Grid And Many-Core Computing @ Virgo

Gergely Debreczeni
<Gergely.Debreczeni@rmki.kfki.hu>
(KFKI RMKI, Budapest)

on behalf of the Virgo Collaboration

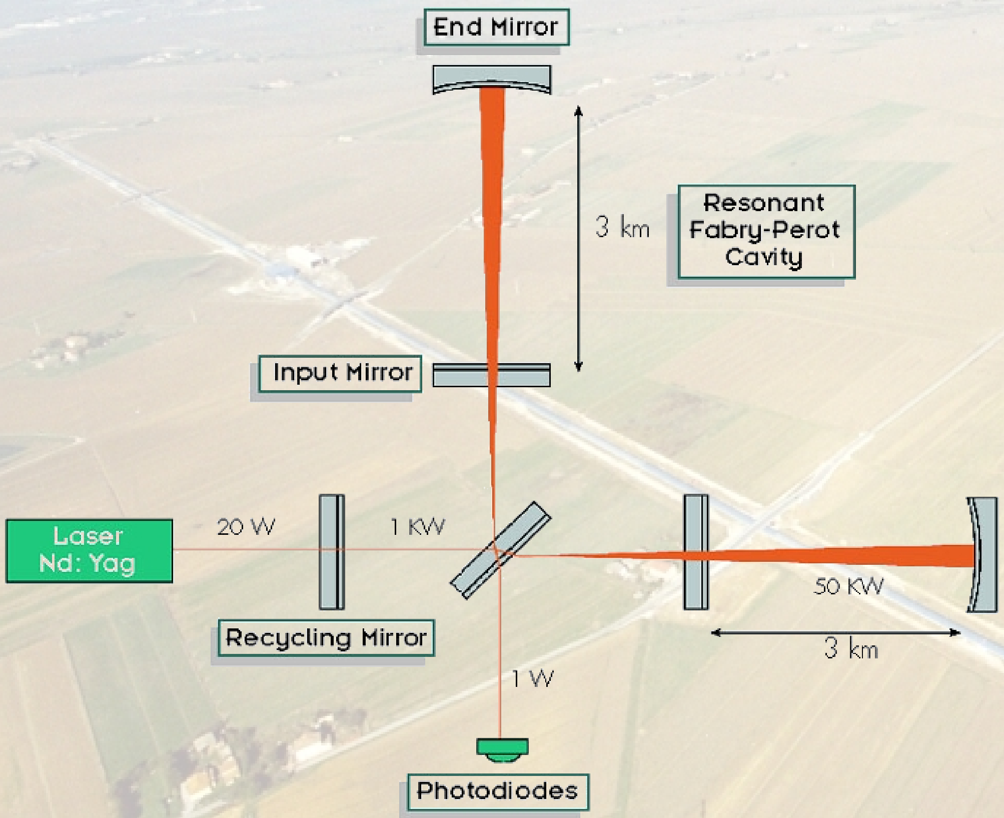
VIRGO

About this talk

- The Virgo experiment
- Storage and computing needs
- Data transfer solutions
- Analysis on dedicated clusters
- Analysis on the Grid
- The Virgo pilot pool
- Many Core computing @ Virgo
- HPC for Virgo ?
- The Macgo Project
- Future plans

The Virgo experiment

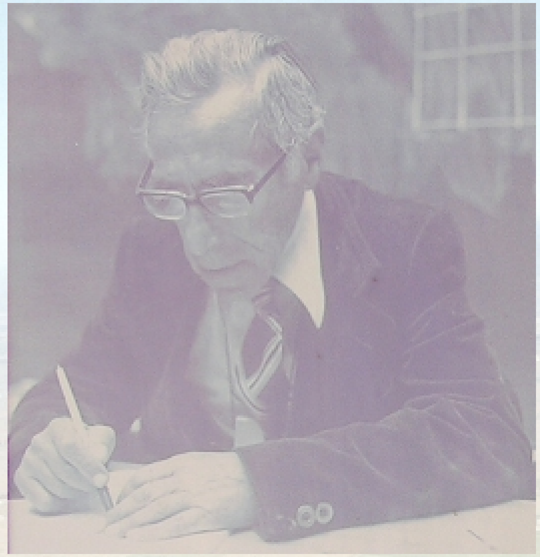
- Cascina, Italy
- 3 km-es arm length
- Fabry-Perot cavities 50/150 finesse
- 6800 m^3, 10^{-10} mbar vacuum!!
- 20 W laser
- solid concrete base, 20-50 m deep
- 1 MW power consumption
- excellent seismic isolation (10^{-10})

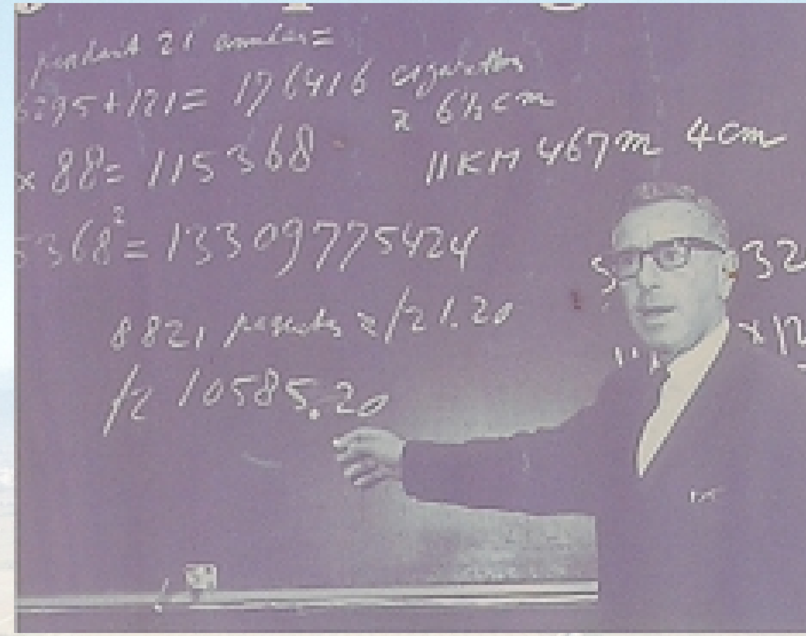


At the beginning

Wim Klein

Dutch mathematician
(1919-1986)





The last man who was faster than computers

He was able to calculate the 19th root of a 133 digit number in head. He was much faster than the computers of his era regarding basic arithmetic calculations.

After having worked many years at CERN Wim Klein gives a last demonstration of his amazing talent for mental calculations

Data and computing needs

- ITF (interferometer) output sampling rate 20 kHz
- Hundreds of auxiliary channels
- Various analysises on 4 and 16 kHz
- cc. 150 day science time/year
- cc. 10.4 MB/sec data flow rate
- cc. 160 TB data / year / IF

This amount of data has to be transferred, recalibrated, downsampled, processed, analyzed (several time)

The amount of data is not overwhelming for storages available today, but the **arithmetic density** of algorithms used for the analysis varies from less significant (online analysis) to practically infinity (CW searches).

No uniqe/optimal computing solution for all, various approaches has to be used: such as Cluster, Grid, HPC and Many Core computing

Data transfer – solution

• **Goal:** Transfer ITF output and auxiliary channel data to permanent storage **"in-time"**.

• **Currently:** Various solutions for different destinations:

- SRB (Storage Resource Broker, Lyon)
- BBFTP (RFC1323 implementation, CNAF)
- LDR(Ligo Data Replicator, Ligo sites).

Some of them is obsolete other not supported any more. Do the job but miss several feature, for ex. data bookkeeping.

• **Use of grid tools:** FTS (File Transfer Service) was tried, but was found to be unnecessarly complicated/rigid for this task.

• **Solution:** lcg client tools with custom transfer and command queues (TQ/CQ), called the DT server. Under final realistic stress test: Will be used soon for Cascina – Bologna transfer.

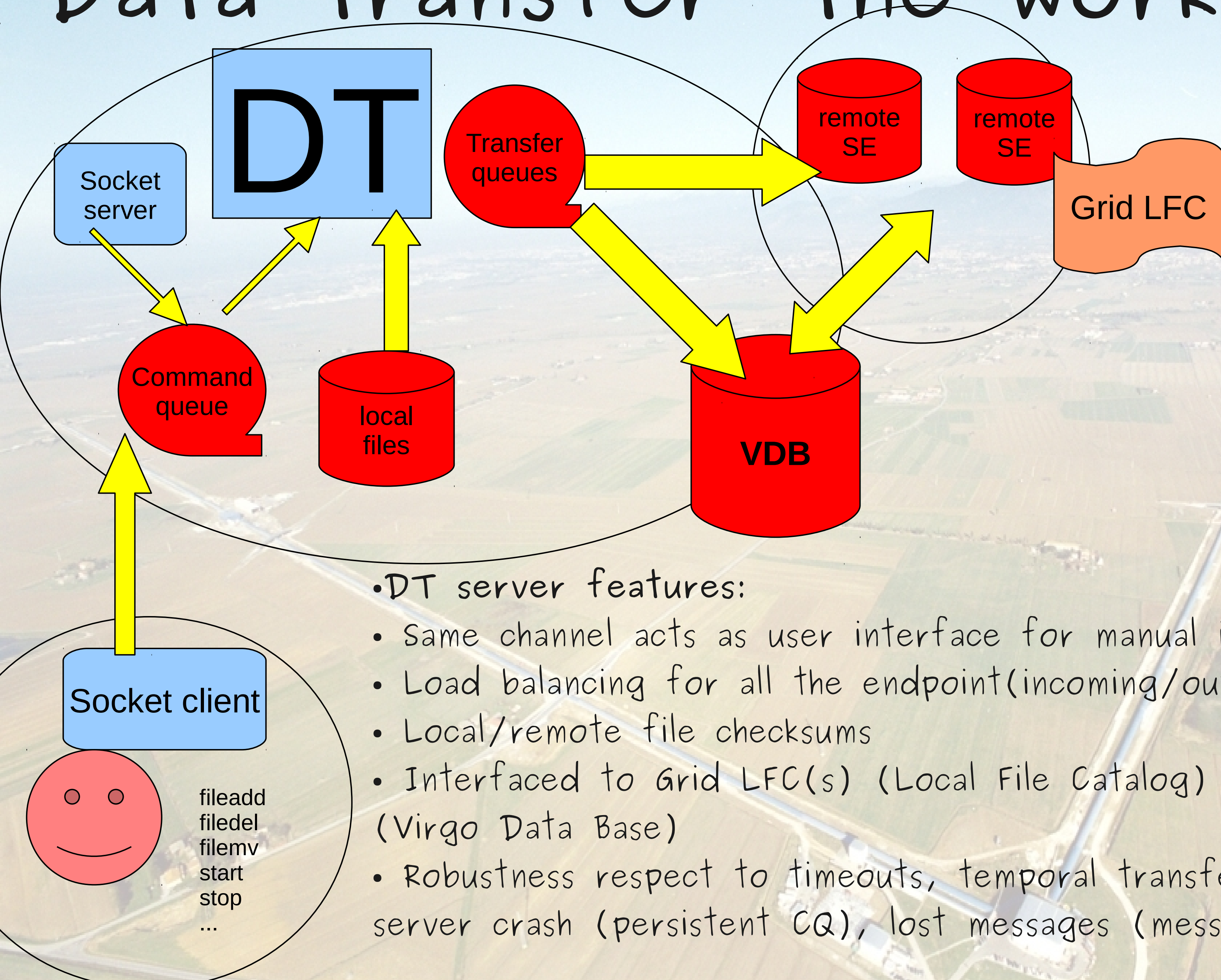
DT server features:

• Standard interface among the most important CCs worldwide.

• LFCs provide transparent access to distributed data

• Implements communication channel with DAQ via socket server/client, new files are automatically added.

Data transfer-the workflow



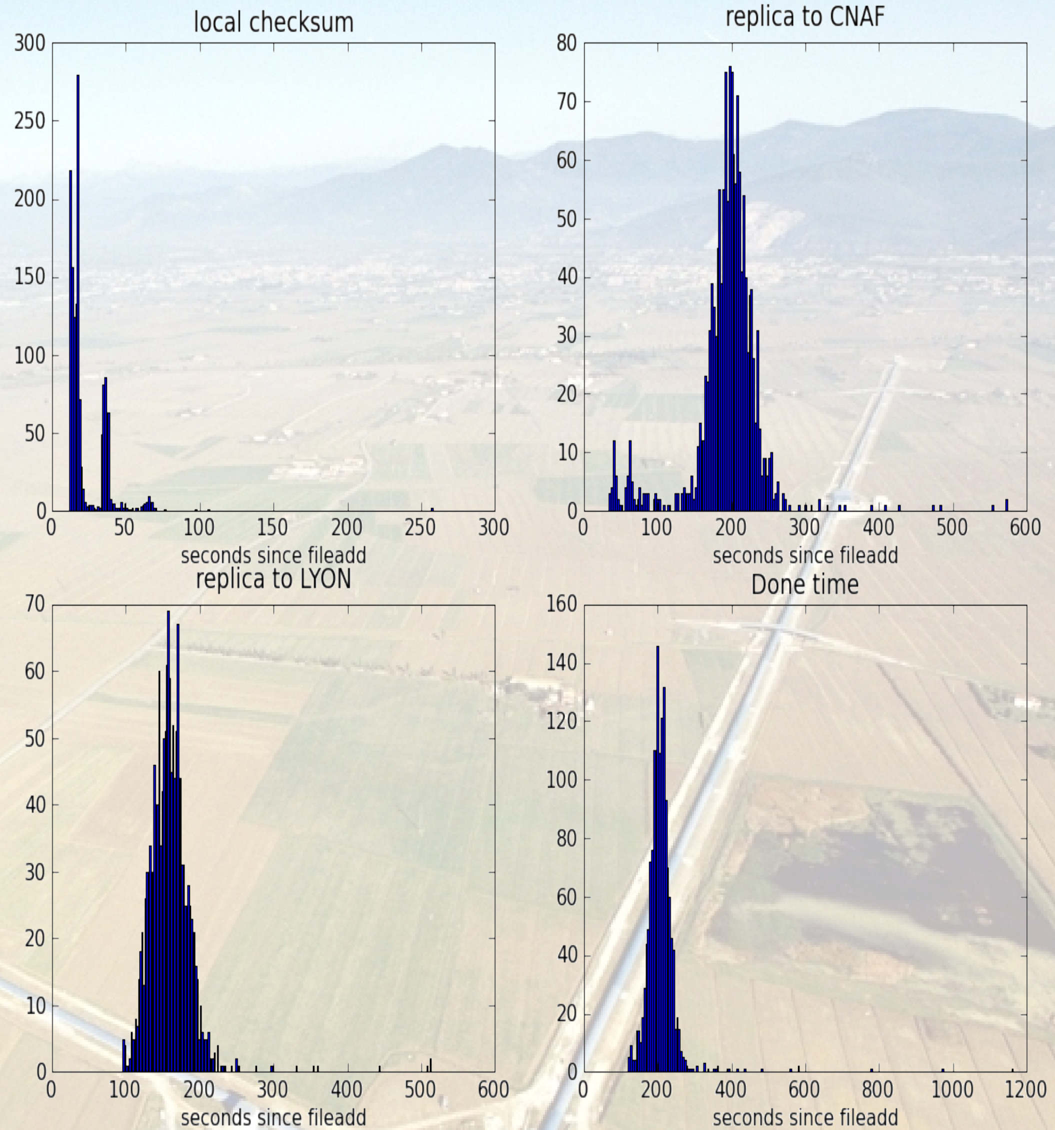
•DT server features:

- Same channel acts as user interface for manual intervention
- Load balancing for all the endpoint(incoming/outgoing)
- Local/remote file checksums
- Interfaced to Grid LFC(s) (Local File Catalog) and VDB (Virgo Data Base)
- Robustness respect to timeouts, temporal transfer failures, DT server crash (persistent CQ), lost messages (message register).

Data transfer – results

Transfer results:

- Data can be replicated to any number of remote storage with an SRM interface.
- Average completion time from "fileadd" to Done is ~ 200 sec.
- Stable operation over time



Analysis on clusters

• The Ligo and Virgo experiements closely working together, collaborations share and exchange knowledge, software, data and computing resources.

• Various analysis pipelines are running on dedicated – and so far – isolated LDG (Ligo Data Grid) clusters.

• Interconnection of clusters just happened.

• These pieces of softwares is strongly bounded to one specific batch scheduler (Condor) and require locally shared file systems which makes the port difficult to other non-Condor, non-local architectures.

The solution is

• either detach the software from Condor and make it possible to use non-local non-shared file systems with various batch schedulers (LSF, Torque, PBS, SGE, etc..)

• Create a Condor pool – EGI Grid gateway

• or set up a Pilot Pool in order to provide a homogen interface towards this hierarchical, relational workflows

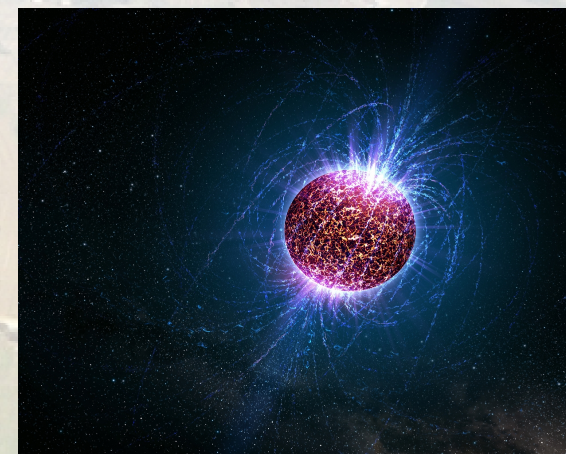
The Virgo Virtual Organisation

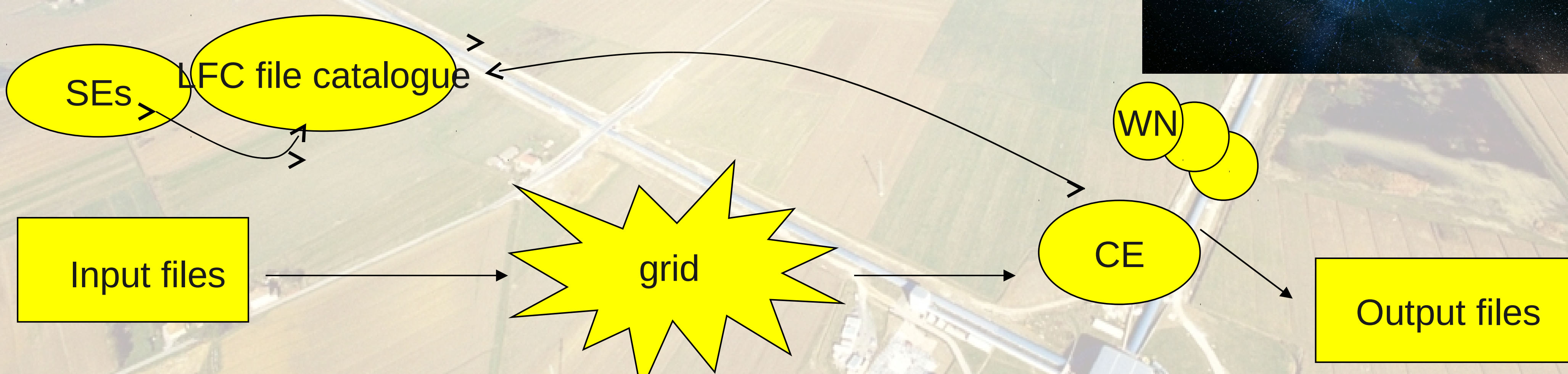
- Virgo is one of the 211 Virtual Organisation of the EGI Grid.
- Some Virgo analysis is running smoothly on the Grid since quite a long time !
- Ideal for CPU intensive jobs crunching "small" amount of data.

- Resources: ~34 sites, ~9000 CPU, ~50 user, several hundred TB storage

CW analysis on the Grid

- **CW (Continous Wave)** analysis searches periodic signals emitted by pulsars, isolated NS (neutron stars). In the all-sky searches we want to explore a portion of the source parameter space as large as possible (position, signal frequency and its derivatives)
- Such kind of search is computationally bound, even if very efficient (altought non-optimal in sensitivity) methods have been developed
- In particular, the Hough transform stage of the hierarchical procedure developed in Virgo is the most computationally expensive.
- The analysis scheme:





CW analysis on the Grid

• Members of the Virgo CW (Continous Wave) group ported their software for the use on Grid.

•*In this search computing power directly translates to sensitivity ... and the problem is "embarassingly paralellizable".*

Advantages:

1) More resources
2) Independent of local batch systems (BQS, LSF)
3) LFNs (Logical File Name)s make it possible to forget physical storage places and architectures.
4) Experiment software installation becomes easier.

Workflow:

1) Jobs submitted from a native Grid UI (User Interface) to sites supporting the Virg VO, where
2) pre-uploaded, replicated and registered input files are processed
3) Results and various outputs are also registered on Grid storage and downloaded t the UI on completition of the job group.

Features

1) A typical analysis run consist tens of thousand jobs running in several sites
2) A ***Supervisor*** program has been developed to manage the analysis:
    1) Basic check of Gird status and site, WMS availability
    2) Creation of submission scripts on the fly
    3) Job submission and job status monitoring
    4) Output file management
    5) Re-submission, failure handling
    6) Checkpointing in case of Supervisor crash

CW analysis on the Grid

- In the past we have successfully analyzed C6 and C7 Virgo Engeneering Run data over a 1000 Hz band (CQG, 24, S491, 2007)

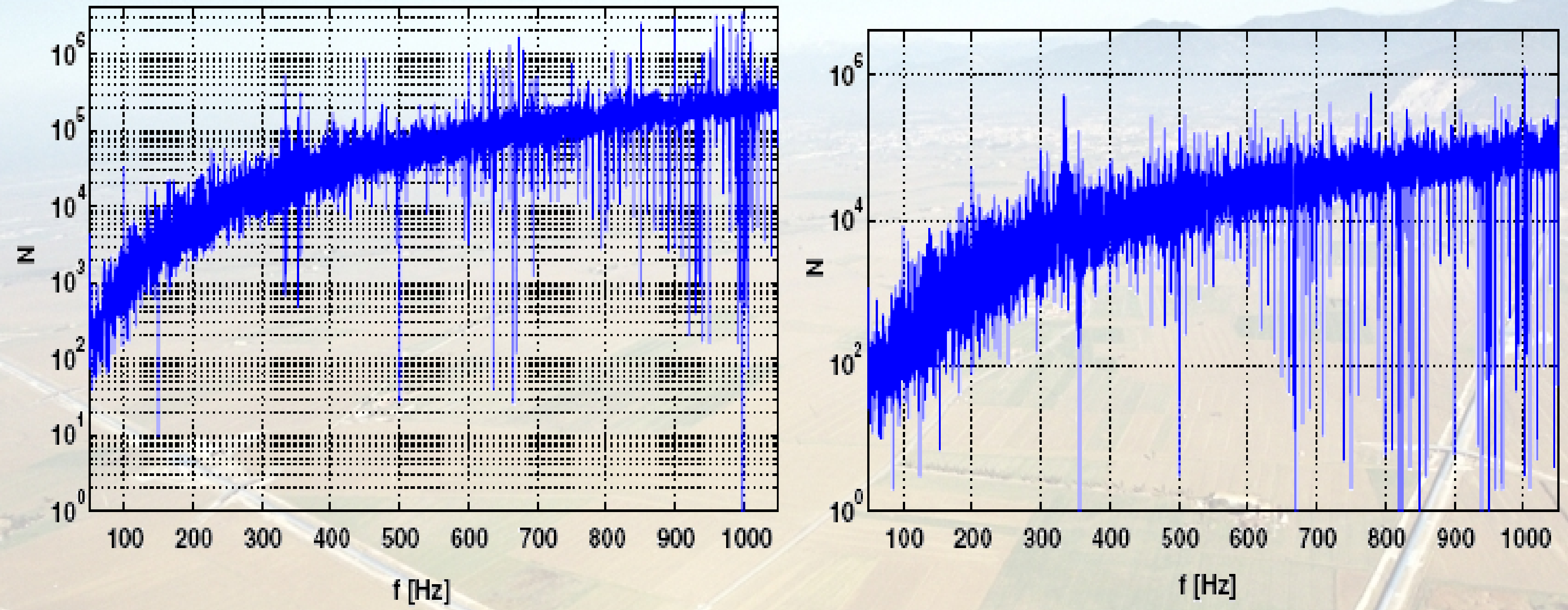


Figure 4. Frequency distribution of candidates *of the first order* with null spin-down for C6 (left) and C7 (right).

- Also VSR1 data has been analyzed, mainly for testing the Supervisor program.
- The Grid has also been used for Monte Carlo simulations of the population of NS (MNRAS, 359,1050 (2005))

CW analysis on the Grid

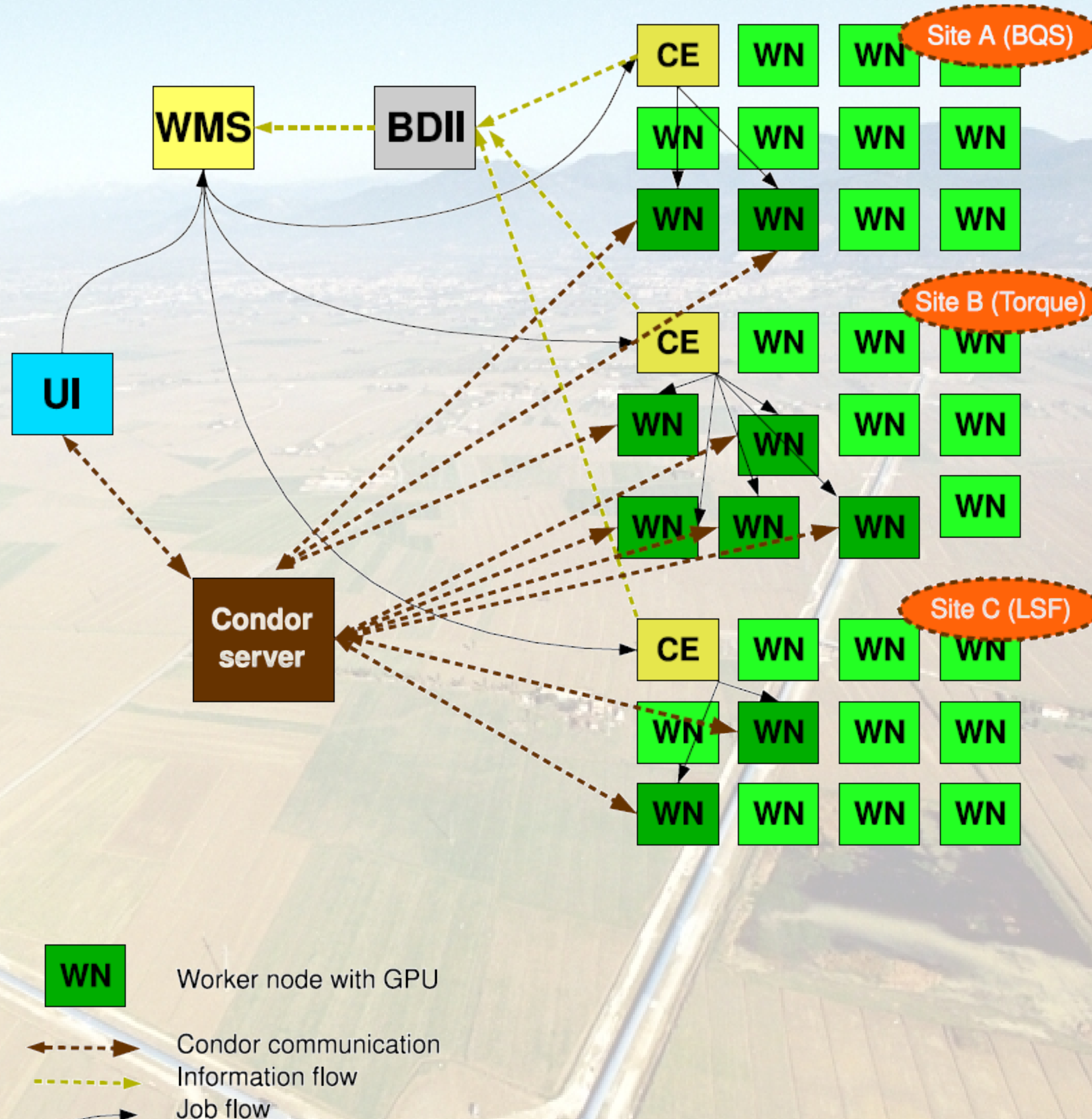
- We are working on a more efficient implementation of the Hough-transformation.
- When ready will be ported to Grid
- VSR2 and VSR3 data is going to be analyzed
- The port of the algorithms to GPU is also started

There is still a lot of candidate to analyize:

The Virgo Pilot Pool – I

The Virgo Pilot Pool properties:

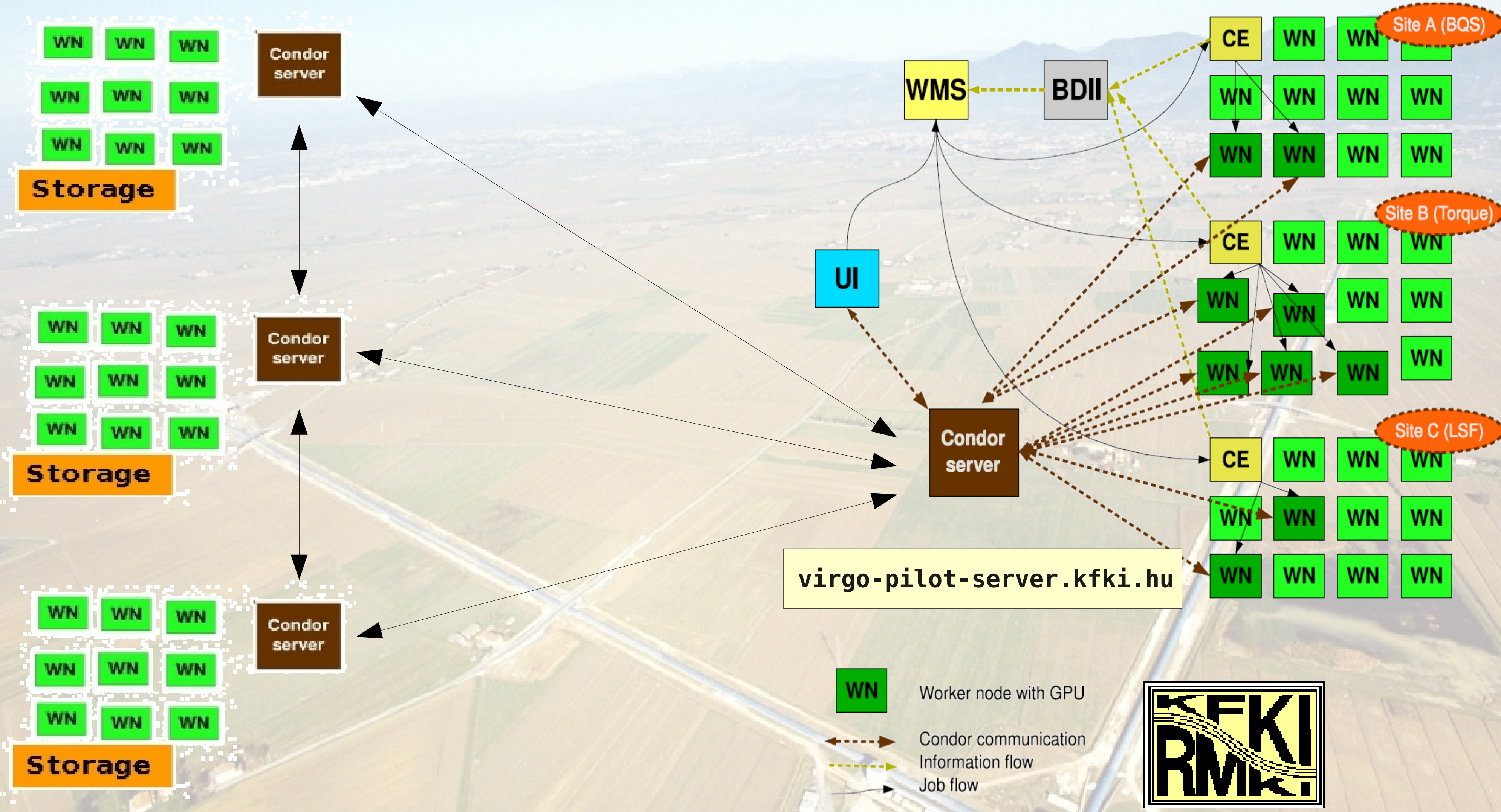
- Homogen infrastructure over the inhomogen Grid
- Less administrative interaction/delay
- User transparent mechanisms
- Low latency submission
- Global priorities
- Late-binding to resources
- No stucked-in jobs
- Improved job failure rate due to pilot prechecks
- Interactive login
- Smooth interaction interoperability with LDG/OSG.



The Virgo Pilot Pool – II

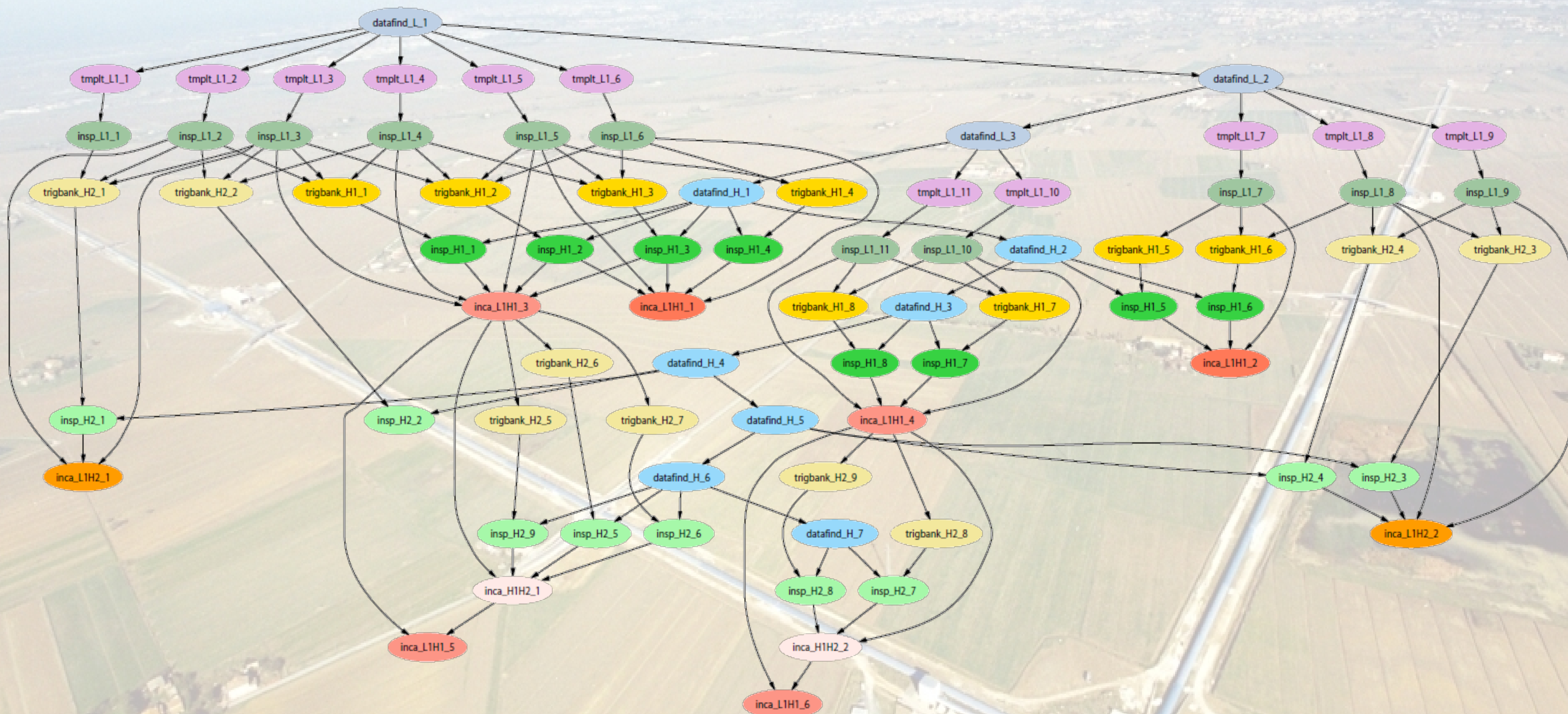
Ligo clusters

Virgo EGI sites



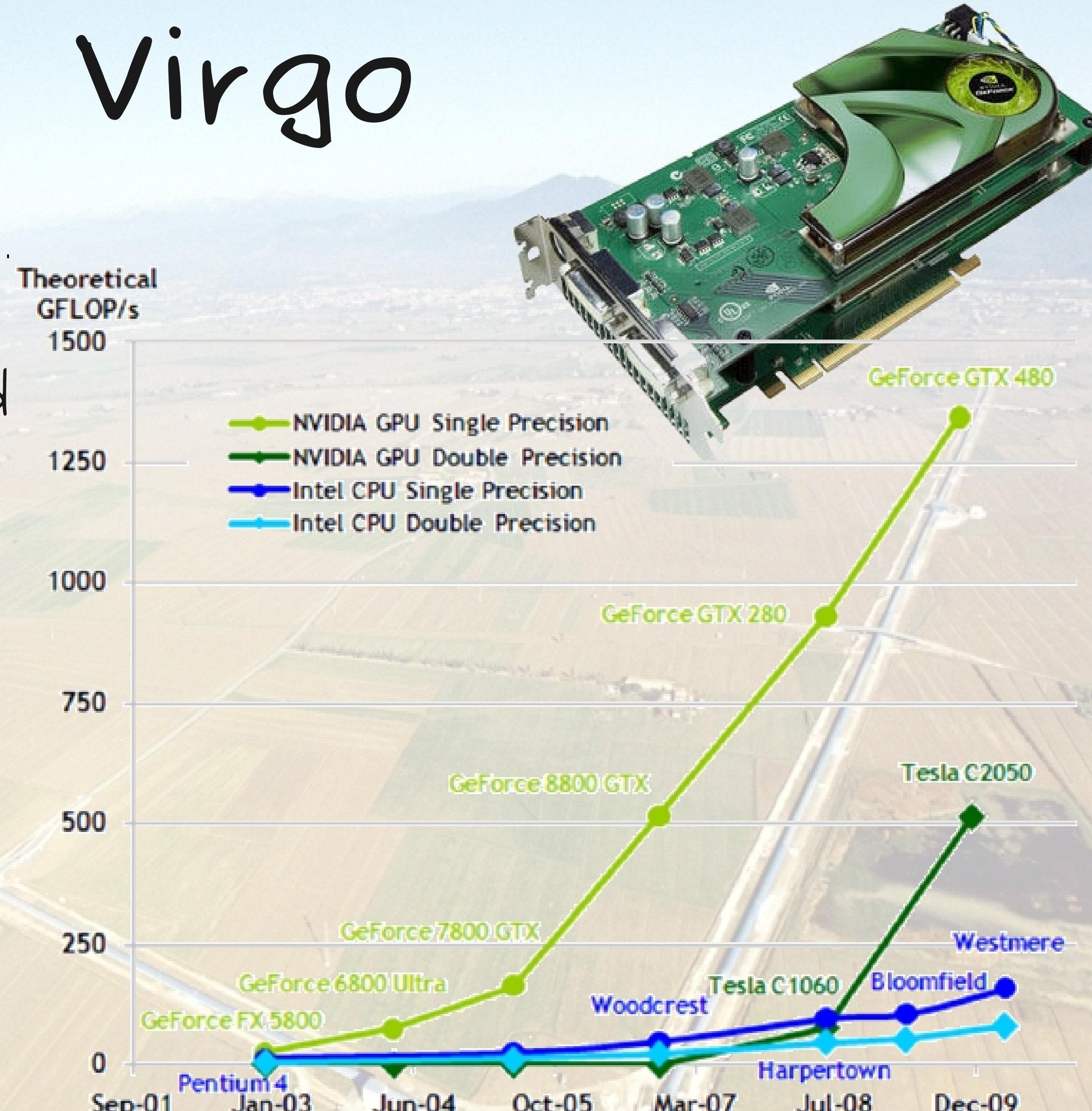
The Virgo Pilot Pool – III

- Mapping of abstract workflows like DAGs/DAXes to the Grid is now easily possible with the Virgo Pilot server.
- Complex and relational workflow handling is missing from Cream/WMS.



Many Core Computing @ Virgo

- Motivation:
- A lot of algorithms used in gravitational wave analysis can be massively parallelized and/or a lot of independent calculations can be performed at the same time.
- **BUT:** Many Core computing developes drastically, protocols, programing languages, standards emerging slowly -> Platform independece can be crucial
- **Ongoing works:** CBC pipeline, CW searches like F-statistics, Hough method

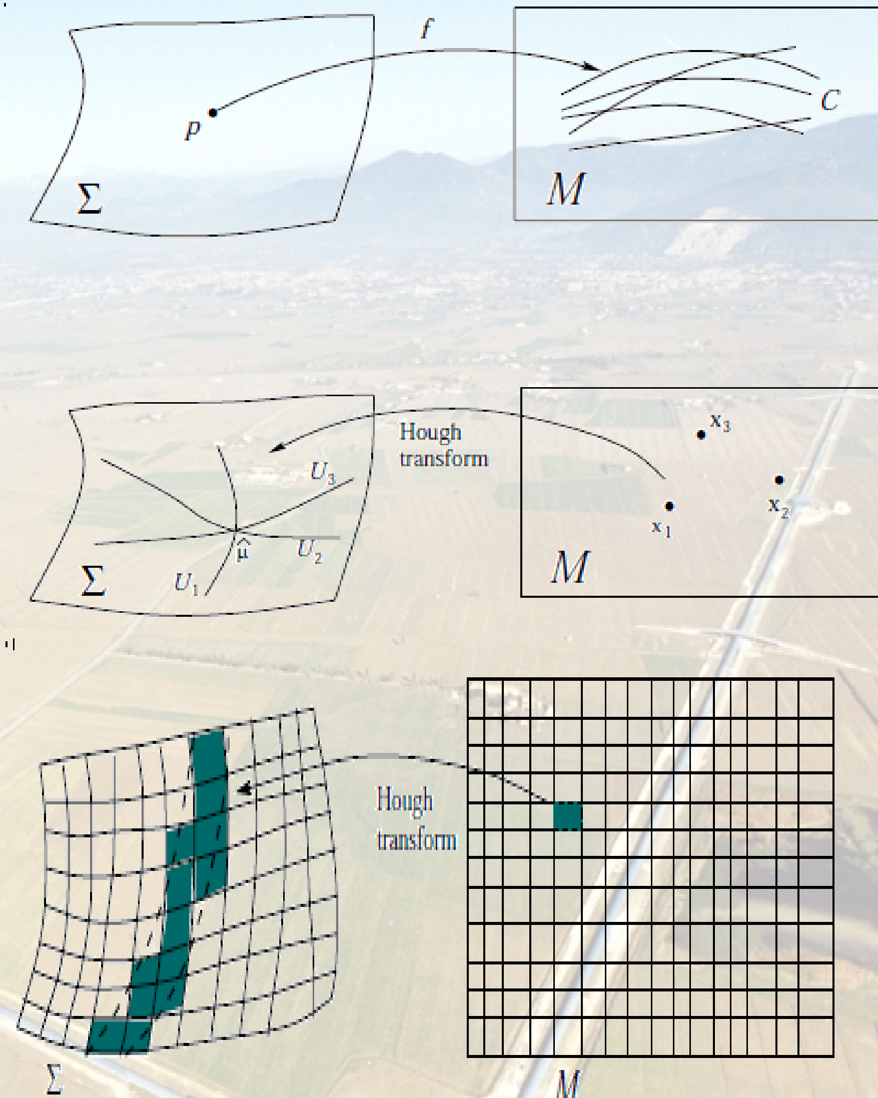


MCC @ Virgo –Hough trans.

The Hough – method

- is a robust parameter estimator method widely used – among others – in pattern recognition, artificial vision and in gravitational wave searches
- Used in the incoherent step of some CW search pipeline.
- The algorithms contains several massively paralelliazble steps, such as
  - Fast Fourier Transformation
  - thresolding on 2D surfaces
  - multi dimensional independent parameter scan

- Ideal for GPU!
- OpenCL implementation under development by RMKI Virgo Grop

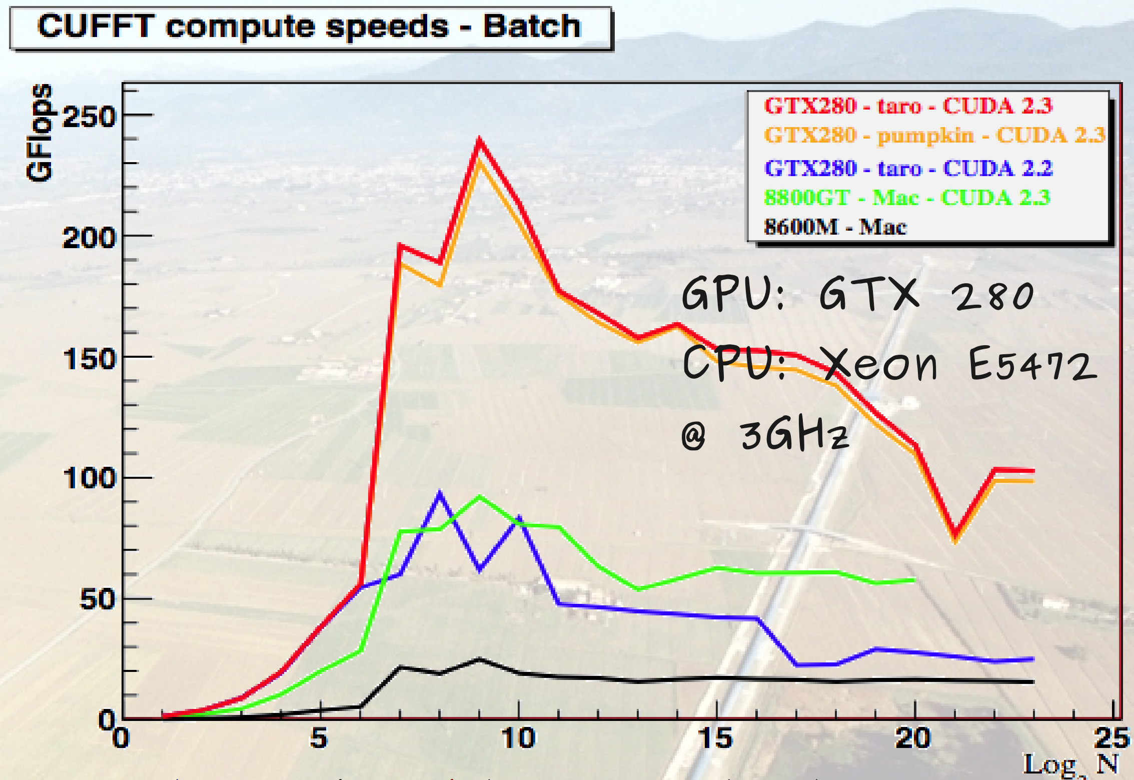


MCC @ Virgo
Polinomial CW search

University of New Hampshire

NIKHEF

Code block:
- FFT of data – done once, insignificant time cost
- Create Filter – fast
- FFT of Filter – few needed
- Complex multiply FilterxData
- Inverse FFT of result (slow, many needed)
- Reduction – find maximums
- Shift filter (very fast)
- Next filter
- Next data segment



GPU: GTX 280
CPU: Xeon E5472 @ 3GHz

Blue and Red loop completely on GPU, key to speed-up: batch mode processing of multiple operation.

30x gain

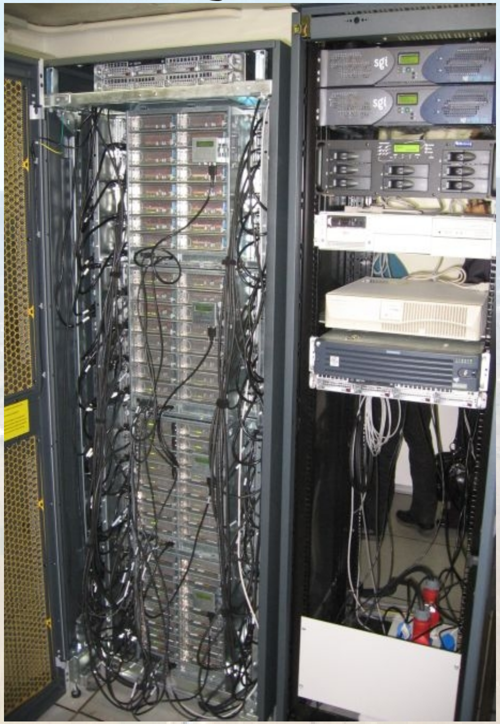
HPC @ Virgo ?

KFKI RMKI
HPC machine

- The machine (SGI Altix ICE)
- 512 CPU (64 x 8)
- Intel Xeon, X5365 @ 3 GHz
- 16 GB mem/8 core
- 10 TB Raid5 disk storage
- Infiniband (40 Gb/s) blade connectvity

- Software
- RedHat Enterprise Linux
- Condor and PBS batch scheduler, SGI MPI, Tempo cluster management, OpenMPI, SGI ProPack, etc...

- Power
- Water cooling
- 37 KW peak power
- 7 TeraFlop computing capatcity

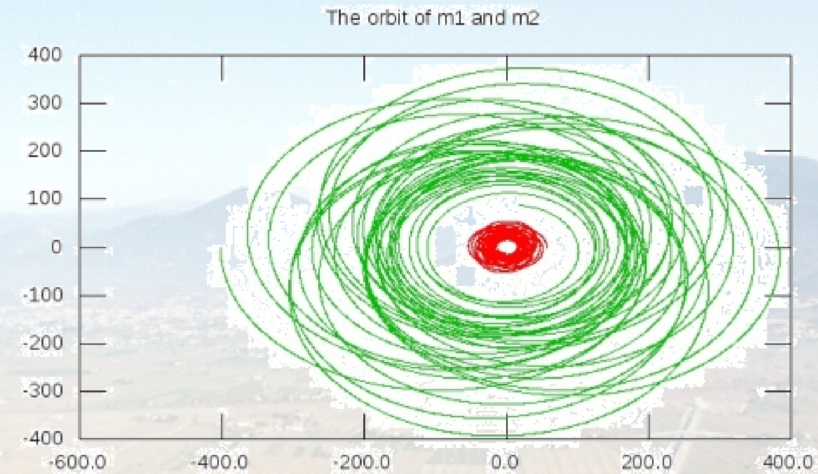


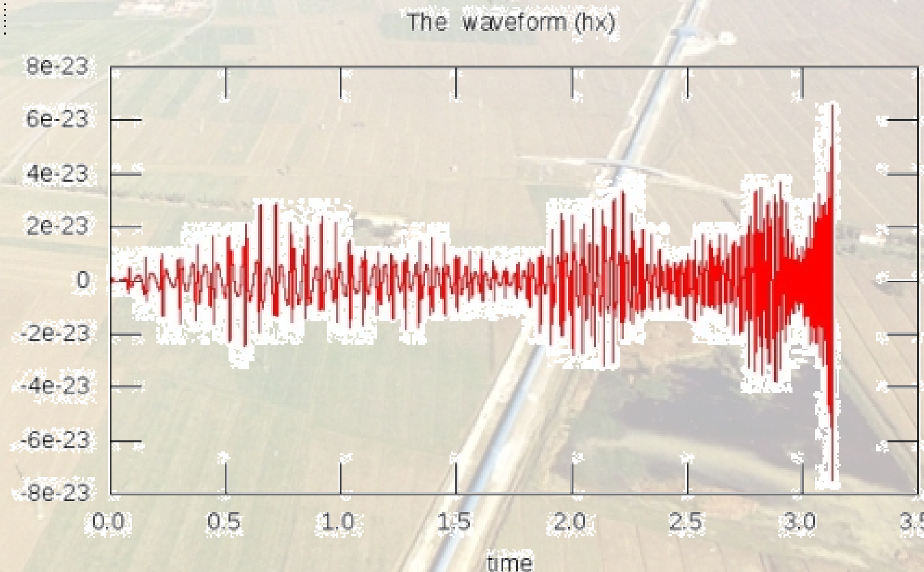
- Purchased for another (medical, GenaGrid) project, but
- available for Virgo Grid on temporal basis
- Used as a simple PC cluster
- No real advantage found over PC cluster for gravitational wave applications

MCC @ Virgo – Num. rel.

The RMKI num. rel. group

- working on various numerical relativity problems such as
    - spinning,ecc. grav. wave templates
    - radiation backreaction on Kerr background
    - neutron star stability models
    - radiating processe highly non-linear Einstein-matter systems
- mathematical problems like
    - determinant calculation
    - vector (field) multiplication
    - BLAS routines
    - multi (N > 1000) dimensional matrix vector contraction



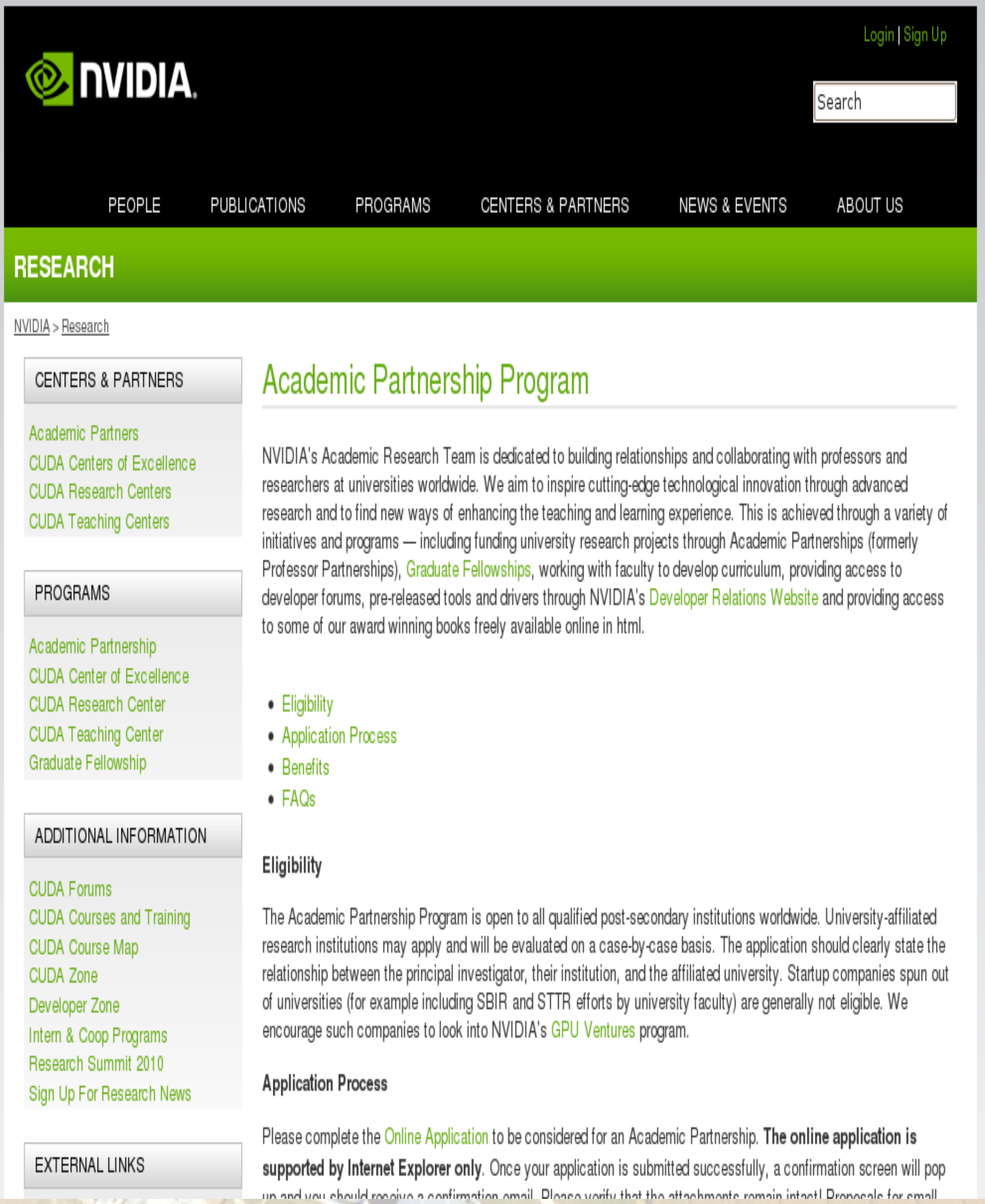


Work has just began to port GridRipper (http://www.kfki.hu/~cspeter/gridripper) code to GPUs.

MCC @ Virgo – nVidia PPP

The nVidia Professor Partnership Program

- aimed to provide research groups with hardware for evaluation studies of various computational problems to GPU architectures.
- The RMKI Virgo Group successfully applied for this program
- got 2 Tesla C2060 card
- will be used for benchamark
    - experimental gravitational wave search
    - and numerical relativity application

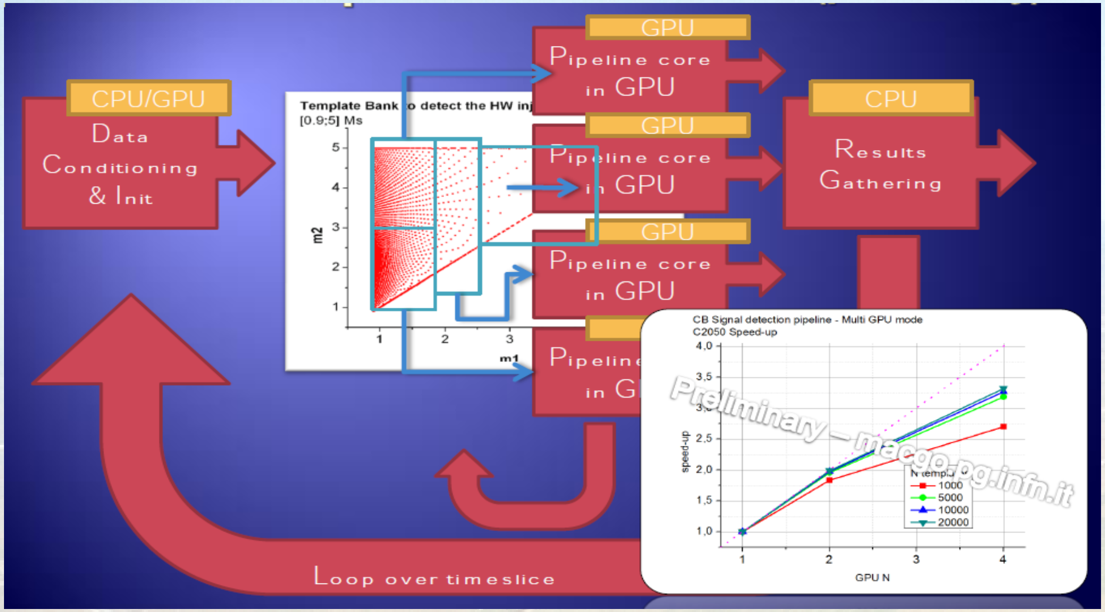


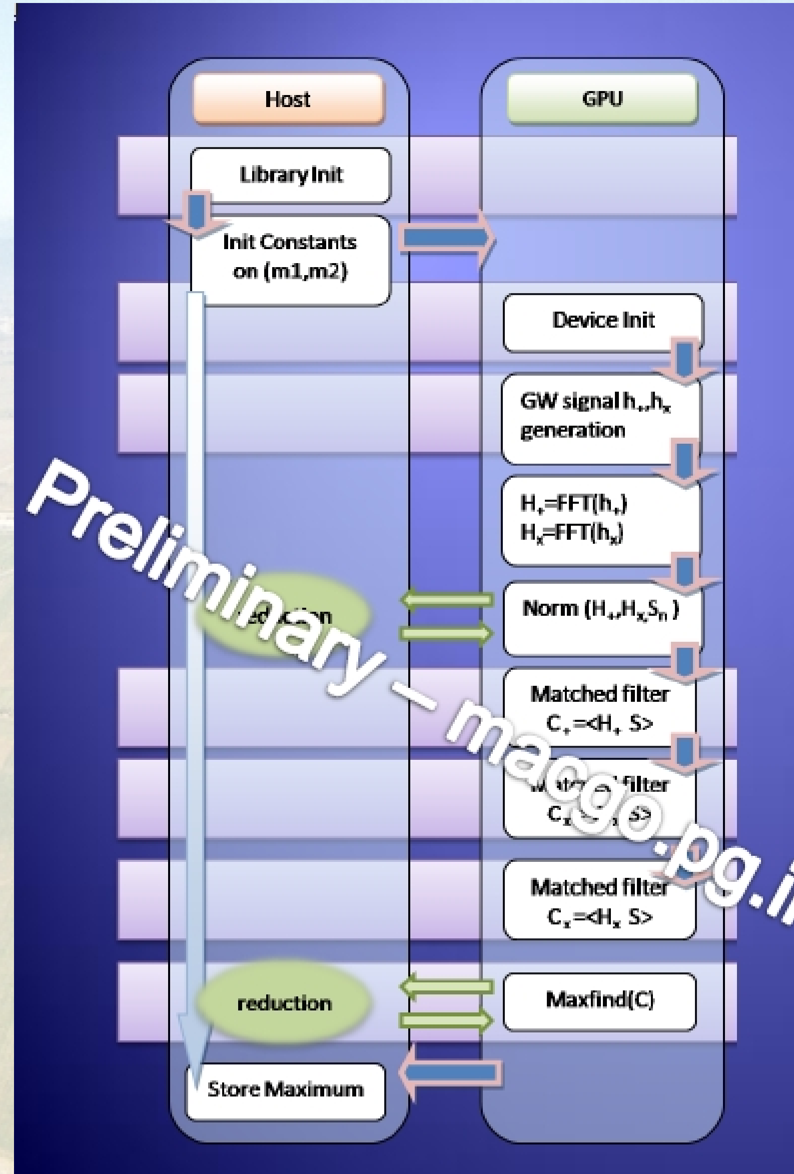
MCC @ Virgo/ET – MacGo

The MaCGO project (Manycore Computing for future Gravitational Observatories) (http://macgo.pg.infn.it)

- is linked to the Einstein Telescope activities it's about mapping forecasted computational needs to manycore architectures
- exploring and developing dedicated algorithms for single/multi GPU device architectues
- exploring and use of OpenCL (manycore programing standard)
- production of the first release of numerical library for GW-DA based on many-core architectures
- and signal analysis pipeline for Compact Binary sources

MCC @ Virgo/ET – MacGo





Preliminary results:

- Price normalized gain of the same algorithm respect to CPU is **x50.** (GTX 275 and C1060) (30 template/sec 2^20 length matched filter)
- New GPUs with Fermi Core contribute with an additional factor of **x4.** (120 template/sec)
- Multi-GPU pipelines an other factor of **x3.** (360 templates/sec)

Summary

- Gravitational Wave searches are extensively using a wide range of various computing resources
- Different algorithms have different optimal mapping to computing architectures
- **Grid tools** help in
  - data transfer issues
  - and in running non-relational, relatively simple analysis workflows
- **Dedicated clusters** with shared file system for user jobs and hierarchical, relational workflows
- **Pilot pool** provides a solution for mapping complex workflows to **Grid**.

- Highly paralellizable algorithms are very well suited to **GPUs**
- Various future **manycore architectures** and programing languages will be of great help for GW-DA !

Thank you for your attention !

Extra slides – Grid: what is missing

- **Data transfer:** FTS is far too heavy for Virgo – size collaborations. Lack of flexibility and support for Postgres/MySQL backend. Steep learning curve.
- **Matchmaking:** A *client side* matchmaking plugin would allow easier pilot approacs and late bind to resources without having to reinvent/rewrite matchmaking for the pilot job dispatcher.
- **POSIX file access** to Grid Storage (Classic SE) and global transparent file system. – (DPMfs or NFS4 interface, etc..)
- Large **scheduling delay** / status update in some circumstances.
- **Poor user interface**, no middleware provided job submission, managment tool.