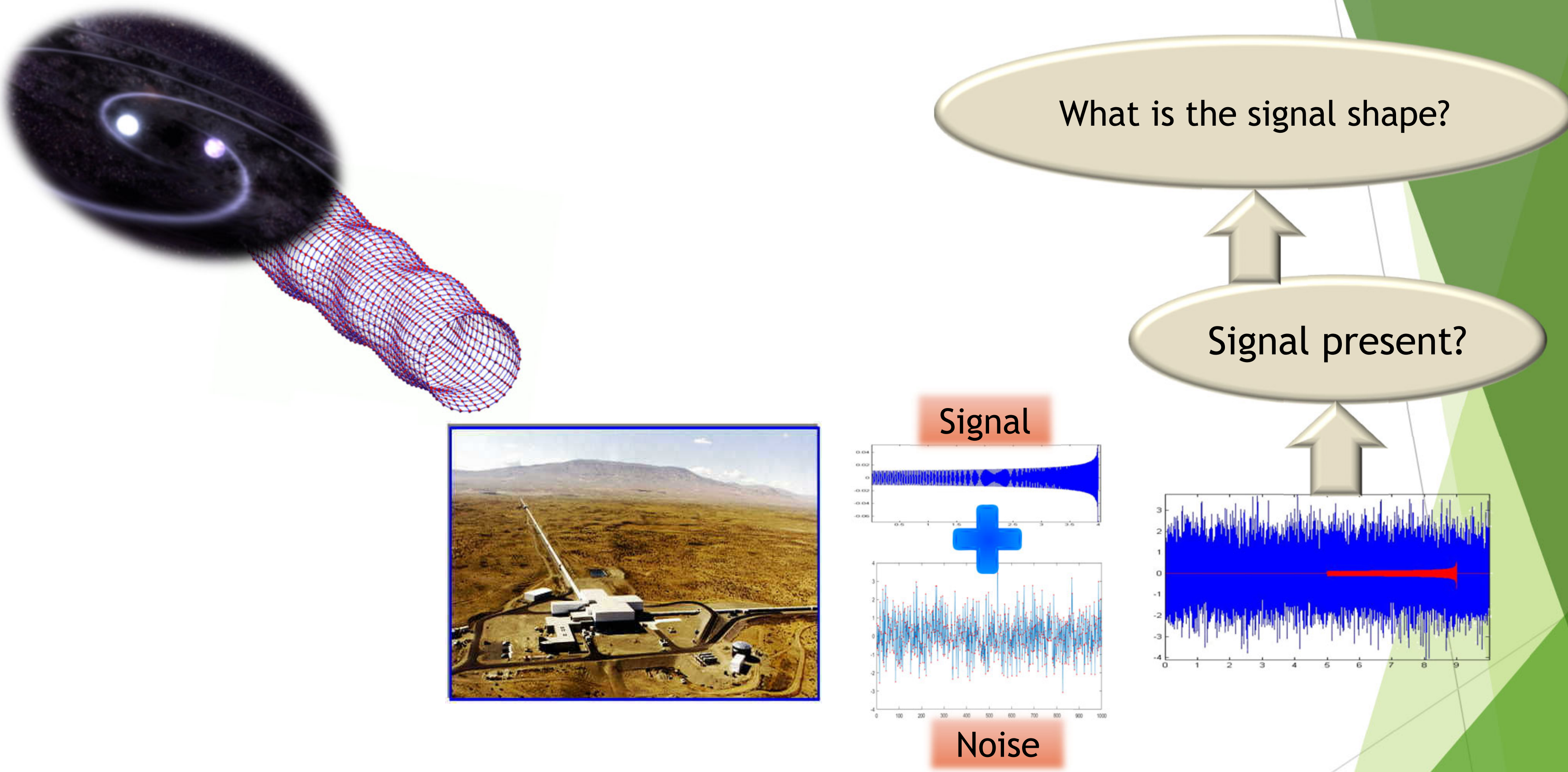
# DETECTION AND ESTIMATION OF DETERMINISTIC GW SIGNALS

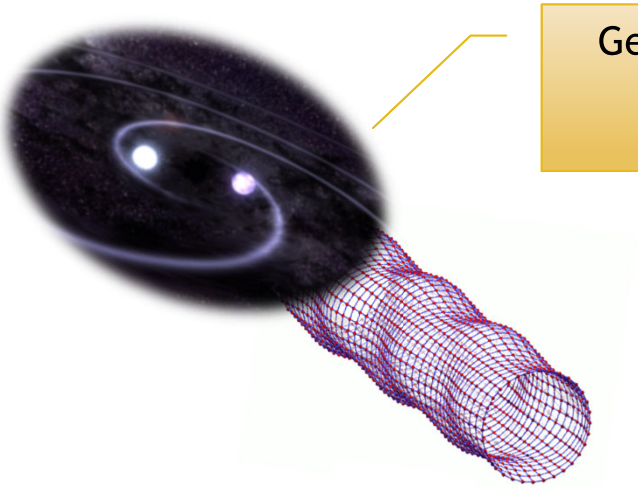
Soumya D. Mohanty
Dept. of Physics and Astronomy
The University of Texas Rio Grande Valley

UTRGV

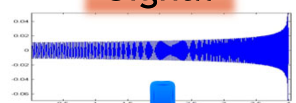


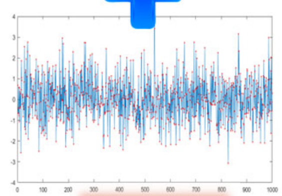

# Gravitational Wave (GW) Astronomy

General theory of relativity + Astrophysics

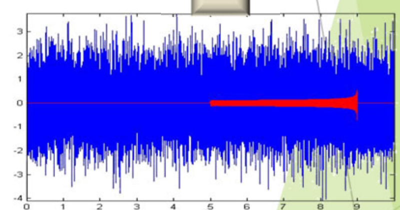


Signal



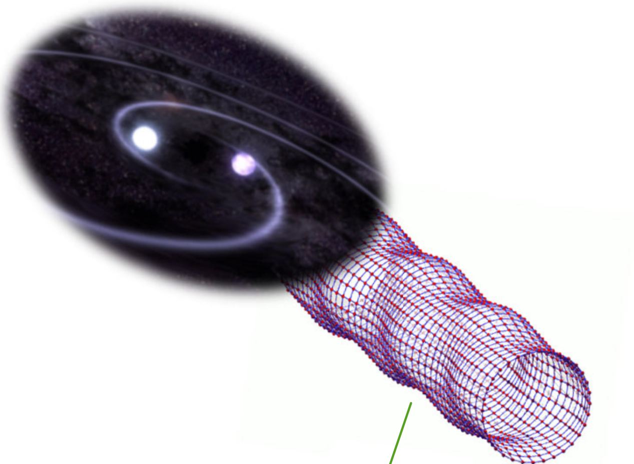


Noise



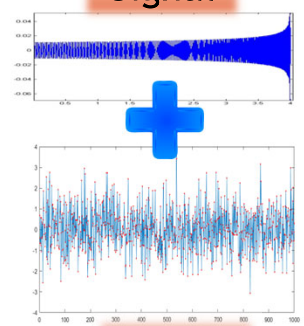
Signal present?

What is the signal shape?

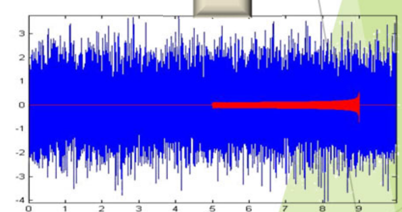


GW as a
Plane tensor wave

Signal

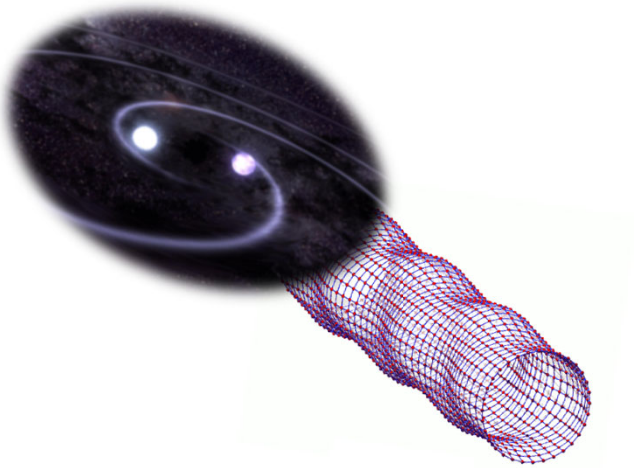


Noise



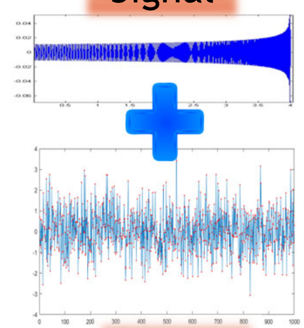
Signal present?

What is the signal shape?

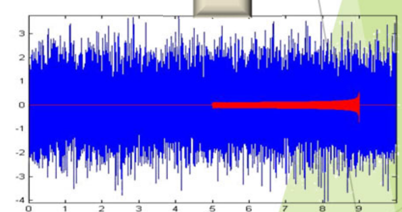


Instrumentation details

Signal

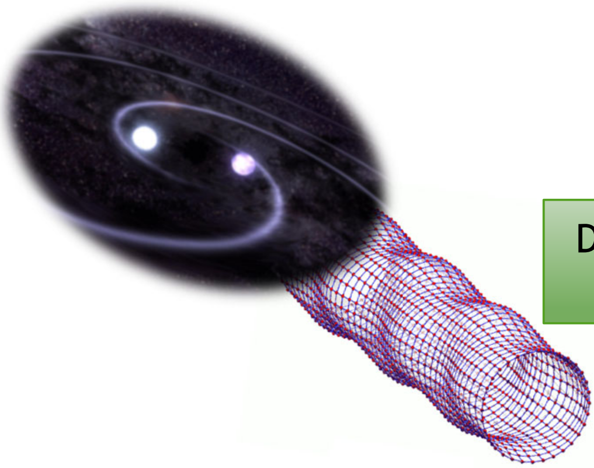


Noise



Signal present?

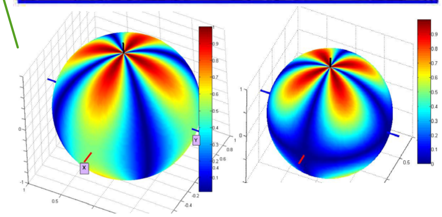
What is the signal shape?



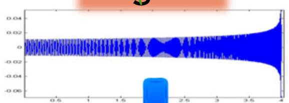
Detector response:
GW strain

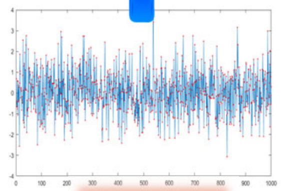
Instrumentation details

- Antenna patterns
- Calibration

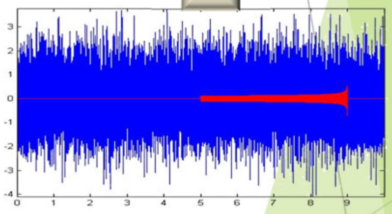


Signal



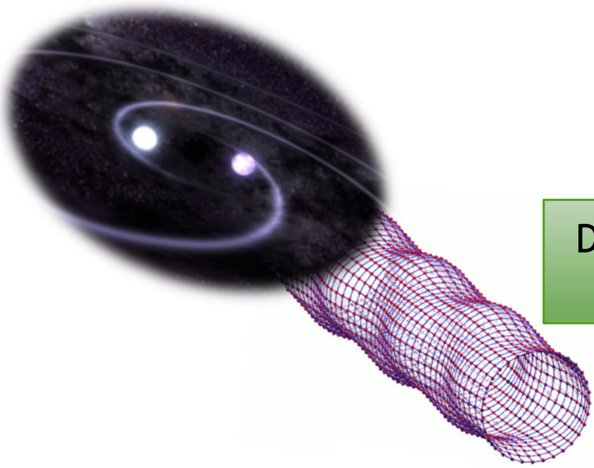


Noise



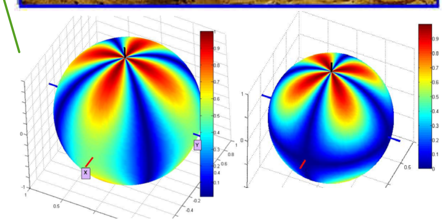
Signal present?

What is the signal shape?

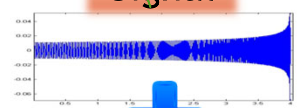


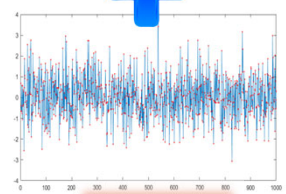
Detector response:
GW strain

Instrumentation details
- Antenna patterns
- Calibration



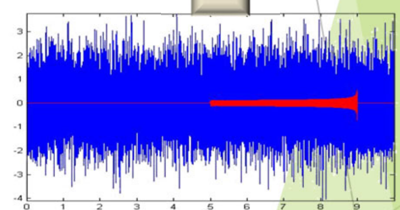
Signal





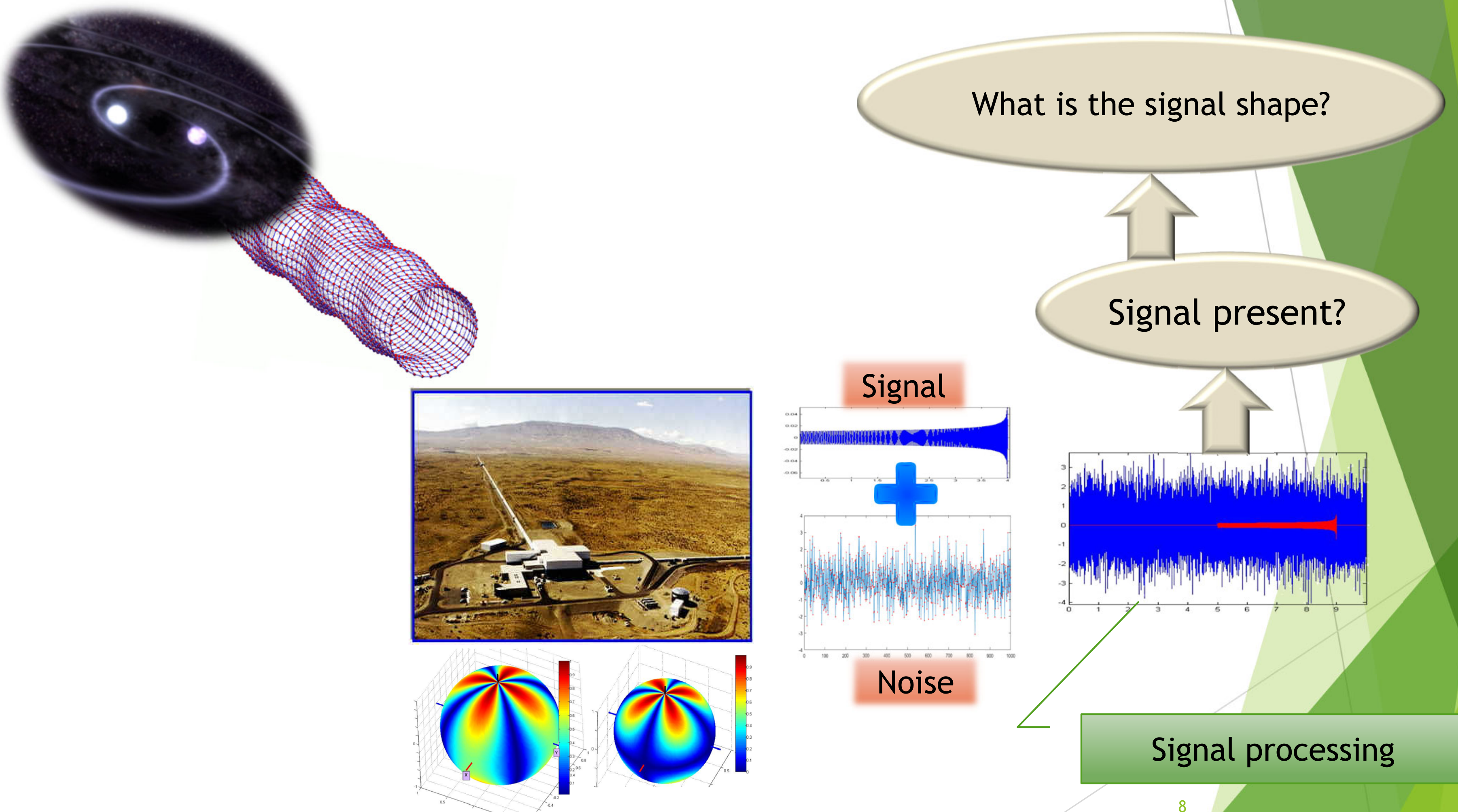
Noise

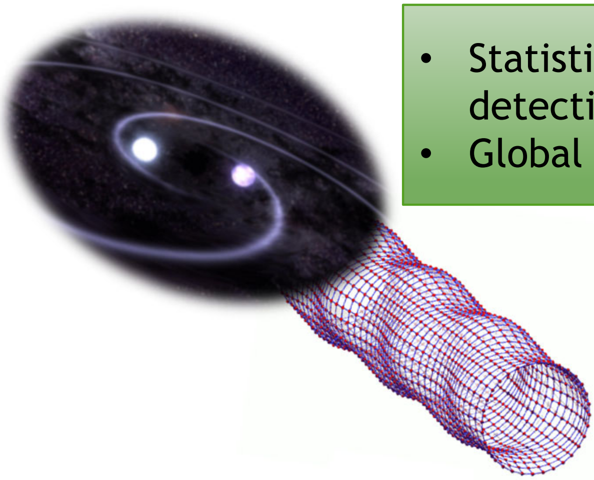
Noise model



Signal present?

What is the signal shape?

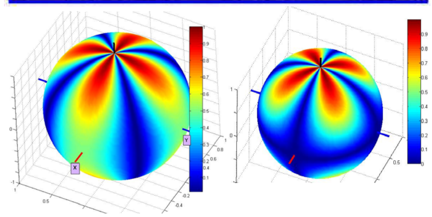

What is the signal shape?
Signal present?
Signal
Noise
Signal processing

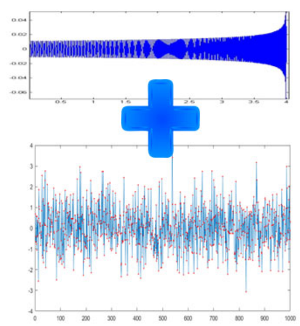


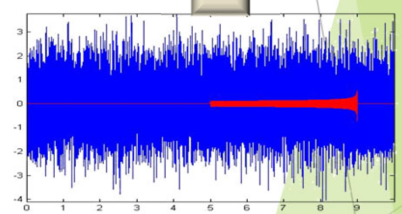
- Statistical theory of signal detection and estimation
- Global optimization

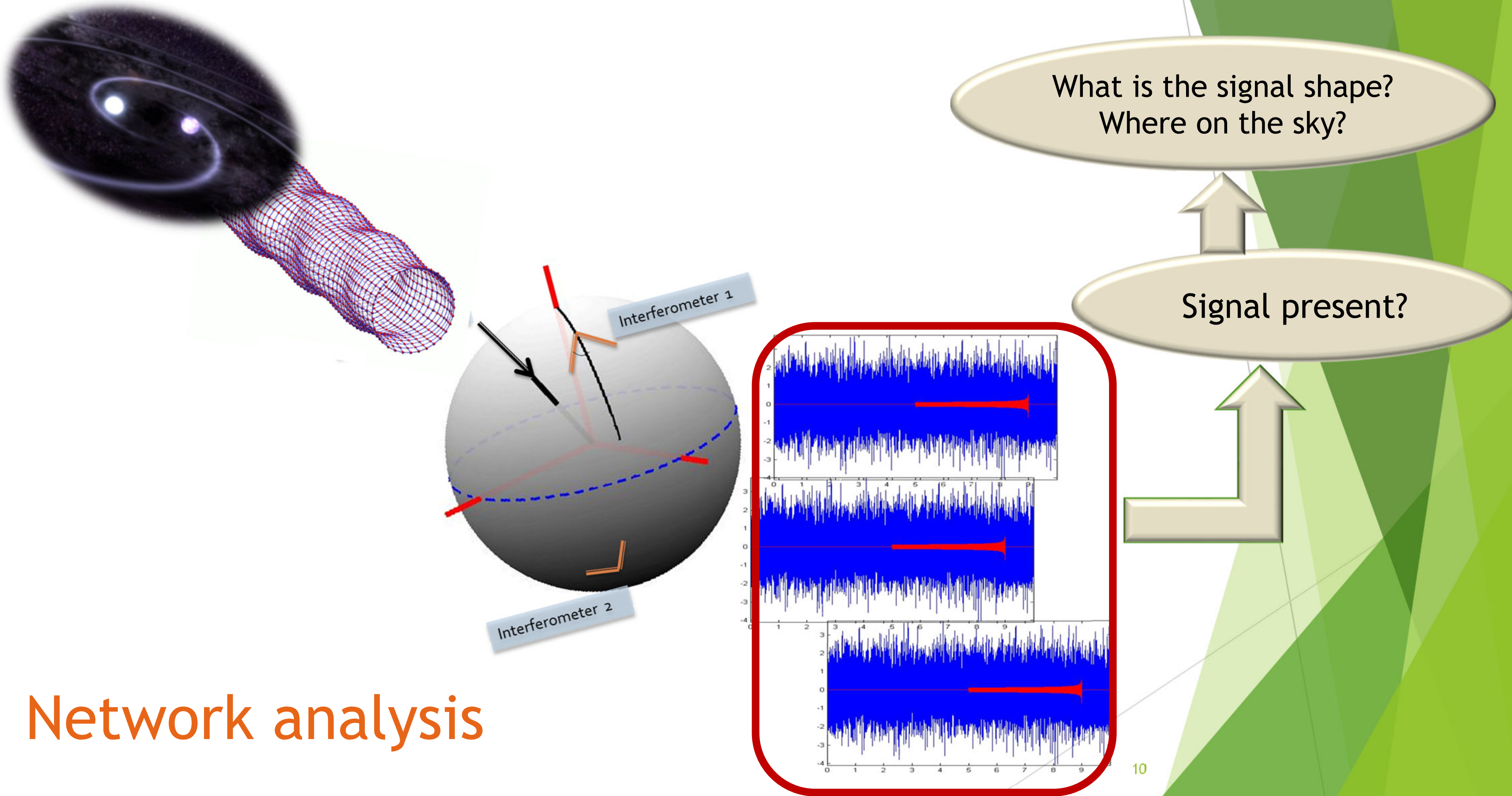
What is the signal shape?

Signal present?

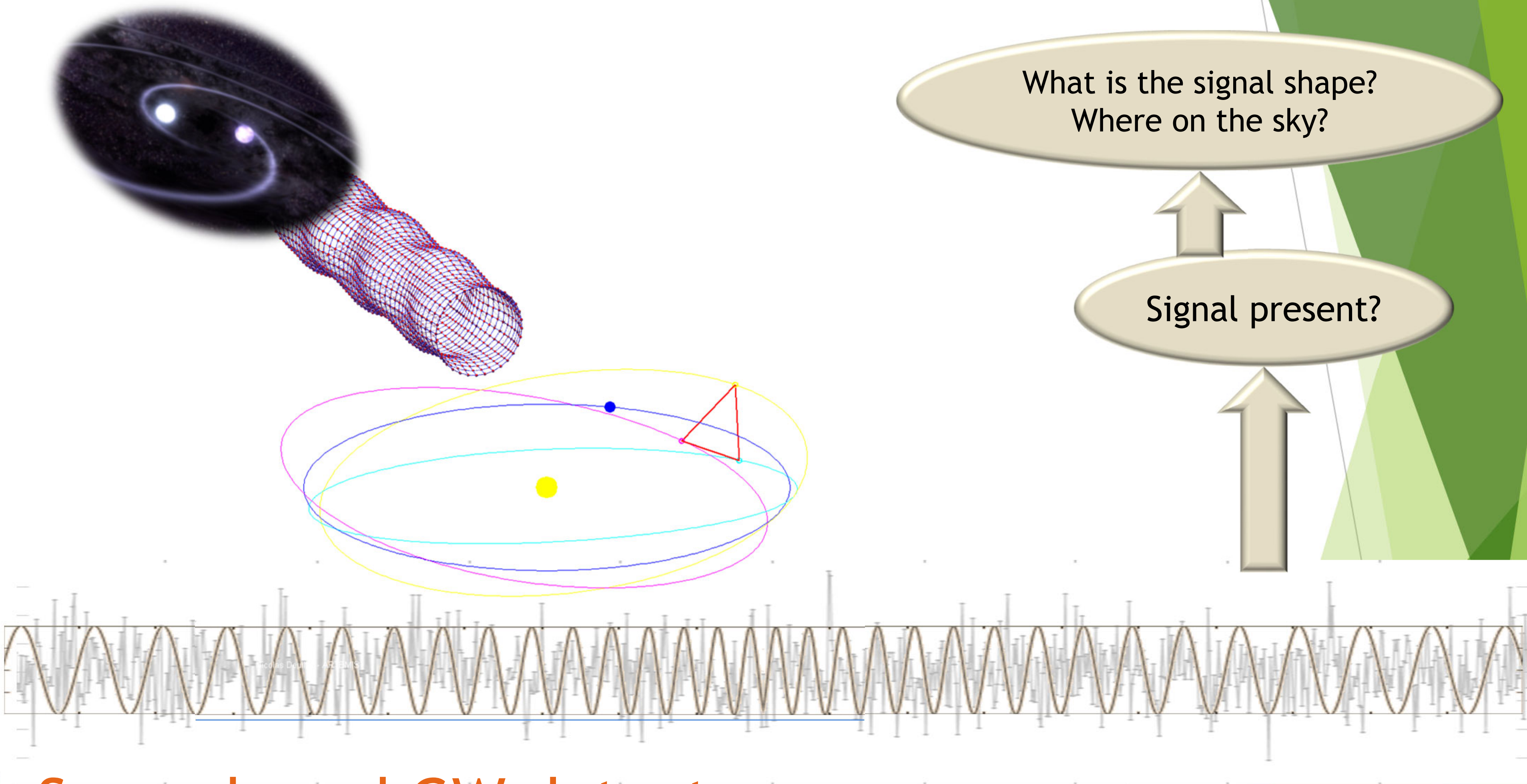









# Network analysis




# Space-based GW detector

# PLAN

- Signal processing basics: Fourier transform, Filtering
- Signal and noise
  - Antenna pattern functions
  - Stationary noise, Power Spectral Density, Gaussian noise
- Statistical theory (deterministic signals): Estimation and Detection
- Global optimization: Particle Swarm Optimization
- Real data issues

# RESOURCES

- Github: mohanty-sd / DATASCIENCE_COURSE
- Github: mohanty-sd / SDMBIGDAT19
- Github: mohanty-sd / GWSC
- Books:
  - S. D. Mohanty, "Swarm intelligence methods for statistical regression," CRC Press
  - Prandoni and Vetterli, "Signal Processing for communications,"Type equation here. Chapters 1 through 5.

# Basic signal processing

Gravitational Wave Data Analysis School in China

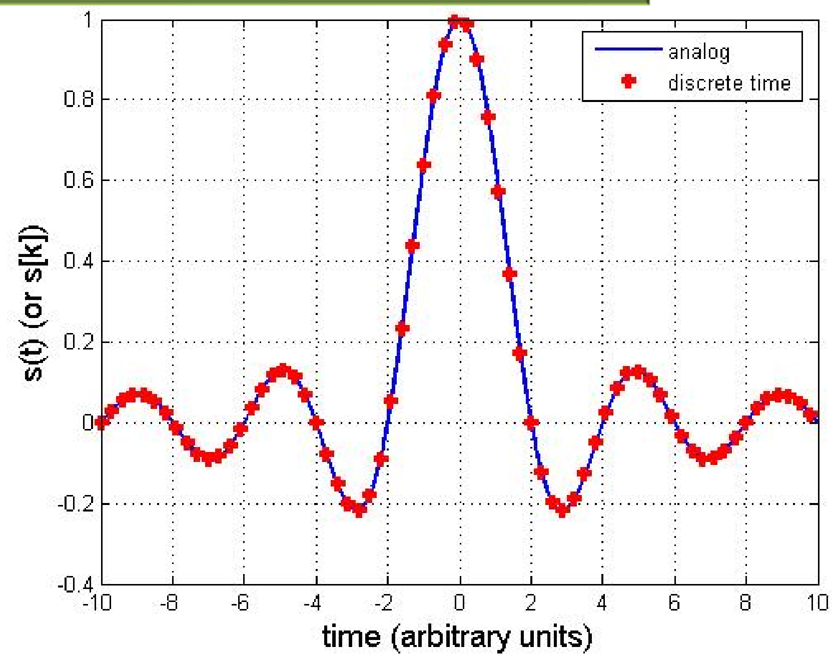
Soumya D. Mohanty

UTRGV

# Analog vs. Digital

- **Analog** signal: a continuous function of "time" $s(t)$
  - Example: The sound signal coming from your instructor to your ear
  - "time": A convenient name for the X-axis
- **Discrete time** signal: sequence of values ("samples") of an analog signal

$$s[k] = s(t_k), \text{k=...}, -1, -2, 0, 1, \ldots$$



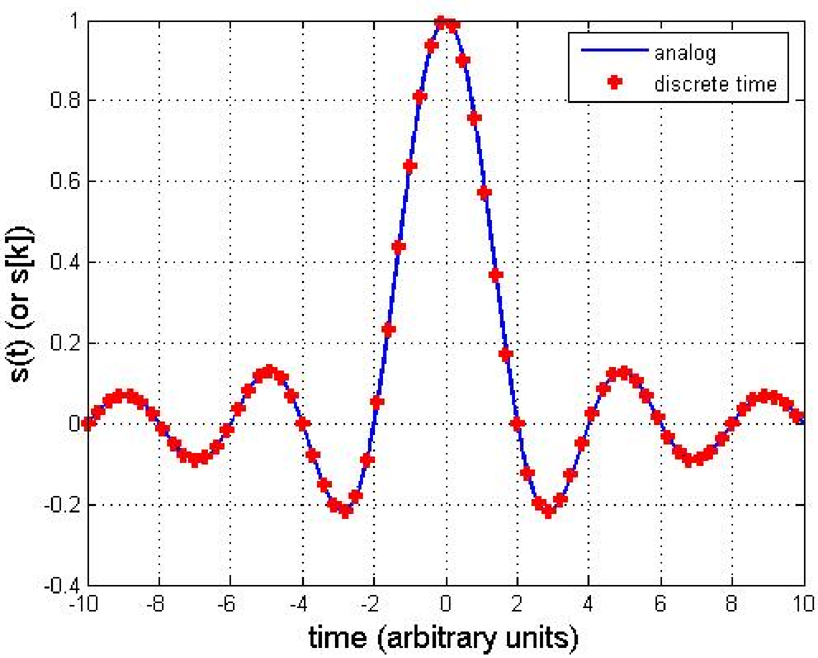
# Analog vs. Digital

- **Analog** signal: a continuous function of time $s(t)$
  - Example: The sound signal coming from your instructor to your ear
- **Discrete time** signal: sequence of values ("samples") of an analog signal

$$s[k] = s(t_k), k = \ldots, -1, -2, 0, 1, \ldots$$

- **Digital** signal: $s[k]$ **quantized** due to representation by a finite number of bits in the binary system

$$5 \rightarrow 101 = 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$$
$$6 \rightarrow 110 = 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0$$
$$7 \rightarrow 111 \text{ but } 8 \rightarrow ?,\ 6.3 \rightarrow ?$$



# Discrete time signal

A discrete time signal: Digital signal where quantization effects are negligible

- GW data is a discrete time signal (**time series**) [*Data* and *signal* used interchangeably]
- Finite length signal: $\boldsymbol{x}$ or $\bar{x} = (x[0], x[1], \cdots, x[N-1])$ or $\bar{x} = (x_0, x_1, \ldots, x_{N-1})$
- $x[n] = x(t_n)$

1D signal and samples spaced equally in time: **Uniformly sampled time series**

- $t_n = n\Delta$, where $\Delta$ is called the sampling interval (or sampling period)

The number of samples per second: **Sampling frequency** (Hz)$= \frac{1}{\Delta}$ (where $\Delta$ is in seconds)

- LIGO data sampling frequency is typically $16384$ Hz ($= 2^{14}$ samples/sec)
- LISA sampling frequency will be $\approx 2$ Hz
- PTA data sampling frequency is on average $\approx 7 \times 10^{-7}$ Hz

# Fourier transform

Continuous and Discrete

# (Analog) Fourier transform

- Converts a "time" domain signal to Fourier-domain (or frequency domain) signal
- For an analog signal

$$\tilde{s}(f) = F[s(t)] = \int_{-\infty}^{\infty} s(t) e^{-2\pi i f t}\, dt$$

- Fourier transform is **invertible** (Inverse Fourier transform)

$$s(t) = F^{-1}[\tilde{s}(f)] = \int_{-\infty}^{\infty} \tilde{s}(f) e^{2\pi i f t} df$$

- Existence condition: $\int_{-\infty}^{\infty} |s(t)|$ must be finite (Note that this excludes a pure sinusoidal signal!)

# Discrete Fourier Transform (DFT)

- Transformation of finite length discrete time signal (i.e., sequence) that is a discretized version of the Fourier transform
- Consider a sequence $(x_0, x_1, \ldots, x_{N-1})$ with $N$ samples

**Discrete Fourier Transform (DFT)**

$$\tilde{x}_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i k n/N}$$

- $k = 0,1, \ldots, N-1$
- DFT is a sequence: $(\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_{N-1})$
- Resembles:
- $\tilde{x}(f) = F[x(t)] = \int_{-\infty}^{\infty} x(t) e^{-2\pi i f t}\, dt$

**Inverse DFT**

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{x}_k\, e^{2\pi i n k/N}$$

- $n = 0,1, \ldots, N-1$
- Resembles:
- $x(t) = F^{-1}[\tilde{x}(f)] = \int_{-\infty}^{\infty} \tilde{x}(f) e^{2\pi i f t} df$

Homework: DFT is an approximation to the Fourier Transform (more precisely the Fourier series)

# Discrete Fourier Transform: Matrix form

$$\tilde{x}_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i k n/N}$$

Matrix form:

$$\begin{pmatrix} \tilde{x}_0 \\ \tilde{x}_1 \\ \vdots \\ \tilde{x}_{N-1} \end{pmatrix} = F \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{pmatrix} \rightarrow \tilde{x}^T = F\, \bar{x}^T$$

$$\text{where } F = \begin{pmatrix} e^{2\pi i 0\times 0/N} = 1 & e^{2\pi i 0\times 1/N} = 1 & \dots & e^{2\pi i 0\times (N-1)/N} = 1 \\ e^{2\pi i 1\times 0/N} = 1 & e^{2\pi i 1\times 1/N} & \dots & e^{2\pi i 1\times (N-1)/N} \\ \vdots & \vdots & \vdots & \vdots \\ e^{2\pi i (N-1)\times 0/N} = 1 & e^{2\pi i\, (N-1)\times 1/N} & \dots & e^{2\pi i (N-1)\times (N-1)/N} \end{pmatrix}$$

- Each row of $F$ is a complex sinusoid $\Rightarrow$ $\tilde{x}_k$ is the dot-product of $\bar{x}$ with sinusoids of frequency $f_k = k \times \frac{1}{N\Delta} = \frac{k}{T}$
- $\Rightarrow$ Sinusoidal signals appear as peaks in a DFT

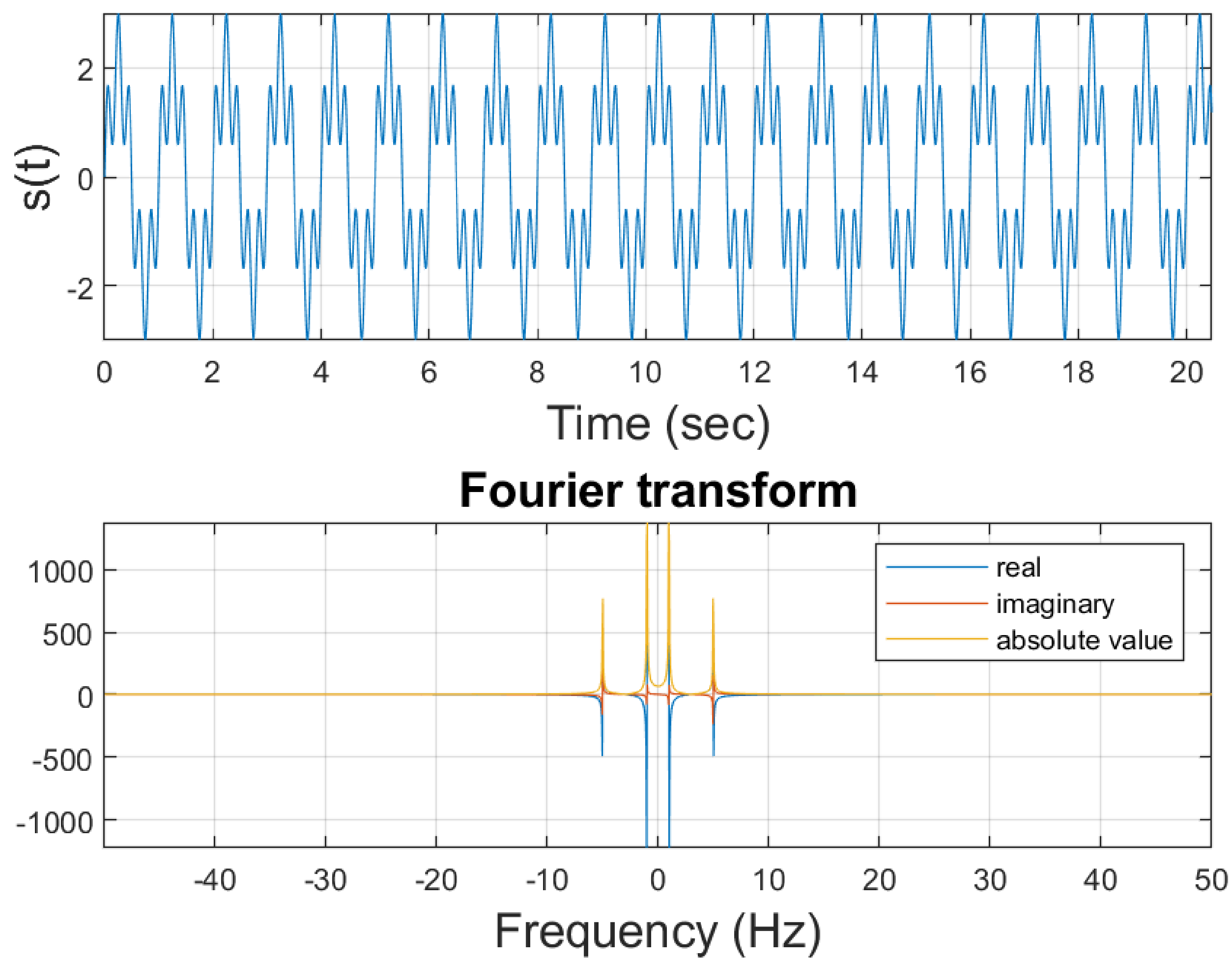
# Fast Fourier Transform

$$\tilde{x}^T = F\,\bar{x}^T$$

- Computing the DFT involves taking the product of an $N \times N$ matrix, $F$, with an $N$ element column vector, $\bar{x}^T$
- $\Rightarrow 2N^2$ multiplications and additions (i.e., floating point operations)
- Fast Fourier Transform (FFT): a clever algorithm for carrying out the same matrix product with $O(N \log_2 N)$ floating point operations
- Example: For $N = 1024$, the FFT is $\approx 200$ times faster!

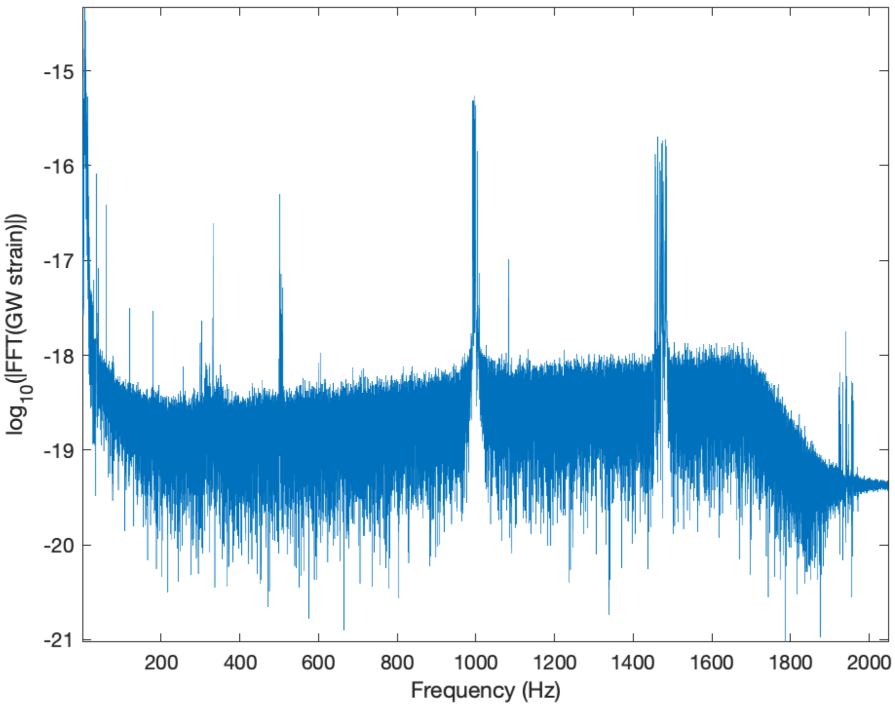
# Example: FFT

- See the script: DATASCIENCE_COURSE / DSP / FTExample.m
- The signal consists of the sum of two sinusoids
- The magnitude of the FFT shows two peaks
- Fourier transform is useful for resolving a sum of sinusoids



# GW data

- Use the script DATASCIENCE_COURSE / GWDATA / readgwoscdat.mlx to plot the FFT magnitude of GW data.



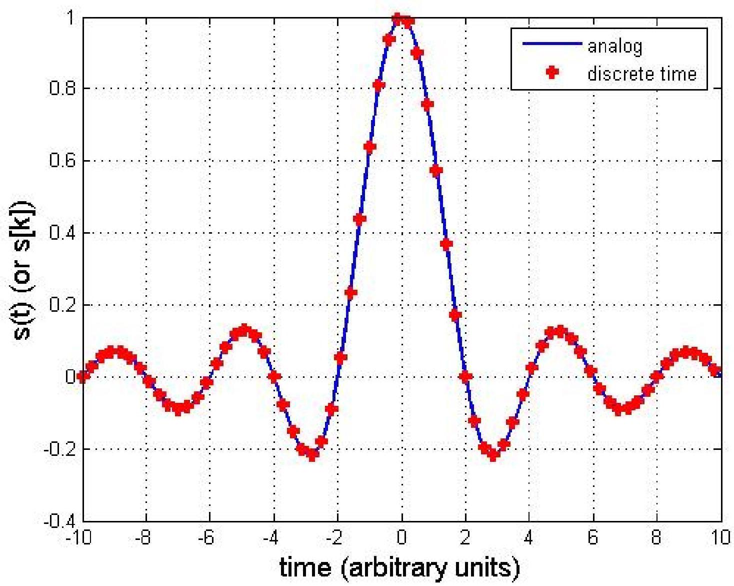
# Nyquist sampling theorem

# Nyquist sampling theorem: Analog to Digital

- Nyquist theorem: An analog signal $s(t)$ can be sampled without any loss of information if
  - The analog signal is band-limited:

$$\tilde{s}(f) = 0 \text{ for } f \notin [-f_B, f_B]$$

  - The sampling frequency $f_s \geq 2f_B$
  - $f_B$ is called the bandwidth
- The critical sampling frequency $f_s = 2f_B$ is called the Nyquist frequency (or Nyquist rate)



# Nyquist sampling theorem: Analog to Digital

- Intuitive explanation of Nyquist theorem:
  - Review Fourier transform
  - [Homework] Any signal, $s(t)$, can be expressed as a sum ("superposition") of sinusoidal signals, $A\sin(2\pi f t + \phi_0)$, with different frequencies, amplitudes, and initial phases
- To reconstruct a sinusoid from its samples, it is enough to reconstruct it in just one period $t \in [a, a + \frac{1}{f}]$
- Two unknowns (Amplitude, initial phase) for a given $f \Rightarrow$ minimum 2 samples needed per period
- Maximum frequency $f_B$ (bandwidth) $\Rightarrow$ 2 samples in $1/f_B \Rightarrow$ Period of sampling $= \frac{1}{2} \times \frac{1}{f_B} \Rightarrow$ Frequency of sampling (=1/sampling period): $f_s = 2f_B$

# Digital filtering

# Digital filtering

- Digital filter: Running weighted sum of a sequence

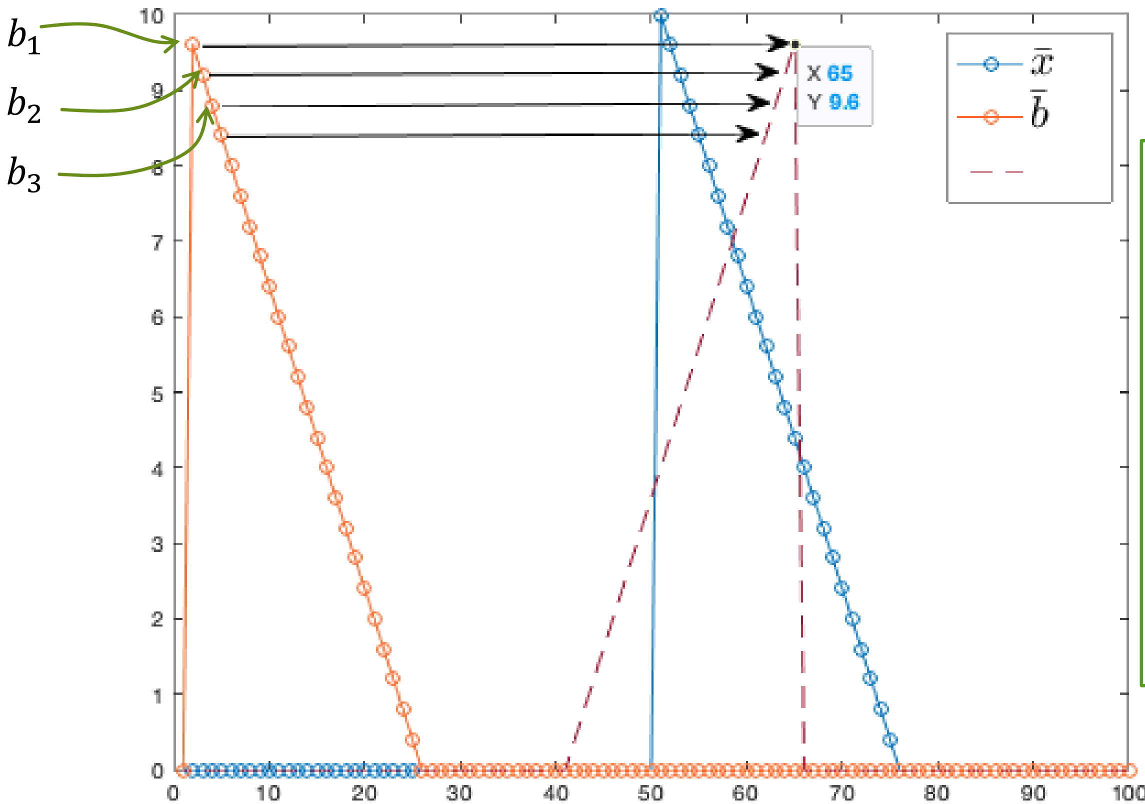
Output sequence $(y_0, y_1, \dots)$

$$y_k = b_1 x_k + b_2 x_{k-1} + \cdots + b_m x_{k-m}$$

Input sequence $(x_0, x_1, \dots)$

- $(b_1, b_2, \dots, b_m)$: Filter coefficients

# Discrete time convolution




- Sequences: $\bar{b}$ and $\bar{x}$

$$y_k = b_1 x_k + b_2 x_{k-1} + \cdots + b_m x_{k-m}$$

- For example

$$\Rightarrow y_{65} = b_1 x_{65} + b_2 x_{64} + \cdots$$

Reflection → Multiplication → Summation

- Filtering is discrete-time convolution operation

Discrete time convolution

- Sequences: $\bar{b}$ and $\bar{x}$
- Definition:

$$y_k = b_1 x_k + b_2 x_{k-1} + \cdots + b_m x_{k-m}$$

- Note how the output sequence is smoother than the input sequence

# Digital filtering

- **Digital filter:** Running weighted sum of a sequence

$$y_k = b_1 x_k + b_2 x_{k-1} + \cdots + b_m x_{k-m}$$

- $(b_1, b_2, \dots, b_m)$: Filter coefficients

- **Generalization (Matlab convention)**

$x_k$: input sequence

$y_k$: output sequence

$$a_1 y_k + a_2 y_{k-1} + \cdots + a_{p+1} y_{k-p} = b_1 x_k + b_2 x_{k-1} + \cdots + b_m x_{k-m}$$

# Filter transfer function

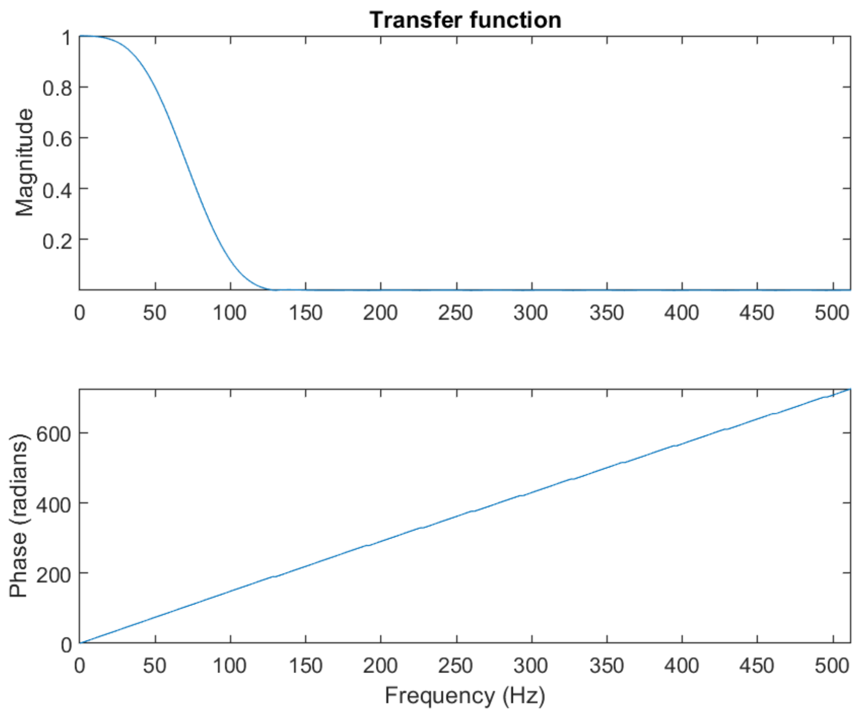
- Convolution theorem:

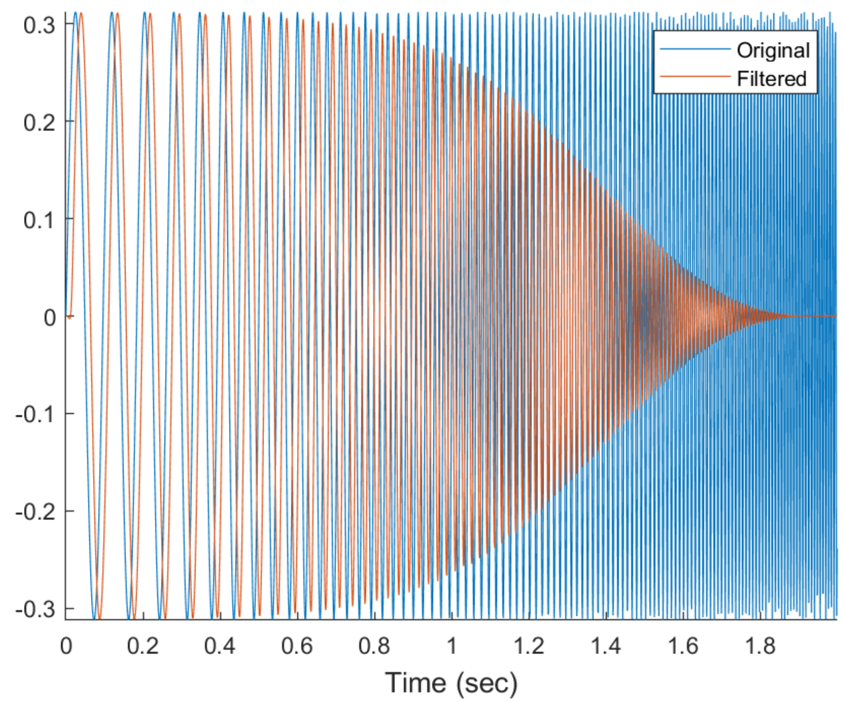
  DFT of the filter output ($\tilde{y}(f)$) = DFT of the input ($\tilde{x}(f)$) $\times$ Filter Transfer function ($T(f)$)

- Filter transfer function: DFT of the filter impulse response
- Impulse response: Output sequence when the input sequence is an impulse sequence, $(0,0,\ldots,0,1,0,\ldots,0)$
  - Finite impulse response (FIR) filter
  - Infinite impulse response (IIR) filter
- Filter design: By choosing an appropriate filter transfer function, we can alter the spectrum of a signal

# Example: Low pass filter

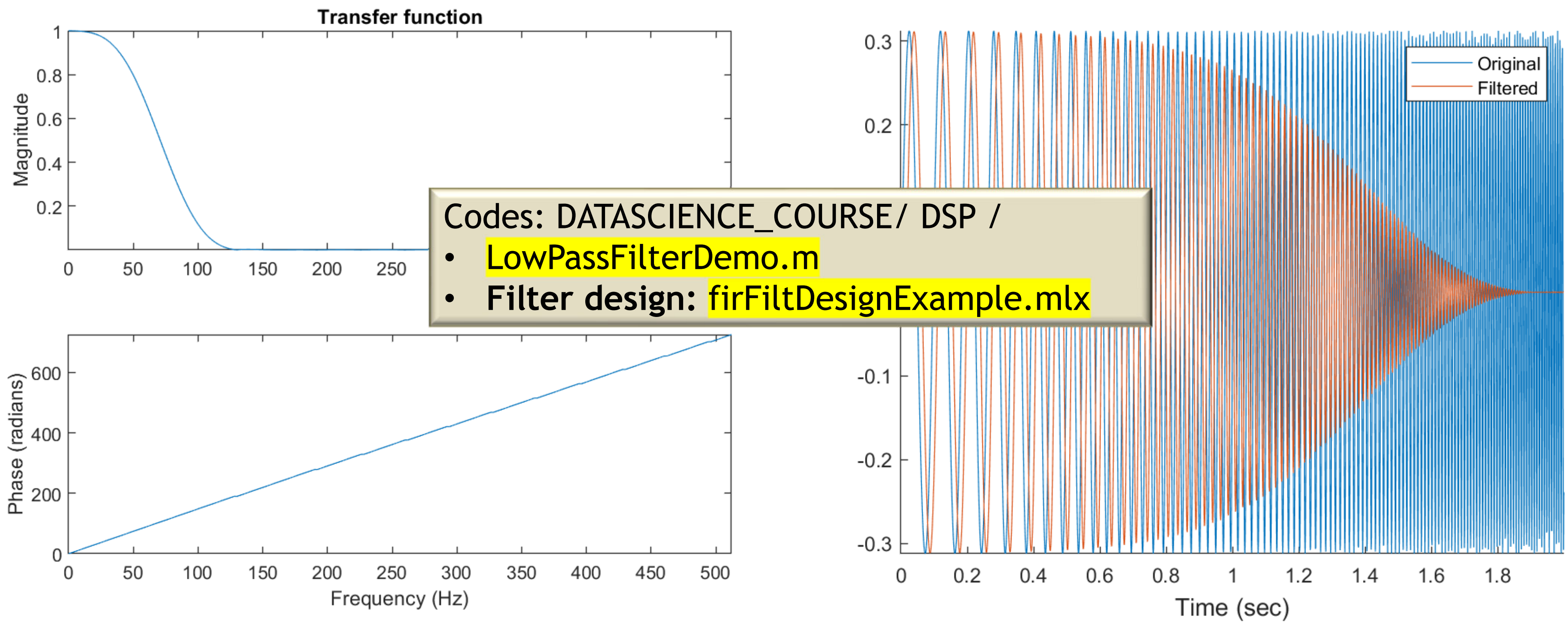
- A low pass FIR filter applied to a  signal of increasing frequency
  - High instantaneous frequency part of the signal is suppressed



# Example: Low pass filter

- A low pass FIR filter applied to a  signal of increasing frequency
  - High instantaneous frequency part of the signal is suppressed




Codes: DATASCIENCE_COURSE/ DSP /
- LowPassFilterDemo.m
- **Filter design:** firFiltDesignExample.mlx

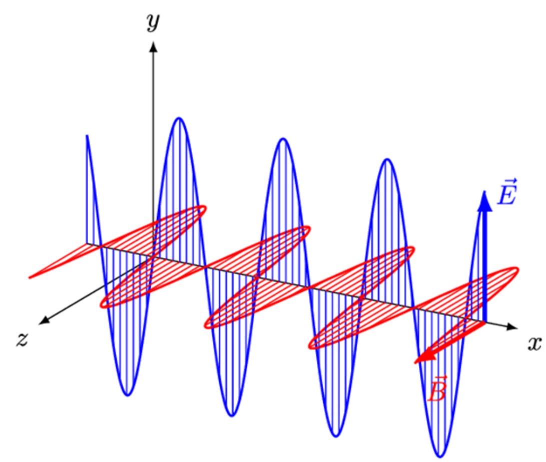
# Gravitational wave strain

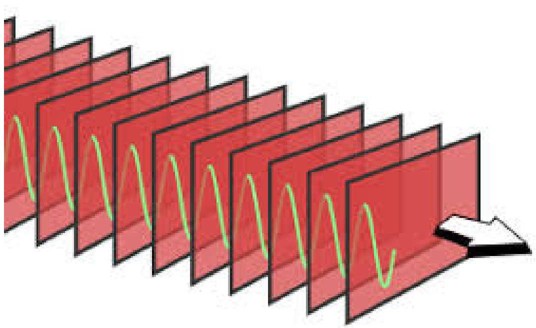
# Gravitational waves for data analysts

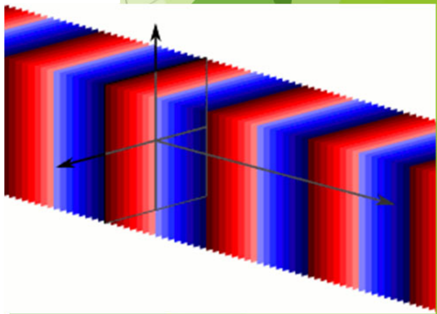
For data analysis, GWs can be treated as just another type of plane wave like those in acoustics or Electromagnetism

## Plane Electromagnetic (EM) wave

- Electric and magnetic field vectors are perpendicular to direction of wave propagation
- Fields are uniform in any plane perpendicular to the wave propagation direction
- Fields vary from one plane to another
- Planes move with speed of light
- Observer sees fields oscillating as the planes move past a point
- Effect: Acceleration of electric charges



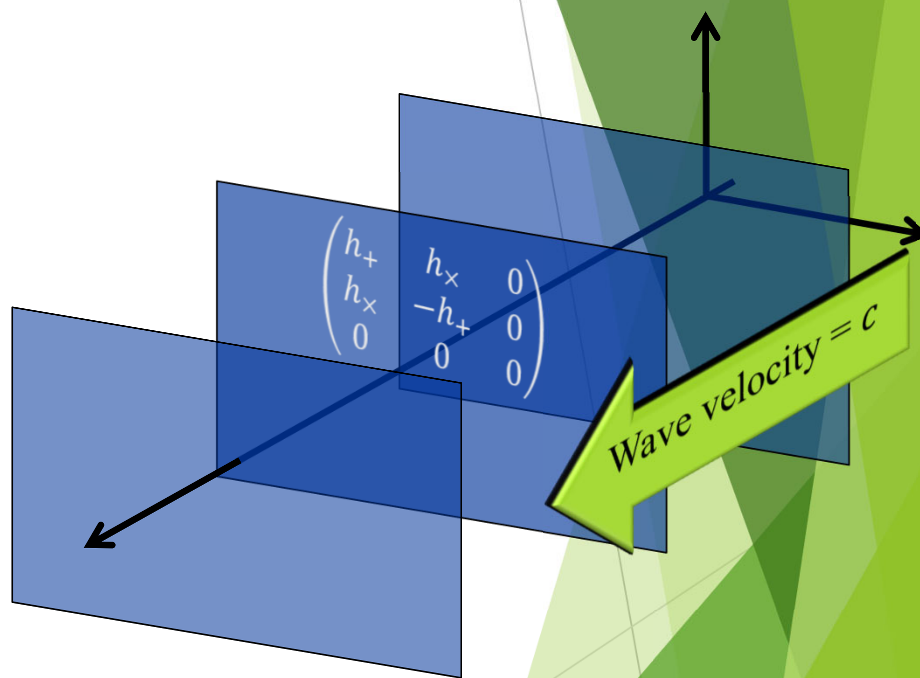





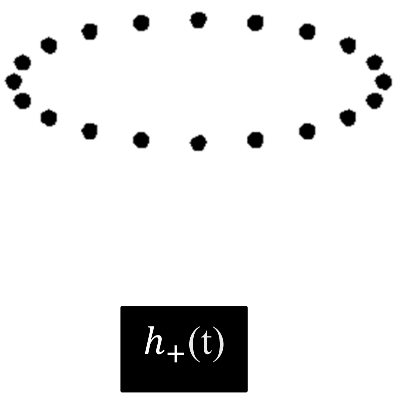
# Gravitational waves for data analysts

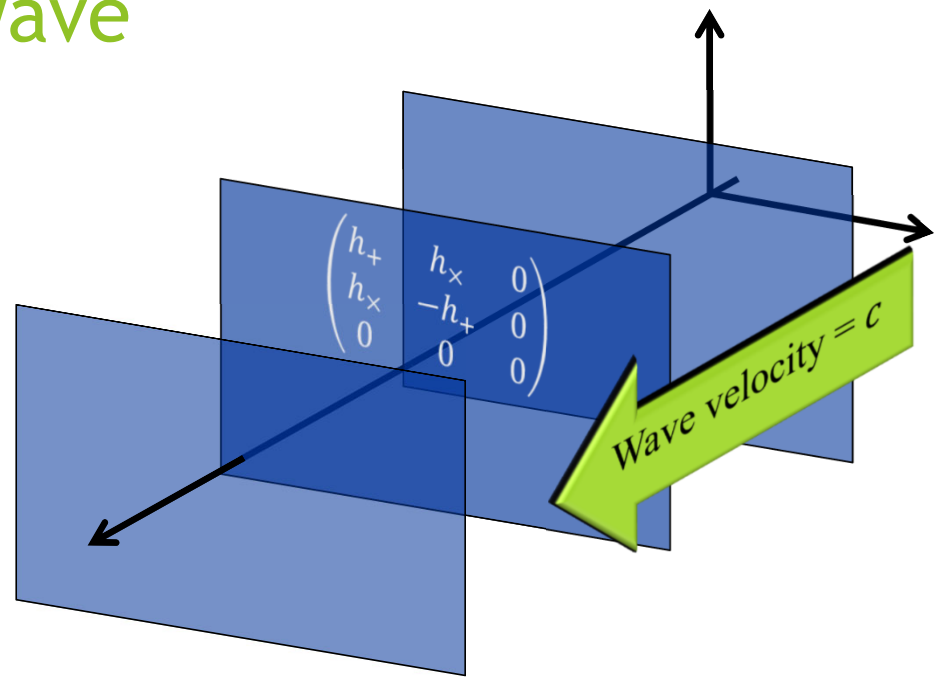
## Plane Gravitational Wave

- **Tensor wave**: oscillating matrix elements at every point
- (Specific gauge choice) Only two independent functions of time $h_+(t)$ and $h_\times(t)$
- Wave speed same as speed of light
- Effect: Space stretches and squeezes as a GW passes by (“Ripples in space-time”)



# Plane gravitational Wave

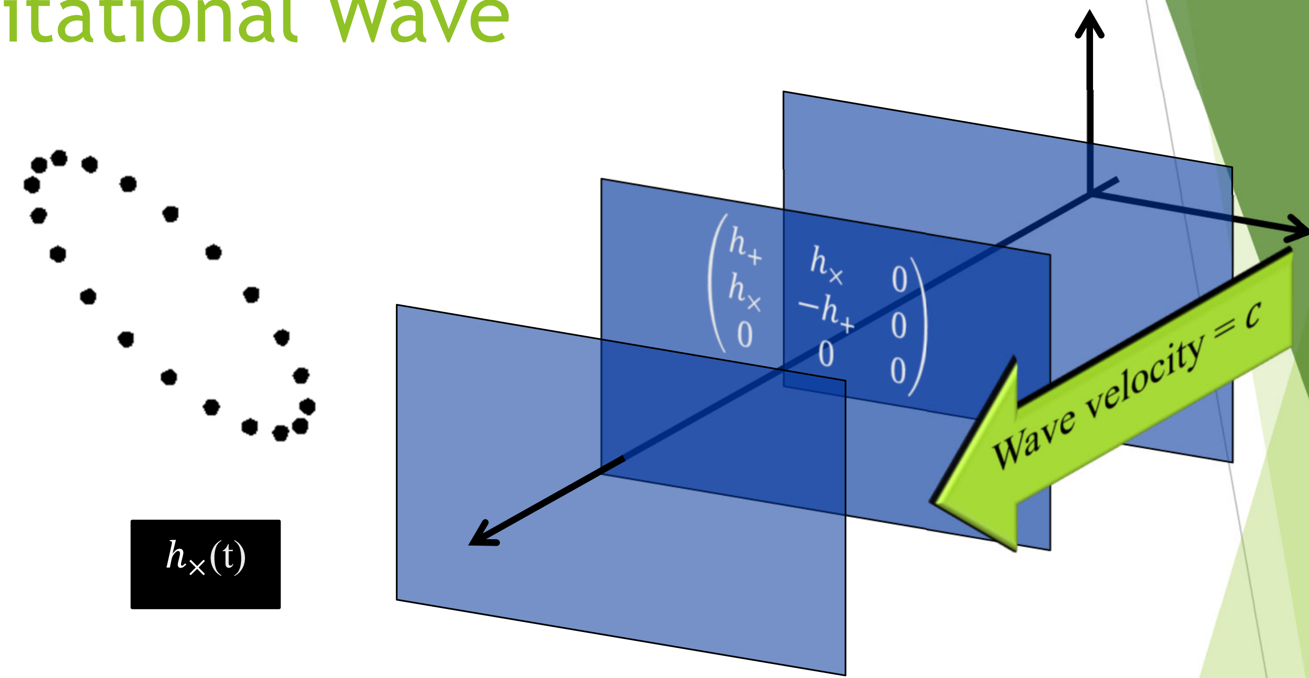



Plane wave crossing a fixed point

$$\left.\begin{aligned} h_{11} = -h_{22} = h_+(t) \\ h_{12} = h_{21} = h_\times(t) \end{aligned}\right\} \text{two } \textbf{polarizations}$$

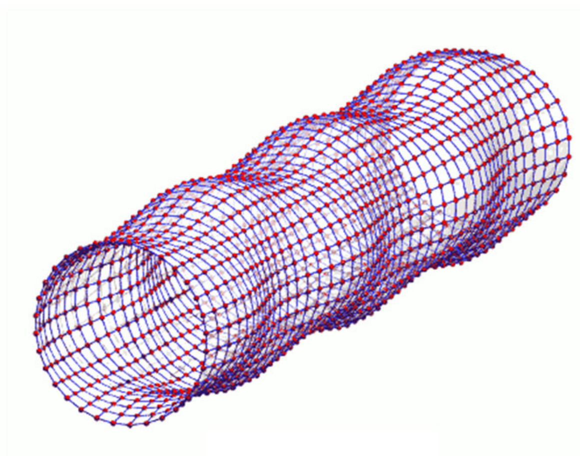
# Plane gravitational Wave

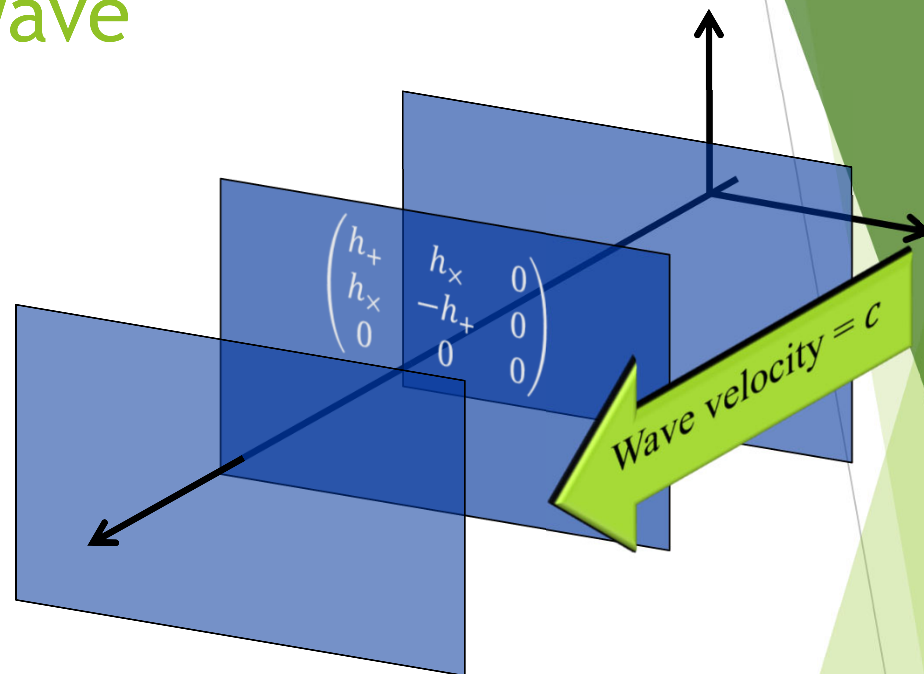



Plane wave crossing a fixed point

$$\left.\begin{aligned} h_{11} = -h_{22} = h_+(t) \\ h_{12} = h_{21} = h_\times(t) \end{aligned}\right\} \text{two } \textbf{polarizations}$$

# Plane gravitational Wave

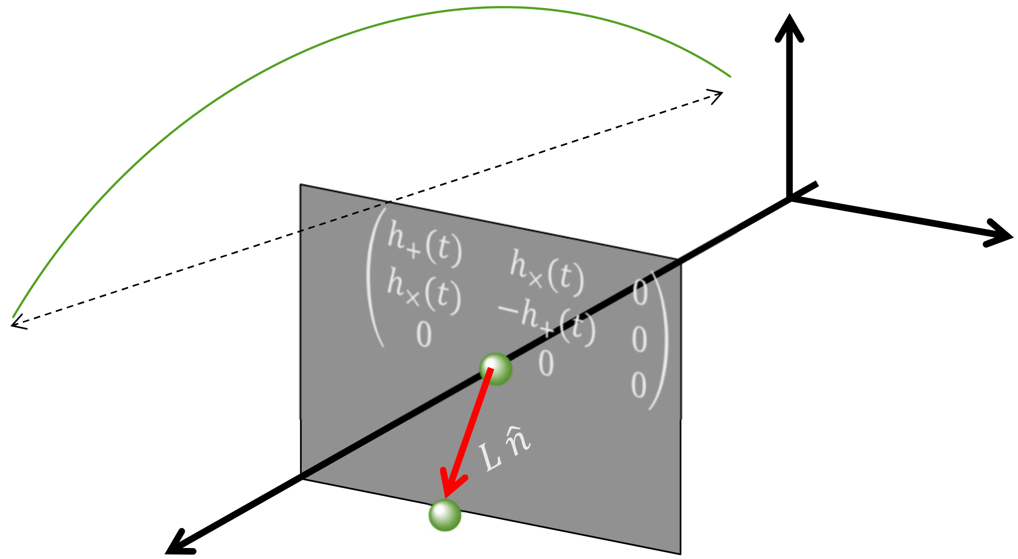





Plane wave crossing a fixed point

$$\left.\begin{aligned} h_{11} = -h_{22} = h_+(t) \\ h_{12} = h_{21} = h_\times(t) \end{aligned}\right\} \text{two } \textbf{polarizations}$$

# Long wavelength approximation




GW Strain (long-wavelength approximation):

$$s(t) = \frac{\Delta L(t)}{L} = \frac{1}{2}\sum_{i,j=1}^{3} h_{ij}(t)\, n_i n_j$$

$$= \underbrace{W_{ij}}_{\substack{Wave\\Tensor}}(t) n^i n^j = W_{ij}\,(t)(\hat{n}\otimes\hat{n})^{ij} = \overleftrightarrow{W} : \overleftrightarrow{D} \text{ (tensor contraction)}$$

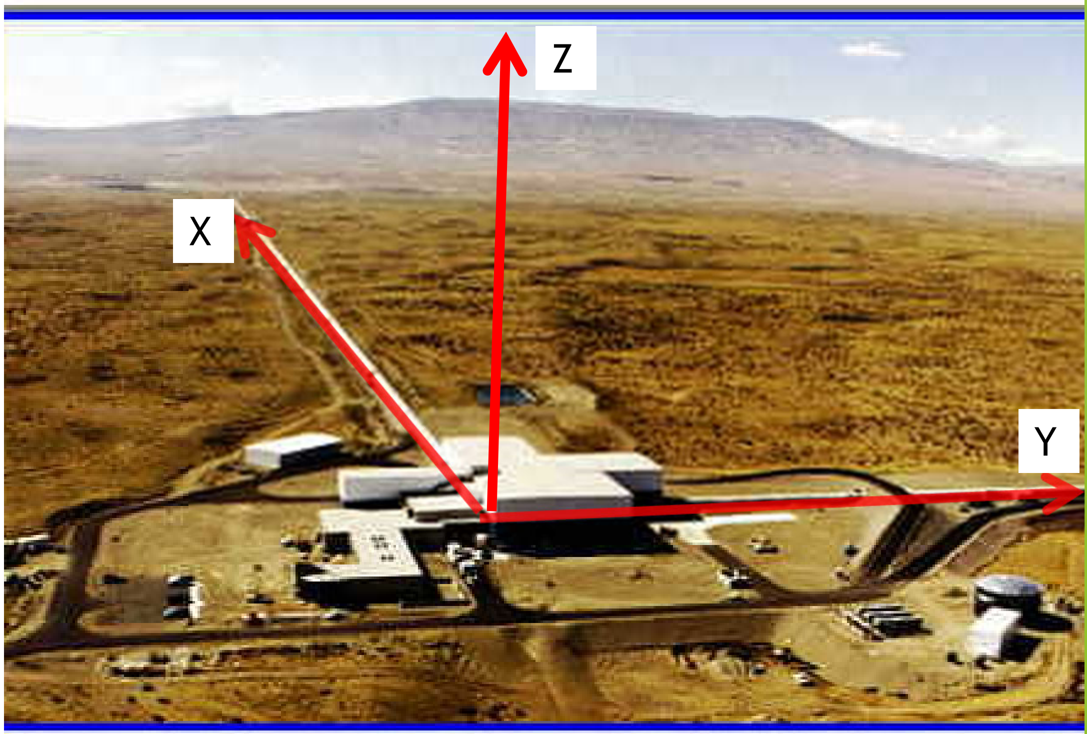
# Strain signal

- GW interferometer: Measured quantity is the difference in arm lengths
- Detector tensor:

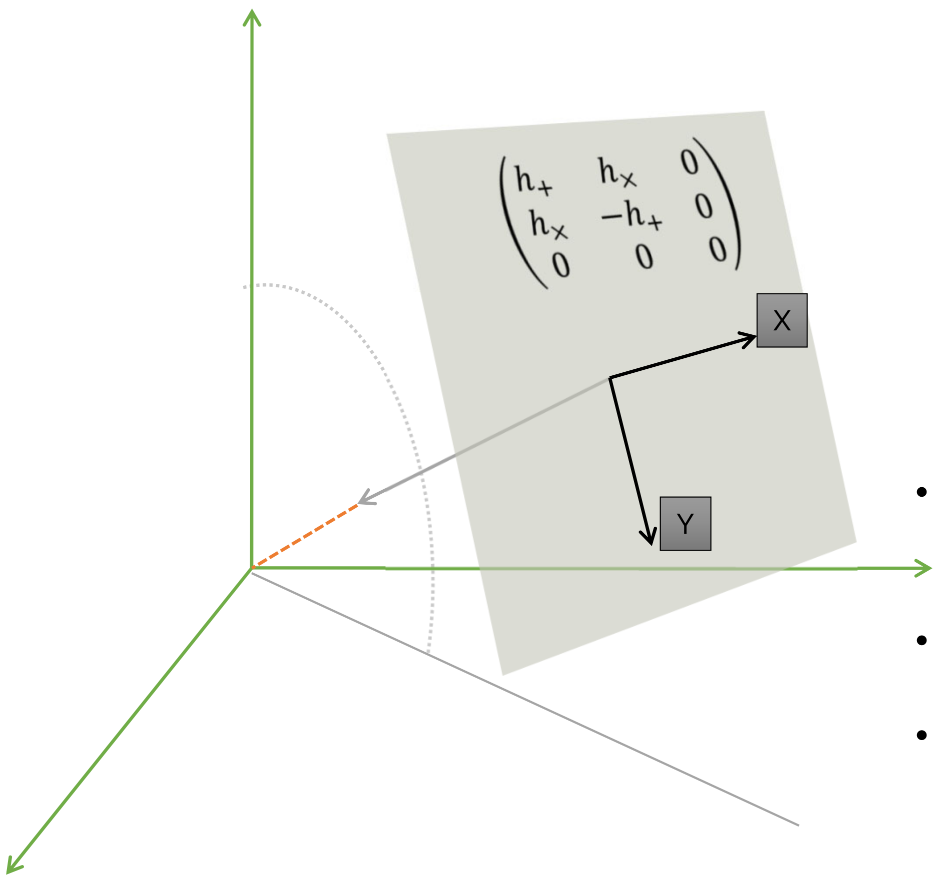
$$\overleftrightarrow{D} = \hat{n}_X \otimes \hat{n}_X - \hat{n}_Y \otimes \hat{n}_Y$$

- Defined purely by the orientation of the detector arms
- Strain signal:

$$s(t) = \overleftrightarrow{W} : \overleftrightarrow{D}$$



# Wave tensor




- The wave tensor is expressed most simply in the "Wave frame": Z-axis along wave propagation direction
- In terms of X and Y axes orthogonal to wave frame Z-axis:

$$\begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix} = h_+(t)\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + h_\times(t)\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\overleftrightarrow{W} = \frac{1}{2}h_+(t)(\hat{x}\otimes\hat{x} - \hat{y}\otimes\hat{y}) + \frac{1}{2}h_\times(t)(\hat{x}\otimes\hat{y} + \hat{y}\otimes\hat{x})$$

- Common notation in papers:

$$\overleftrightarrow{W} = h_+(t)\,\overleftrightarrow{e}_+ + h_\times(t)\overleftrightarrow{e}_\times$$

- $\overleftrightarrow{e}_+$ and $\overleftrightarrow{e}_\times$ are called "plus" and "cross" polarization tensors
- The wave tensor is defined purely in terms of the wave frame

# Strain signal

$$s(t) = \overleftrightarrow{W} : \overleftrightarrow{D} = h_+(t)\overleftrightarrow{D} : \overleftrightarrow{e}_+ + h_\times(t)\overleftrightarrow{D} : \overleftrightarrow{e}_\times$$

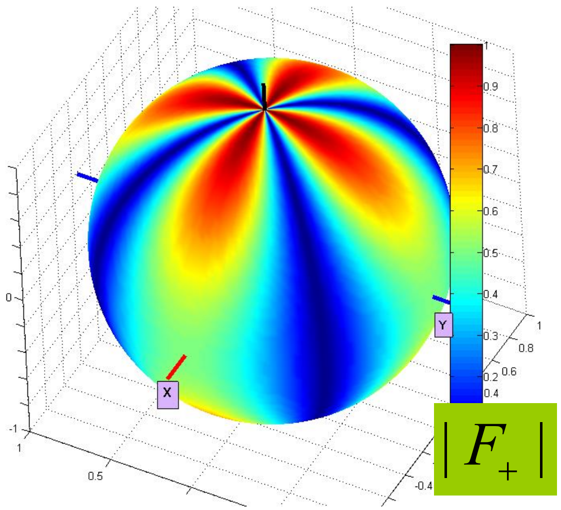
$$s(t) = h_+(t)F_+\left(\hat{k}\right) + h_\times(t)F_\times(\hat{k})$$

- $F_{+,\times}$ are called the antenna pattern functions of the detector and depend on the direction from which the wave is coming
  - $\hat{k}$ is specified by sky angles $\theta$ and $\phi$ (different conventions)
  - $\Rightarrow F_{+,\times}\left(\hat{k}\right) = F_{+,\times}\,(\theta, \phi)$

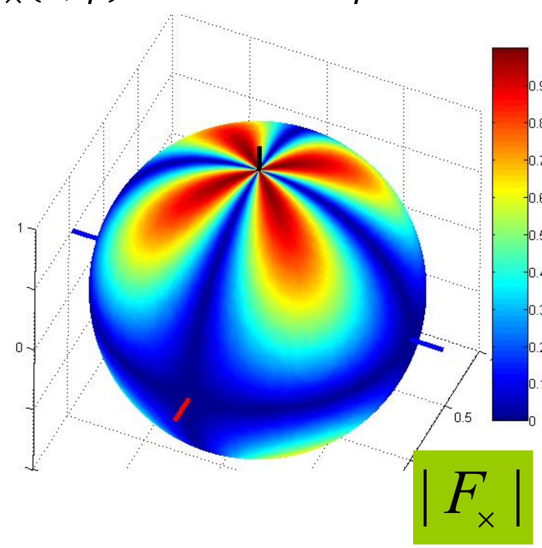
# Antenna patterns: L-shaped interferometer

$$s(t) = F_+(\underbrace{\theta, \phi}_{sky\ angle})h_+(t) + F_\times(\theta, \phi)h_\times(t)$$

$$F_+(\theta, \phi) = \frac{1}{2}(1 + \cos^2\theta)\cos 2\phi$$

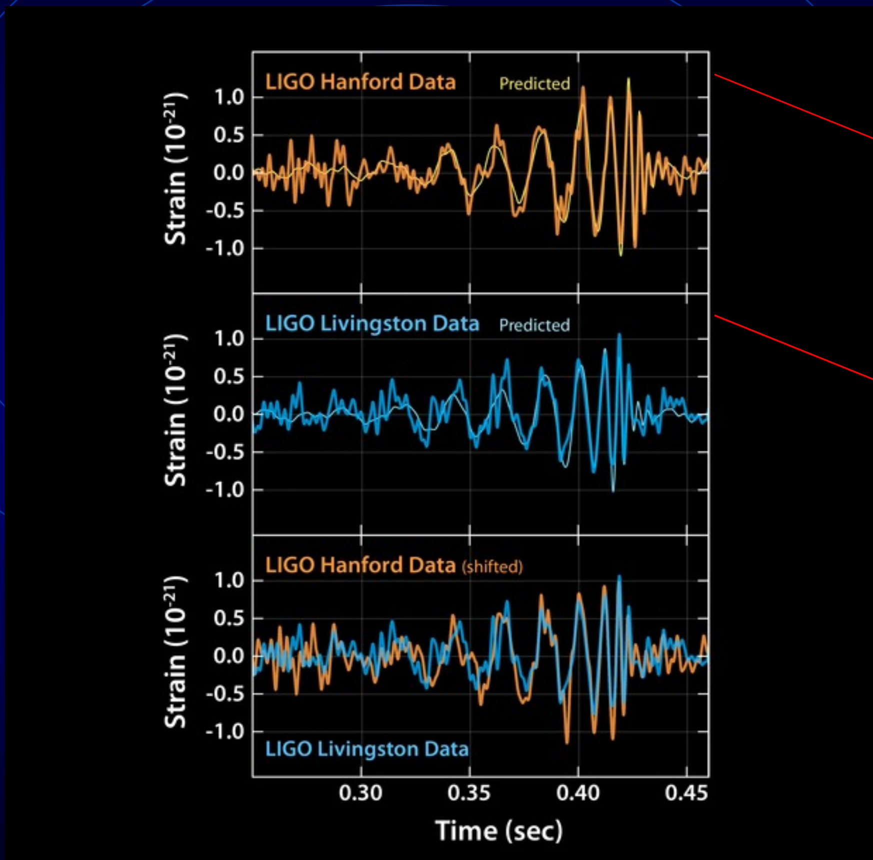



$$F_\times(\theta, \phi) = \cos\theta \sin 2\phi$$

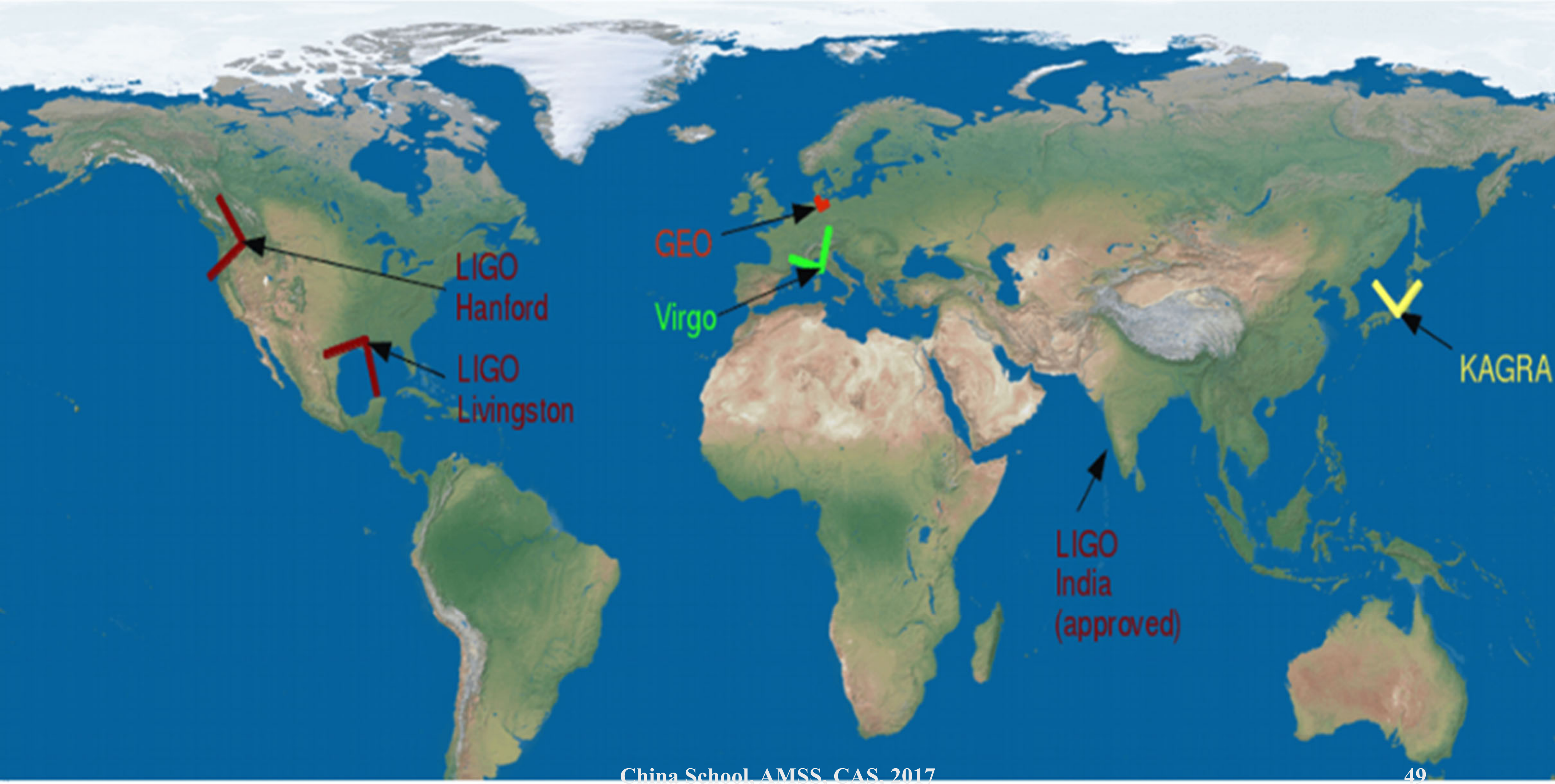



See code: GWSC / GWSIG / skyplot.m; testskyplot.m; testdetframefpfc.m

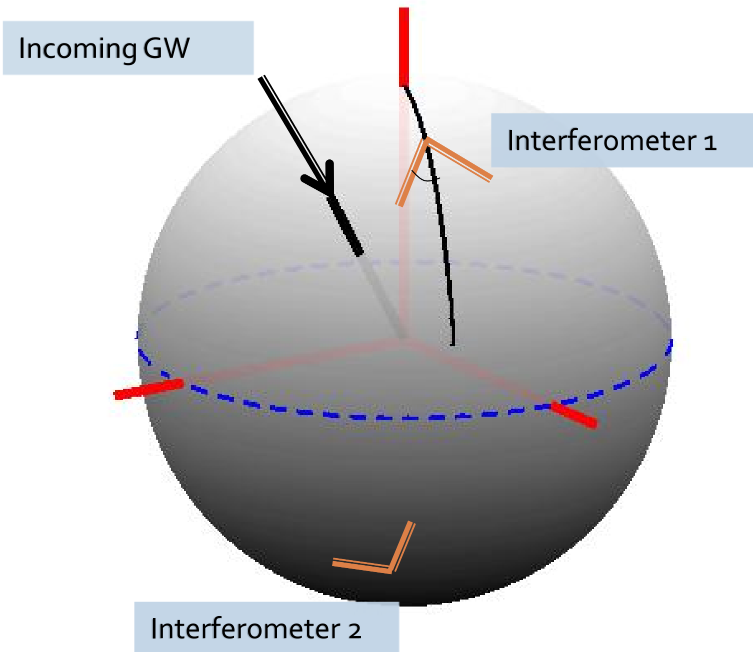
# GW150914: Strain signal (Whitened)




# Network of detectors


GEO
LIGO
Hanford
Virgo
LIGO
Livingston
KAGRA
LIGO
India
(approved)

# Network response



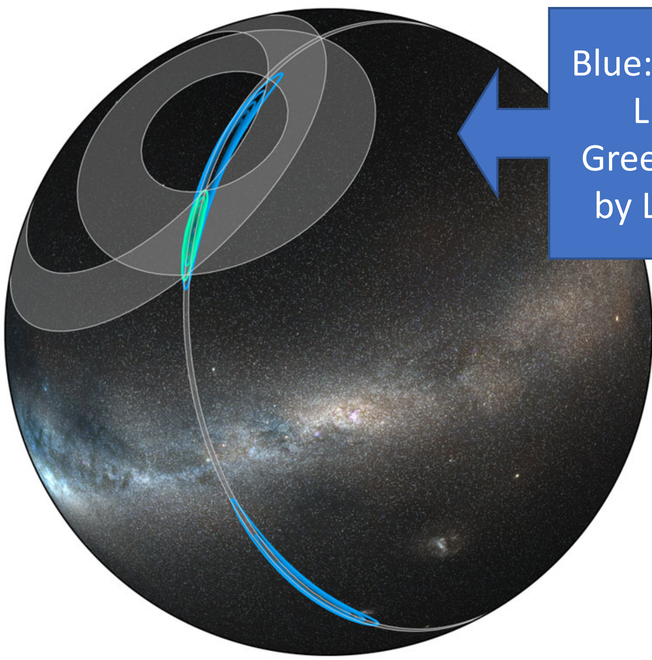

- Incoming GW is a plane wave

- It will hit the different detectors at different times

- Not only will different detectors see the signal at **different times**, but they will also measure **different strain signals** because of detector antenna patterns

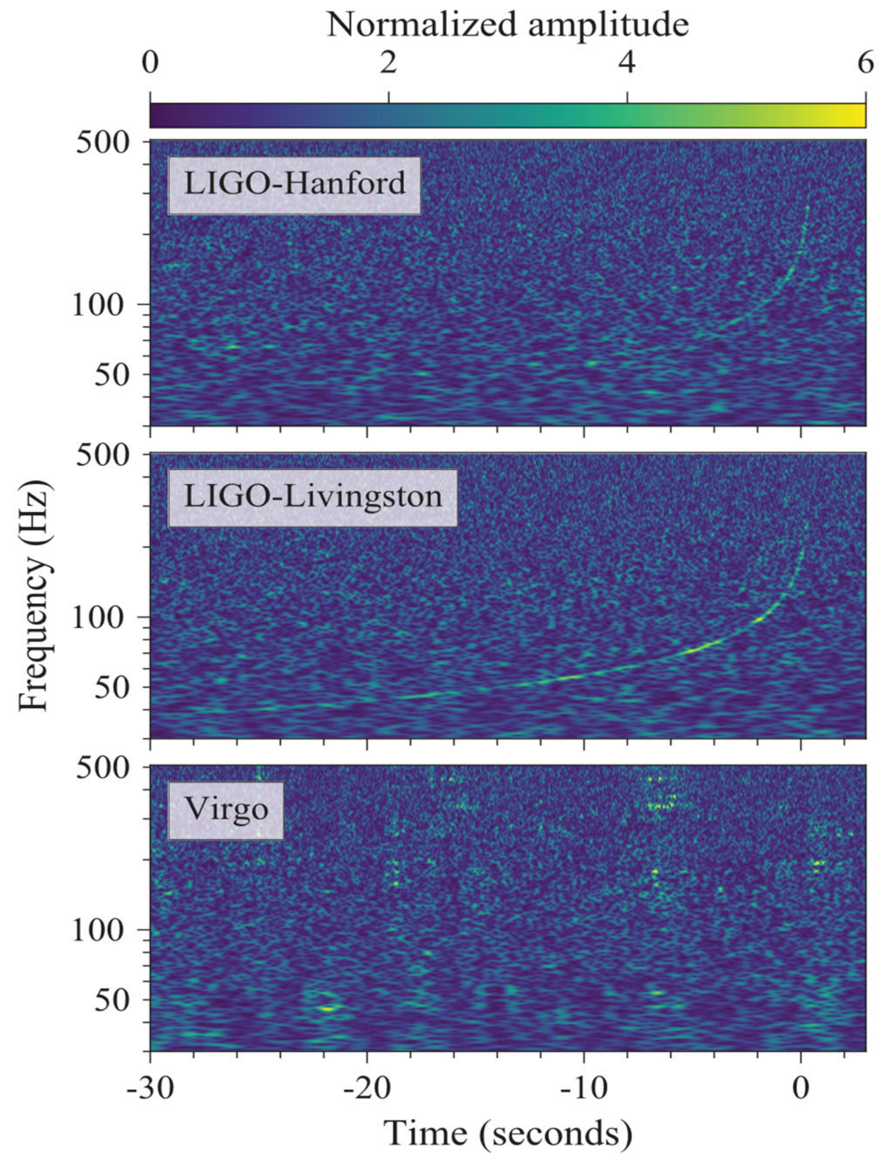
- ⇒ Sky location and polarizations can be estimated using a detector network

# GW170817: a 3-detector event



Credit: LIGO/Virgo/NASA/Leo Singer (Milky Way image: Axel Mellinger)

Blue: Localization by LHO and LLO
Green: Localization by LHO-LLO-Virgo



# Probability theory refresher

# Probability density function: PDF

- Because of noise, GW strain at any time instant is the trial value of a continuous random variable
- Probability density function (pdf) $p_X(x)$ gives us the probability of getting a trial value of $X$ in the infinitesimal range $[x, x + dx]$

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

Probability of getting a trial value from an infinitesimal interval

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

Summing up the probabilities to get the probability of the full event $\mathbb{A}$

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

- Frequentist: Probability of an event is the frequency of occurrence of trial values in the event in a large (approaching infinity) number of trials
  - Frequency: Ratio of the number of times a given event occurs to the total number of trials
  - Example: Throwing a coin multiple times and estimating the frequency of getting heads
- Since the probability for any event is positive: $p_X(x) \geq 0$

# Expectation

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

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

$f(X) = X$

**mean** of $p_X(x)$

$f(X) = (X - E[X])^n$

**$n^{th}$ order central moment** of $p_X(x)$

$n = 2 \rightarrow$ "variance" $\sigma^2$

"standard deviation" : $\sigma = \sqrt{[(X - E[X])^2]})$

Linearity property of expectation

$$E[af(X) + bg(X)] = \int_{-\infty}^{\infty} (af(x) + bg(x)) p_X(x) dx$$
$$= aE[f(X)] + bE[g(X)]$$

# Estimation of expectation

- Estimation using sample average

$$E[f(X)] = \int_{-\infty}^{\infty} f(x)p_X(x)dx \approx \sum_{k=-\infty}^{\infty} f(x_k)[p_X(x_k)\delta x] \approx \sum_{k=-\infty}^{\infty} f(x_k) \times \frac{\#trials\ x \in [x_k, x_k + \delta x]}{N_{trials}}$$

$$\approx \frac{1}{N_{trials}} \sum_{k=-\infty}^{\infty} \left( \sum_{m=1}^{\#trials\ x\in[x_k,x_k+\delta x]} f(x_{m,k}) \right), \text{where } x_k \le x_{m,k} \le x_k + \delta x\ and\ f(x_{m,k}) \approx f(x_k)$$

| … | $[x_{-4}-x_{-3}]$ | $[x_{-3}, x_{-2}]$ | $[x_{-2}, x_{-1}]$ | $[x_{-1}, x_0]$ | $[x_0, x_0 + \delta x]$ | $[x_1 = x_0+\delta x, x_2]$ | $[x_2, x_3]$ | $[x_3, x_4]$ | … |
|---|---|---|---|---|---|---|---|---|---|
| … | **** | ******* | ***********<br>*********** | ***********<br>* | *** | ***** | * | ********** | … |

$$\Rightarrow E[f(X)] \approx \frac{1}{N_{trials}} \sum_{j=1}^{N_{trials}} f(x_j)$$

Example: $f(X) = X$
Estimate of mean is the **sample mean**

$$E[X] \approx \frac{1}{N_{trials}} \sum_{j=1}^{N_{trials}} x_j$$

# Joint probability density function: bivariate pdf

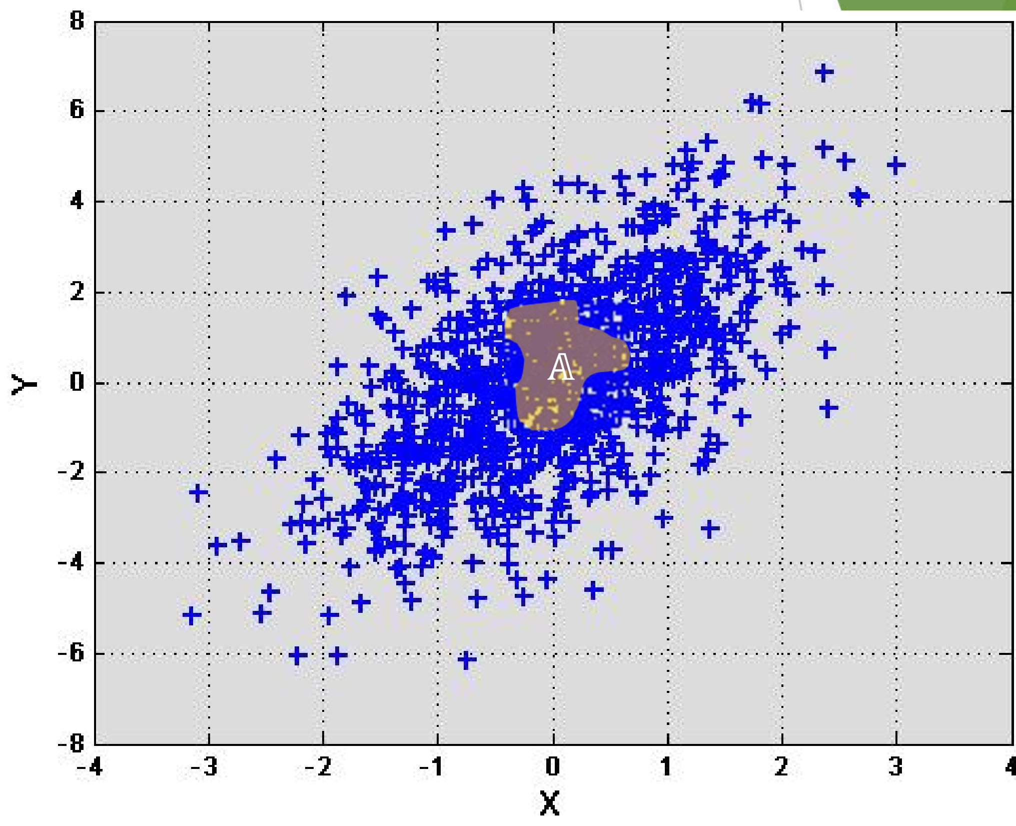
- Consider 2 continuous random variables $X$ and $Y$
- Trial outcome is a pair $(x, y)$ of values
- Joint pdf of two continuous random variables $X$ and $Y$: $p_{XY}(x, y)$
- Gives us:

$$\Pr(X \in [x, x + dx] \text{ AND } Y \in [y, y + dy]\,) = p_{XY}(x, y)dx\,dy$$

$$\Pr\big((X, Y) \in \mathbb{A} \subset \mathbb{R}^2\big) = \iint_{\mathbb{A}} p_{XY}(x, y)dxdy$$

# Geometrical View

- Example: Two random variables $X, Y$
- Sample space: $\mathbb{R}^2$ (plane)
  - A trial outcome $(x, y\ ) \in \mathbb{R}^2$ can be visualized as a point
- The probability of the event, $\mathbb{A}$, shown = the probability of getting a trial point in that area



# Joint probability density function: multivariate pdf

- Joint pdf of $N$ random variables $(X_1, X_2, \dots, X_N)$

$$\Pr\big((X_1, X_2, \dots, X_N) \in \mathbb{V} \subset \mathbb{R}^N\big) = \int_{\mathbb{V}} p_{X_1 X_2 \dots X_N}(x_1, x_2, \dots, x_N)\, d^N x$$

$$\Rightarrow \int_{\mathbb{R}^N} p_{X_1 X_2 \dots X_N}(x_1, x_2, \dots, x_N)\, d^N x = 1$$

- Joint expectations

$$E[f(X_1, X_2, \dots, X_N)] = \int_{-\infty}^{\infty} dx_1 dx_2 \dots dx_N\, f(x_1, x_2, \dots, x_N) p_{X_1, X_2, \dots, X_N}(x_1, x_2, \dots, x_N)$$

# Covariance

<u>$N$ random variables</u>

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

- $C_{ij}$ is called the Covariance of $X_i$ and $X_j$
  - $C_{ii} = E[(X_i - E[X_i])^2]$ is the variance of $X_i$
- Covariance is a measure of how dependent $X_i$ and $X_j$ are on each other (Caution: zero covariance does not always mean independence!)
- $C_{ij}$ can be seen as the elements of an $N$-by-$N$ square matrix: Covariance matrix $C$
- Properties of the covariance matrix:
  - Symmetric matrix: $C^T = C$ (because $C_{ij} = C_{ji}$)
  - Positive definite: $\bar{x}\boldsymbol{C}\bar{x}^T > 0$ for $\bar{x} \neq 0 \in \mathbb{R}^N$

# Stochastic process (NOISE)

# Stochastic process (Noise): definition

- Stochastic Process (or Noise): An ordered sequence of random variables

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

- Changing the ordering of the random variables changes the stochastic process

- GW data is a realization (set of trial values) of a stochastic process
  - The indices of the random variables in GW data are sample numbers
- Finite subsequence of the stochastic process

$$\bar{X} = \left(X_p, X_{p+1}, \dots, X_{p+N-1}\right) \underbrace{\rightarrow}_{Relable\ indices} (X_0, X_1, \dots, X_{N-1})$$

# Wide-sense stationary process

- All first and second order moments are time-translation independent
- First order moment, i.e. the mean, $E[X_k]$, is constant, independent of $k$
- Second order moment, i.e., the covariance $C_{k,k+m}$ is independent of $k$ and depends only on $m$
- $\Rightarrow C_{k,k} = \sigma^2 = constant$: The variance is the same for all random variables
  - For a wide-sense stationary process, it makes sense to talk about the mean and variance of the stochastic process

# Autocovariance

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

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

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

- If follows from the definition that :

$$\phi(0) = E[(X_k - \mu)^2] = \sigma^2 \text{ : variance}$$

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

- It can be proven that:

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

- The definitions are more convenient if one switches to a continuous index ("time")

Autocovariance function:

$$\phi(\tau) = E[(X(t) - \mu)(X(t + \tau) - \mu)]$$

# Power spectral density (PSD)

# Power Spectral Density

Power spectral density: Fourier transform of autocovariance function

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

$S_n(f)$ is real and symmetric as $\phi(\tau)$ is an even function

Physical interpretation

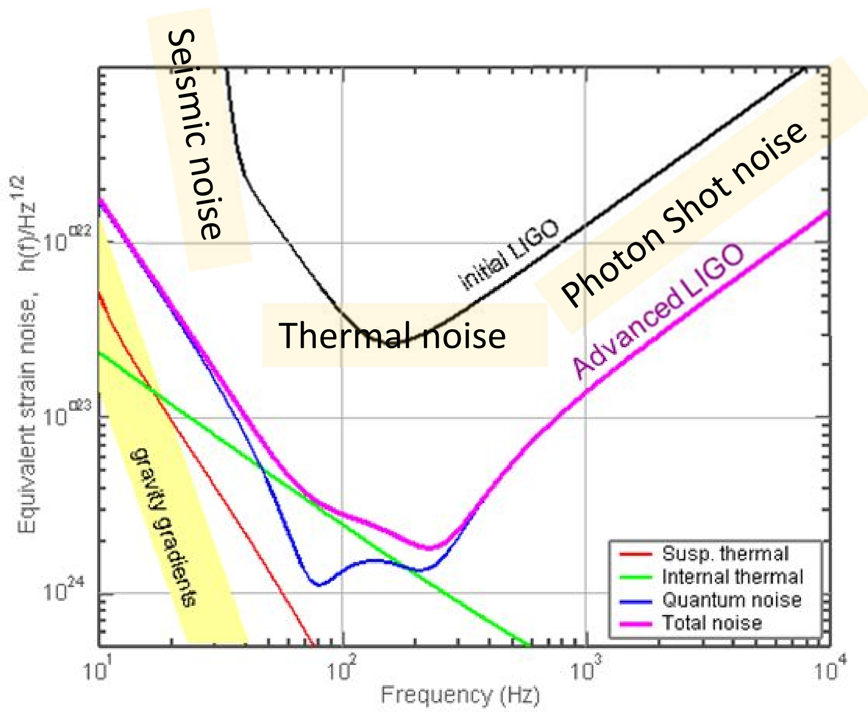
The variance of the noise process

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

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")
$\Rightarrow$ the name "Power Spectral Density"

# Power Spectral Density

- The PSD of noise carries information about the origin of noise in different frequency bands
- Example: Design sensitivity curves for LIGO detectors ($\sqrt{S_n(f)}$; units: $Hz^{-1/2}$)



# White noise

- Wide-sense stationary noise with autocovariance:

$$\phi(m) = \begin{cases} \sigma^2, & m = 0 \\ 0, & m \neq 0 \end{cases} \Rightarrow S_n(f) = const.$$

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

  - Analogous to white light: Equal intensity (or "power") of light at all visible frequencies
- Simple example of white noise sequence: iid noise
  - Independent and Identically distributed noise
  - iid Gausian noise realization in Matlab: call `randn` repeatedly

# Wiener-Khinchin theorem

Input noise process with PSD $S_{in}(f)$ → Filter with transfer function $T(f)$ → Output noise process with PSD $S_{out}(f)$

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

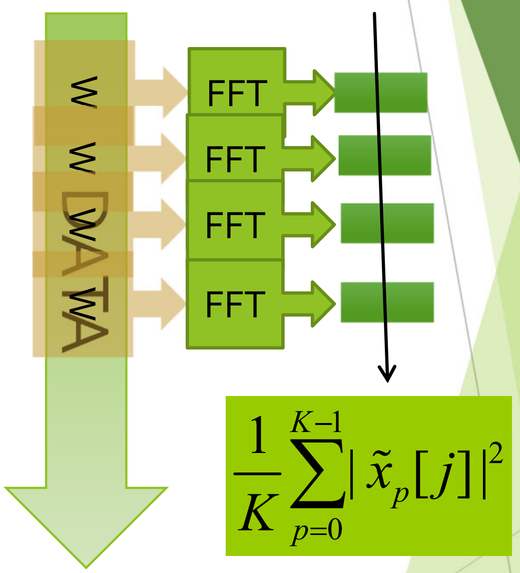
- Passing white noise through a filter yields colored noise with $\mathrm{PSD} \propto |T(f)|^2$
- Can be used to simulate colored noise or flatten ("Whiten") the PSD of real noise
- See GWSC / NOISE / gengwnoise.m, test_gengwnoise.m

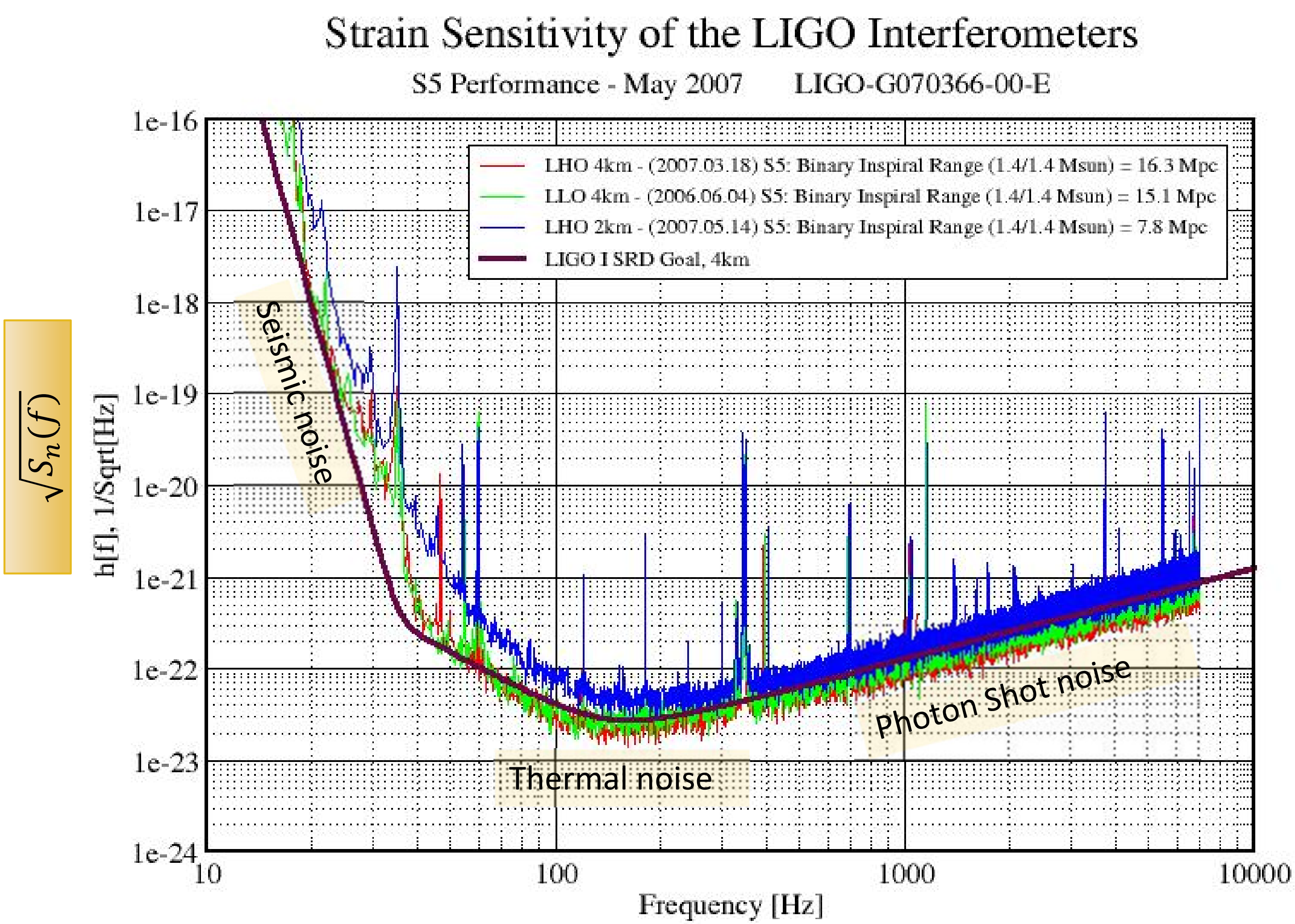
# PSD estimation: Welch's method

- PSD is not a random quantity: based on expectations
- Just like expectations can be *estimated* from given data, we can *estimate* the PSD from given data
- Welch's method of overlapping windows
  - Already a built-in function in Matlab: `pwelch`
  - Compute FFTs of $K$ short overlapping windowed segments
- Calculate modulus squared of the FFT

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

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





$\sqrt{S_n(f)}$
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]
1e-16
1e-17
1e-18
1e-19
1e-20
1e-21
1e-22
1e-23
1e-24
10
100
1000
10000
Frequency [Hz]
Seismic noise
Thermal noise
Photon Shot noise

# Homework

# Fourier transform

- Linearity: $F[s(t)+g(t)] = F[s(t)] + F[g(t)] = \tilde{s}(f) + \tilde{g}(f)$

- Hermiticity property: For $s(t)$ real (i.e., $s^*(t) = s(t)$ )

$$s^*(t) = \int_{-\infty}^{\infty} \tilde{s}^*(f) e^{-2\pi i f t} df = \int_{-\infty}^{\infty} \tilde{s}^*(-f) e^{2\pi i f t} df = s(t) = \int_{-\infty}^{\infty} \tilde{s}(f) e^{2\pi i f t} df$$

$$\tilde{s}(-f) = \tilde{s}^*(f) \Rightarrow \begin{cases} Re\left(\tilde{s}(f)\right) \to \text{Even function} \\ \mathrm{Im}\left(\tilde{s}(f)\right) \to \text{Odd function} \end{cases}$$

# Fourier transform

- Rewriting,

$$s(t) = \int_{-\infty}^{\infty} \tilde{s}(f) e^{2\pi i f t} df = \int_{-\infty}^{\infty} \Big(\underbrace{a(f)}_{even} + i\, \underbrace{b(f)}_{odd}\Big) e^{2\pi i f t} df$$

$$= \int_{-\infty}^{\infty} \left[\Big(\underbrace{a(f)\cos(2\pi f t)}_{even\times even=even} \underbrace{-b(f)\sin(2\pi f t)}_{odd\times odd=even}\Big) + \cancel{i\,\Big(\underbrace{a(f)\sin(2\pi f t)}_{even\times odd=odd} + \underbrace{b(f)\cos(2\pi f t)}_{odd\times even=odd}\Big)}\right] df$$

$$= 2\int_{0}^{\infty} \sqrt{a^2(f) + b^2(f)}\, \cos\left(2\pi f t + \tan^{-1}\frac{b(f)}{a(f)}\right) df$$

- Thus, the meaning of the Fourier transform is that any $s(t)$ can be expressed as a sum ("superposition") of sinusoids of the form

$$A(f)\cos(2\pi f t + \Phi(f))$$

over a range of frequencies

- The Fourier transform of a signal is called its **Spectrum** and decomposing a signal into a superposition of sinusoids is called **Spectral decomposition**

# Fourier transform of a sinusoid

- If $s(t) = A\sin(2\pi f_0 t + \phi)$, its Fourier transform is

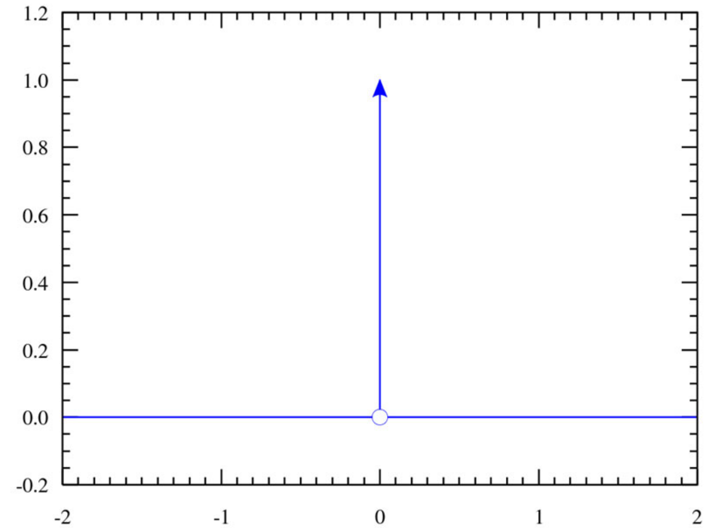
$$\int_{-\infty}^{\infty} A\sin(2\pi f_0 t + \phi)\, e^{-2\pi i f t} dt = A\delta(f - f_0)e^{i\phi} + A\delta(f + f_0)e^{-i\phi}$$

where $\delta(x)$ is the Dirac-delta function

$$\delta(x) = \int_{-\infty}^{\infty} dy\, e^{-2\pi i y x} = \begin{cases} \infty; & x = 0 \\ 0; & \text{otherwise} \end{cases}$$

$$\int_{-\infty}^{\infty} dx\ f(x)\delta(x - x_0) = f(x_0)$$

- The Fourier transform of a sinusoid exists only at the frequency of the sinusoid




Schematic diagram of the Dirac delta function by a line surmounted by an arrow. The height of the arrow is usually used to specify the value of any multiplicative constant, which will give the area under the function. The other convention is to write the area next to the arrowhead

# Discrete Fourier Transform (DFT)

DFT can be seen as an approximation to the Fourier transform

- $x(t) = 0$ for $t \notin [0, T] \Rightarrow$

$$\tilde{x}(f) = \int_{-\infty}^{\infty} x(t) e^{-2\pi i f t}\, dt = \int_{0}^{T} x(t) e^{-2\pi i f t}\, dt$$

- Let $\Delta$ be the sampling interval and $T \approx N\Delta$ be the duration of the signal
- Sample times: $t_n = n\Delta$

$$\tilde{x}(f) = \int_{0}^{T} x(t) e^{-2\pi i f t}\, dt \approx \Delta \sum_{n=0}^{N-1} x(t_n) e^{-2\pi i f t_n} = \Delta \sum_{n=0}^{N-1} x_n\, e^{-2\pi i f (n\Delta)}$$

- Let: $f_k = \frac{k}{N\Delta} = \frac{k}{T}$

$$\frac{\tilde{x}(f_k)}{\Delta} \approx \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i k n}{N}} = \tilde{x}_k$$

- $\Rightarrow$DFT is an approximation to the FT at Fourier frequencies $f_k$, $k = 0, 1, \ldots, N-1$

# Properties of DFT

$$\tilde{x}_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i k n/N}$$

- For N even:

$$\tilde{x}_{\frac{N}{2}+p} = \sum_{n=0}^{N-1} x_n e^{-2\pi i p n/N} e^{-\pi i n} = \sum_{n=0}^{N-1} (-1)^n x_n e^{-2\pi i p n/N}$$

$$\tilde{x}_{\frac{N}{2}-p} = \sum_{n=0}^{N-1} (-1)^n x_n e^{2\pi i p n/N}$$

$$\therefore \tilde{x}_{\frac{N}{2}+p} = \tilde{x}^*{}_{\frac{N}{2}-p}$$

- Hermiticity property: DFT frequencies above $N/2$ are negative frequencies in the Fourier transform

| $\overline{x}$ | DFT |
|---|---|
| 1 | 21.0000 |
| 2 | -3.0000 - 5.1962i |
| 3 | -3.0000 - 1.7321i |
| 4 | -3.0000 |
| 5 | -3.0000 + 1.7321i |
| 6 | -3.0000 + 5.1962i |

# Properties of DFT

$$\tilde{x}_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i k n/N} \Rightarrow \tilde{x}_0 = \sum_{n=0}^{N-1} x_n e^0 = \sum_{n=0}^{N-1} x_n \Rightarrow \text{Real}$$

- For N even:

$$\tilde{x}_{\frac{N}{2}} = \sum_{n=0}^{N-1} x_n e^{-\pi i n} = \sum_{n=0}^{N-1} (-1)^{np} x_n \Rightarrow \text{Real}$$

| $\bar{x}$ | DFT |
|---|---|
| 1 | **21.0000** |
| 2 | -3.0000 - 5.1962i |
| 3 | -3.0000 - 1.7321i |
| 4 | **-3.0000** |
| 5 | -3.0000 + 1.7321i |
| 6 | -3.0000 + 5.1962i |

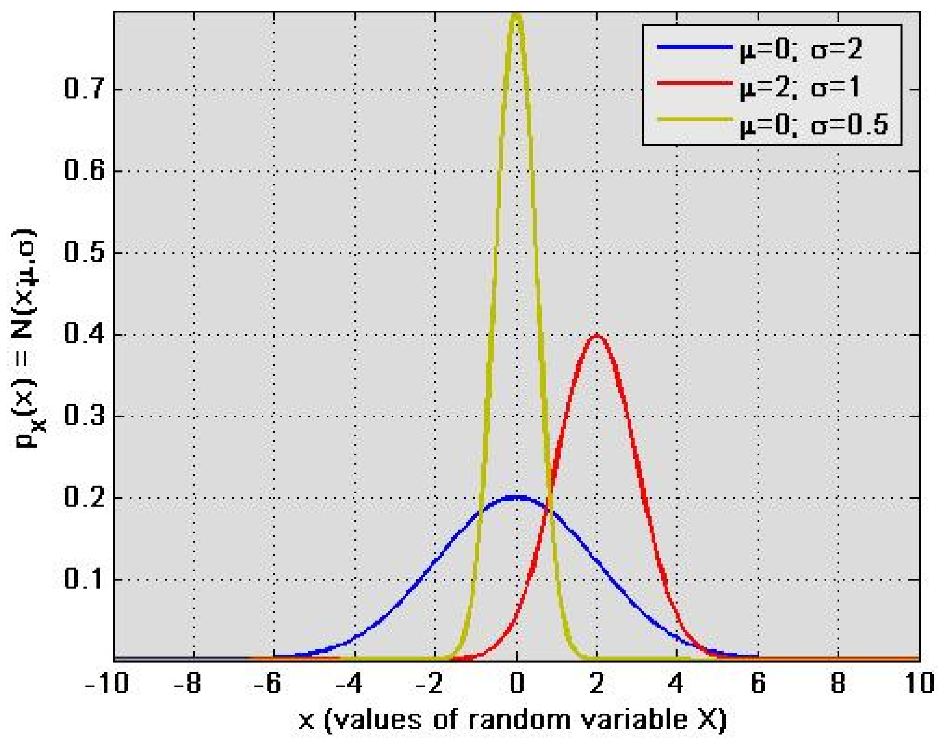
# Exercise: Antenna Patterns

- Using the Earth's center as the origin of your reference frame, obtain the components of the arm unit vectors of a LIGO-like interferometer
- Write code to compute the detector tensor in this frame
- For a given sky location, write code to compute the components of the wave frame X and Y unit vectors
- Write code to contract the polarization tensors and the detector tensor to obtain antenna pattern values for this sky location
- Repeat on a grid of sky locations and plot the resulting values on a sphere to visualize the antenna pattern functions
- Use code with actual coordinates and arm orientations of the two LIGOs and Virgo

# Normal pdf

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

- Also called “Gaussian” pdf in Physics
- $\mu = E[X]$ : Mean and also location of peak
- $\sigma = \sqrt{[(X-E[X])^2]}$ : standard deviation and spread of pdf around the mean
- Trial values can be drawn using the `randn` function in Matlab



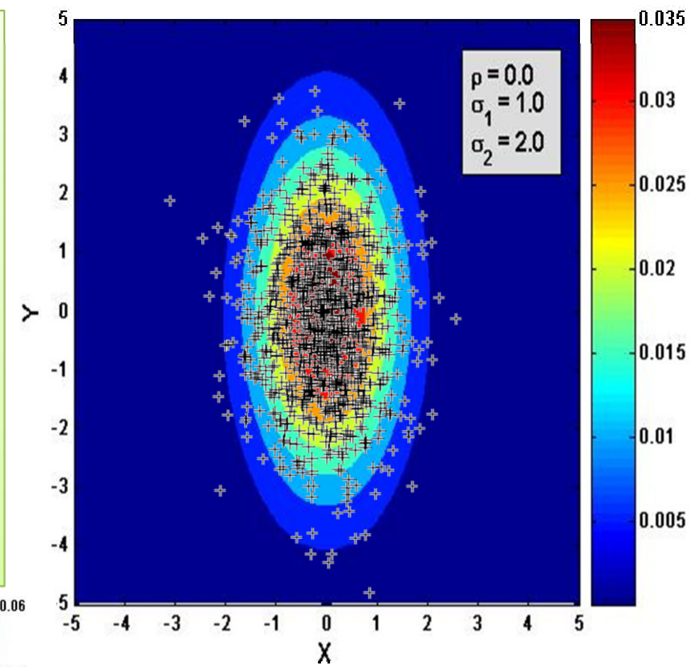
# Bivariate Normal pdf

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

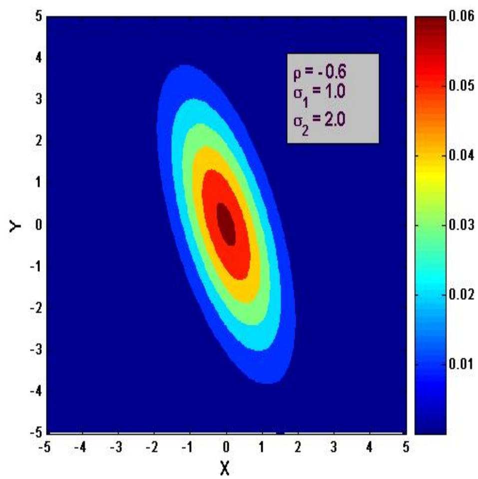
- $\mu_x = E[X], \mu_y = E[Y]$ are the mean values of $X$ and $Y$
- $\mathbf{C}$ is the covariance matrix of $X$ and $Y$
- $|\mathbf{C}|$ is the determinant of $\mathbf{C}$

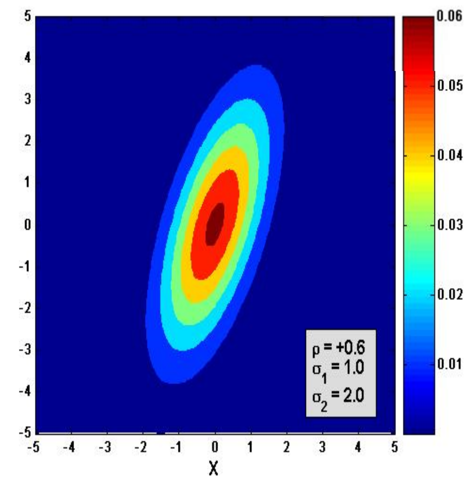
Positive definite $\Rightarrow |\mathbf{C}| > 0 \Rightarrow \mathbf{C} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}; \; -1 < \rho < 1$

$\rho$ : Correlation coefficient; measures dependence between $X$ and $Y$




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



# Multivariate Normal pdf

- Learn more using DATASCIENCE_COURSE / NOISE / bivarnorm.mlx
- The script shows how to generate trial values from any bivariate Normal pdf
- **Understand:** the meaning of covariance by looking at how the distribution of values for one random variable depend on the values taken by the other
- **Exercise:**
  - Generate $N$ trial values from bivariate normal pdf
  - Compare the sample mean and the covariance matrix with their true values for increasing $N$

# Multivariate Normal pdf

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

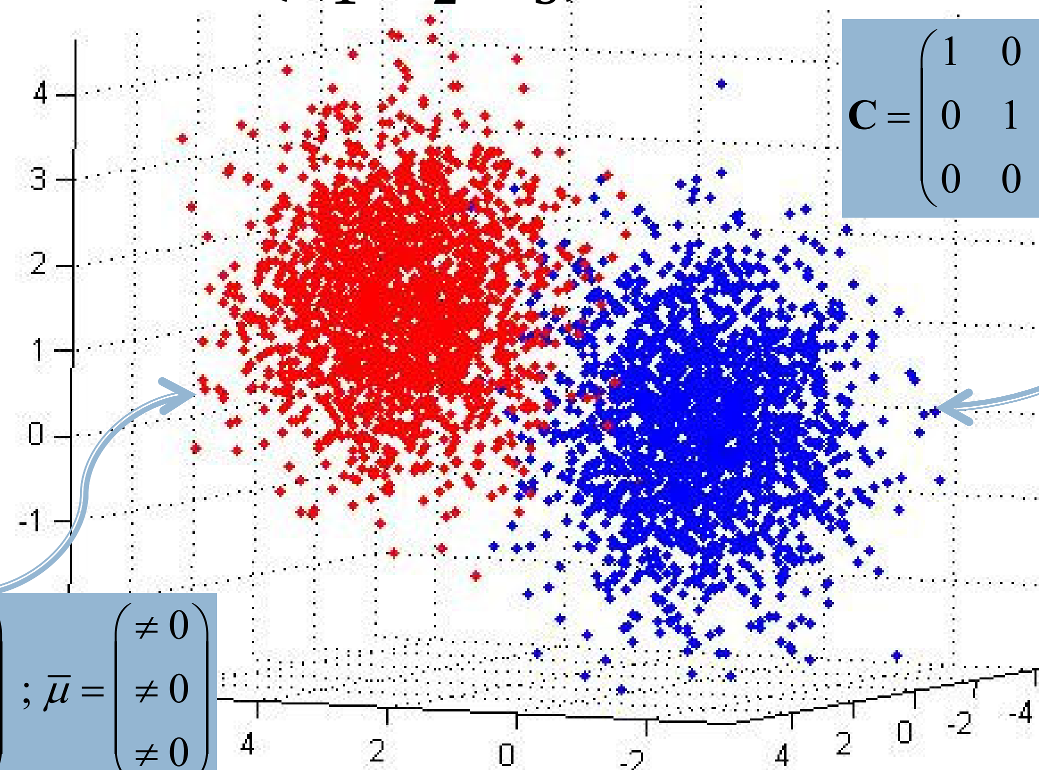
- $\bar{\mu} \in \mathbb{R}^N$ : Mean vector, $\mu_i = E[X_i]$
- $\boldsymbol{C}$ is the covariance matrix $\rightarrow$ $\mathrm{C}_{ij} = E[(x_i-\mu_i)(x_j-\mu_j)]$
- $\boldsymbol{C}$ is **positive definite** and **symmetric** $\Rightarrow$ Ellipsoid in $\mathbb{R}^N$
- If $\boldsymbol{C}$ is a diagonal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

Supplementary reading: https://fabiandablander.com/statistics/Two-Properties.html

# Geometrical View

- A trial outcome $(x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ can be visualized as a point in N-dimensional space
- Joint pdf tells us how the points from an infinity of trials are going to be distributed
- Example: Trial outcomes from Tri-variate normal pdf

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




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

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

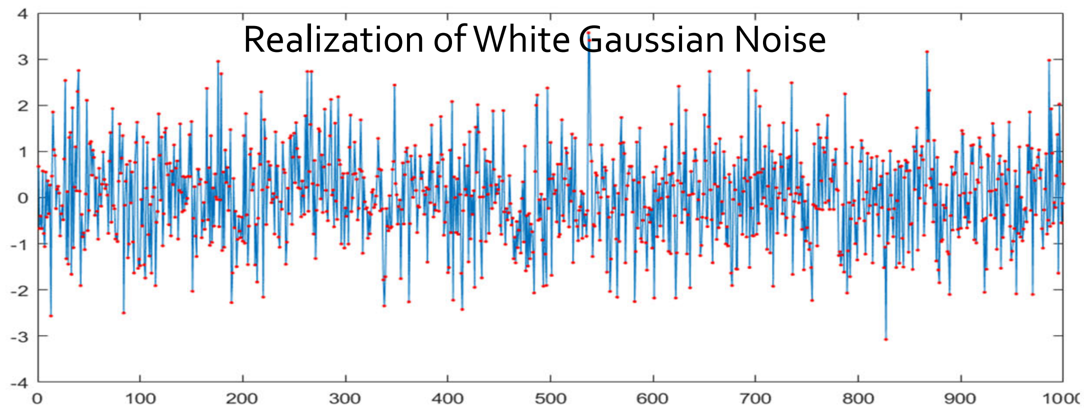
# Gaussian noise: a common noise model

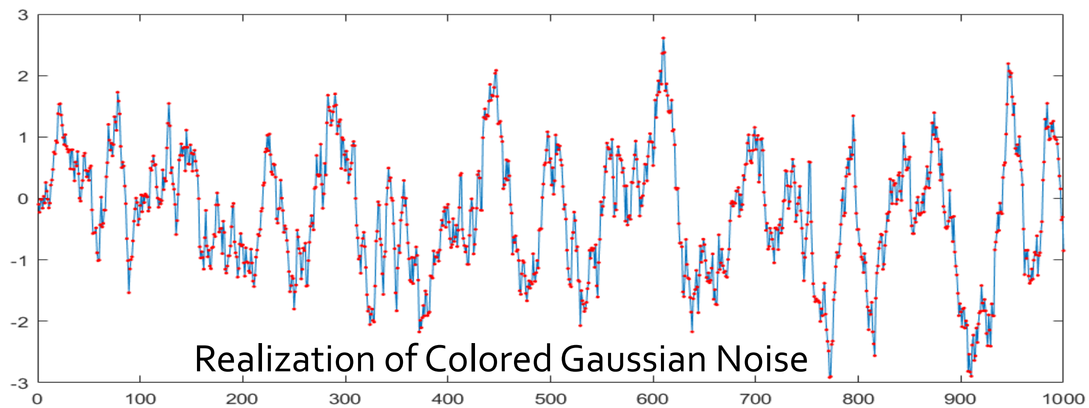
- Definition: The joint pdf of **any** subsequence of the stochastic process is a multivariate normal pdf

$$
\begin{aligned}
p_{\bar{X}}(\bar{x}) &= \frac{1}{(2\pi)^{N/2}|\boldsymbol{C}|^{1/2}} \exp(-\frac{1}{2}(\bar{x}-\bar{\mu})\boldsymbol{C}^{-1}(\bar{x}-\bar{\mu})^T) \\
&= \frac{1}{(2\pi)^{N/2}|\boldsymbol{C}|^{1/2}} \exp(-\frac{1}{2}\langle \bar{x}-\bar{\mu}, \bar{x}-\bar{\mu}\rangle) \\
&= \frac{1}{(2\pi)^{N/2}|\boldsymbol{C}|^{1/2}} \exp(-\frac{1}{2}\|\bar{x}-\bar{\mu}\|^2)
\end{aligned}
$$

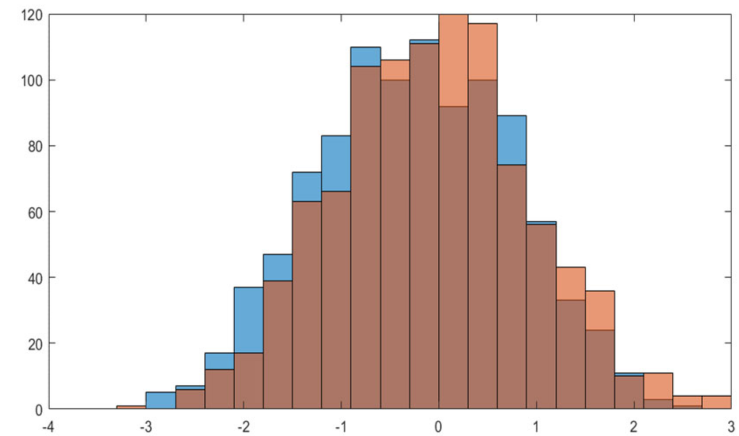
- $\bar{x} = (x_0, x_1, \ldots, x_{N-1}) \in R^N$ (row vector)
- $\bar{\mu} = (\mu_0, \mu_1, \ldots, \mu_{N-1}) \in R^N$ (row vector)
- $E[X_i] = \mu_i$
- $C_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$: Covariance matrix
- $|\boldsymbol{C}|$ : Determinant of $\boldsymbol{C}$
- $\langle \bar{x},\ \bar{y}\rangle = \ \bar{x}\boldsymbol{C}^{-1}\bar{y}^T$ is an inner product (“dot product”) between vectors
- $\|\bar{x}\|^2 = \langle \bar{x}, \bar{x}\rangle$ is the squared norm (“squared length”) of a vector

# 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



# Exercise

- The script DATASCIENCE_COURSE / NOISE / simplecolnoise.mlx demonstrates the simulation of colored noise
- Alter the relevant parameters involved and generate noise realizations with different PSDs