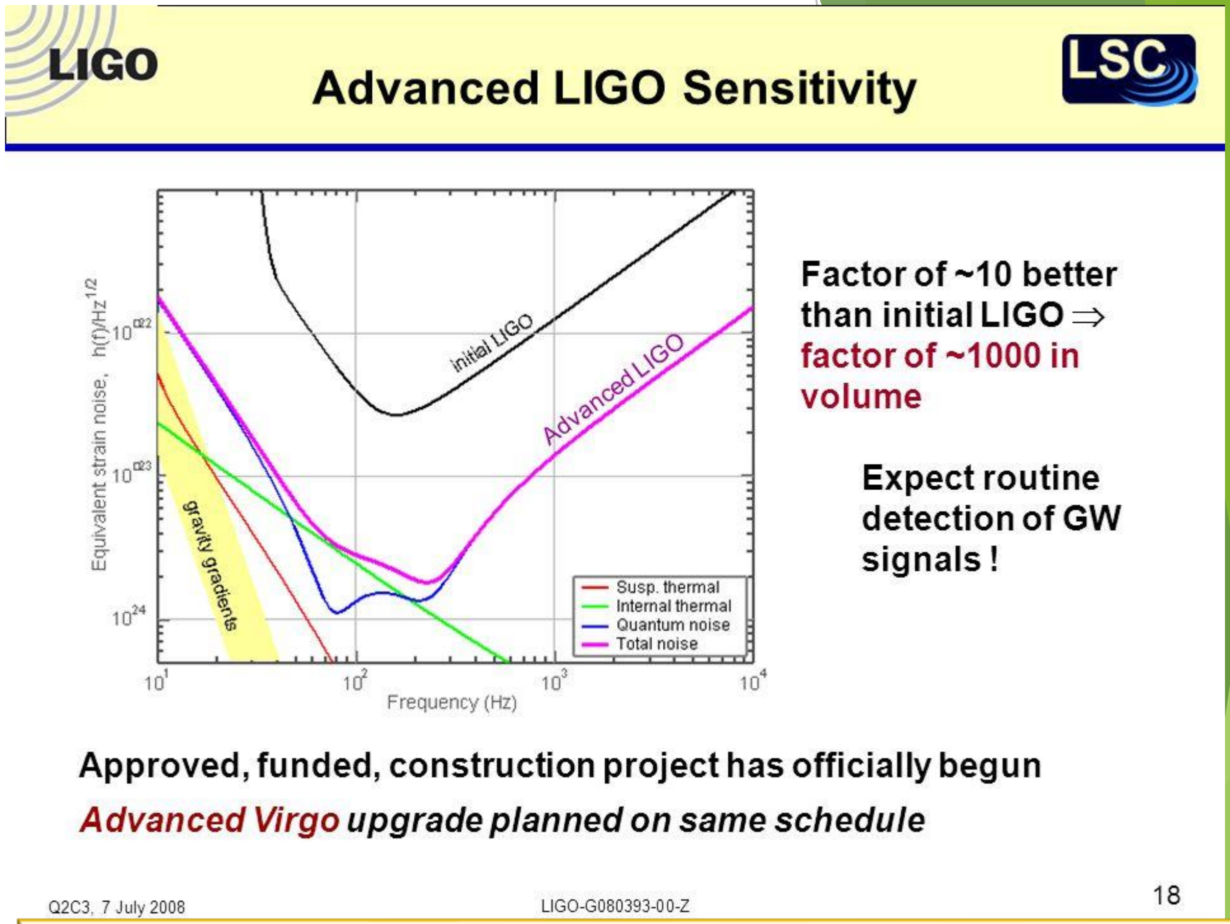
Simulating GW detector noise

Download design sensitivity curve file

Interpolate to uniformly space frequency values: $S_n(f)$

Matlab built-in filter design functions: match transfer function to $\sqrt{S_n(f)}$

Pass WGN through the filter



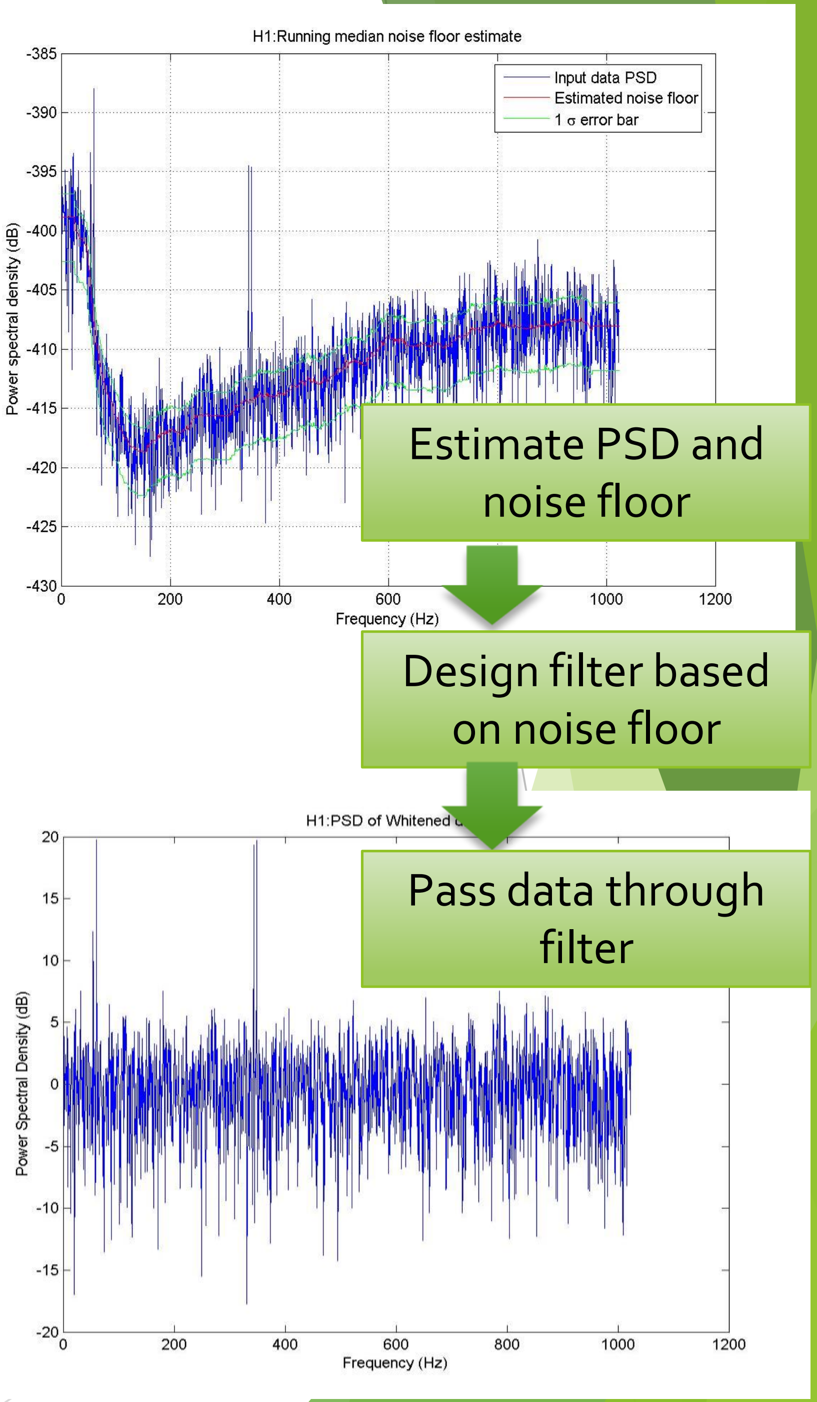
- Different conventions:
  $\sigma^2 = \int_{-\infty}^{\infty} S_n^{(two)}(f)df$, Or, $\sigma^2 = \int_0^{\infty} S_n^{(one)}(f)df$
- Two-sided or one-sided PSD: factor of 2 difference in definition
- **Typically, sensitivity curves are $\sqrt{S_n(f)}$ with one-sided convention**

Whitening

- Given data with colored noise, convert to data with white noise
- Solution: pass the given data through a filter whose transfer function is

$$T(f) = \frac{1}{\sqrt{S_n(f)}}$$

- For safety, the estimate is obtained from some separate **training data**



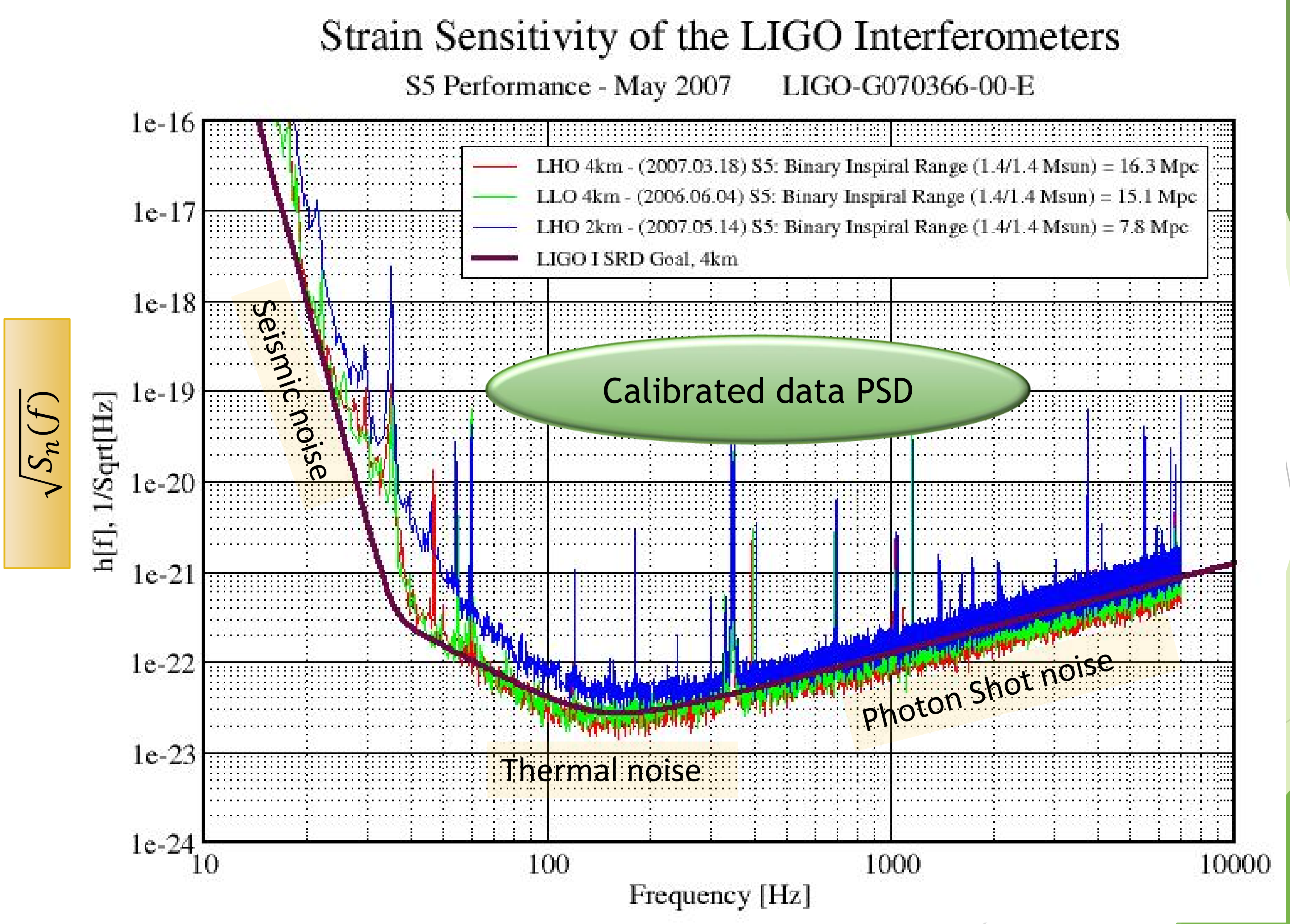
Noise in GW detectors

Noise models and real noise

The development of GW data analysis methods is often based on the Gaussian and stationary noise model with some **design** (smooth) PSD

This is especially true when the detector does not exist (e.g. LISA) but data analysis methods must be developed and tested

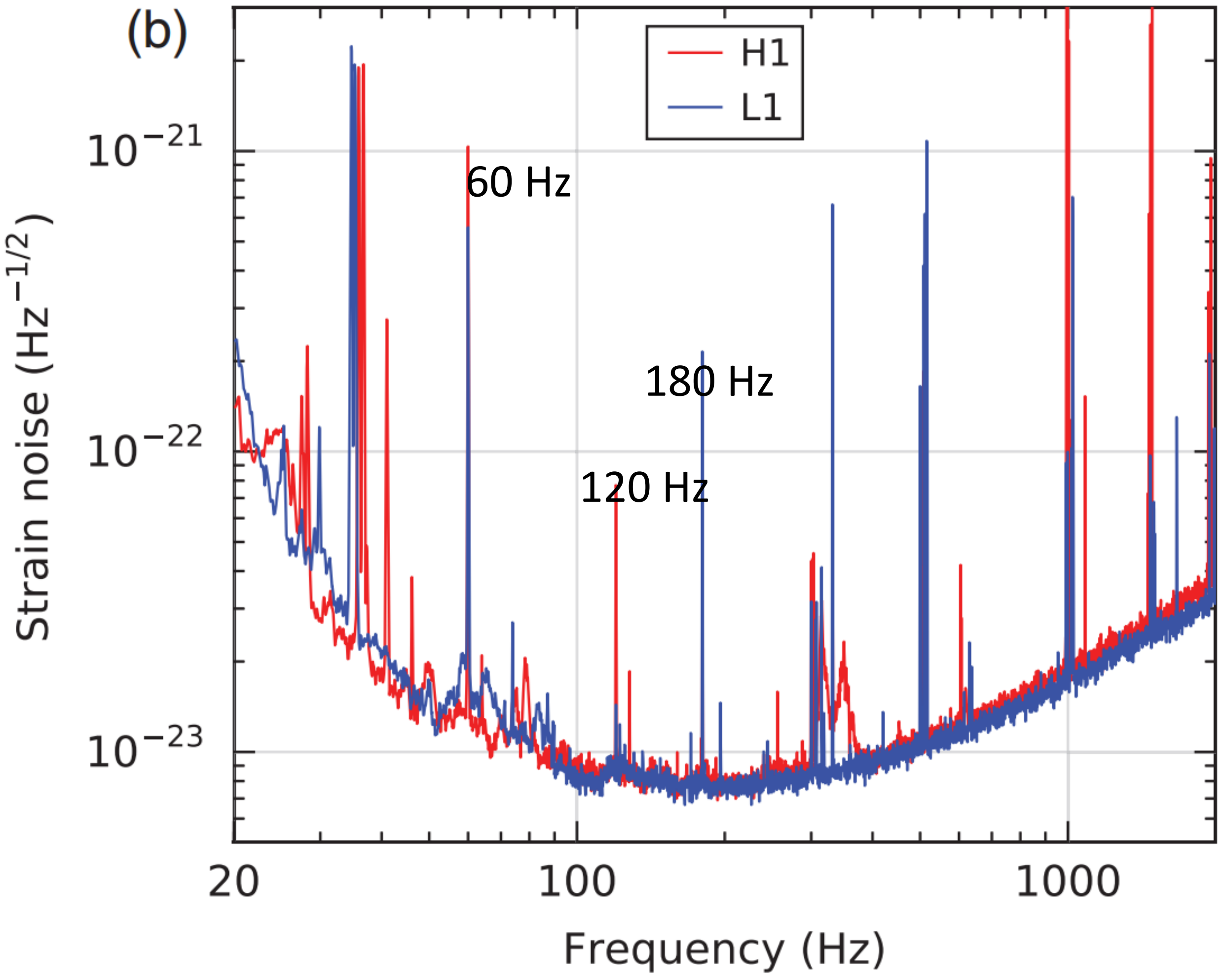
Real detector noise is often very far from this assumption

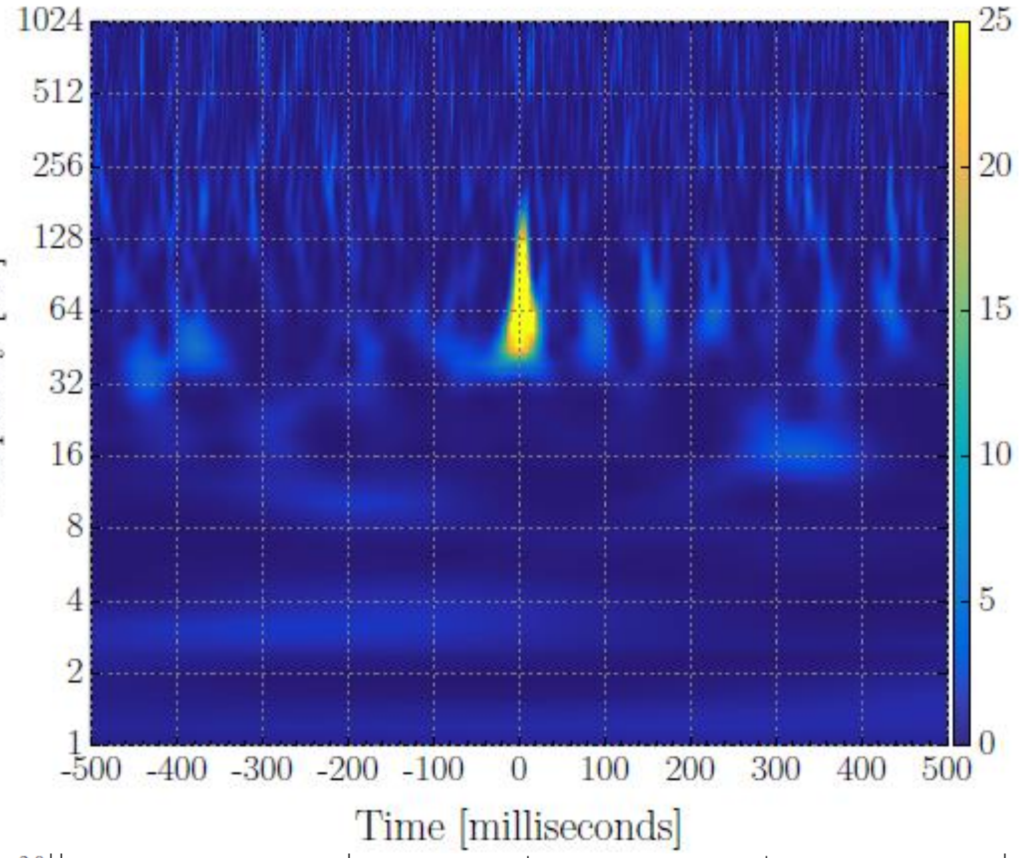


Strain Sensitivity of the LIGO Interferometers
S5 Performance - May 2007 LIGO-G070366-00-E
LHO 4km - (2007.03.18) S5: Binary Inspiral Range (1.4/1.4 Msun) = 16.3 Mpc
LLO 4km - (2006.06.04) S5: Binary Inspiral Range (1.4/1.4 Msun) = 15.1 Mpc
LHO 2km - (2007.05.14) S5: Binary Inspiral Range (1.4/1.4 Msun) = 7.8 Mpc
LIGO I SRD Goal, 4km
h[f], 1/Sqrt[Hz]
$\sqrt{S_n(f)}$
Frequency [Hz]
Calibrated data PSD
Seismic noise
Thermal noise
Photon Shot noise

LIGO

- Data conditioning of real data required before performing any kind of GW search
- Typical data conditioning tasks:
  - Whitening the noise floor
  - Removal of line noise (e.g., using notch filters)





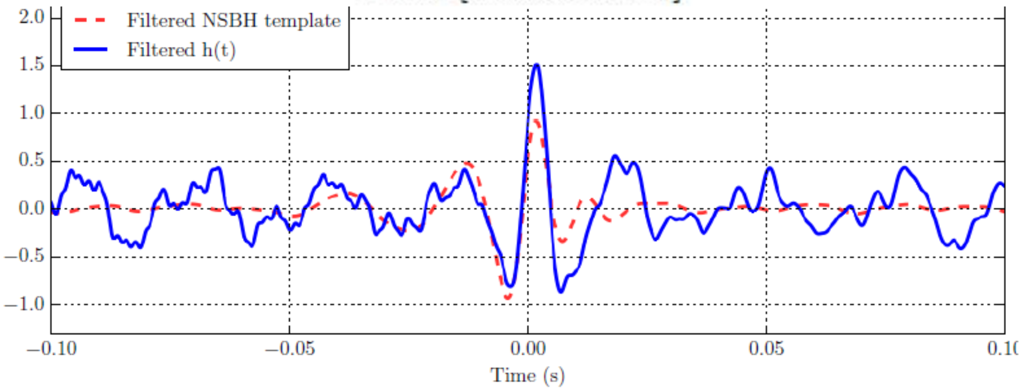
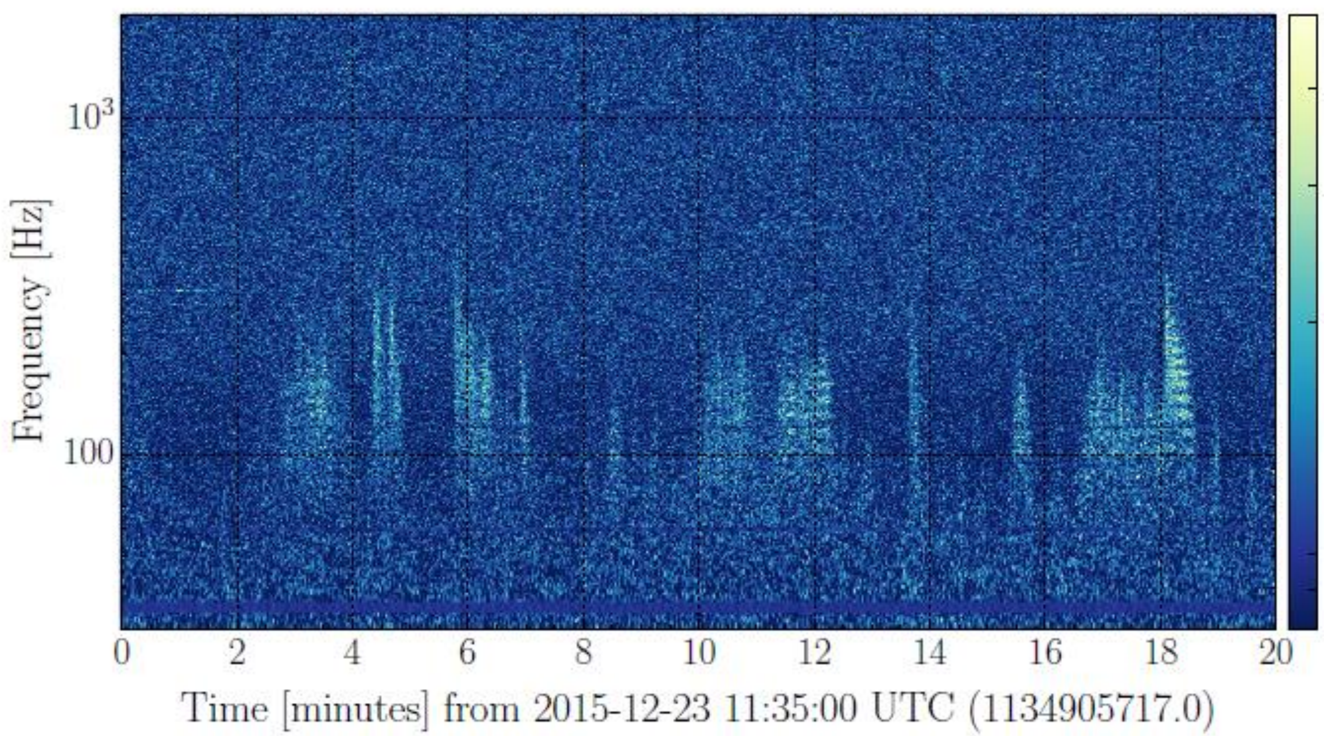


Figure 13: A filtered time-domain representation of the Livingston strain channel $h(t)$ at the time of a blip transient. Overlaid on the strain plot is a filtered CBC waveform that reported a high re-weighted SNR value at the time of



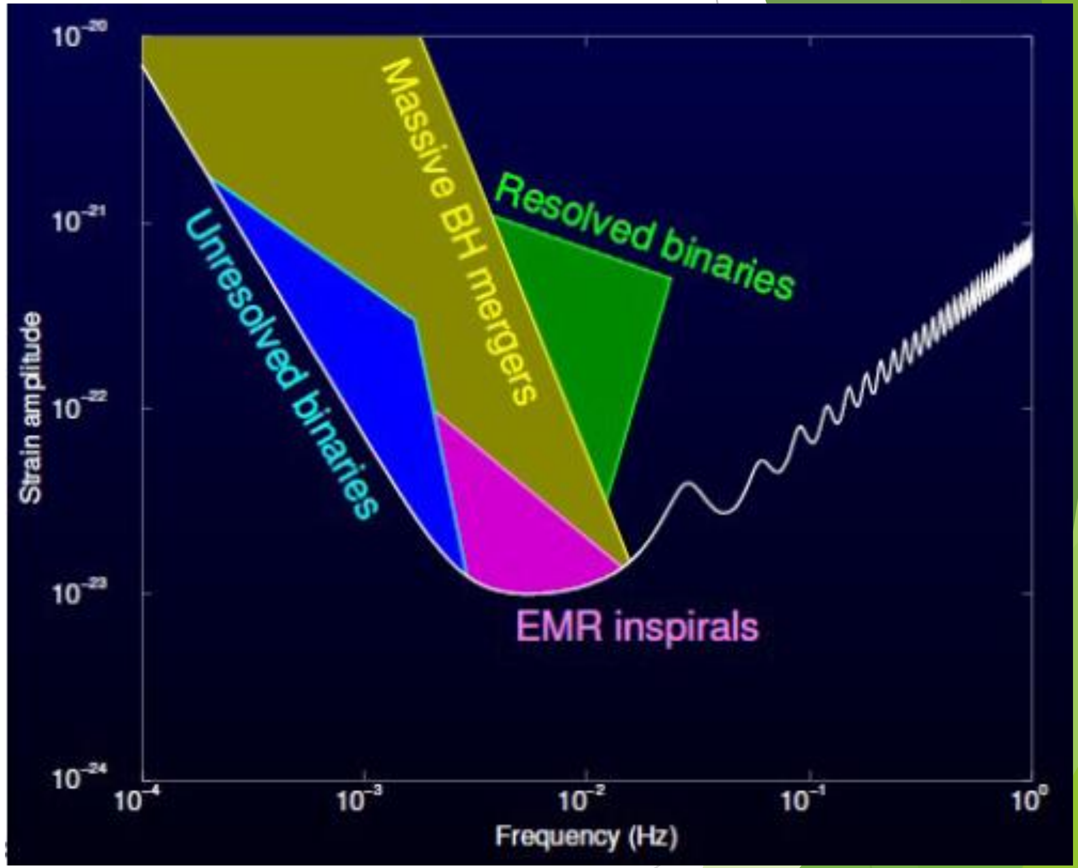
ure 16: A time-frequency spectrogram of the 60-200 Hz noise. This noise in storms that often last for many minutes. This time scale and fr

Glitches and non-stationarity

arXiv:1710.02185v3 [gr-qc]

LISA

- LISA does not exist yet!
- All LISA data analysis work assumes Gaussian stationary noise
- Refer to Mock LISA Data Challenge papers/documents to obtain commonly used PSDs



LISA noise sources

- LISA Path Finder mission provided important information about what real LISA noise may look like
- LISA will be signal-dominated: It will be harder to measure and correct problems with real noise

Noise	Sources
OMS sensing	shot noise, readout electronics, phase meter distortions, frequency noise, etc
electrostatic actuation	suspension loops on certain degrees of freedom, voltage stability, cross-talk coefficients
gas damping	residual gas around TM
magnetic	interplanetary, sc fields, gradient fluctuations
laser radiation pressure	momentum transfer to TM
thermal	radiometer, thermal radiation, outgassing

and so on…

https://www.aei.mpg.de/~hewitson/presentations/presentations_2015/files/lpf_da_elisa.pdf

Pulsar Timing Array noise

- PTA data is not sampled uniformly in time
- Simulated noise in timing residuals is typically assumed to be WGN
    - WGN generation does not care for uniform or non-uniform sampling
- However, real timing residuals from some pulsars exhibit White Noise + Red Noise behavior
    - Red noise:

$$S_n(f) \propto f^{-\gamma} \text{ for some index } \gamma > 0$$

Class. Quantum Grav. **30** (2013) 224007 G

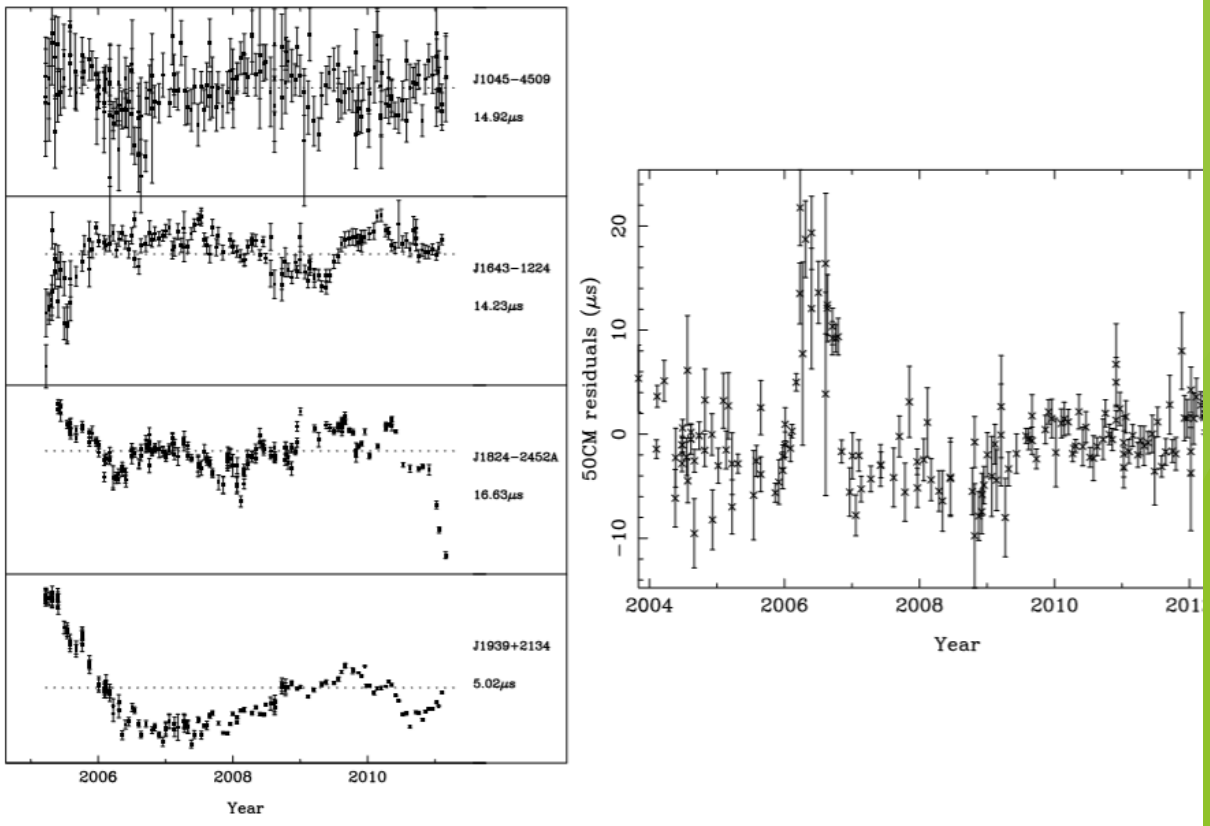


Figure 4. Left panel: a sample of 20 cm data sets that have been corrected for dispersion m variations, but are still significantly affected by low-frequency noise. Right panel: the residuals in the 50 cm band for PSR J1603−7202.