



A Wavelet Based Analysis System for Monitoring Information



ALICE

Iosif Legrand
Caltech - CERN

ACAT 2013 Beijing



Monitoring Information is necessary for System Design, Control, Optimization, Debugging and Accounting

**Computing Models
Modeling & Simulations**

**Optimization
Algorithms**



MONITORING

**~ REAL TIME
Information**

ACCOUNTING

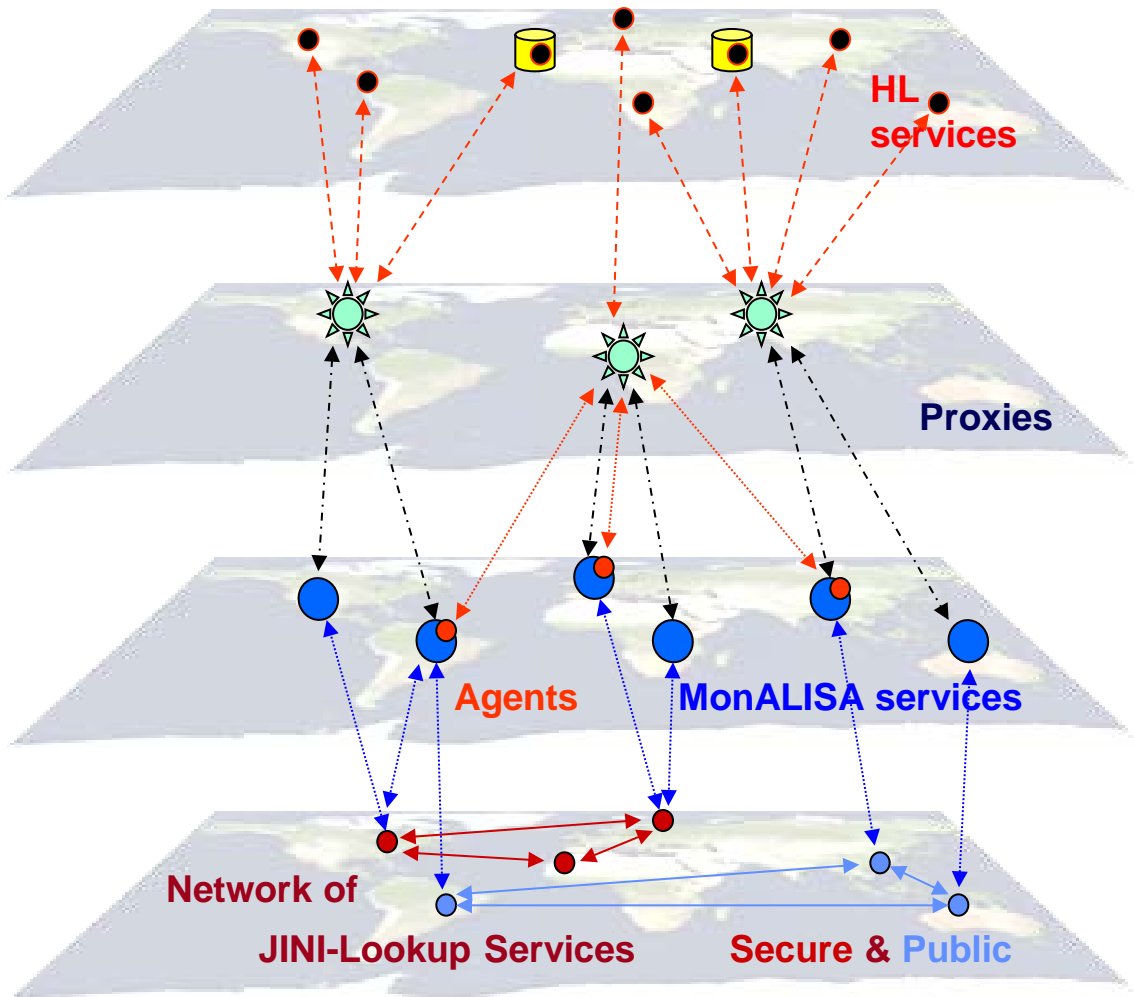
**Create resilient
Distributed Systems**

**Control and
Operational support
ALARMS**

DEBUGGING



The MonALISA Architecture



Regional or Global High Level Services, Repositories & Clients

**Secure and reliable communication
Dynamic load balancing
Scalability & Replication
AAA for Clients**

**Distributed System for gathering and analyzing information based on mobile agents:
Customized aggregation, Triggers, Actions**

Distributed Dynamic Registration and Discovery-based on a lease mechanism and remote events

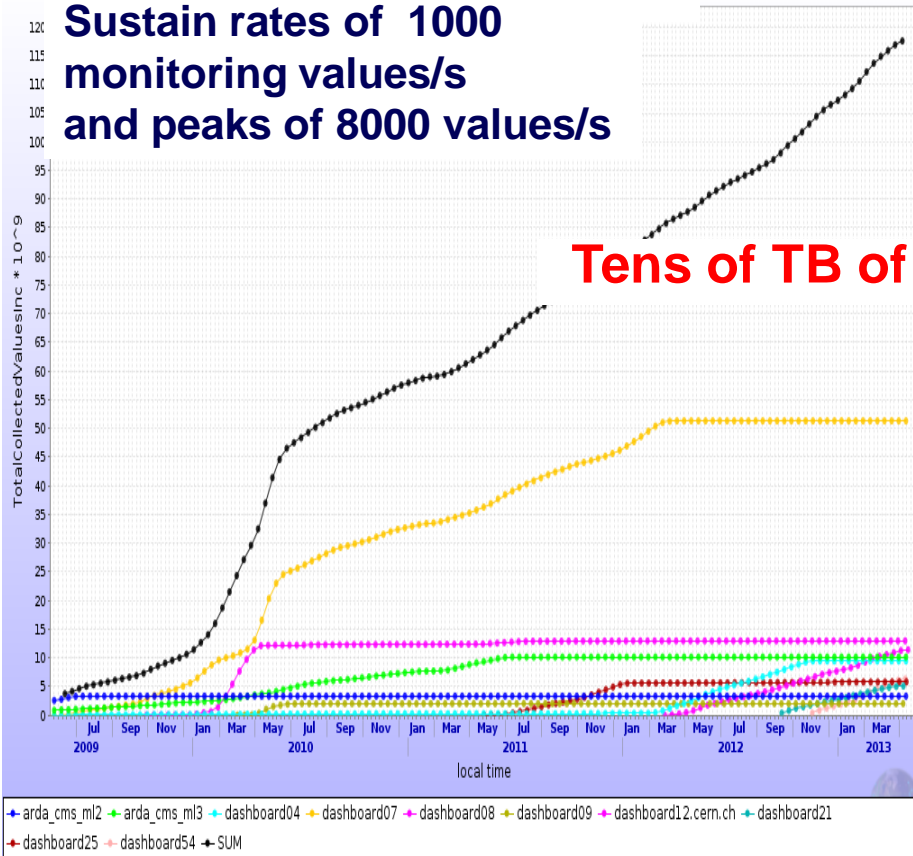
Fully Distributed System with no Single Point of Failure



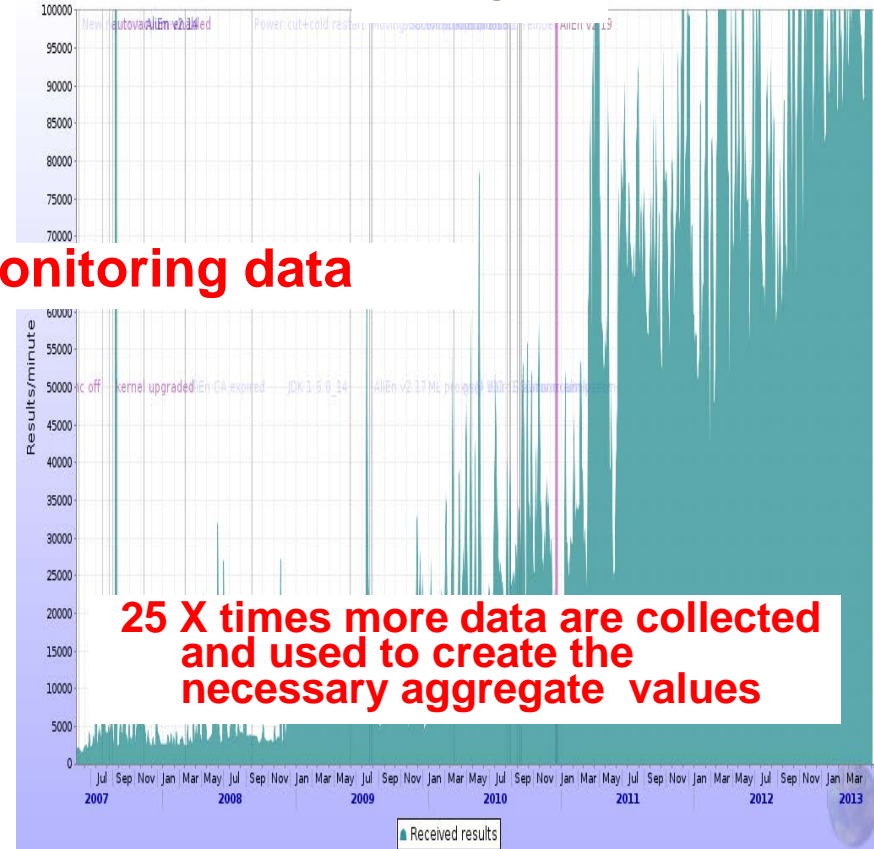
Collecting and Storing Very Large Amounts of Monitoring Information

CMS

Sustain rates of 1000 monitoring values/s and peaks of 8000 values/s

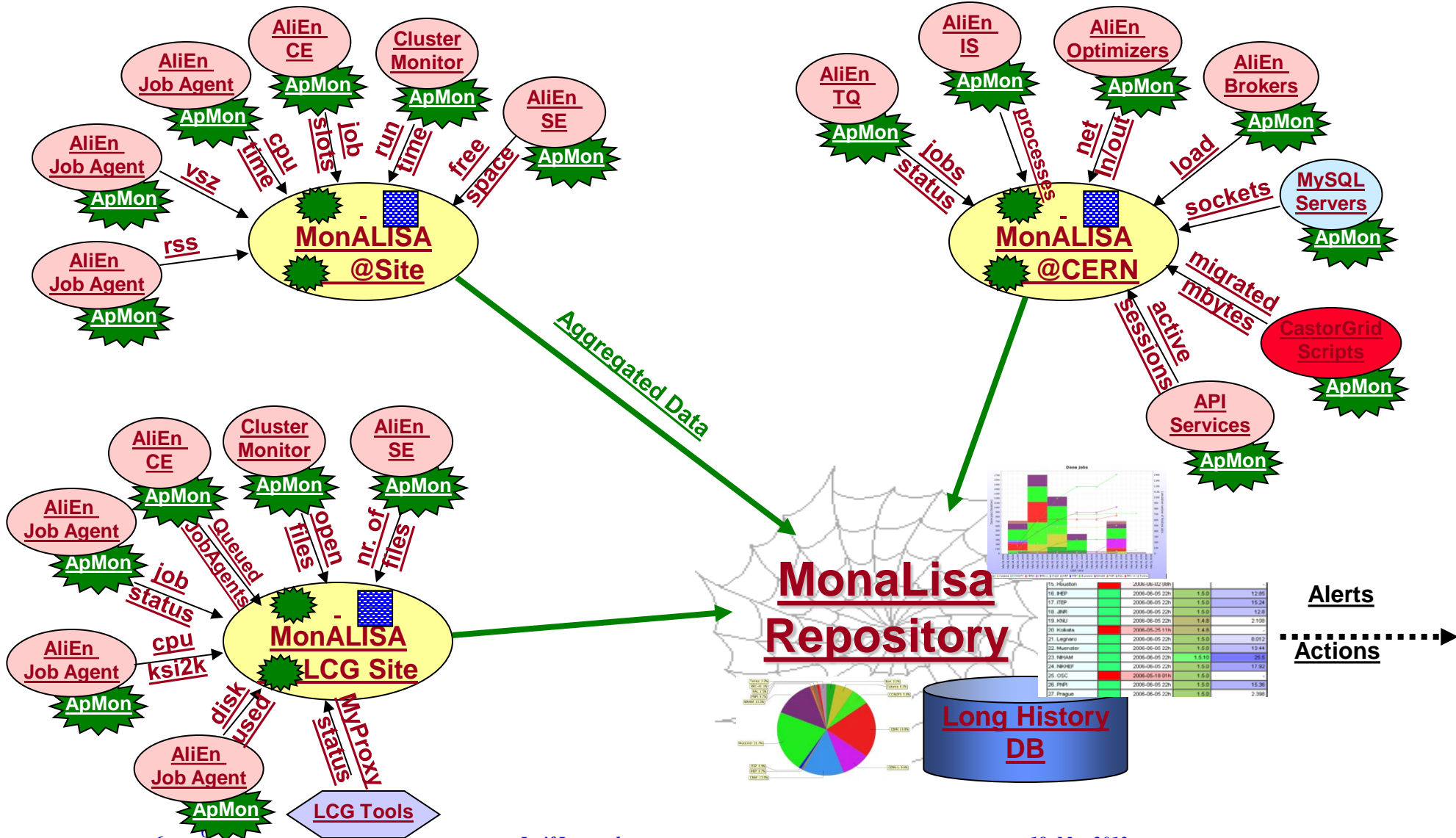


ALICE





Current Monitoring architecture in ALICE





Trying to find a better way to store and analyze large sets of monitoring information

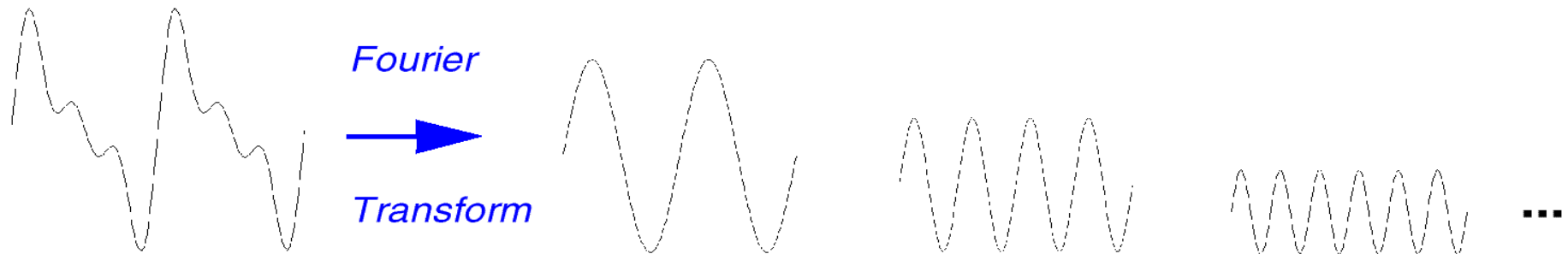
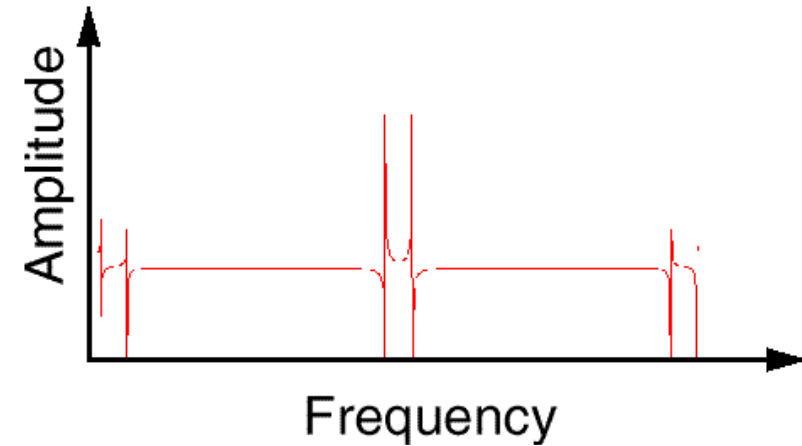
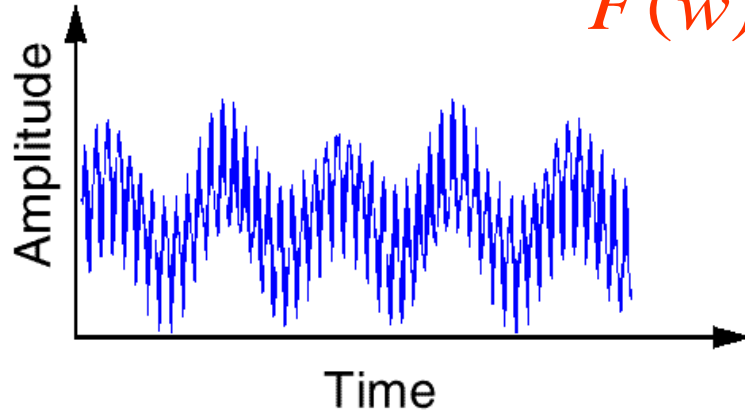
- **Very large amount of monitoring information is currently collected.**
- **The users want more and more monitoring information, but is really difficult to analyze all the data we collect.**
- **Deleting older data or keeping only long term mediated values is not really a solution.**
- **Wavelets seems to provide an effective way to compress monitoring information and to analyze large, complex time series data.**



The Fourier Transform

Represents a signal into constituent sinusoids of different frequencies

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$



However it does not indicate when different frequencies occur



THE WAVELET TRANSFORM

$$\text{CWT}_x^\Psi(\tau, s) = \Psi_x^\Psi(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t) \bullet \Psi^* \left(\frac{t - \tau}{s} \right) dt$$

Translation

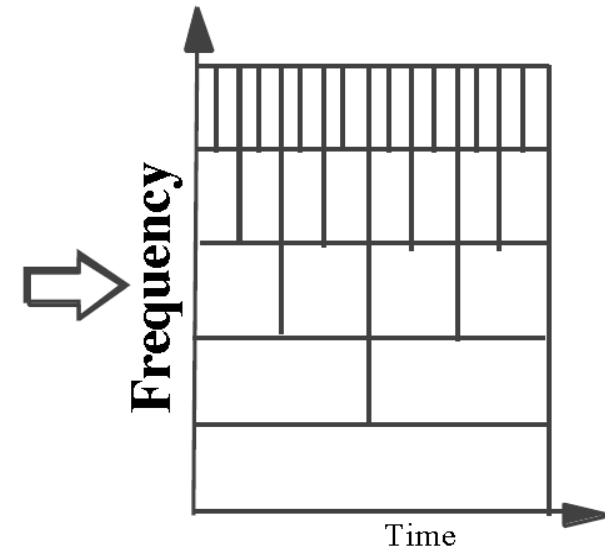
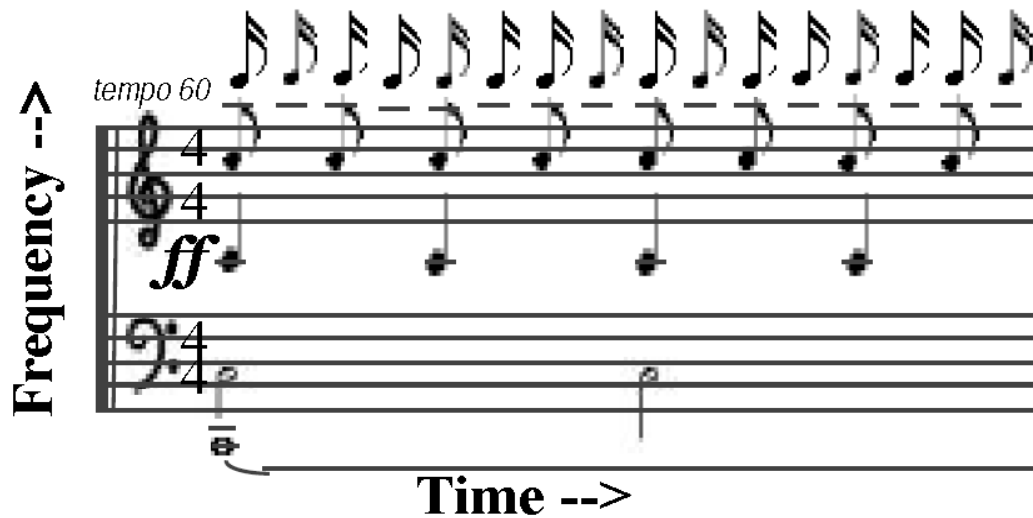
(The time location of the window)

Scale

Mother Wavelet

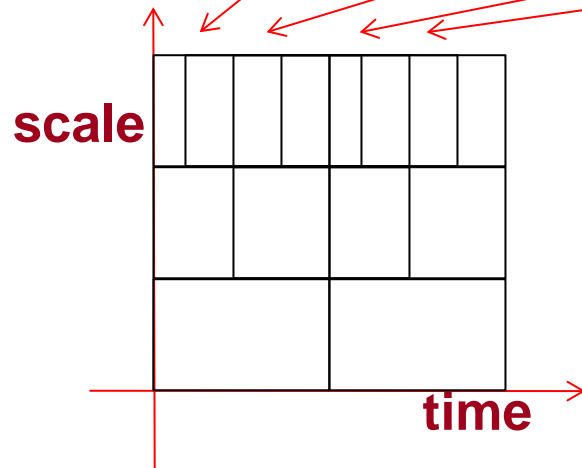
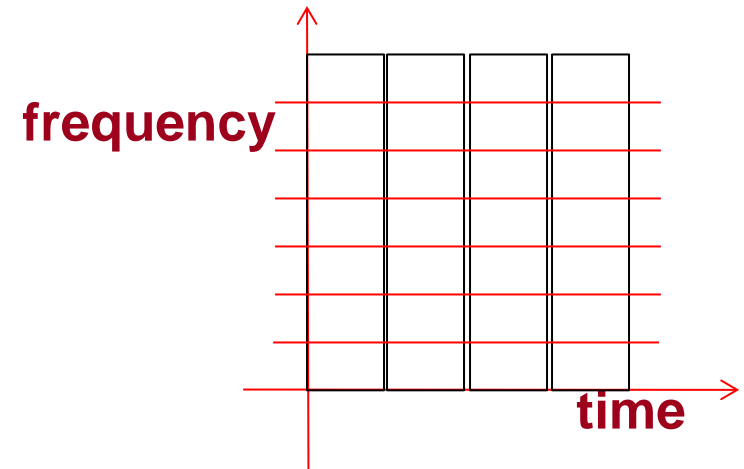
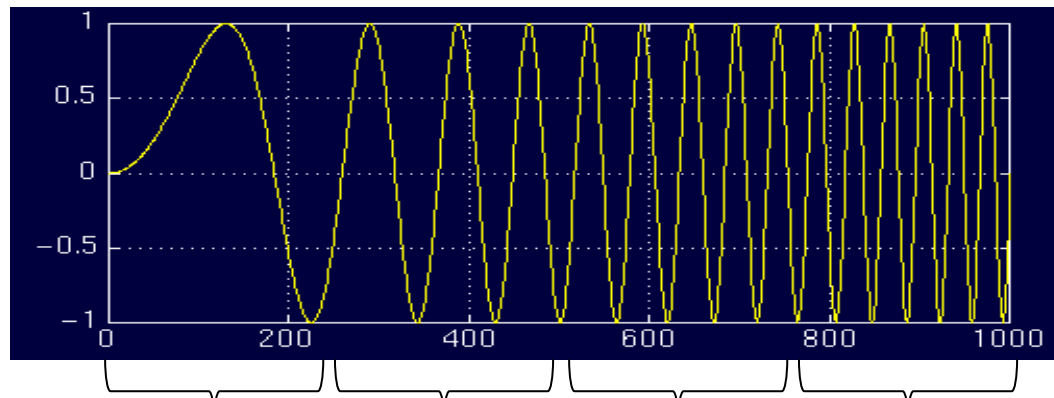


Scale is generalized local frequency





Wavelet Time frequency analysis



Wavelet is Small wave

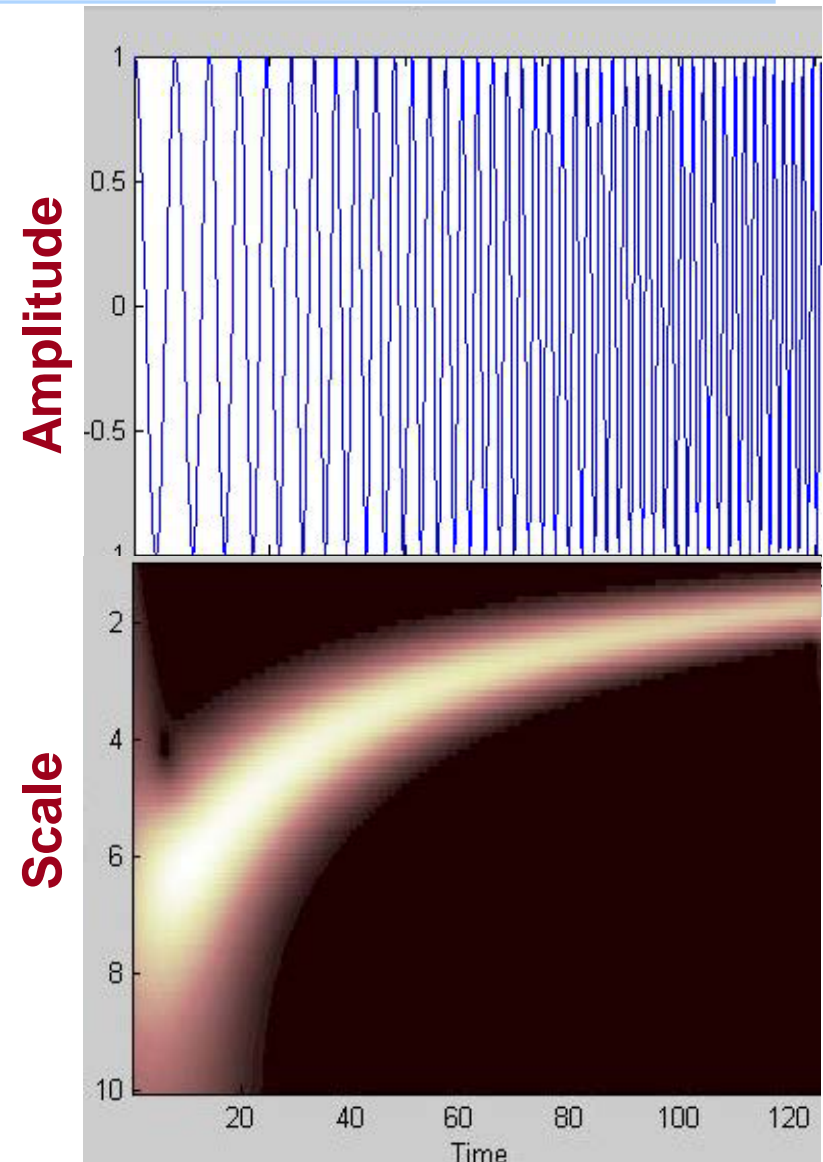
Means the window function is of finite length

All the used windows are its dilated or compressed and shifted versions



Wavelet Transform

- Provides the time-frequency representation
- Capable of providing the time and frequency information simultaneously
- WT was developed to overcome some resolution related problems of the STFT
- We pass the time-domain signal from various highpass and low pass filters, which filters out either high frequency or low frequency portions of the signal. This procedure is repeated, every time some portion of the signal corresponding to some frequencies being removed from the signal

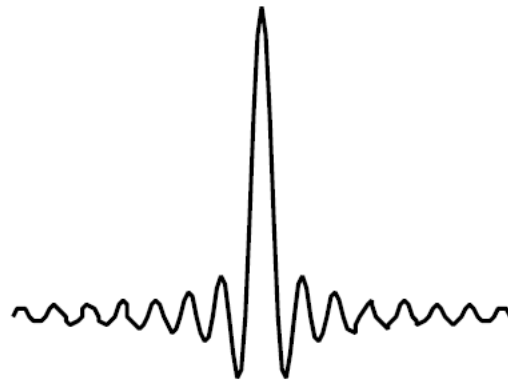




Several Wavelet Functions



Haar



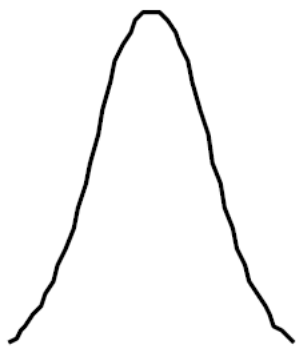
Shannon or Sinc



Daubechies 4



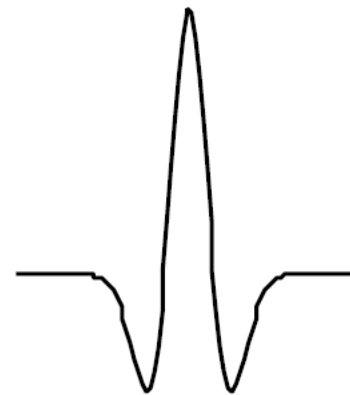
Daubechies 20



Gaussian or Spline



Biorthogonal



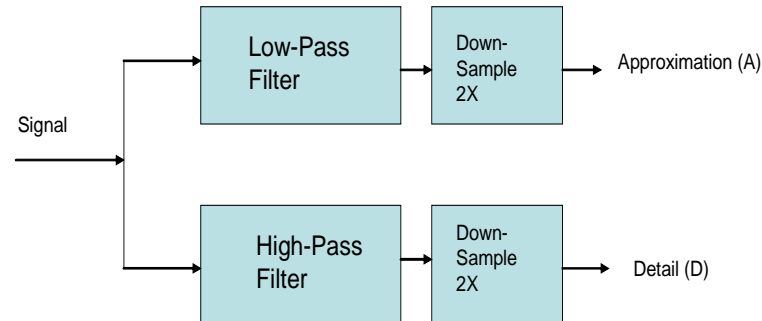
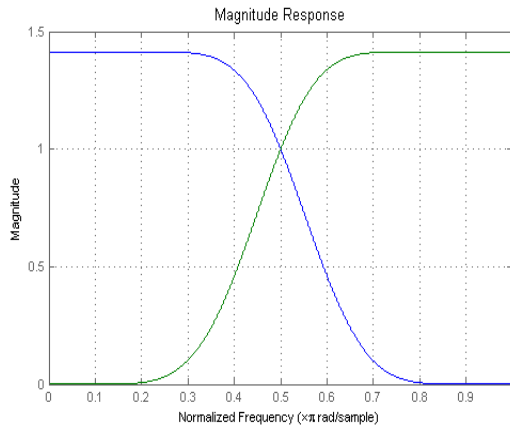
Mexican Hat



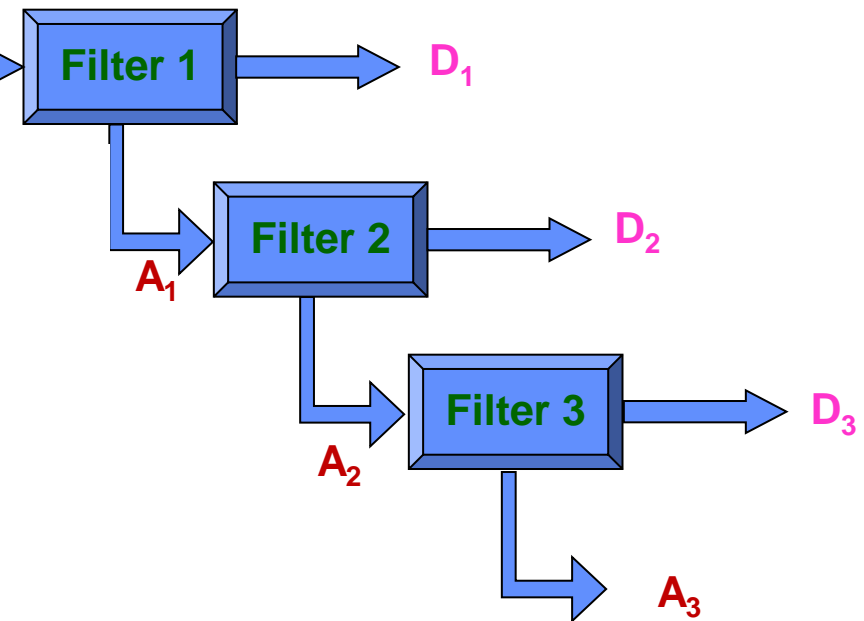
Coiflet



Discrete Wavelet Transform



Signal



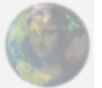
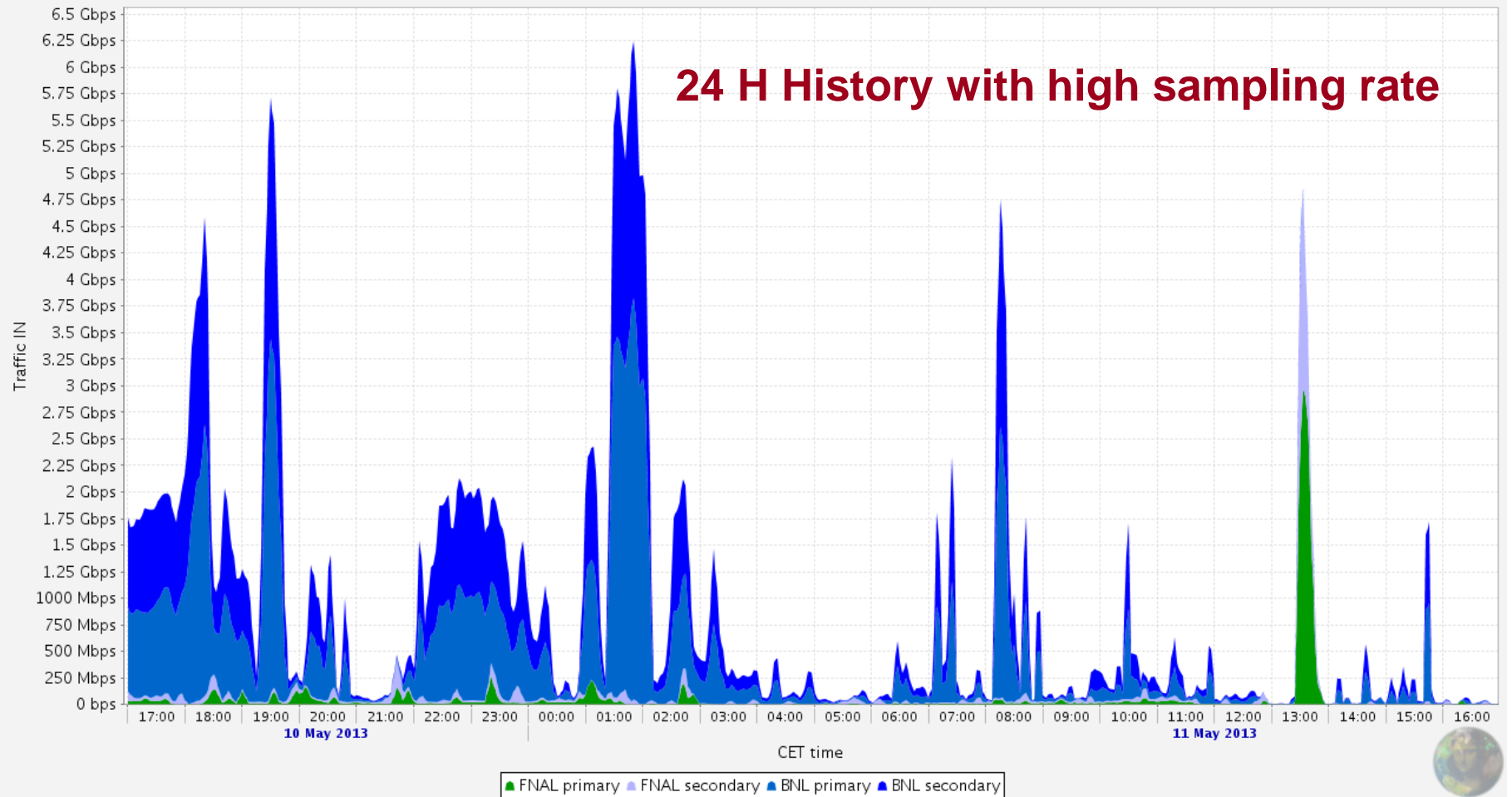
A single level decomposition puts a signal through 2 complementary low-pass and high-pass filters

The output of the low-pass filter gives the approximation (A) coefficients, while the high pass filter gives the detail (D) coefficients



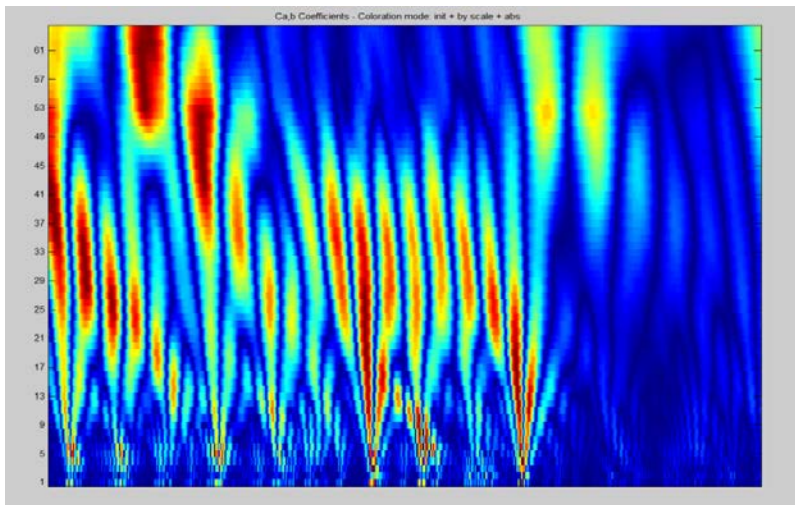
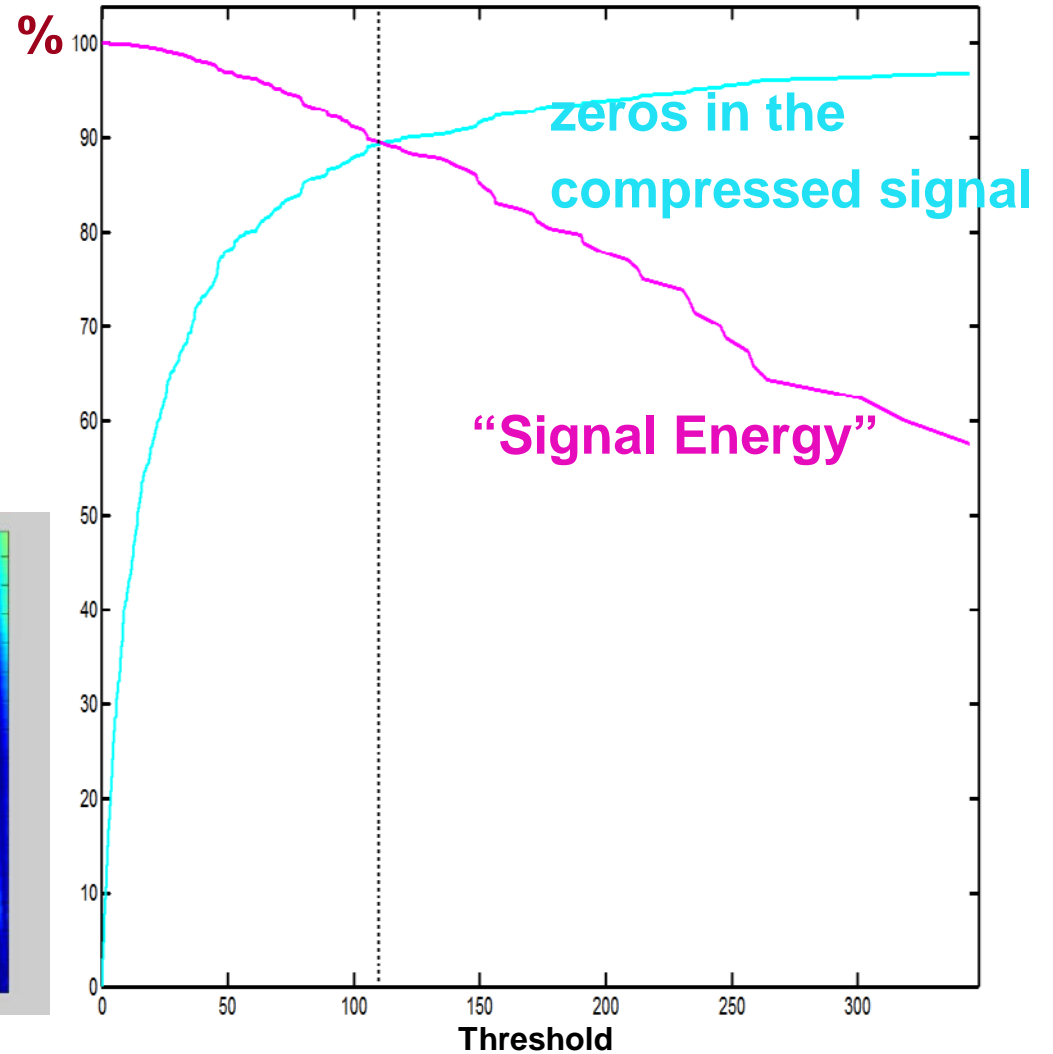
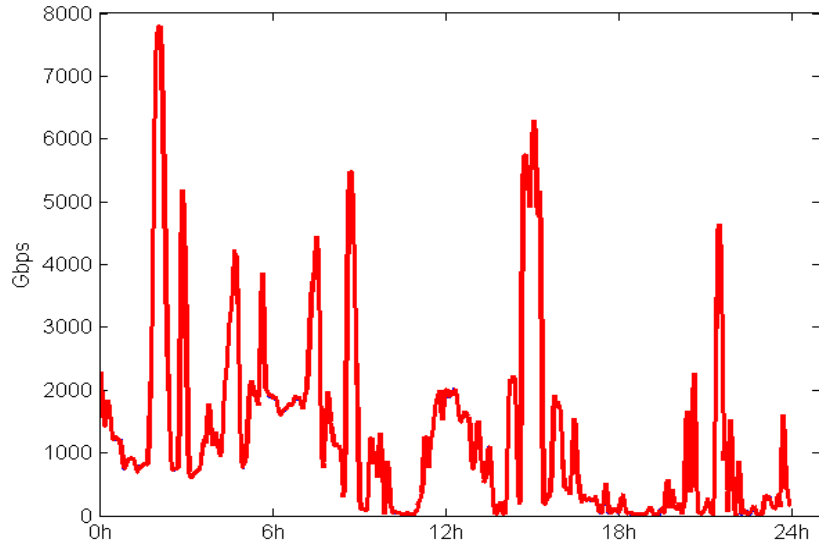
Network Traffic in USLHC net

CERN Tier0-US Traffic



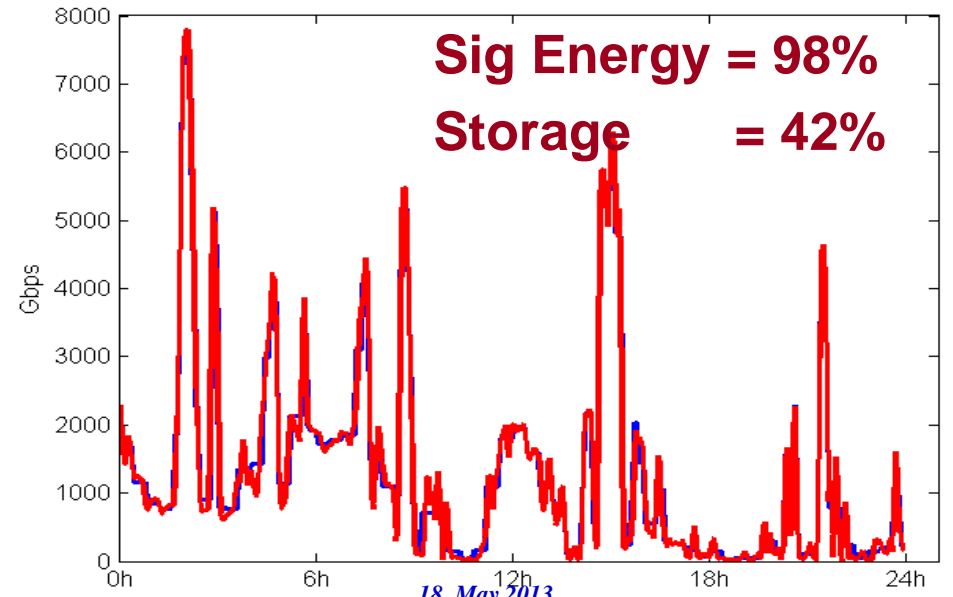
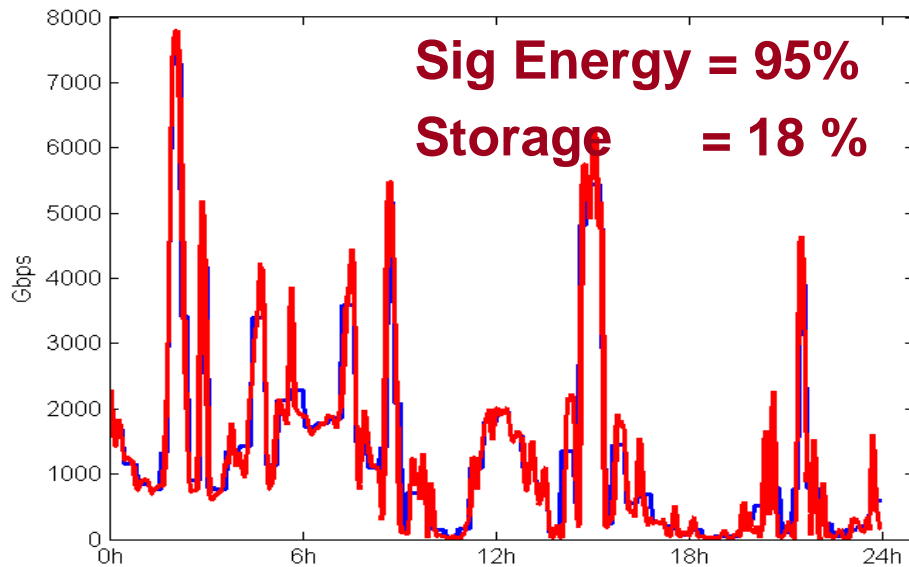
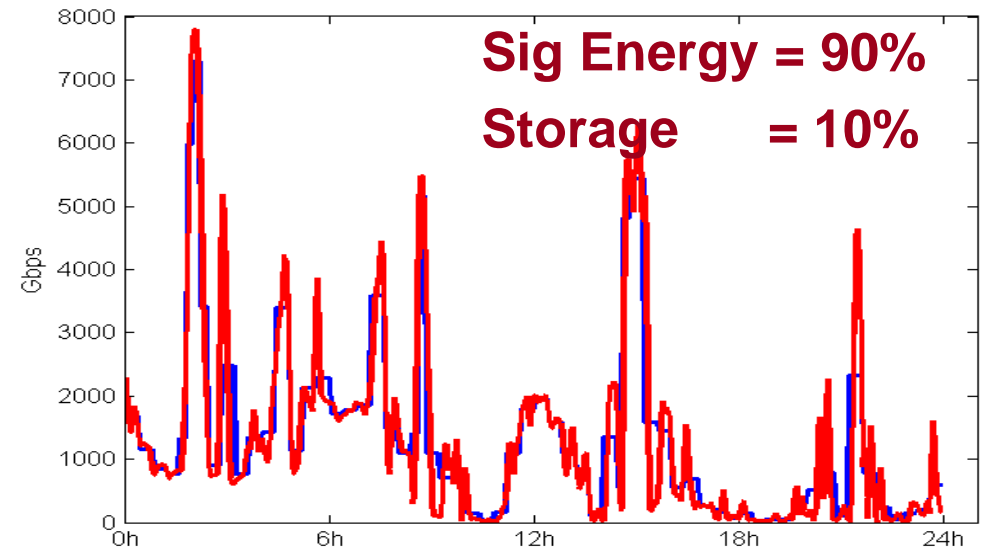
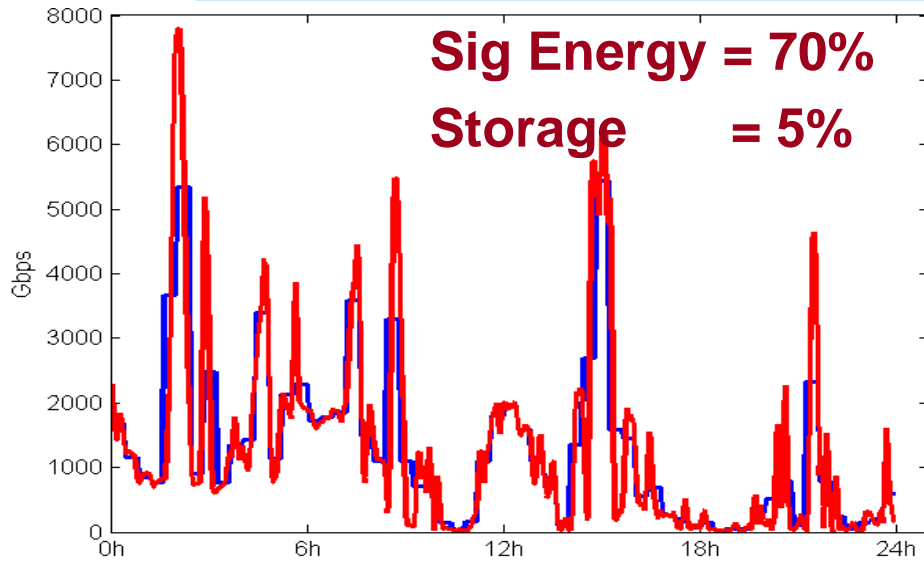


Compressing Time Series Data



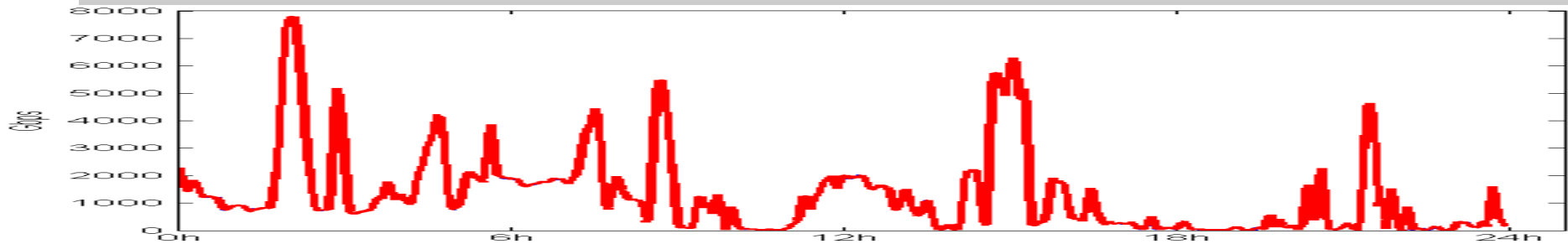
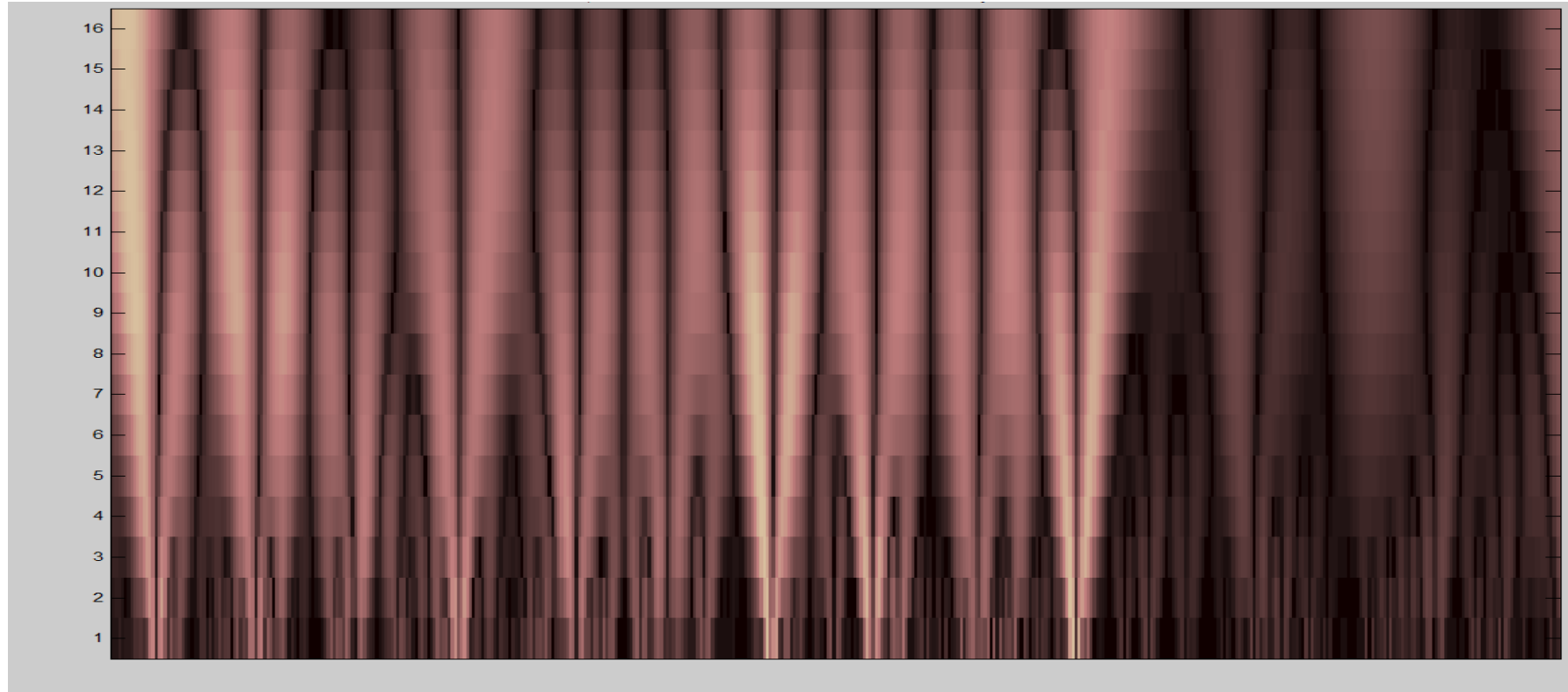


Example of Compression for high sampling network traffic data



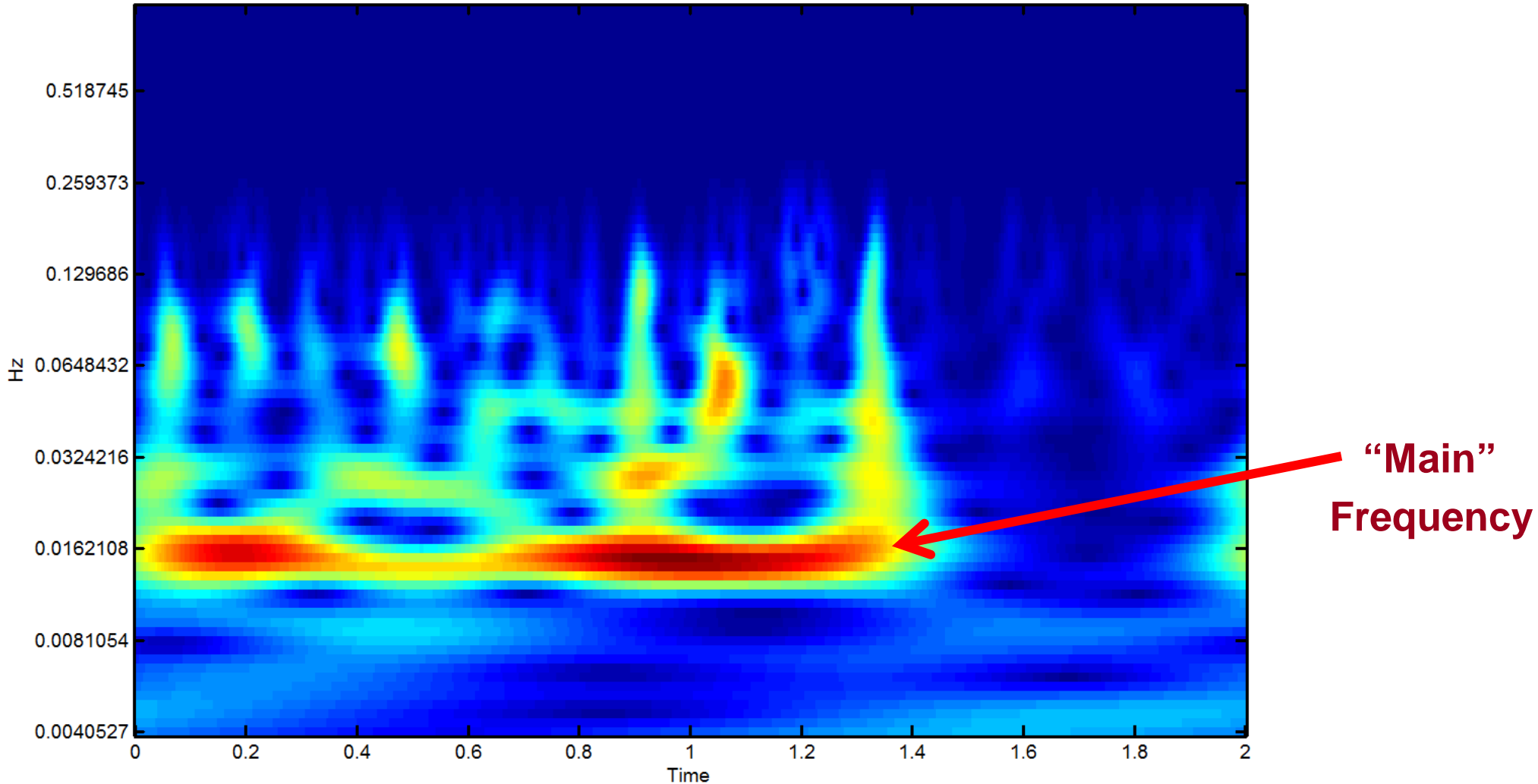


Self Similarity Structure in the Transformed Space





Similar data transfers operations that overlap in the total traffic pattern





Detecting Clusters in Multiple Time Series Data Using Wavelets

- ❑ Use a multidimensional grid structure onto data space
- ❑ These multidimensional spatial data objects are represented in an n-dimensional feature space
- ❑ Apply wavelet transform on feature space to find the dense regions in the feature space
- ❑ Apply wavelet transform multiple times which result in clusters at different scales from fine to coarse

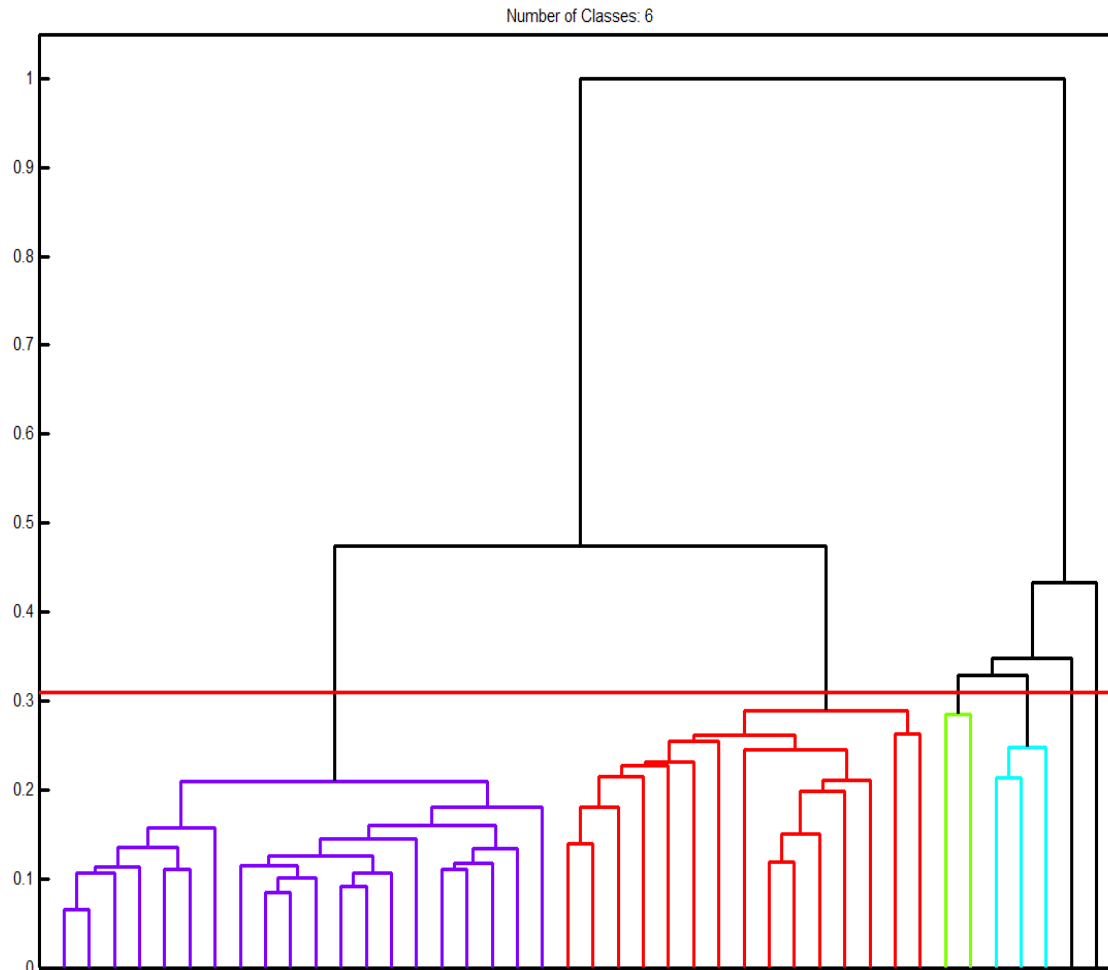


Two months history for the efficiency on all Alice grid sites





Acceding hierarchical clusters for the site efficiency plot



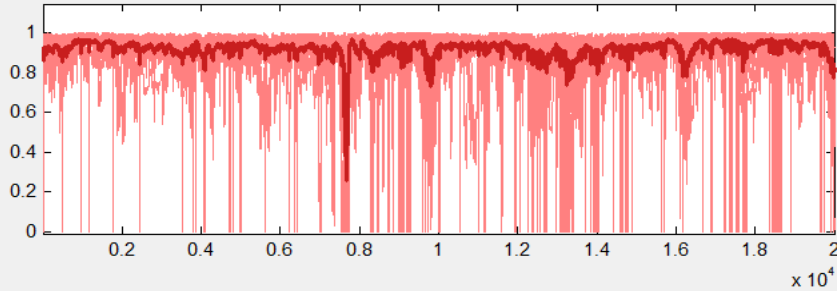
Major Advantages:

- **Complexity $O(N)$**
- **Detect arbitrary shaped clusters at different scales**
- **Not sensitive to noise, not sensitive to input order**

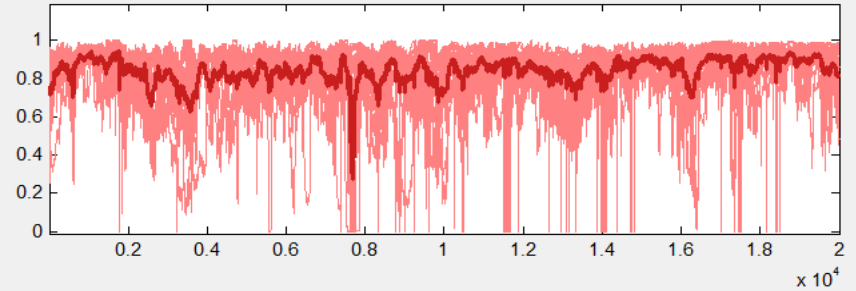


Results for cluster classification for sites

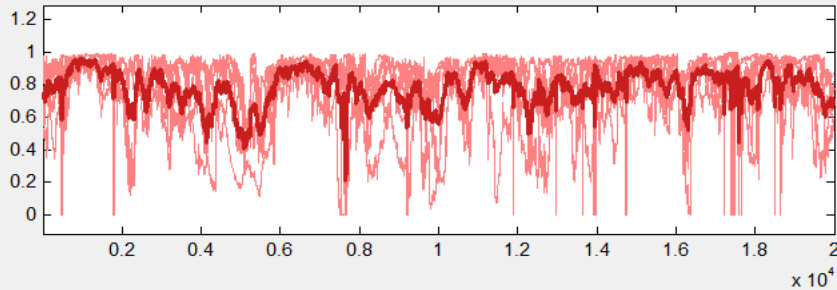
Class 1 - Nb 16 - 38.10% -- D = 0.082
Q1 = 1 - D / max = 0.918 -- Q2 = D / (max-min) = 0.082



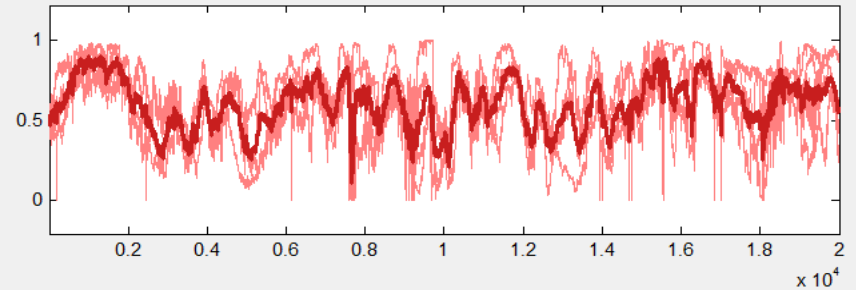
Class 2 - Nb 14 - 33.33% -- D = 0.117
Q1 = 1 - D / max = 0.883 -- Q2 = D / (max-min) = 0.117



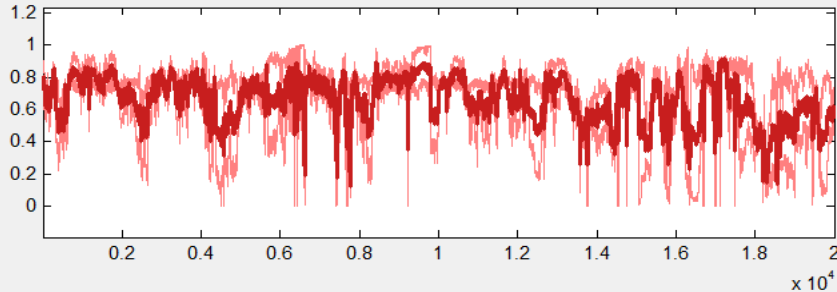
Class 3 - Nb 5 - 11.90% -- D = 0.158
Q1 = 1 - D / max = 0.842 -- Q2 = D / (max-min) = 0.158



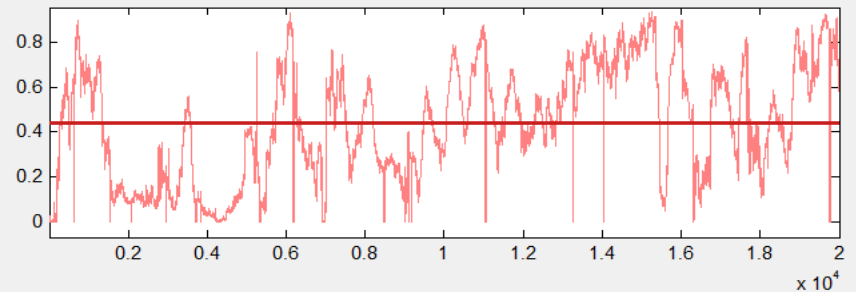
Class 4 - Nb 4 - 9.52% -- D = 0.150
Q1 = 1 - D / max = 0.850 -- Q2 = D / (max-min) = 0.150



Class 5 - Nb 2 - 4.76% -- D = 0.152
Q1 = 1 - D / max = 0.848 -- Q2 = D / (max-min) = 0.152



Class 6 - Nb 1 - 2.38% -- D = 0.251
Q1 = 1 - D / max = 0.732 -- Q2 = D / (max-min) = 0.268



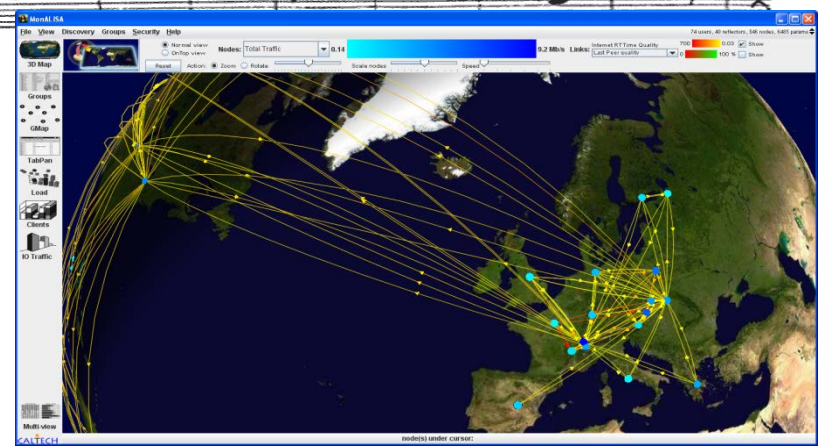
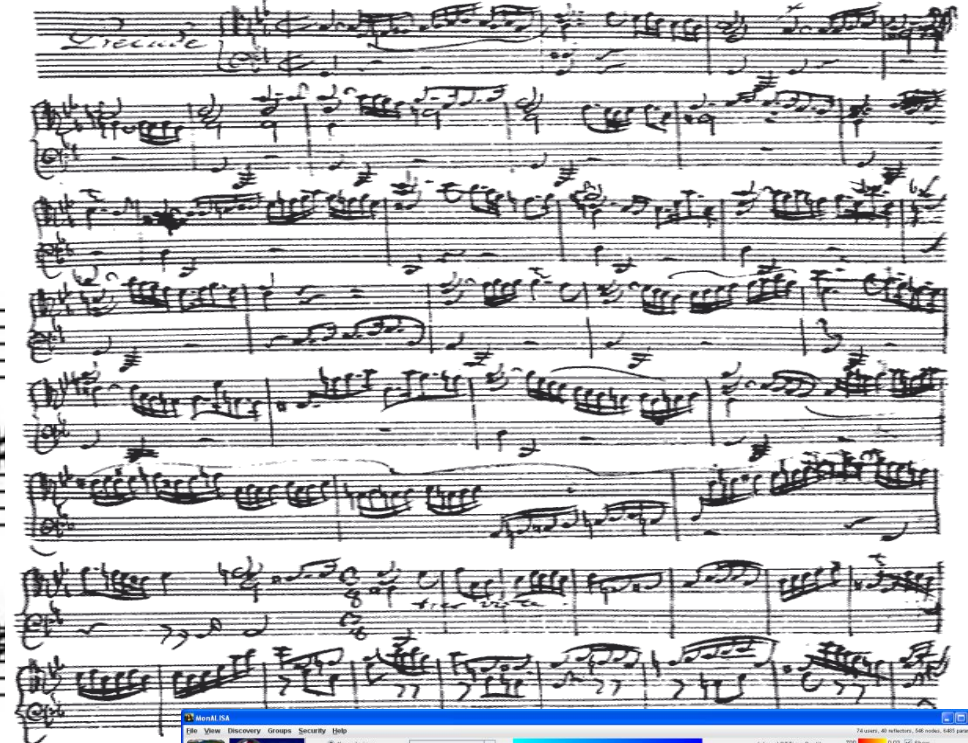
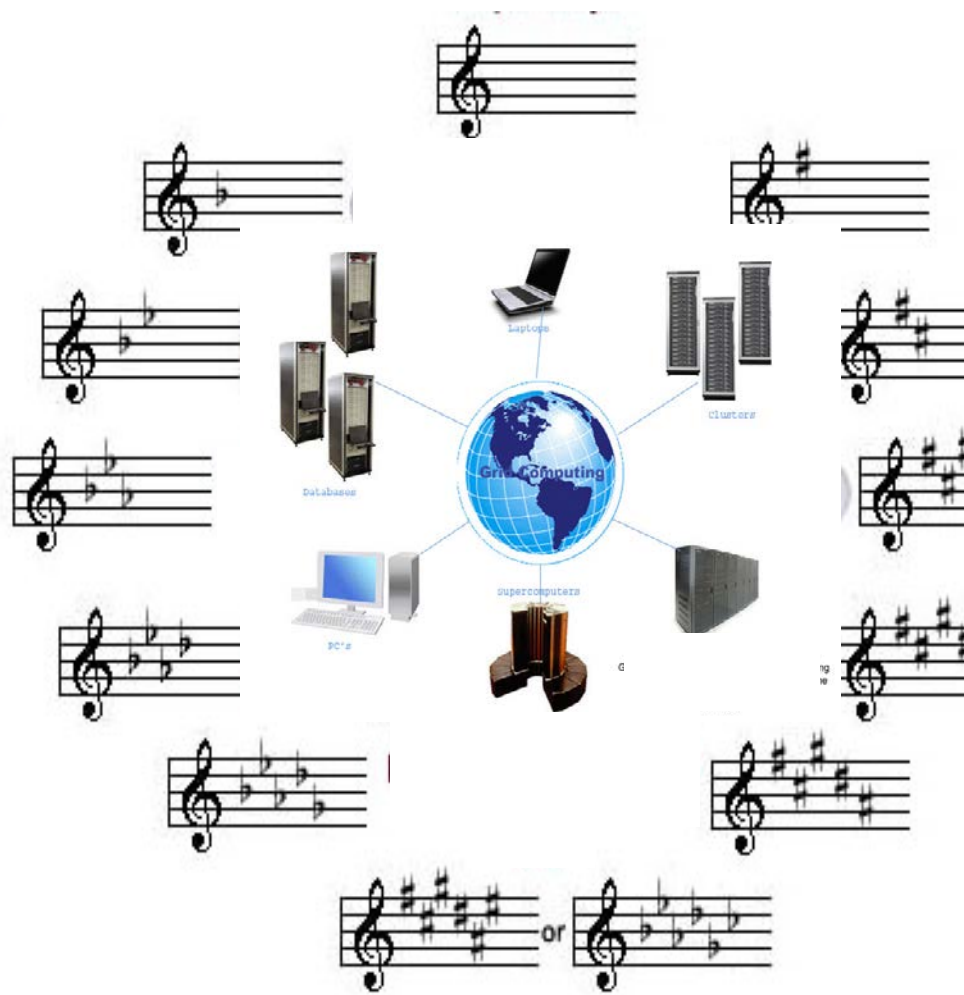


SUMMARY

- **The Wavelet transformation is really effective in compressing monitoring information and preserving details for signals with rapid variations.**
- **The implementation for DWT is fast and works multi-threaded**
- **The storage mechanism for monitoring information should be adapted to keep both uncompressed and compressed data.**
- **The wavelet approach is a powerful tool to analyze complex monitoring information. The time-frequency representation can help to understand the dynamic of complex systems.**
- **Dedicated modules (agents in the MonALISA system) can be used to automatically detect unexpected behavior from different systems and possible to take appropriate actions.**



Monitoring: A listener to the grid orchestra ?



THANK YOU !