Data analysis

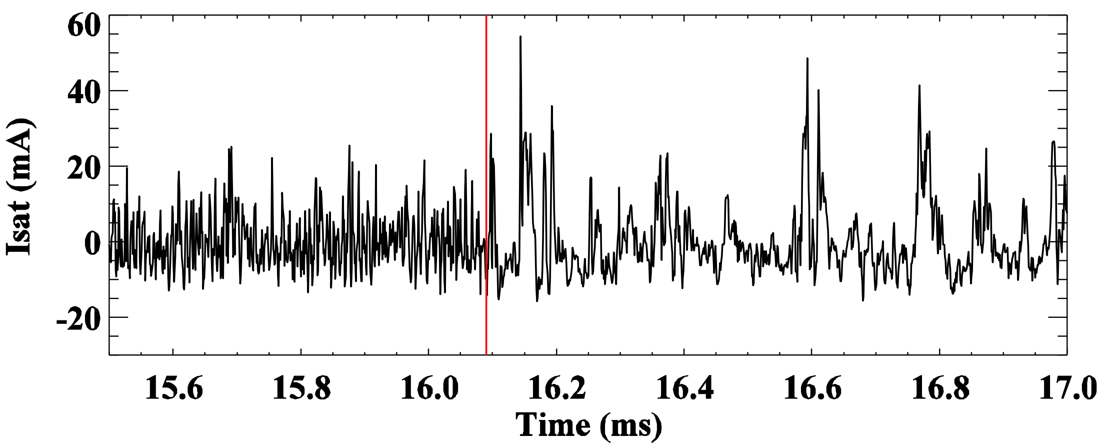
Instituto de Plasmas e Fusão Nuclear
Instituto Superior Técnico
Lisbon, Portugal
http://www.ipfn.tecnico.ulisboa.pt

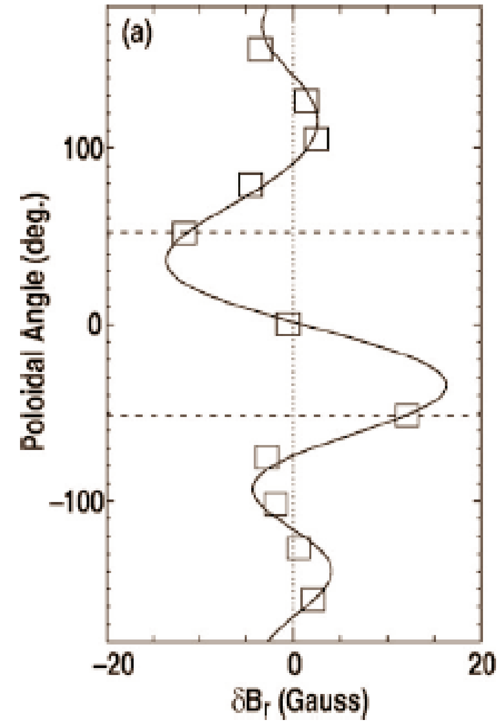
Signal

Any quantity varying in time (or other independent variable, ex. space).
Electrical signal are the most frequent

- ❑ Continuous – Analogic
- ❑ Discrete - Digital

Data analysis: spectral analysis, correlation

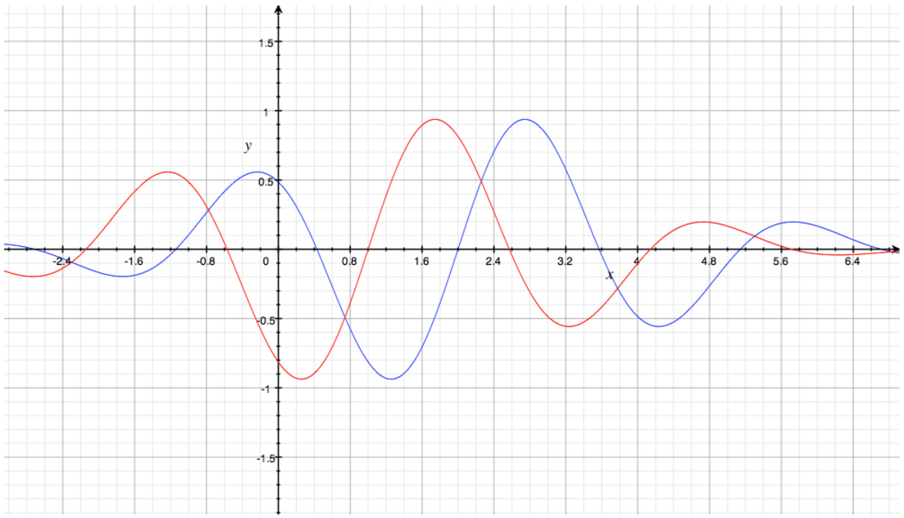


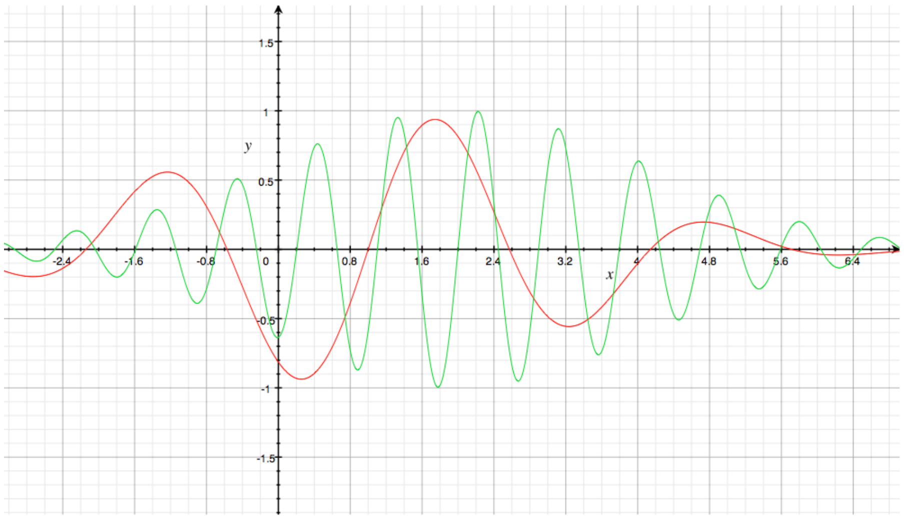


Analysis in time (frequency) and space (wavelength)

Cross-correlation

Correlation ←→ Similarity between two sets of data





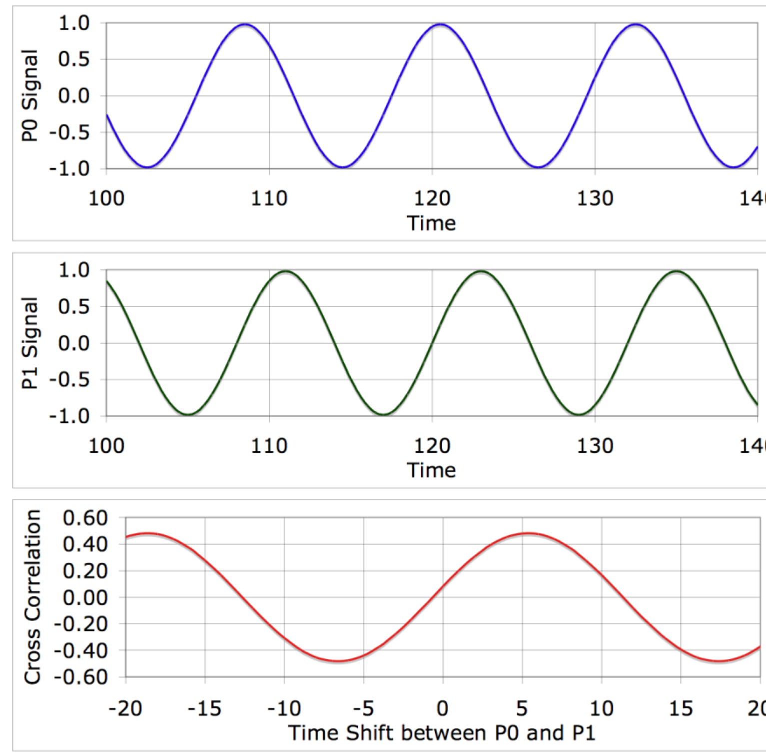
- ❑ Useful to find trends for physics studies or propagation of events
- ❑ Correlation vs causality

Cross-correlation

$$(f \star g)(\tau) = \int_{-\infty}^{+\infty} f^*(t)\, g(t+\tau)\, dt$$

Sliding inner product between two functions (continuous)

- Maximum when functions have similar, synchronous pattern
- Minimum when functions have similar but out-of-phase pattern
- Zero for random signals
- Accounts for time shifts in datasets
- Periodic



Correlations as a function of the time lag

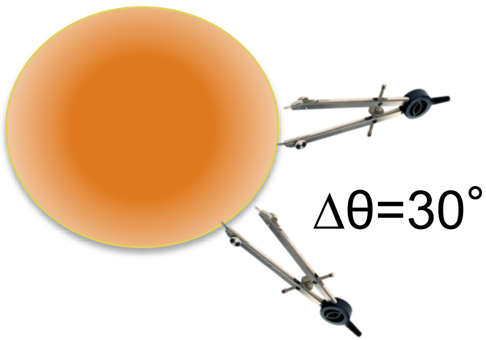
Cross-correlation - discrete functions

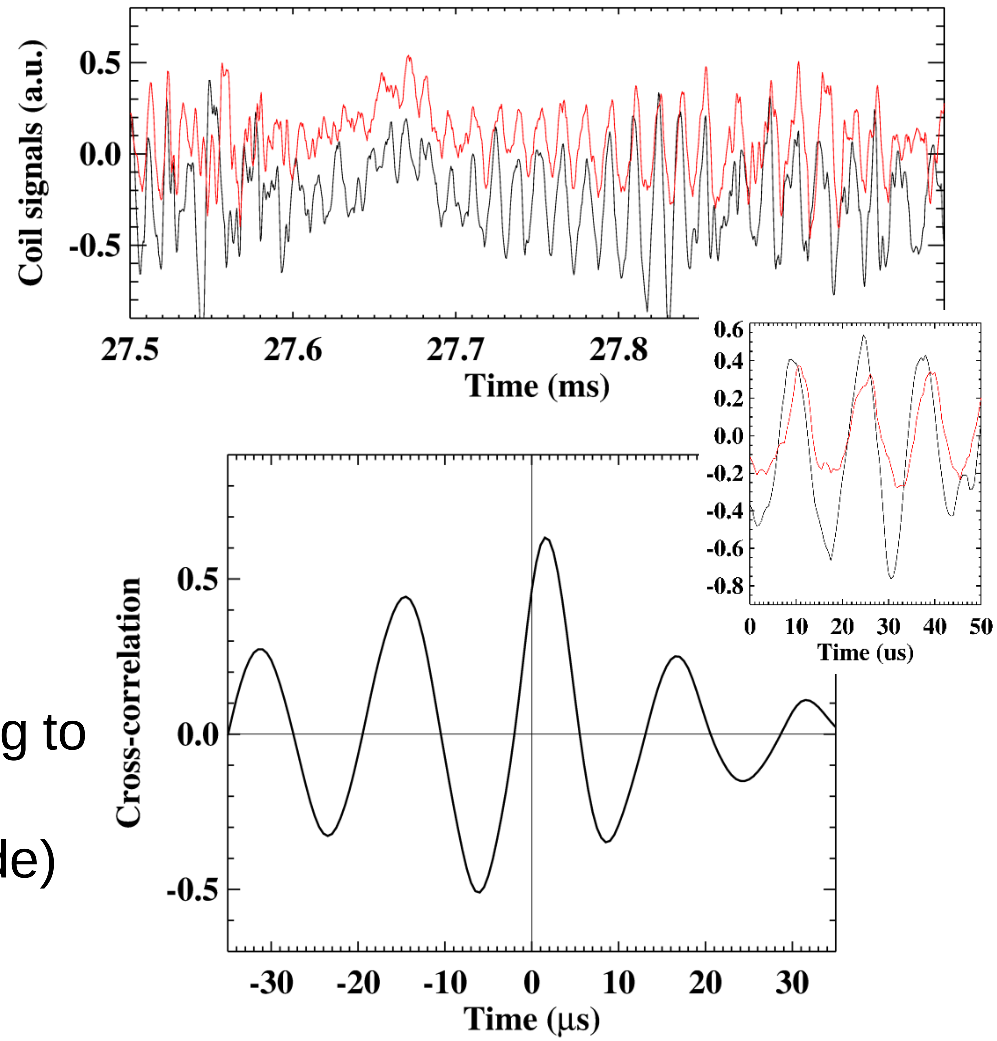
Cross-correlation of “real life” finite sampled signals

- Real life signals have only limited number of samples
- Cross-correlation easier to interpret when bounded in [-1,1]
- Lag vector (index-n) finite, integer

$$P_{xy}(L) = \begin{cases} \dfrac{\displaystyle\sum_{k=0}^{N-|L|-1} (x_{k+|L|} - \bar{x})(y_k - \bar{y})}{\sqrt{\left[\displaystyle\sum_{k=0}^{N-1} (x_k - \bar{x})^2\right]\left[\displaystyle\sum_{k=0}^{N-1} (y_k - \bar{y})^2\right]}} & \text{For } L < 0 \\[4ex] \dfrac{\displaystyle\sum_{k=0}^{N-L-1} (x_k - \bar{x})(y_{k+L} - \bar{y})}{\sqrt{\left[\displaystyle\sum_{k=0}^{N-1} (x_k - \bar{x})^2\right]\left[\displaystyle\sum_{k=0}^{N-1} (y_k - \bar{y})^2\right]}} & \text{For } L \geq 0 \end{cases}$$

Cross-correlation – time delay

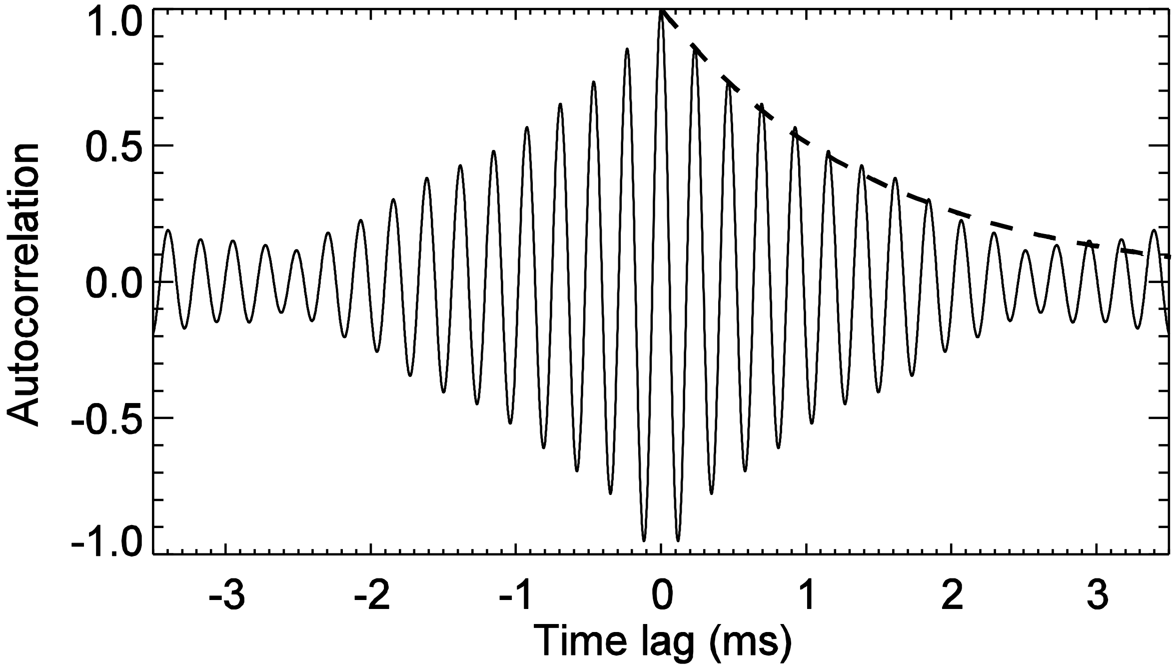




- ❑ Maximum value (similarity) + corresponding delay (time shift)
- ❑ Time delay analysis: $v = \Delta x/\Delta t$
- ❑ Oscillation (period corresponding to dominant frequency) + decay (noise, finite duration of the mode)
- ❑ Useful when signals have a dominant frequency

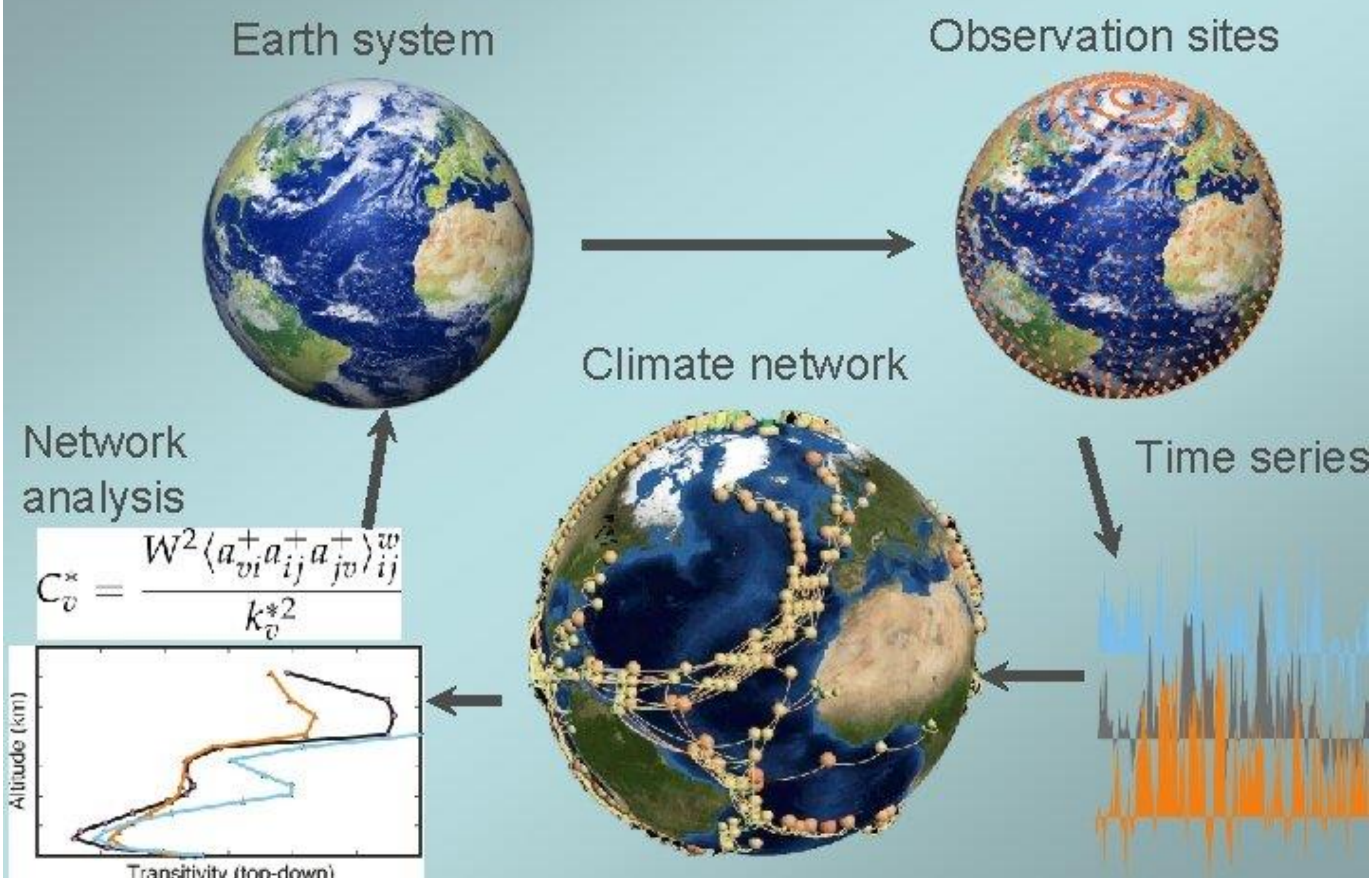
Cross-correlation

- Oscillation (period corresponding dominant frequency) + decay

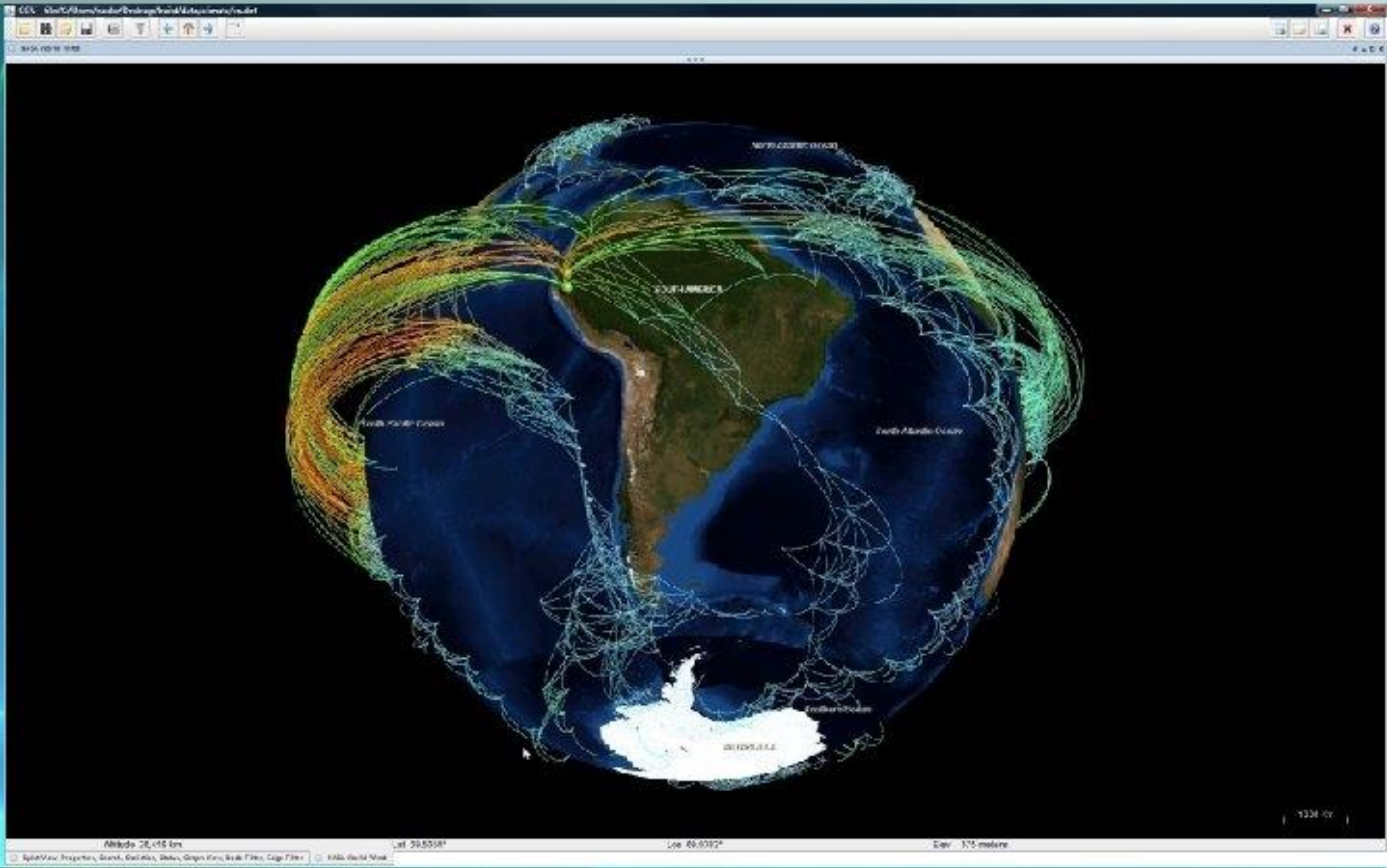


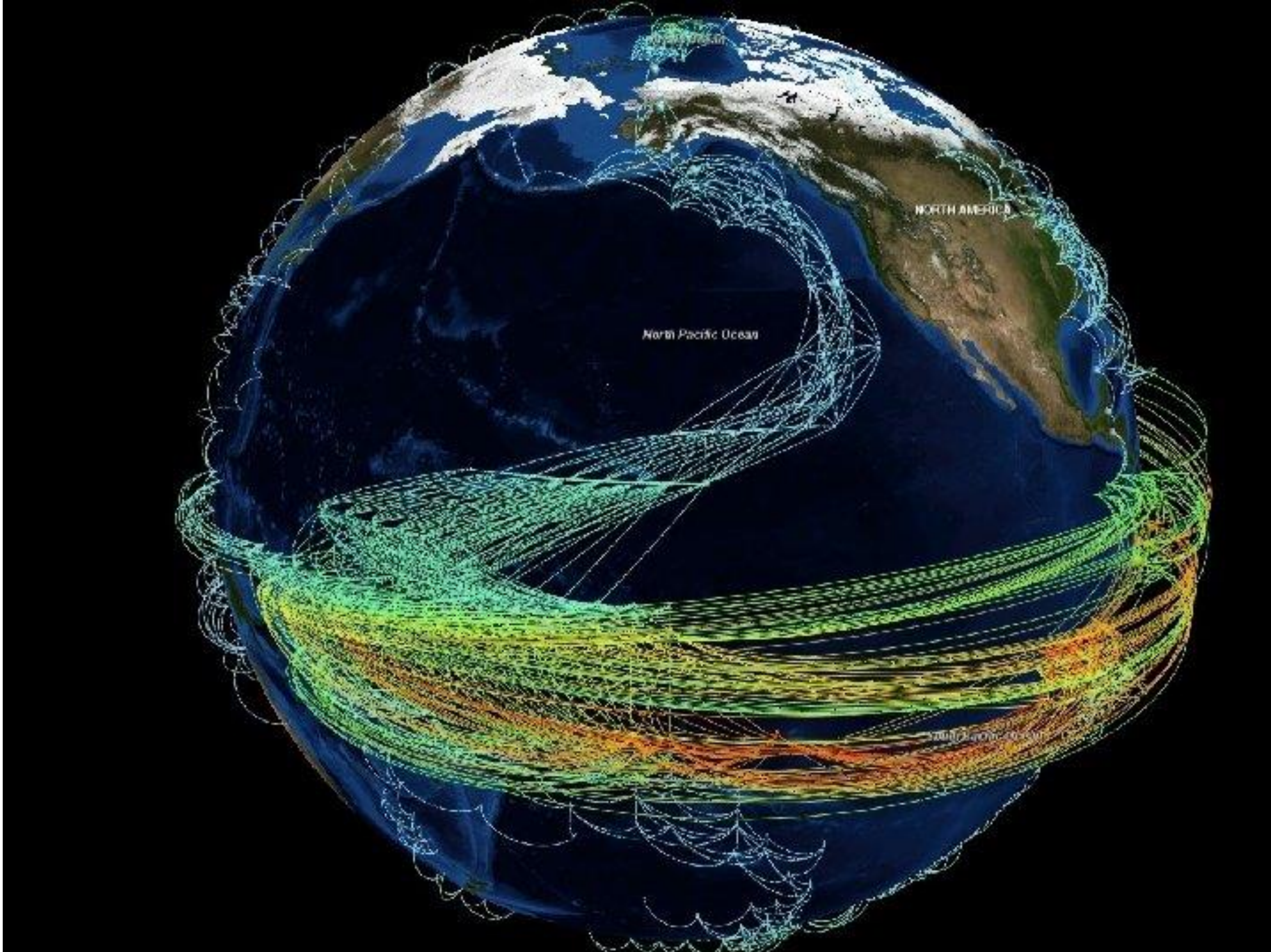
The Nobel Prize in Physics 2021 was awarded to Syukuro Manabe, Klaus Hasselmann, and Giorgio Parisi for their “groundbreaking contributions to our understanding of complex systems,” including major advances in the understanding of our climate and climate change.

Climate Networks



Complex network approach to climate system




NORTH AMERICA
North Pacific Ocean

Visual Analytics tools

temperature climate networks
scalable for > 100.000 edges
graphics card implementation

2D node layout (360 degree circular projection) avoiding edge clutter at the equator

Thomas Nocke, PIK

https://slidetodoc.com/climate-networks-extreme-events-jrgen-kurths-potsdam-institute/

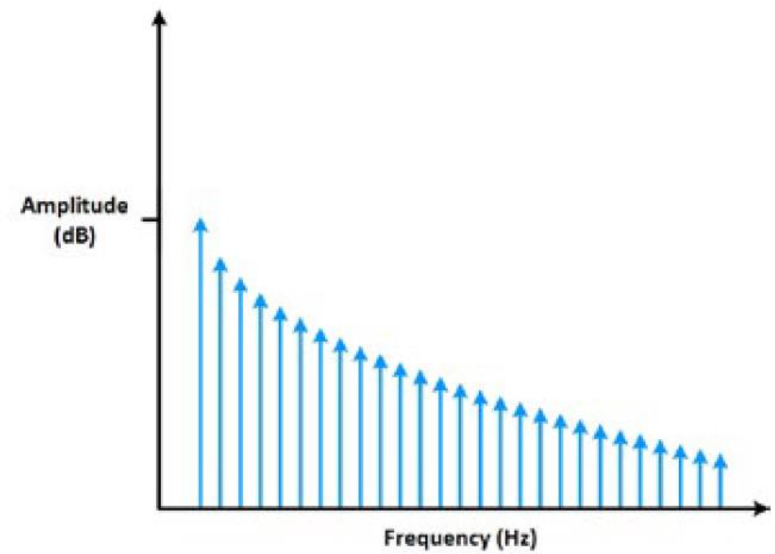
Fourier analysis - basics

- Since a set of $\cos(\omega_i t)$ or $\sin(\omega_i t)$ form an orthogonal basis, any signal can be decomposed on such a basis. We obtain the projections of a function s(t) on this basis…. Amplitude of the Fourier components

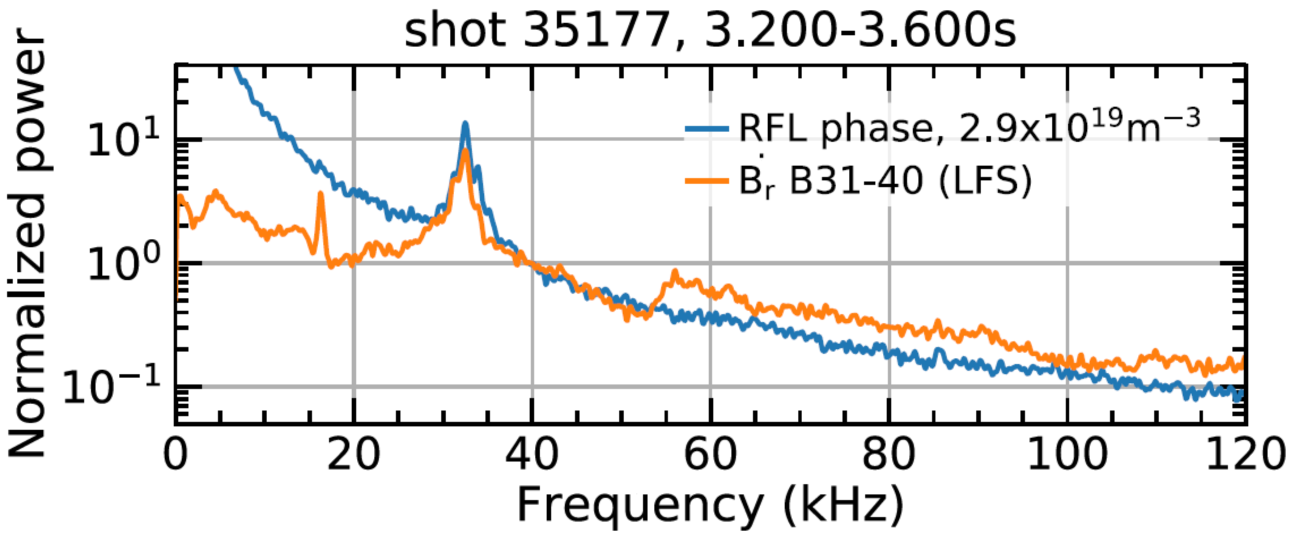
$$S(m) = \sum_{n=1}^{N} s[n] e^{-i2\pi f[m]t[n]} = \sum_{n=1}^{N} s[n] e^{-i2\pi \frac{m}{N} n} \qquad \text{DFT}$$

where $f_s = 1/\Delta t$, m=0..N/2 and $f(m)=m/N\Delta t$ and f(N/2) is the Nyquist frequency ($f_s/2$)
Frequency resolution: $\Delta f = 1/N\Delta t$
Frequency: $0 – 1/2\Delta t$

In practice…the *Fast Fourier Transform (FFT)* is the best implementation



Frequency spectra



$f_s = 1/\Delta t$, Nyquist frequency ($f_s/2$)

Frequency resolution: $\Delta f = 1/N\Delta t$

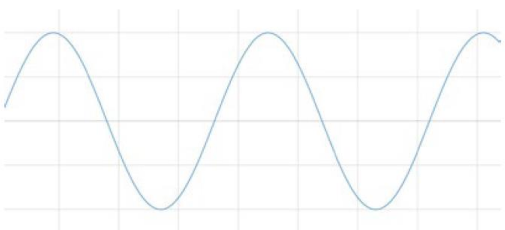
Frequency: $0 – 1/2\Delta t$

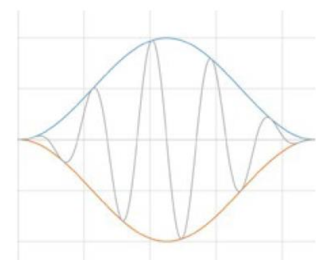
Power spectra: FFT(X) x FFT(X)*

Windowing

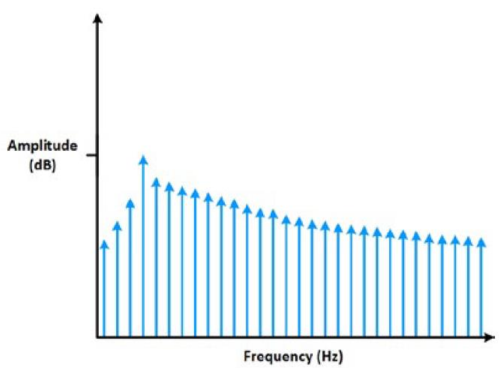
- Windowing of a signal causes its Fourier transform to develop non-zero values (commonly called spectral leakage) at frequencies other than ω
- Effects can be minimized by using a technique called windowing. Windowing reduces the amplitude of the discontinuities at the boundaries by multiplying the time series by a window with an amplitude that varies toward zero at the edges
- Hanning window is generally used

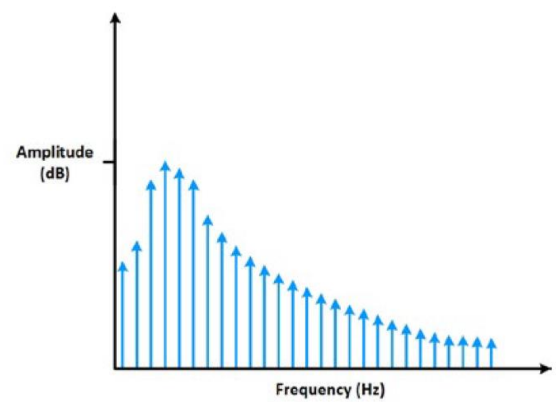
A





B

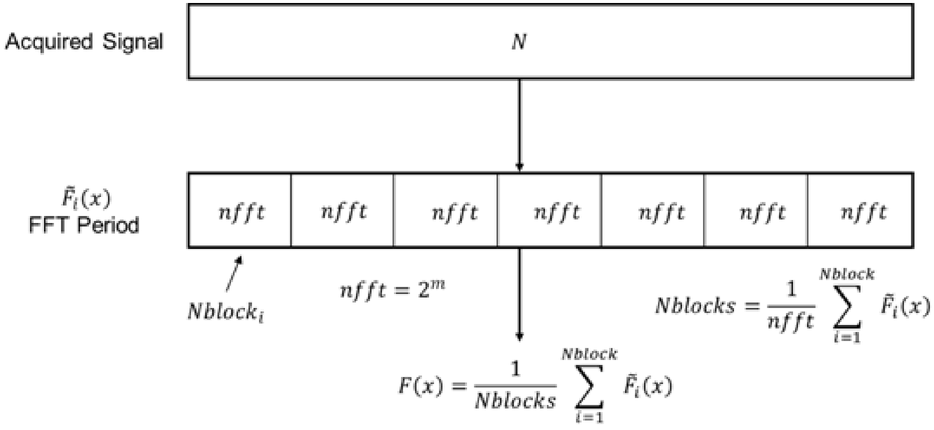


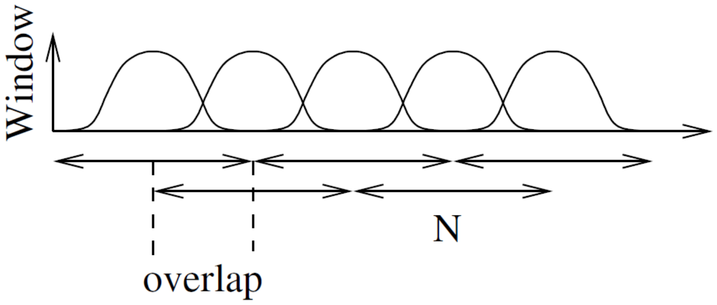


Welch method

Individual power spectra have an associated variance, estimate can be improved

- The signal is split up into overlapping segments (typical overlap of 50%)
- Segments are windowed (overlap mitigate lower influence of points at the edge)
- Compute the discrete Fourier transform and power spectra
- The individual power spectra are then averaged, reducing the variance of the individual power measurements
- Frequency resolution decreases for the same window size

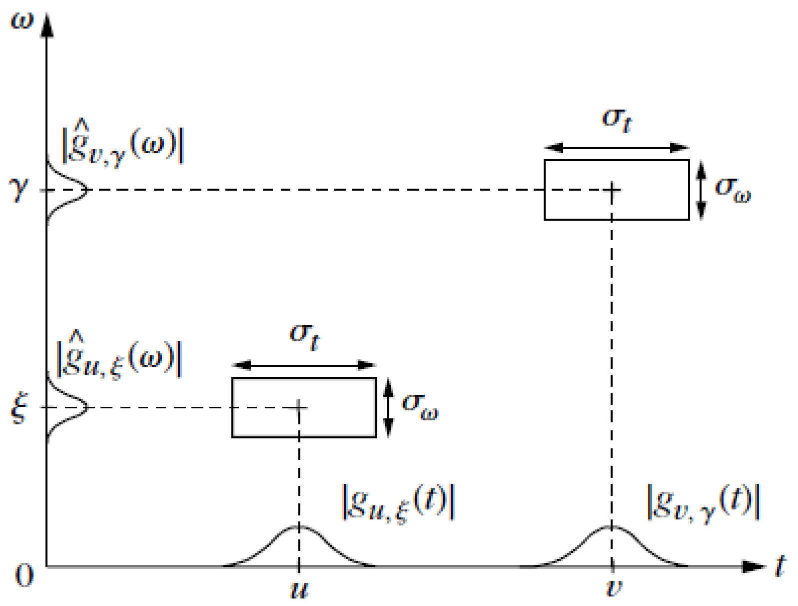


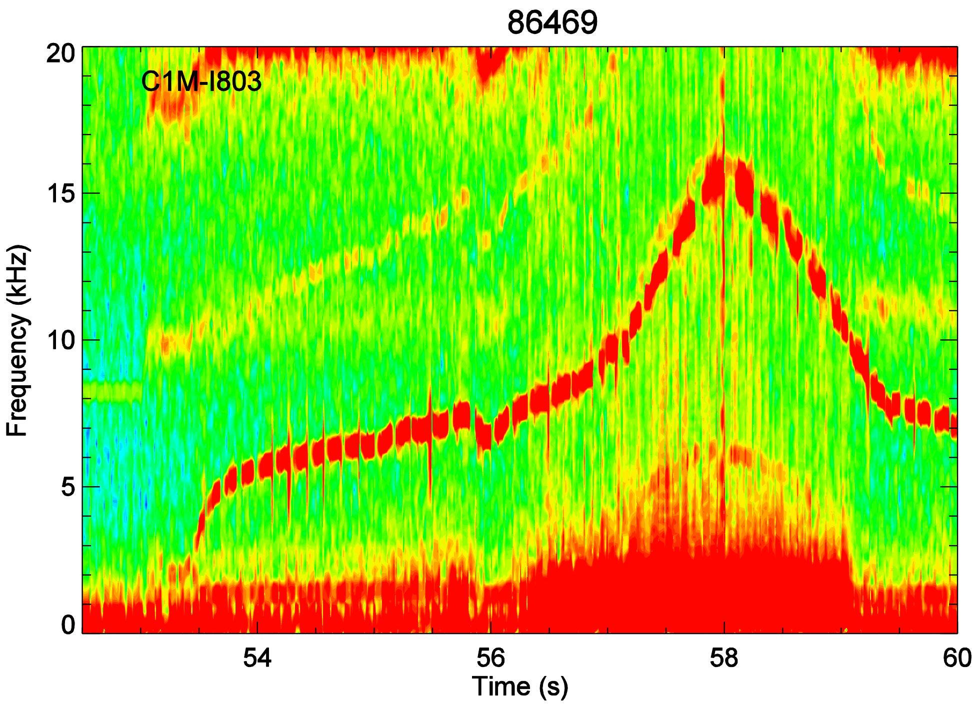


Requires stationary signals

Spectrogram

Evolution of the frequency components of a signal in time: non-stationary signals





Frequency resolution : Δf = 1/NΔt

Temporal resolution : ΔT = NΔt

Balance between time and frequency resolution

Example of a sin
with variable
frequency

Home assignment

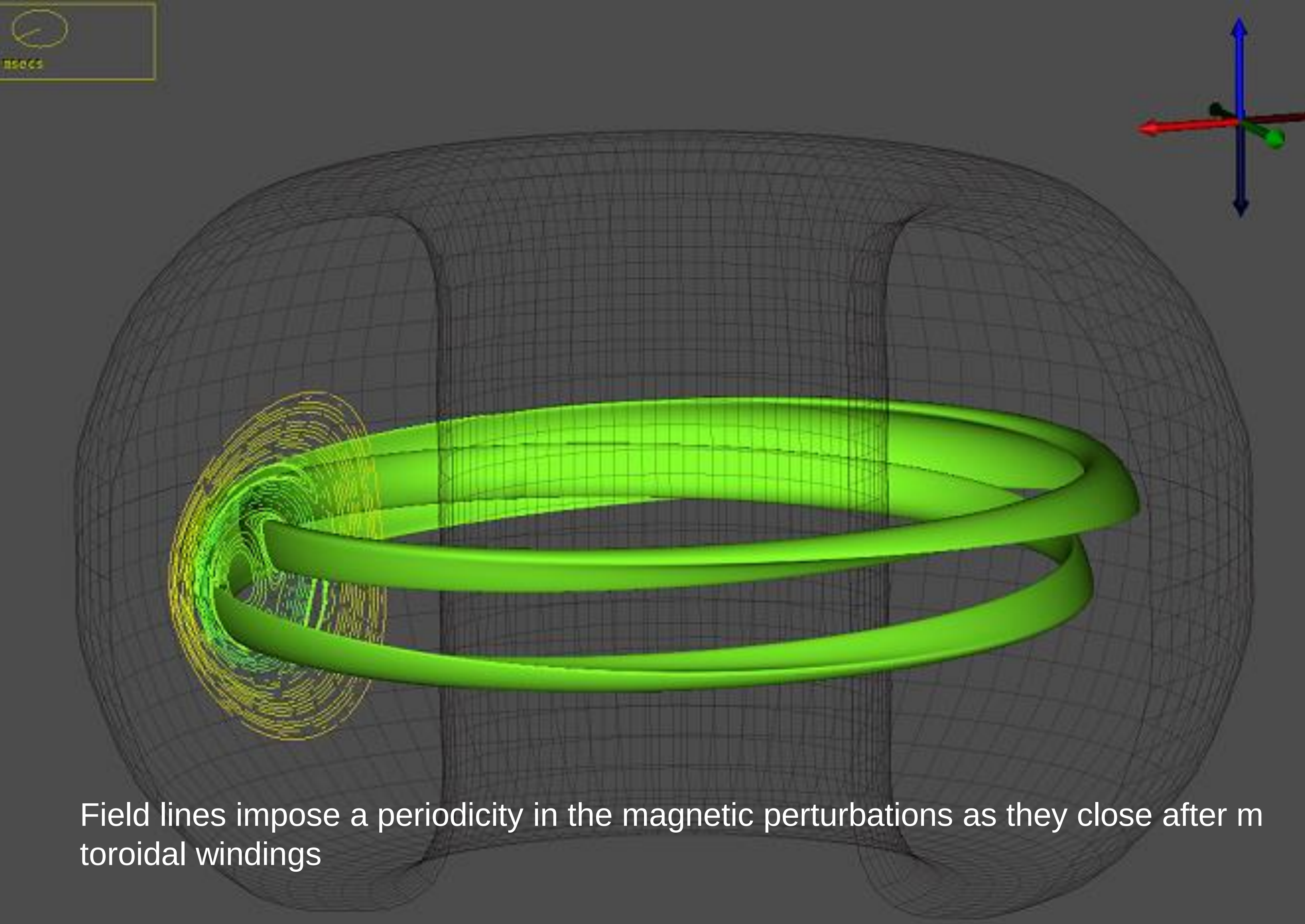
Analysis of magnetic coil signals from ISTTOK

- Spectrogram and power spectra for one coil
- Determine frequency and poloidal mode number of the magnetic instabilities as well as the plasma poloidal rotation velocity (including direction) using cross-correlation

Steps

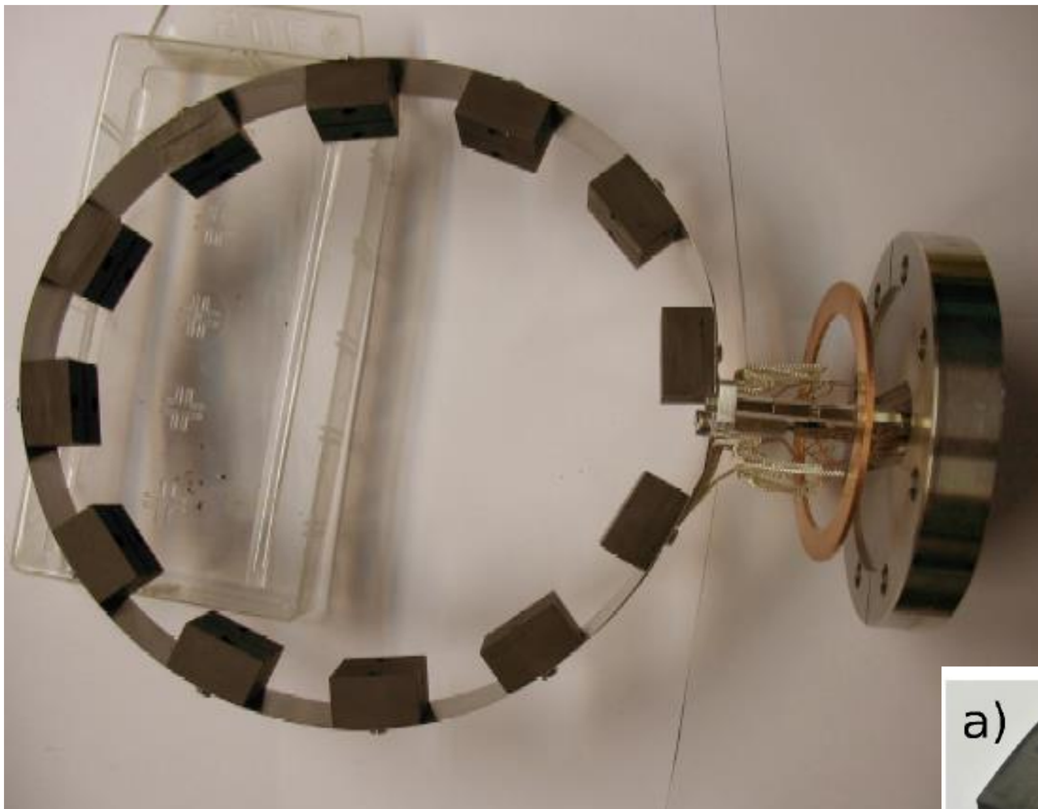
1. Estimate the spectrogram to characterize the evolution of the signal and to identity a suitable period for the analysis
2. Estimate the power spectra in the selected period using the Welch method
3. Compute the cross-correlation between coils to determine the rotation velocity and mode number

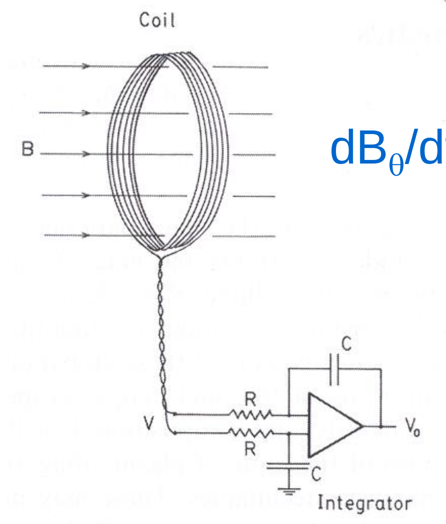
Shots: 35054, 35055, 35056, 35057, 35058, 35059, 35062



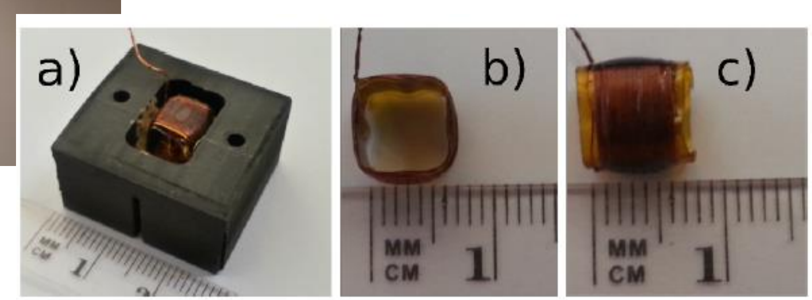
Field lines impose a periodicity in the magnetic perturbations as they close after m toroidal windings

Magnetic probes on ISTTOK





dB_θ/dt (a.u.)

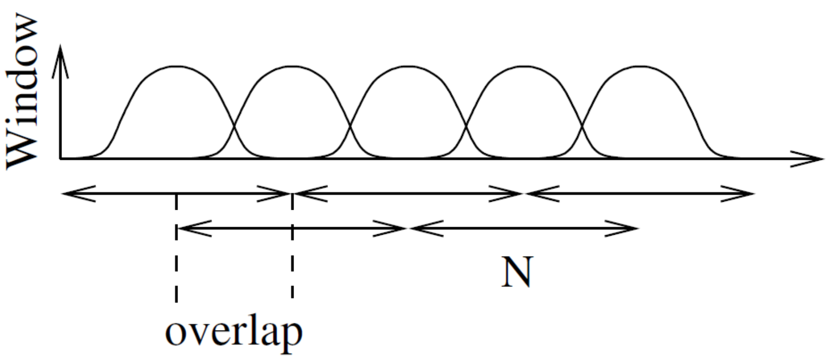


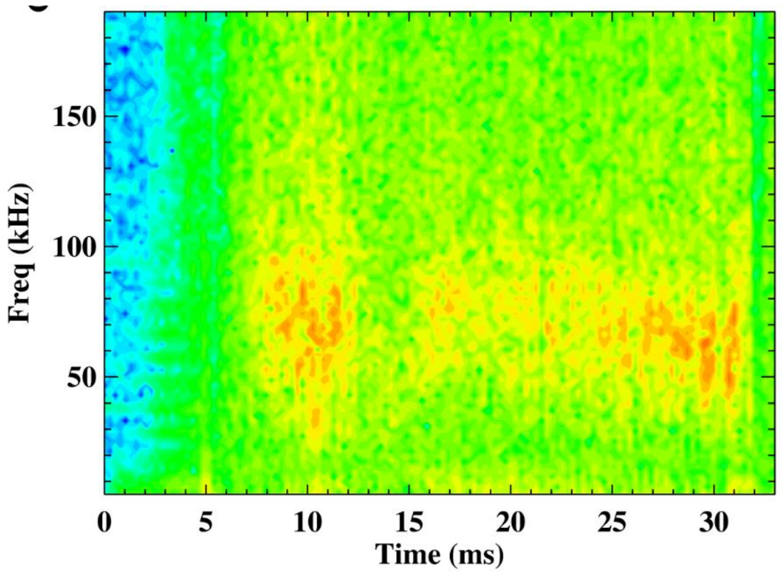
Spectrogram – one coil

- The signal is split up into overlapping segments (typical overlap of 50%)
- Segments are windowed to reduce leakage (overlap mitigate lower influence of points at the edge)
- Compute the discrete Fourier transform and power spectra
- Spectrogram function available
- Important to understand the method to be applied properly

  N, overlap, windowed?

- Try different combinations, test the limits
- Reference: 2k points, ΔT=1 ms, Δf=1 kHz

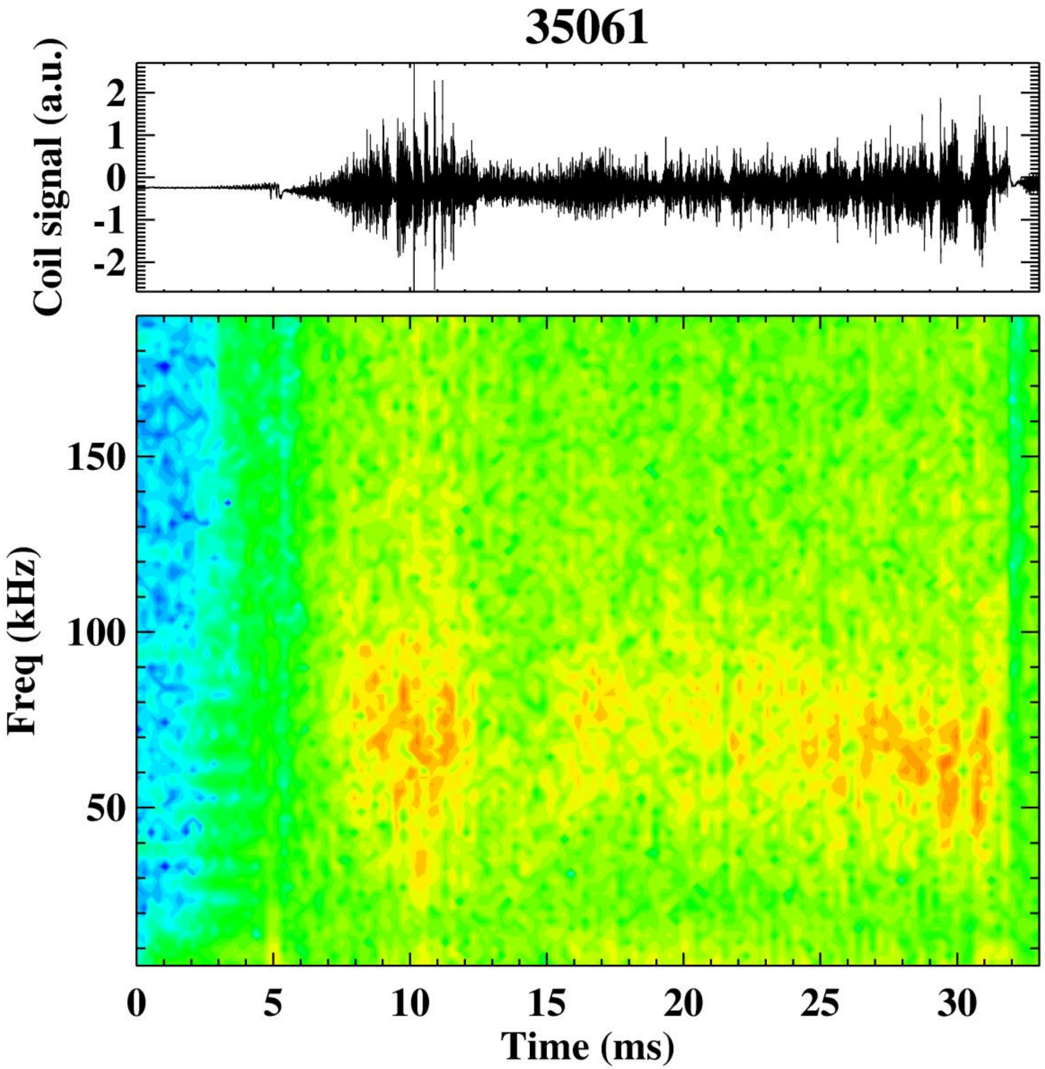




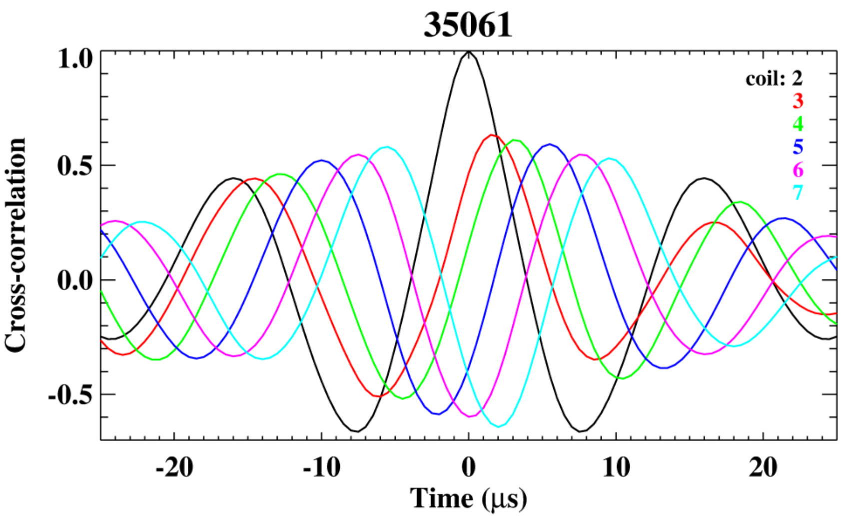
Select time window to be used in the power spectra and cross-correlation

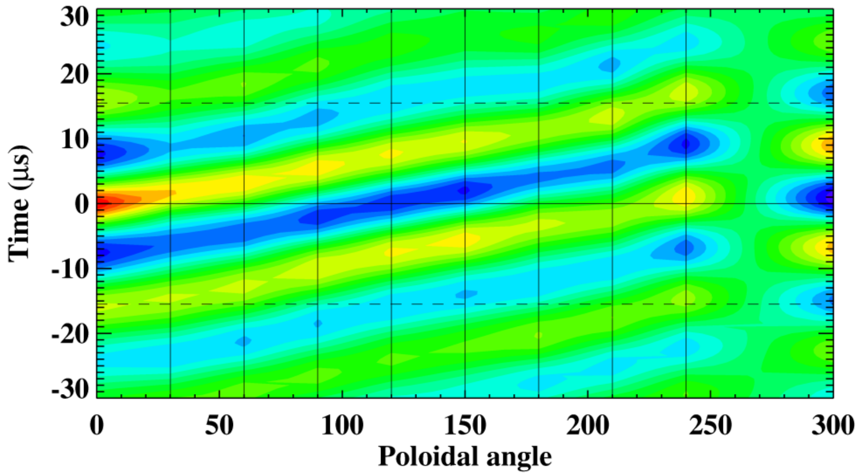
For one coil:

- Look for modes in the spectrogram
- Time window: good statistics ↔ steady state period: <5 ms
- Welch power spectra: ~ 5 ms
- Correlation: 1 – 3 ms



Poloidal structure of the magnetic fluctuations : cross-correlation





- Probe #*1*, 11 not operational
- Window: 1-3 ms
- Lag ~±100 points
- Attention to the sign convection (test with synthetic signal)
- Plot: contour or probe position vs time delay
- If plasma not centered, measured structure and velocity not poloidally constant
- Justify different analysis steps and decisions and indicate analysis details (shot number, time interval, input parameters of the functions used, …)