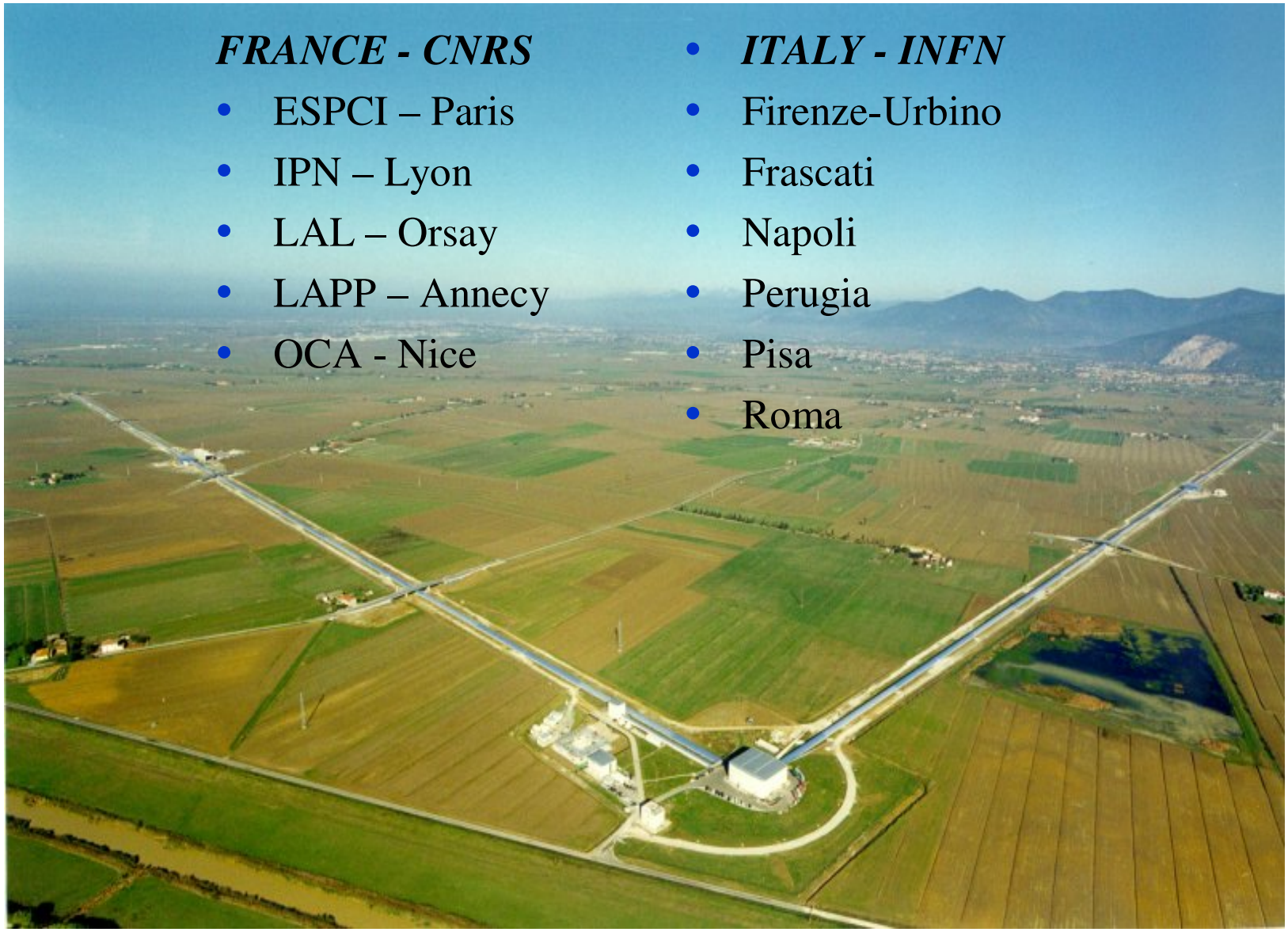
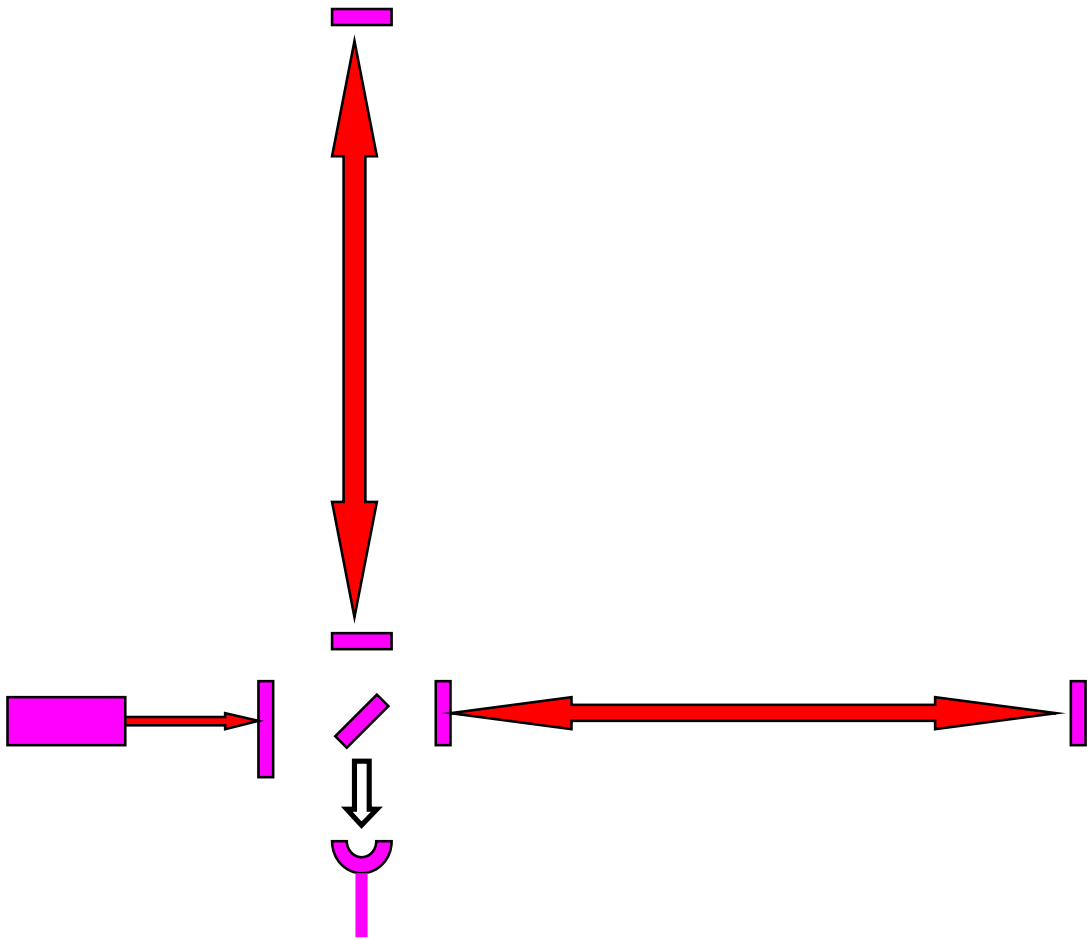
Virgo Data Analysis

Andrea Viceré
Università di Urbino and INFN Firenze
e-mail: vicere@fis.uniurb.it

Virgo Aerial View

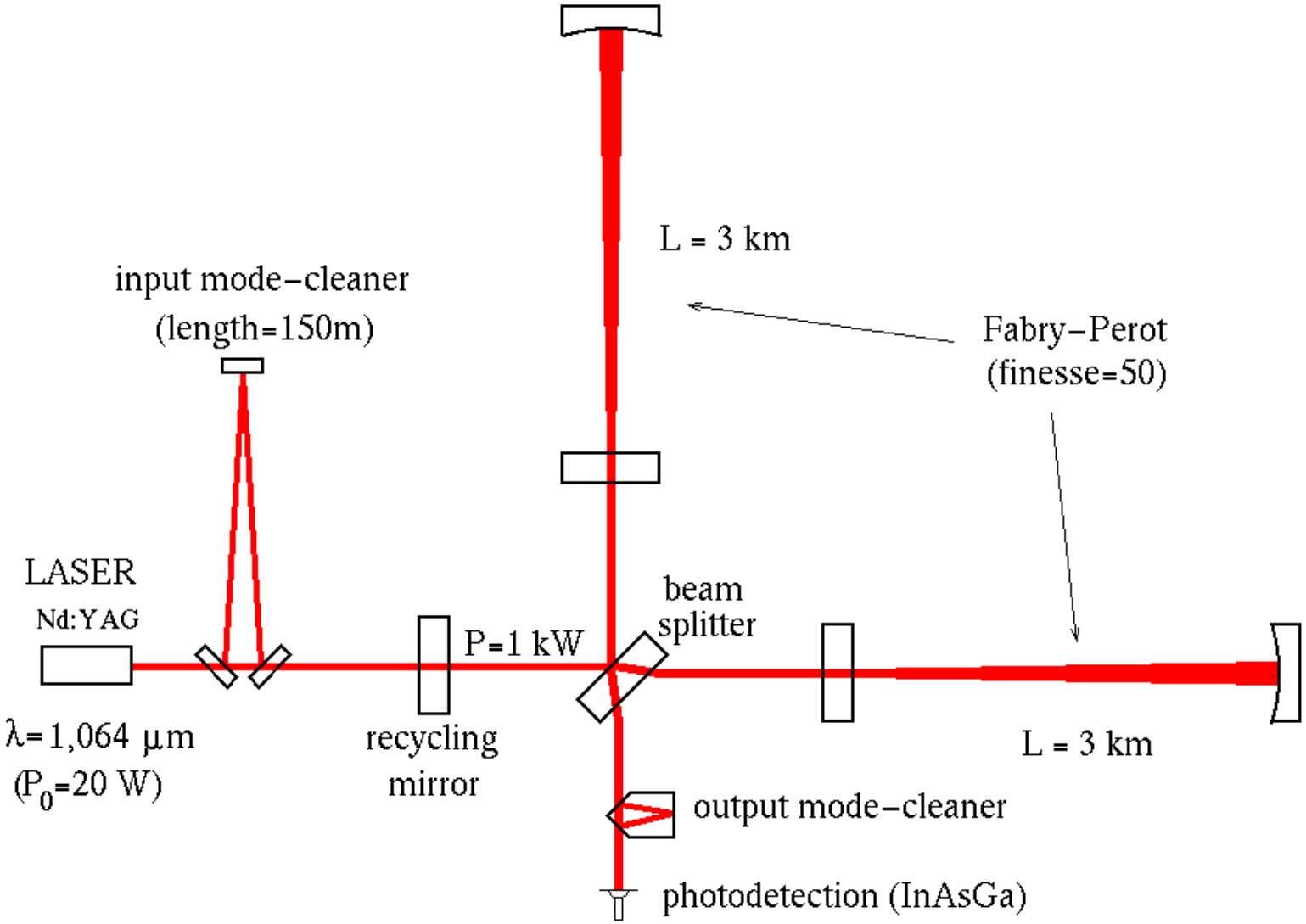


What is VIRGO in one slide

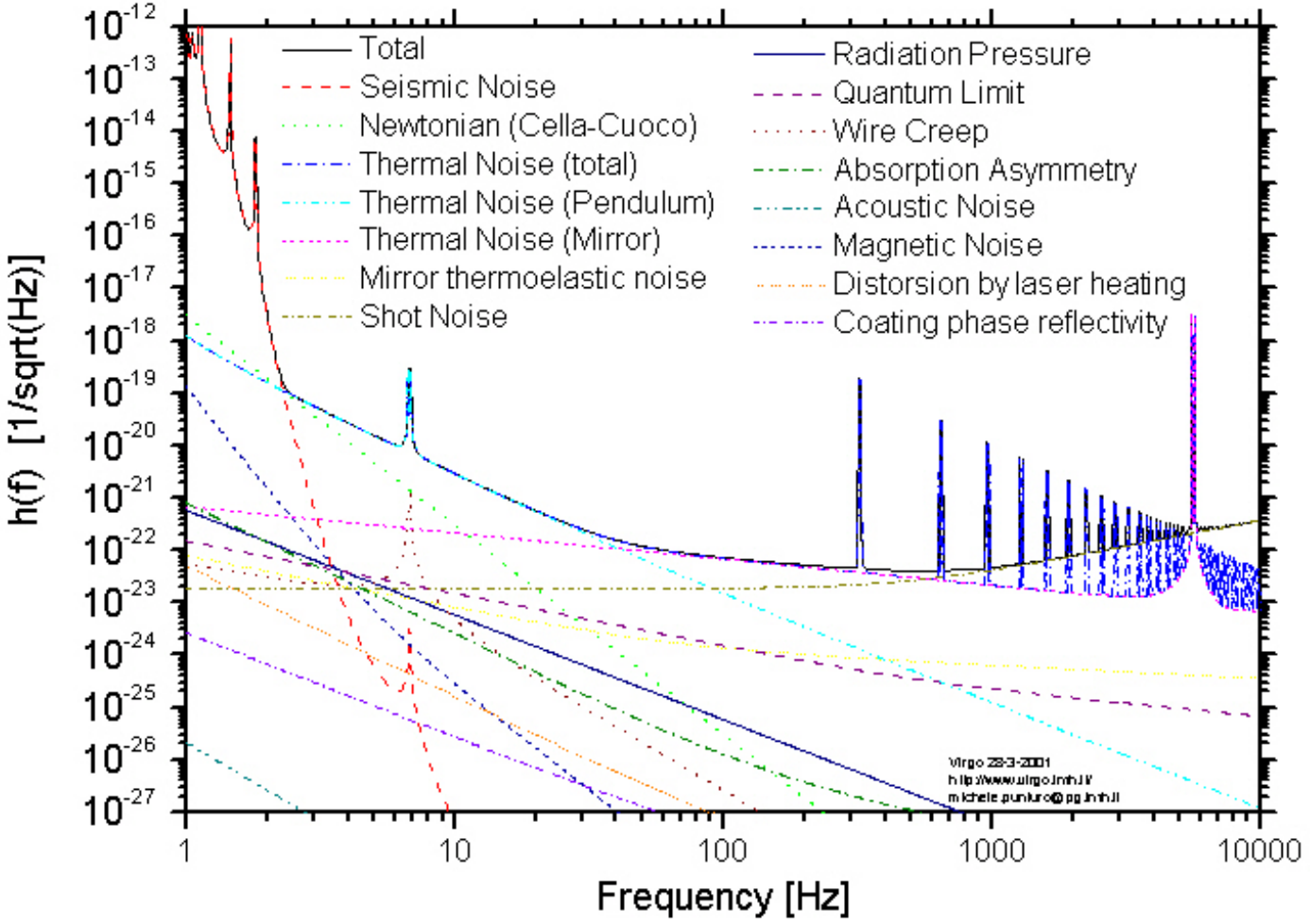


- Michelson Interferometer
- Fabry-Perot cavities in the arms, to extend the optical length
- Translates the metric distortion *h* into modulations of the signal at the dark port
- Bandwidth: from 4Hz up to 10 kHz.
- Audio signals!

Virgo Optical Scheme

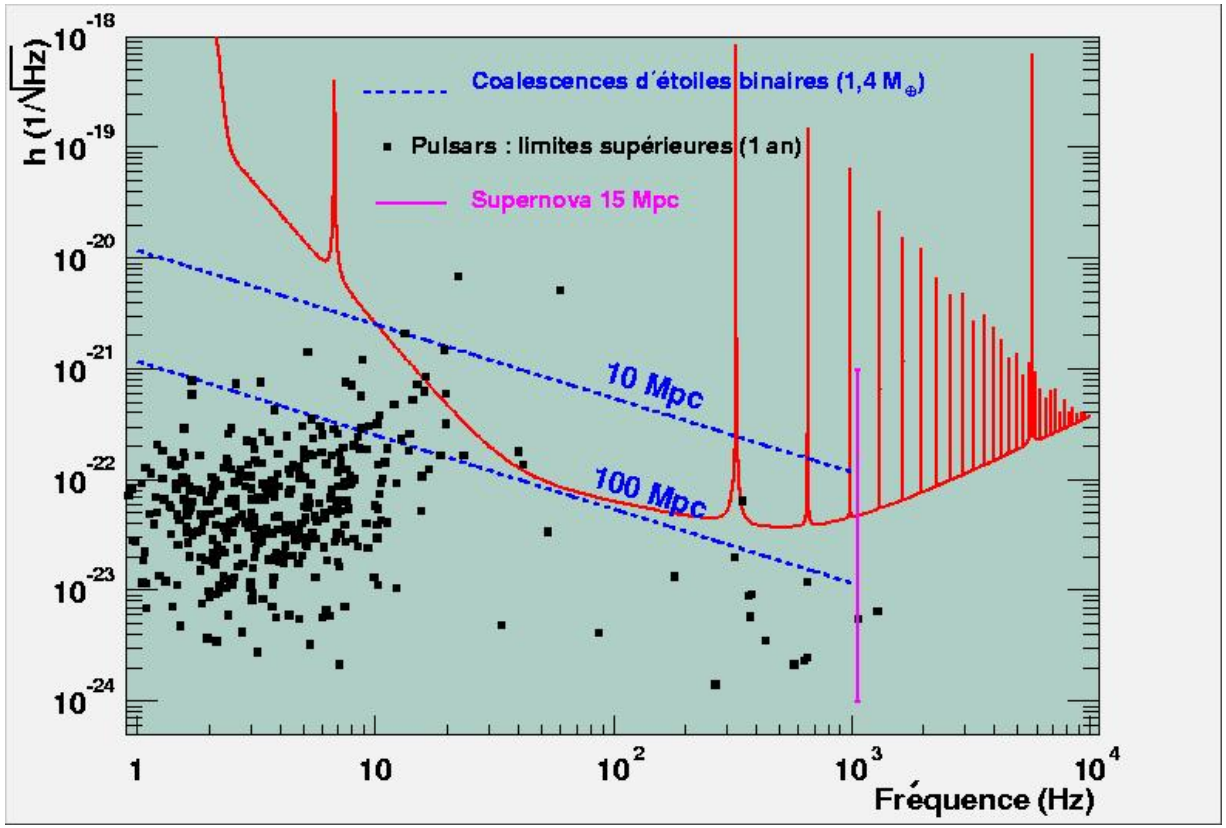


Virgo Design sensitivity



$$h(f) = \left\{ \frac{9 \cdot 10^{-37}}{f^5} + \frac{4.5 \cdot 10^{-43}}{f} + 3.24 \cdot 10^{-46} \left[1 + \left(\frac{f}{500 Hz} \right)^2 \right] \right\}^{1/2}$$

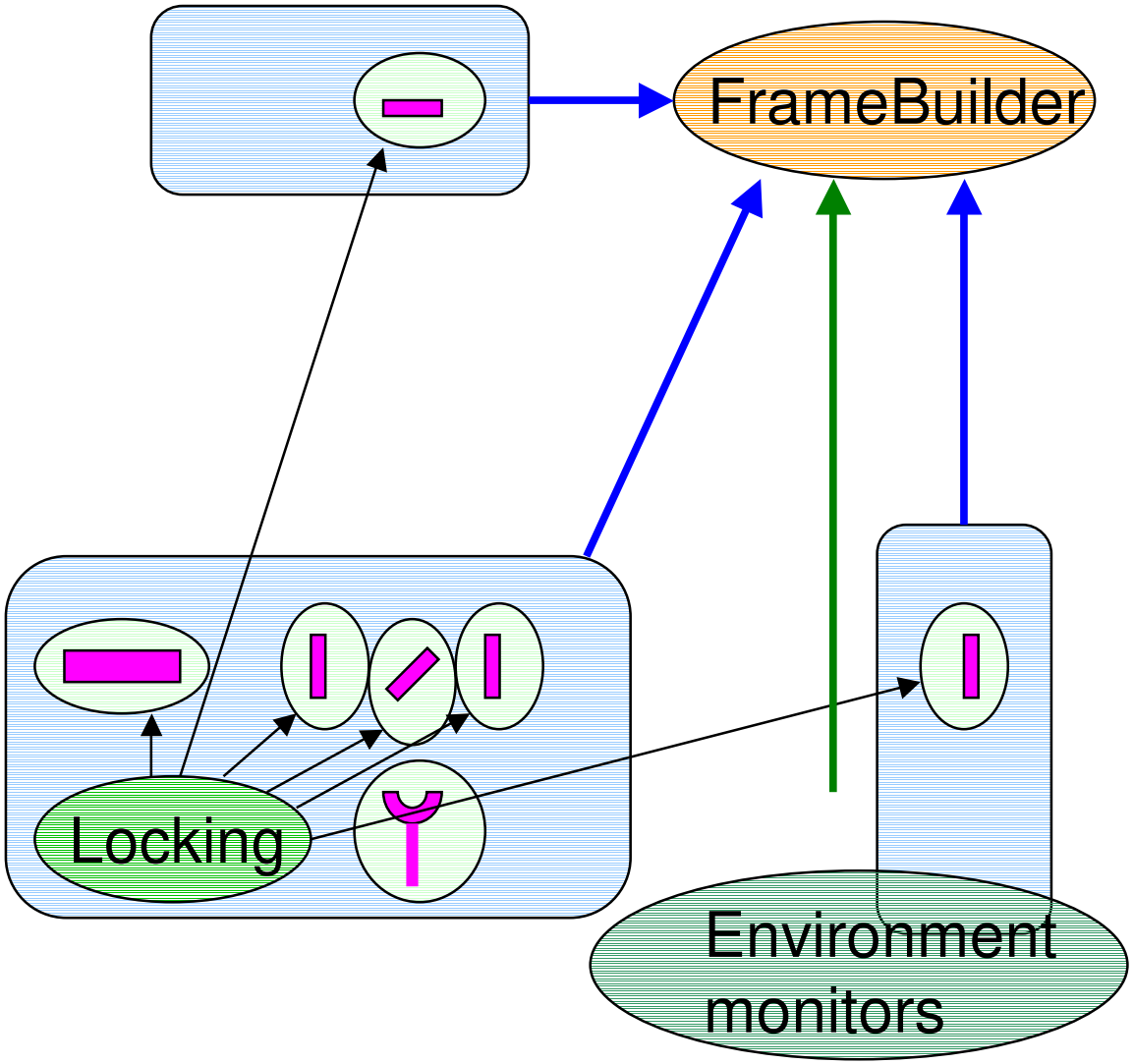
VIRGO Physics



- Goal
  - **signals of astrophysical origin**
- SNR low
  - **No "triggering"**

- Two kind of sources
  - **Impulsive, of various duration, emitted by coalescences of binary stars, or supernova explosions**
  - **Continuous, emitted by distorted rotating neutron stars, or as a background of cosmological origin.**

Data Acquisition

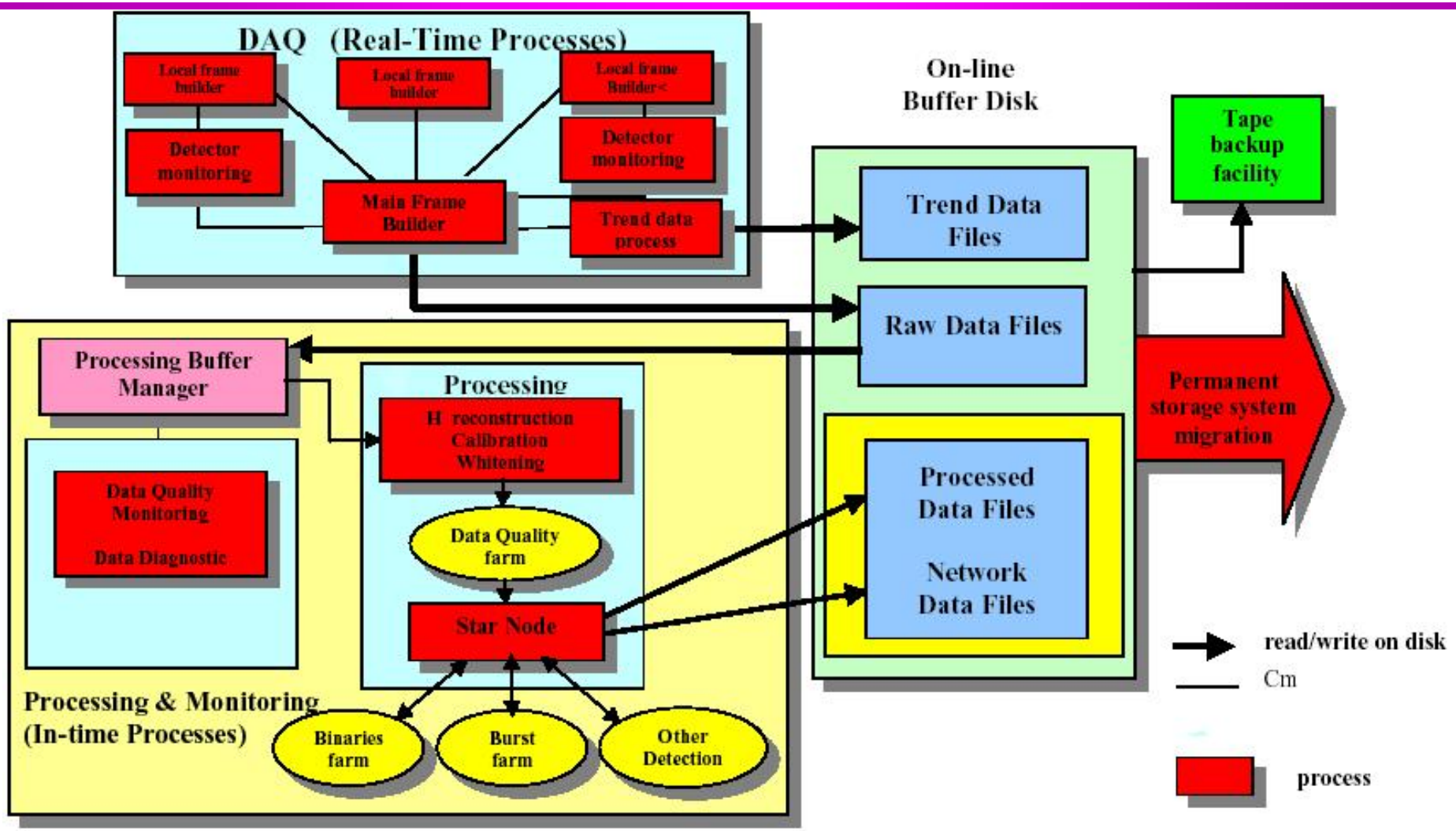


- Time series, not events.
- Multiple sources
  - **Locking signals**
  - **Environmental monitors**
  - **Triggering data**
  - **Data quality parameters**
- Slow (few Hz) and fast (20 kHz) stations
- Large data flow: range 1-5 Mbyte/sec, that is 30-150 Tbyte.year

Data analysis stages

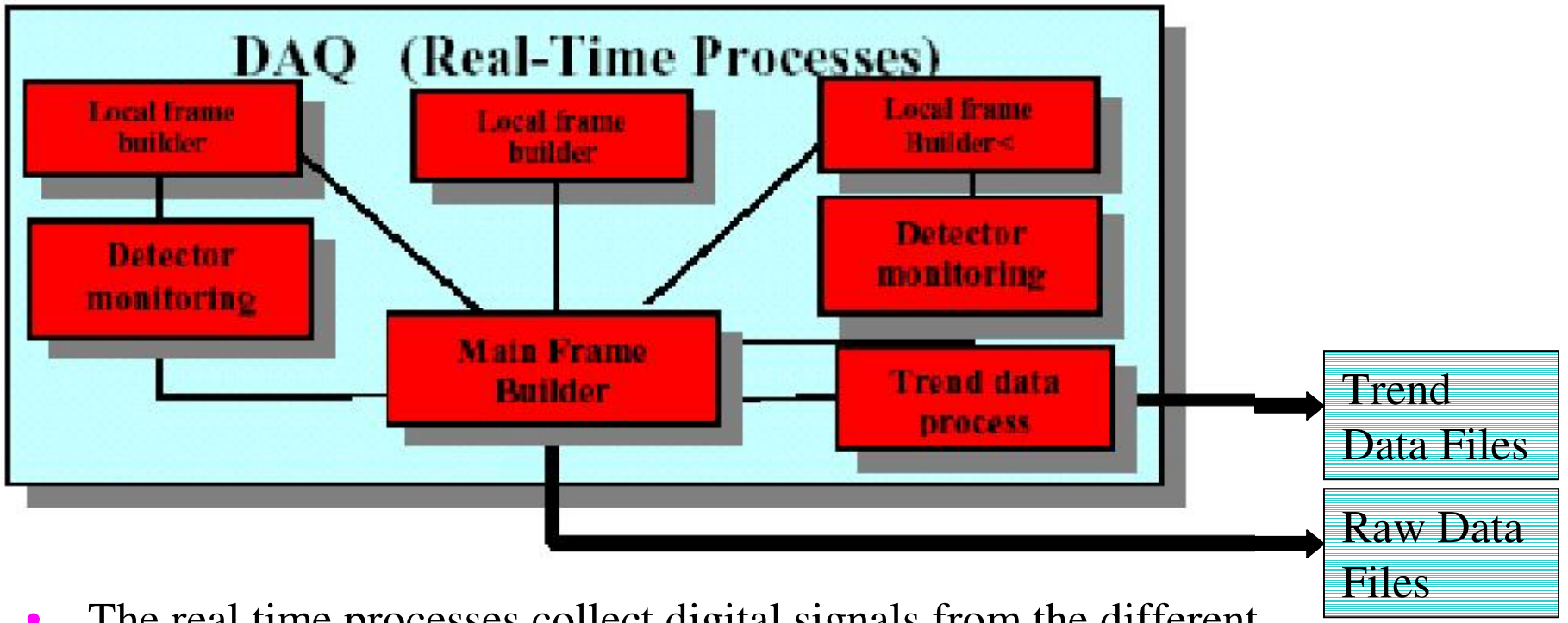
- *Real Time* [Cascina]
  - **Interferometer control and monitoring**
  - **Data acquisition and archiving (on buffering disks)**
- *In Time* [Cascina]
  - **Data conditioning: calibration, re-sampling, subtraction of the instrumental artefacts (the so called *h* reconstruction)**
  - **First stage of the search for impulsive signals (resulting from coalescing binaries or supernova events)**
  - **Data migration, from the disk buffer to a tape storage system**
- *Off-line* [Cascina, Laboratories, Computing centers]
  - **Search for periodic signals from rotating neutron stars**
  - **Second stage of the impulsive signal search**
  - **In-depth study of the instrumental noise**
  - **Re-processing (if needed), simulation and Monte Carlo studies**

Data flow on the Cascina site



- Interferometer control and data acquisition isolated by disk buffers.
- No interference on the DAQ due to the DA algorithms, which get inputs from different server processes.

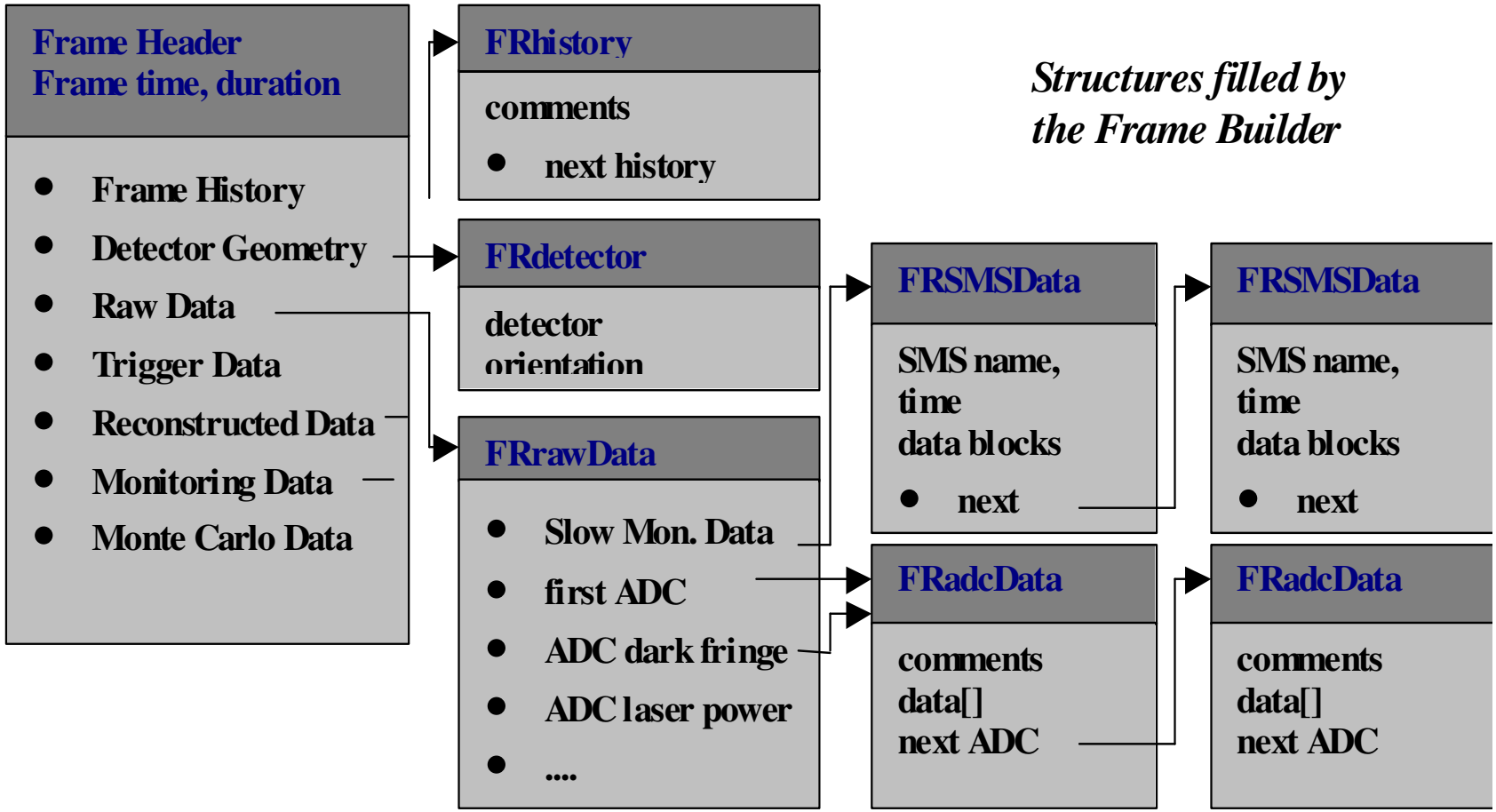
Real time analysis section



- The real time processes collect digital signals from the different sensors and organize them in *frames*, each containing 1 seconds worth of instrumental output
- Statistics useful to monitor the trends are continuously computed and saved, duplicated in separated files for easy of access
- Also the control signals are saved in the frames, to make it possible reproducing off-line the calibration procedure

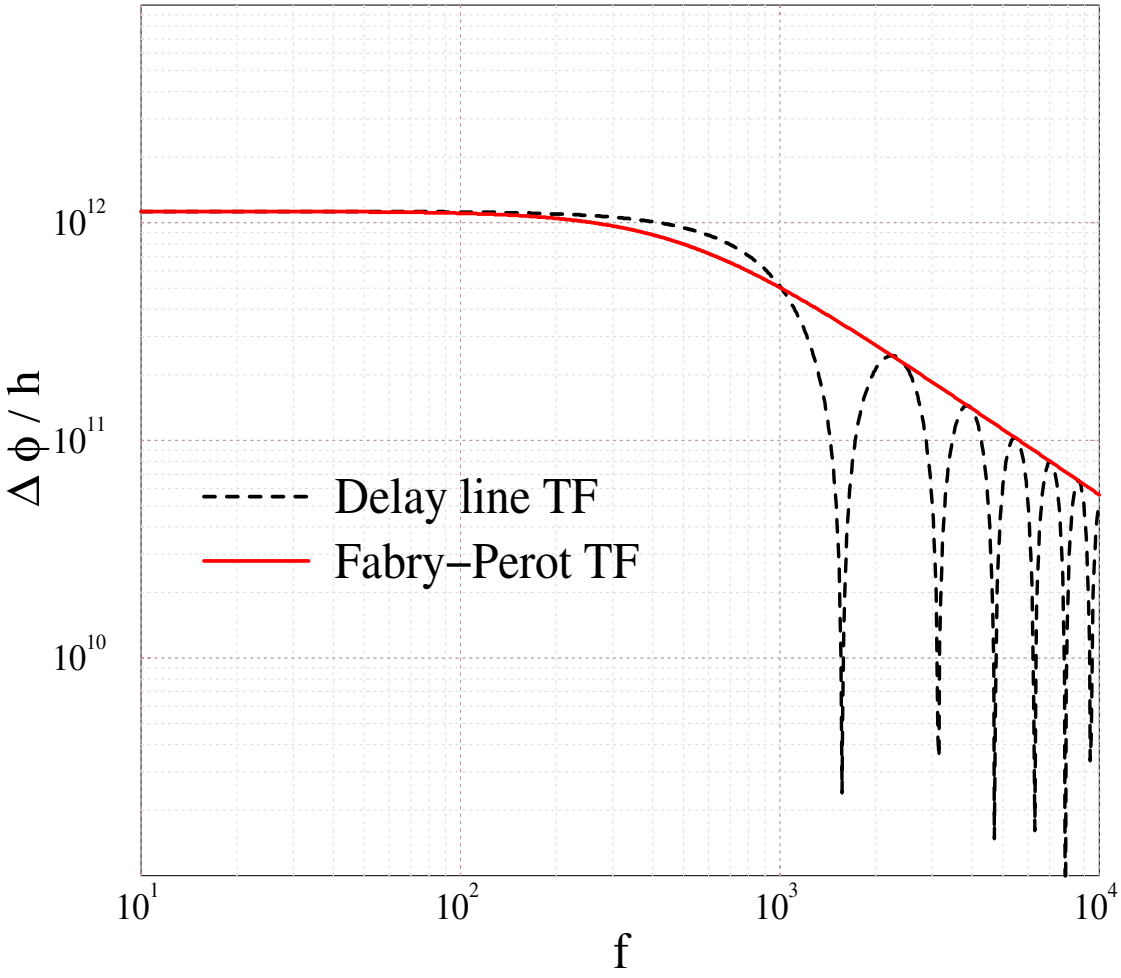
Data Format

- LIGO/VIRGO standard



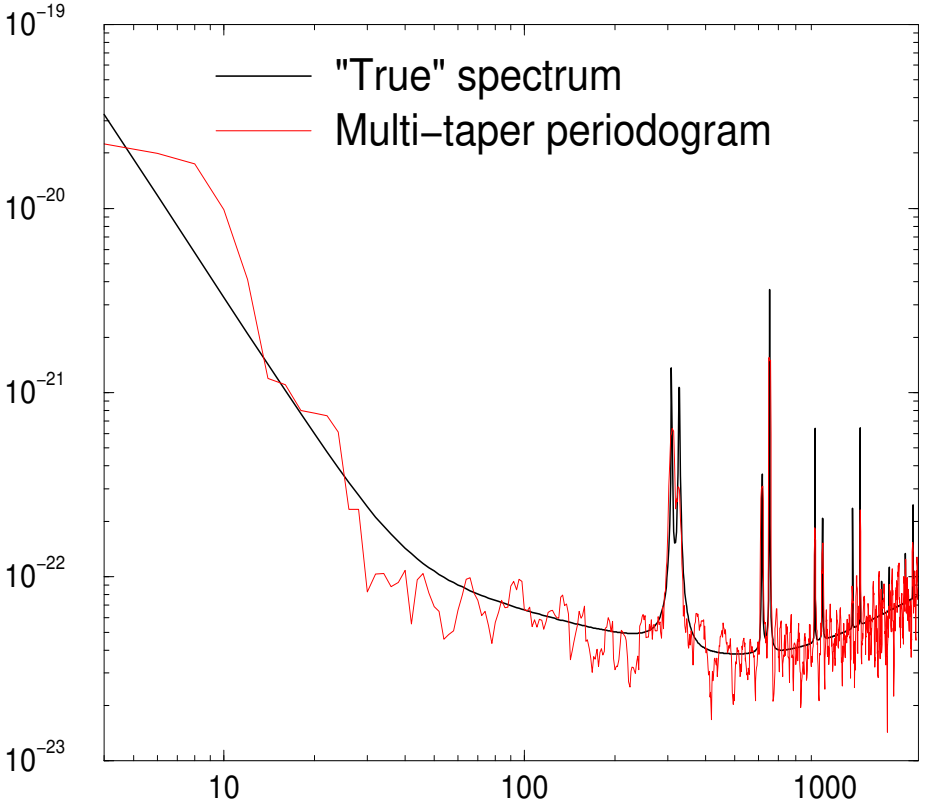
h Reconstruction concept

- To unfold the signal transfer function
  - **At low frequencies, recover h from control signals**
  - **Diode read-out sensitive to offsets from working point**
  - **Requires off and/or on line calibration**
  - **Requires in time operation**



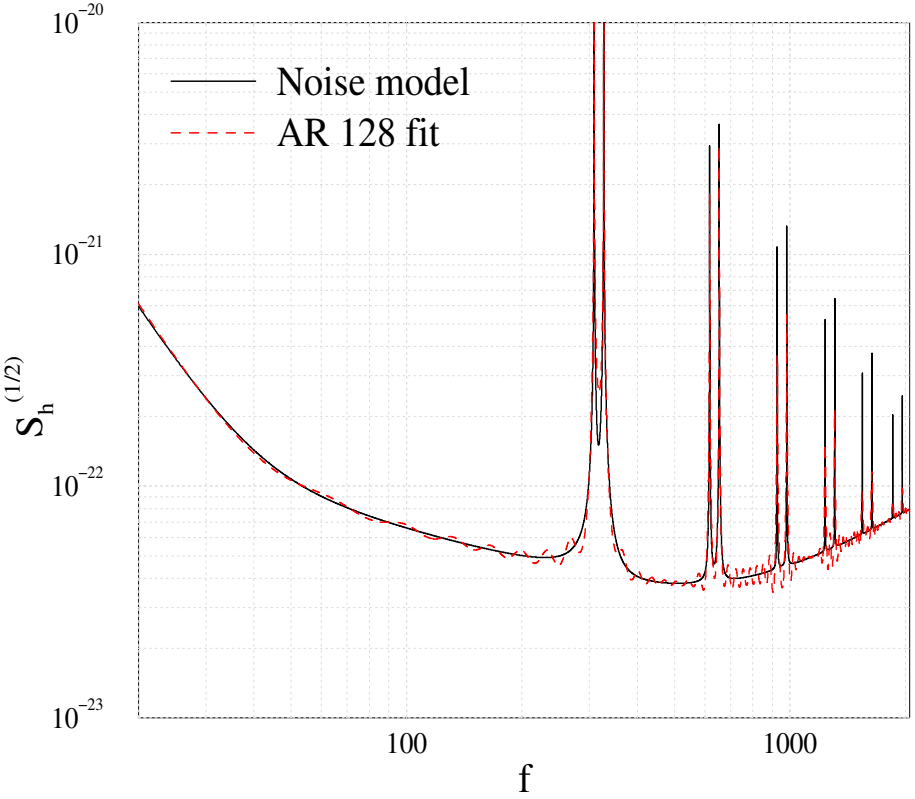
$$\Delta\phi = h\tau_s \frac{8\pi c}{\lambda} \frac{1}{\sqrt{1+(4\pi f\tau_s)}}; \quad \tau_s = \frac{LF}{c\pi}$$

Internal Noise characterization

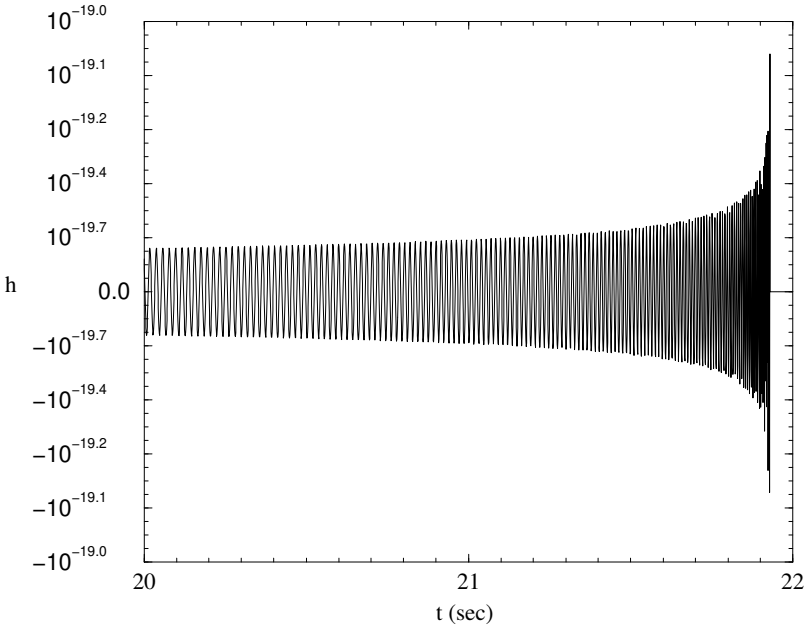


- Non parametric estimation: multi-tapering periodograms.

- Parametric: auto-regressive, moving average models.



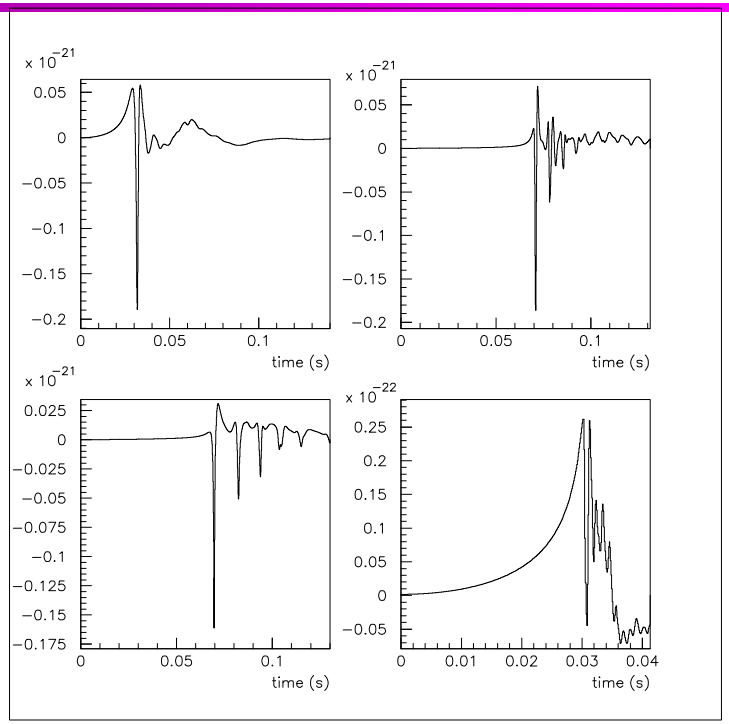
Coalescing binaries search strategy



- Events: “chirps”
  - **Locally sinusoidal signal, with increasing frequency**
  - **Details depend on physical parameters**
  - **Residence in the detection band: from a few seconds up to a few minutes.**

- Basic method: matched filtering
  - **Computing requirements: O(300 Gflop/s) sustained**
  - **Parallelism: on the filters for different parameters**
  - **Hardware architecture: PC farm in master-slave configuration, equipped with an MPI infrastructure**
  - **Software components: waveform generation, parameter space tiling, filtering of the data on the parallel computing system, event reconstruction**

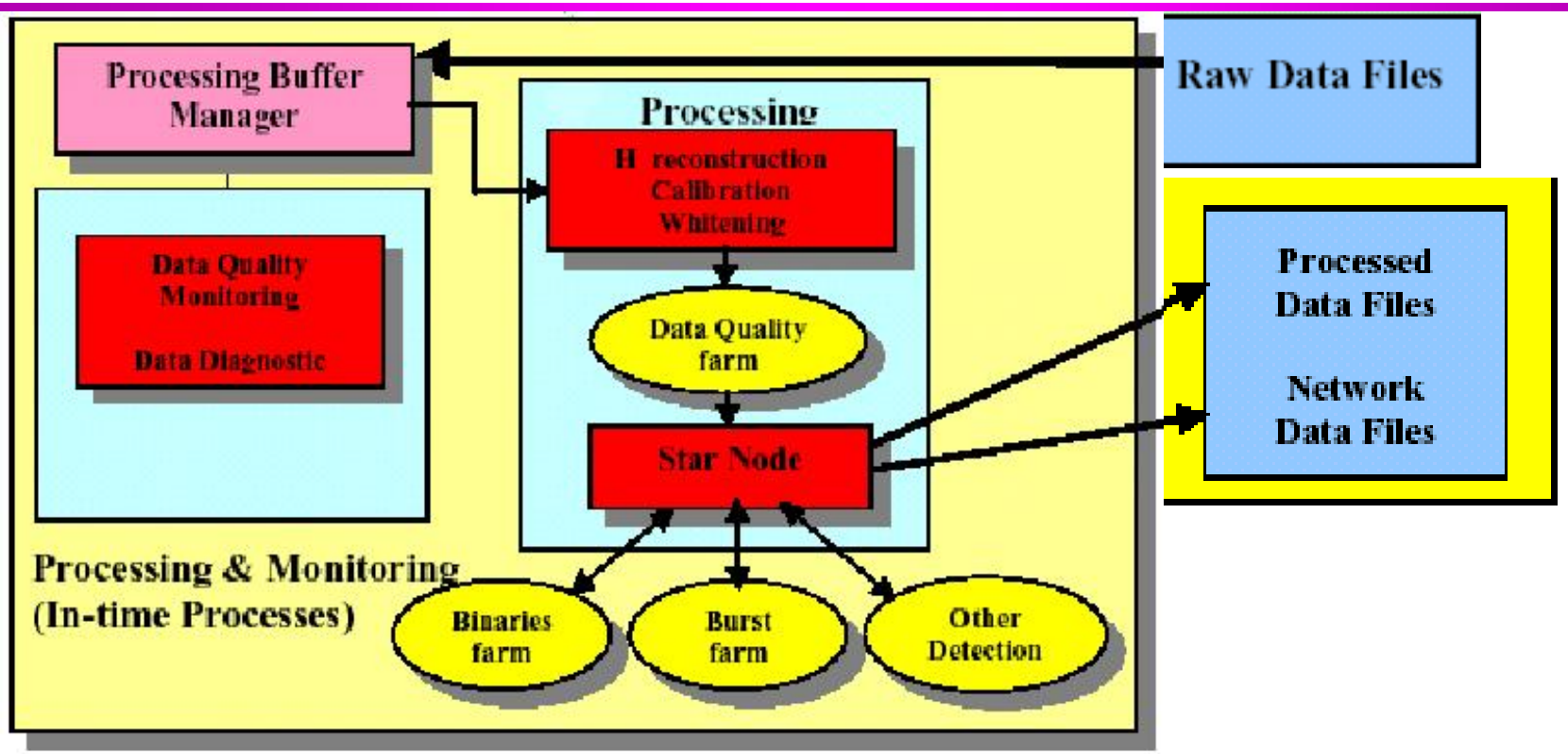
Bursts analysis strategy



- Events: "glitches"
  - **Short duration: few tens of msec, O(10^3) samples**
  - **Minimal knowledge of the waveforms**

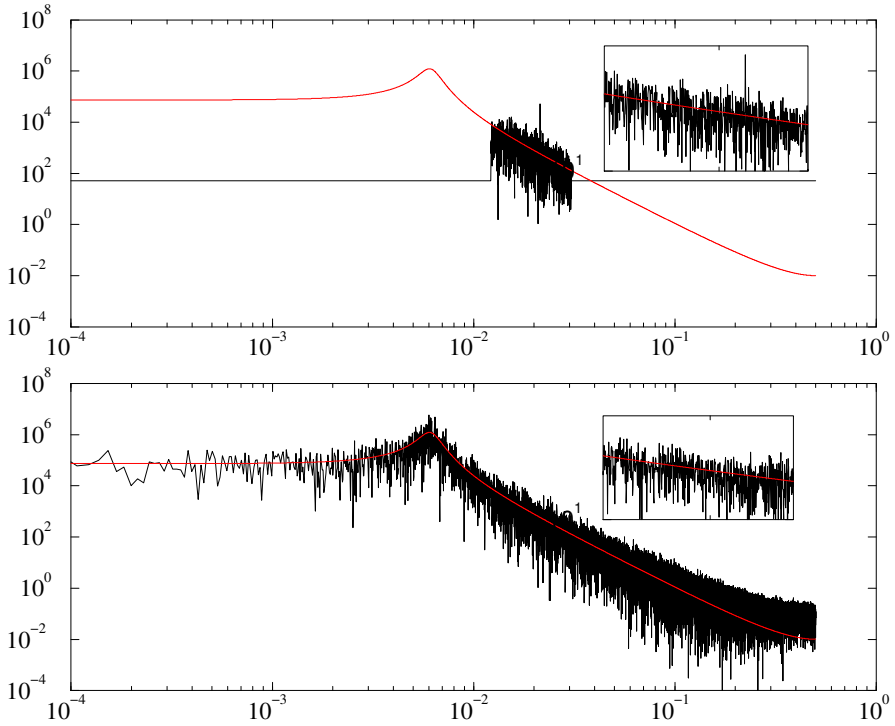
- Unknown waveforms → "blind" search methods
  - **Matched filters for damped sinusoids**
  - **"Excess noise" detectors (in various forms)**
  - **Computing requirements: relatively modest O(1-10 Gflop/s)**
  - **Parallelism: embarassing, on the search methods**
  - **Hardware architecture: master/slave, with different search methods running on different computing units (or partitions of the same hardware)**

In-time analysis section



- Raw Data Files are read by a server process, which hands them to the pipeline of the data conditioning and data quality processes
- Conditioned data distributed by a "Star Node" to specialized computer systems.
- Results sent back to the Star Node and saved in the "Processed Data", including a "Network Data" stream, to be sent to other collaborations for coincidence analysis

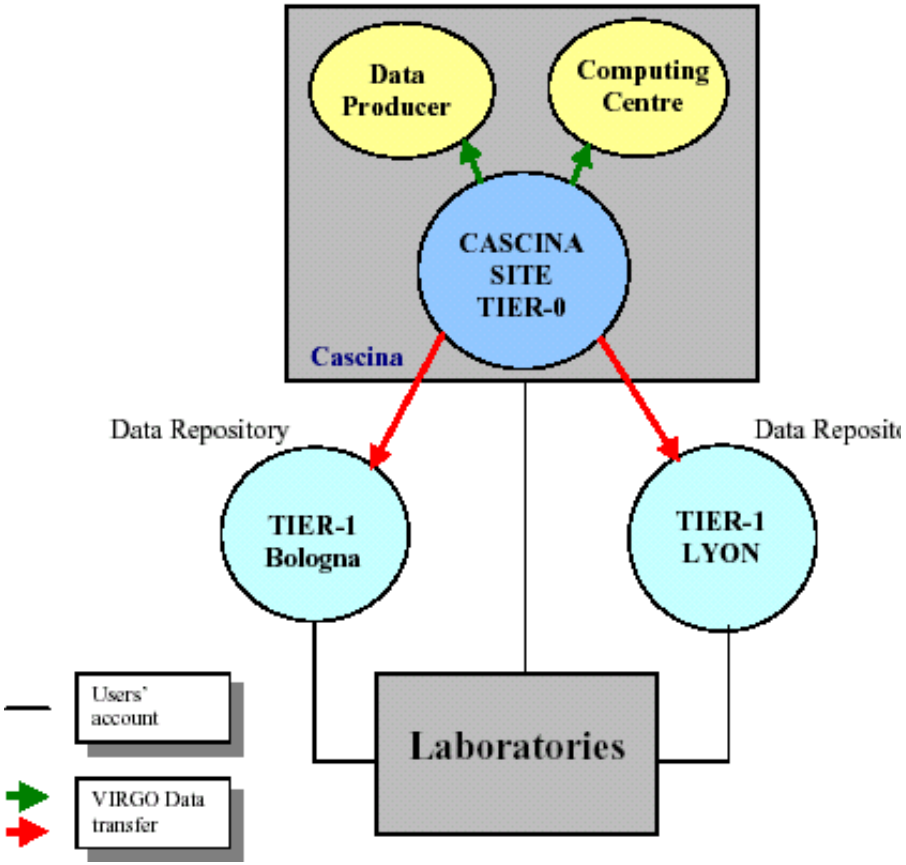
Periodic signals search strategy



- Essentially a periodic signal → FFT methods
  - **But, Earth motion introduces a huge Doppler effect, depending on the source position → large number of parameters in the search**
- Low SNR → long integration time

- Only an *off-line* analysis is feasible
  - **Too many parameters → hierarchical, sub-optimal search**
  - **Partially incoherent analysis based on short FFTs and the Hough transform, supplemented by a coherent follow-up of the best candidates**
- Hardware
  - **A definite parallel architecture still to be finalized, based on PC clusters (for their cost effectiveness) and on a master slave architecture (for its simplicity)**

Data and computing resource access



A centralized file catalogue (the "BookKeeping Data Base") will receive data requests and address the users to the most accessible data repository

- Cascina
  - **Data production and in-time analysis**
  - **Primary archive role**
- Bologna e Lyon
  - **Secondary archives**
  - **Computing centres**
  - **Primary data distributors for the Labs**
- Laboratories
  - **R&D on methods and software**
  - **Offline analysis**
- Data distribution
  - **Based on the GRID toolkit and on data transfer tools (BBFTP, RSYNC)**
- Computing resources sharing
  - **In Italy, GRID is proposed**

Software architecture and environments

- *Real time*
  - **Most of the data acquisition and control software is written in Object Oriented C or in C++**
  - **The inter-process communication is handled by a library developed inside Virgo (the Cm library)**
- *In-time and off-line*
  - **All the libraries share an OO architecture: while no strict rule has been enforced, the tendency is to adopt the C language for basic libraries (frame handling, vector elaboration, signal analysis) and C++ for high level libraries**
  - **Several programming/analysis environment are being experimented**
    - **An environment based on ROOT (VEGA).**
    - **A collection of MatLab extensions (SNAG).**
    - **A scripting environment, based on Tcl/Tk (Dante).**
  - **The inter-process communication software chosen for the coalescing binaries farm is MPI**

Conclusions

- The Virgo collaboration plans to have its first science run in 2004
- The sensitivity will be initially inferior to the design one, but the collaboration will perform a full analysis of the data produced
- To be ready for the analysis of real data, the collaboration is performing Mock Data Challenges, progressively testing more and more elements of the analysis chain
  - **A Monte Carlo is available to produce noise data in Frame format, and to inject physical events.**