Astrophysical Research in INAF and Broadband Networks

Claudio Vuerli, Fabio Pasian [1]
Mauro Nanni, Giuseppe Maccaferri [2]

[1] INAF – Unità Sistemi Informativi (SI)
[2] INAF – Istituto di Radioastronomia (IRA)

Outline

- Astrophysical Applications
- GRID projects in INAF
- GRID and the Virtual Observatory
- New perspectives opened by the GRID: an integrated working environment for astronomers
- GRID and network infrastructure
- New GRID infrastructures

Astrophysical Applications

- Main classification
  - Simulations
    - Simulate various types of astrophysical phenomena (e.g. galaxies/stars formation, evolution and decay) or space missions (e.g. the ESA Planck mission simulation)
    - Limited input data set, massive CPU usage, output data set can be huge
  - Data analysis and data reduction
    - Involve large sets of data and use well consolidated software (e.g. iraf, midas, etc.) or software specifically written for particular space missions (e.g. Planck) or astrphyisical instrumentation (e.g. software for a first calibration of data collected through a specific instrument)
    - Extremely demanding both in terms of I/O and CPU

Astrophysical Applications

- Main classification
  - Data classification
    - Once data have been collected and processed they can be optionally clustered to discriminate among different data sources. Classification is made on the basis of a set of characterization parameters (e.g. the morphology of objects)
    - Input data sets to be processed generally huge, CPU requests and size of output sets generally limited
  - Data presentation
    - Output data are presented to final users. Visualization tools and techniques are largely used in this case
    - Huge input data sets and extremely demanding in terms of CPU power

Astrophysical Applications

- The scenario before the Grid
  - Application runs at home or at supercomputing centers
  - Parallelization on multi-CPU mainframes or CPU clusters (beowulf systems)
  - I/O traffic and data traffic among application threads restricted to local area networks
  - Input and output data transfer made off-line through ancillary batch applications or removable storage media

Astrophysical Applications

- New scenario with the Grid
  - Applications submitted as jobs to the GRID
  - Parallelization still takes place at local area networks only (no parallelization supported by current middleware implementations)
  - I/O traffic now involves the WAN
  - Input and output data transfer made at run-time when the job execution starts
    → **limited network bandwidth introduces severe bottlenecks and sometimes forces to adopt workarounds (i.e. the case of Planck simulations)**

GRID Projects in INAF

- GRID.IT
  - Multidisciplinary project funded through the FIRB
  - INAF participated with three applications…
    - A1: access to astrophysical databases and data reduction
    - A2: porting in GRID the VST pipelines
    - A3: monitoring of astronomical instrumentation via GRID
  - … and three institutes

GRID Projects in INAF

- DRACO
  - A project of the Italian astrophysical community funded through a COFIN
  - A Grid portal thought and designed for astrophysical applications has been developed
  - Various astrophysical applications have been ported in GRID
    - Definition of astronomical data analysis workflow on a service-oriented Grid architecture using BPEL (Manna et al.)
    - Data Mining and Massive computing for GRID at Naples-Salerno Unit
    - Astronomical Data Reduction in Data Grid
    - Various kinds of evolutionary models

GRID Projects in INAF

- Planck simulations in EGEE
  - Planck simulations is an officially supported pilot EGEE application since November 2004
  - The project aims at porting all Planck simulations software in GRID
  - Planck application now in EGEE-II
  - A Planck VO, led by INAF Trieste, is in place. It groups together Grid sites located in six European countries

GRID Projects in INAF

- Future projects and plans
  - Complete the porting of the Planck software in EGEE and possibly on other infrastructures
    - The GRID not only for simulations but also for the processing during the operative phase
  - Bring the G-DSE+QE into production
    - This allow us to bridge the Grid and the Virtual Observatory, a key point to make the GRID really useful and fully exploitable by astronomers

GRID Projects in INAF

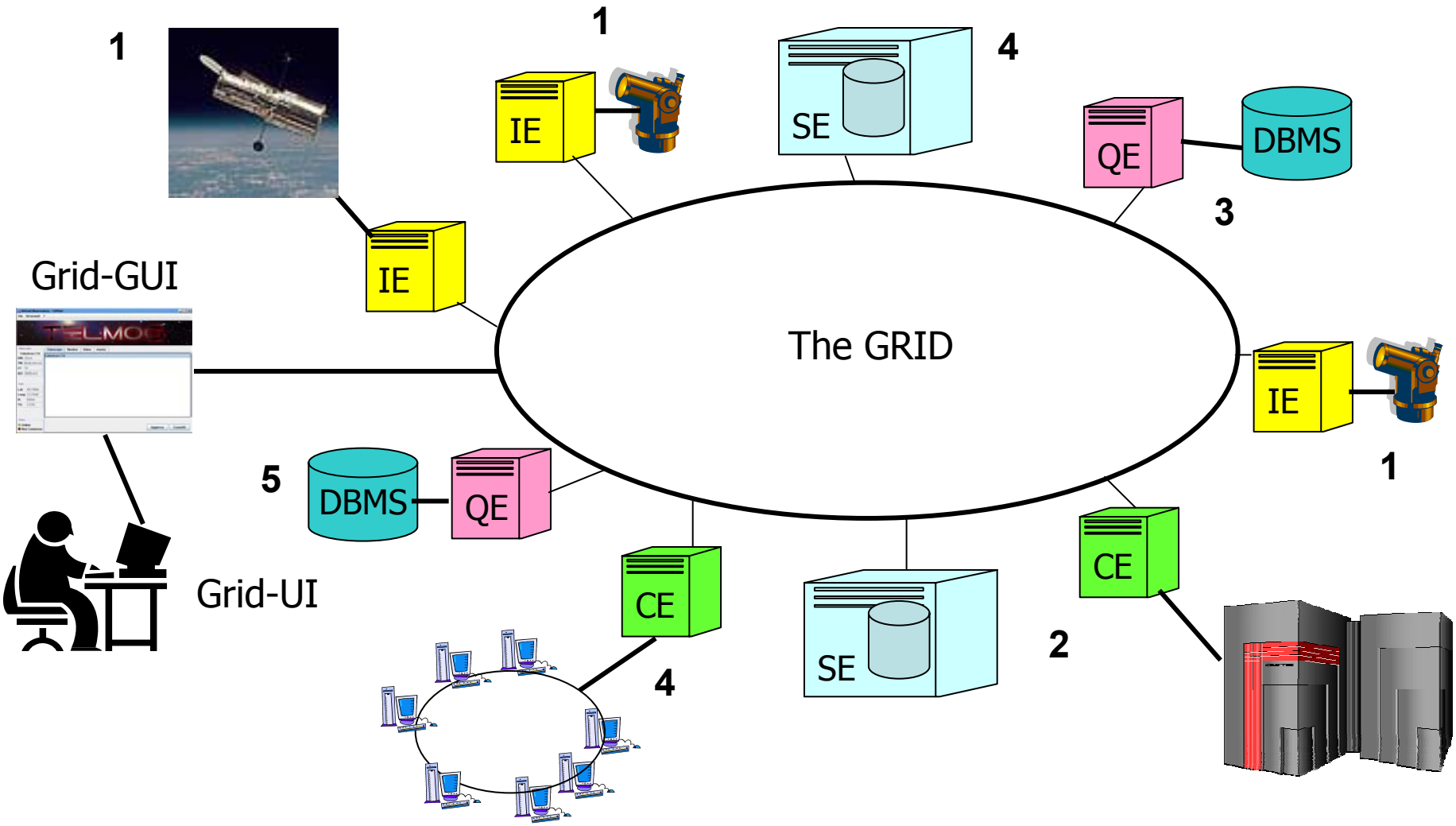
- Future projects and plans
  - Bring the G-ICS+IE into production
    - Telescopes, astronomical instrumentation and sensors in GRID closes the loop and leads to a fully comprehensive integrated working environment system for astronomers
  - Disseminate the GRID among astronomers both at national and European level and train them when necessary
    - Extend the INAF VO to include all INAF institutions
    - Increase the portfolio of GRID-enabled astrophysical applications

GRID and Virtual Observatory

IVOA: formed to enable the international utilization of astronomical archives as an integrated and interoperating Virtual Observatory

- DBs in IVOA expose a standard, uniform, agreed common interface
- IVOA Registries as a DNS for astronomical DBs
  - Location and content information
- Users access the VO without worrying which specific federated DBs they are using
- G-DSE+QE bridges GRID and the VO → VO resources in GRID
  - Users start from the GRID, access one or more VO Databases, use shared GRID resources to process extracted Data, then they optionally feed one or more DBs with the processing results

The Integrated Environment



Demonstration

- Data acquisition by observing radio sources with the IRA radio telescopes located at Medicina (Bologna)
- Data sent to the user immediately (monitoring) and/or fed to a Database for future browsing and exploitation
- The demo will capture only a piece of the future scenario, but well highlights our requirements towards the underlying network infrastructure

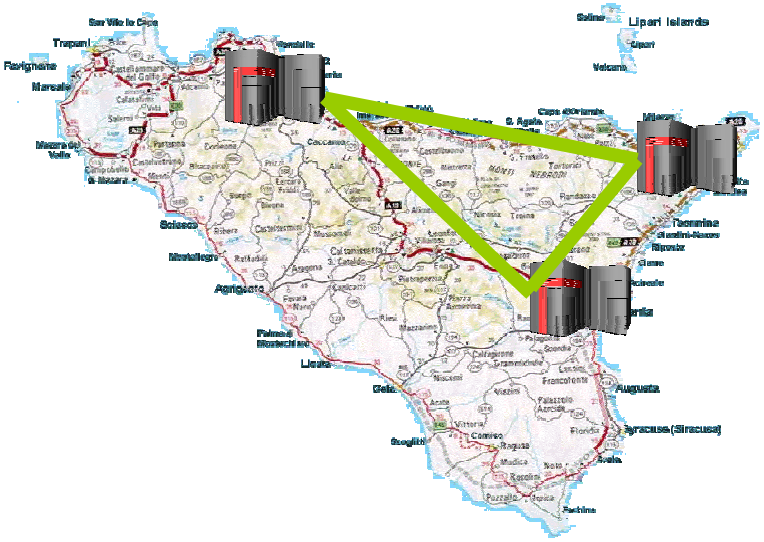
GRID and Net Infrastructure

- The integrated working environment is feasible only if an adequate network infrastructure is in place…
  - To support the flow of data between two grid elements (CE-SE, QE-CE, QE-SE, QE-QE, SE-SE, IE-CE, IE-QE, IE-SE)
  - To support the flow of data between threads when the parallel processing will be supported by the GRID middleware

New GRID Infrastructures

- New GRID infrastructures, shortly in place, represent a new chance for astrophysical applications
  - TRIGRID VL
  - CyberSAR

New GRID Infrastructures



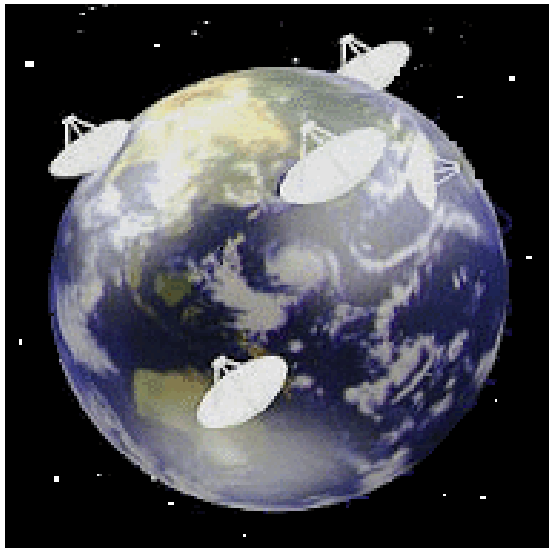
- The example of **Trigrid VL**
- It has a strong impact at national level
- Two INAF institutes involved in it

- **CyberSAR**: Proposal by Cosmolab, a mixed public-private Consortium, it is a strategic project of Sardinia Region.
- Main Objective: To create a hardware/software platform for the High Performance Computing and for simulation of processes of various types

e-Vlbi: Le Antenne in rete

Per fare un radiotelescopio:

- **Grande** come il nostro pianeta
- Funzionante in **tempo reale**

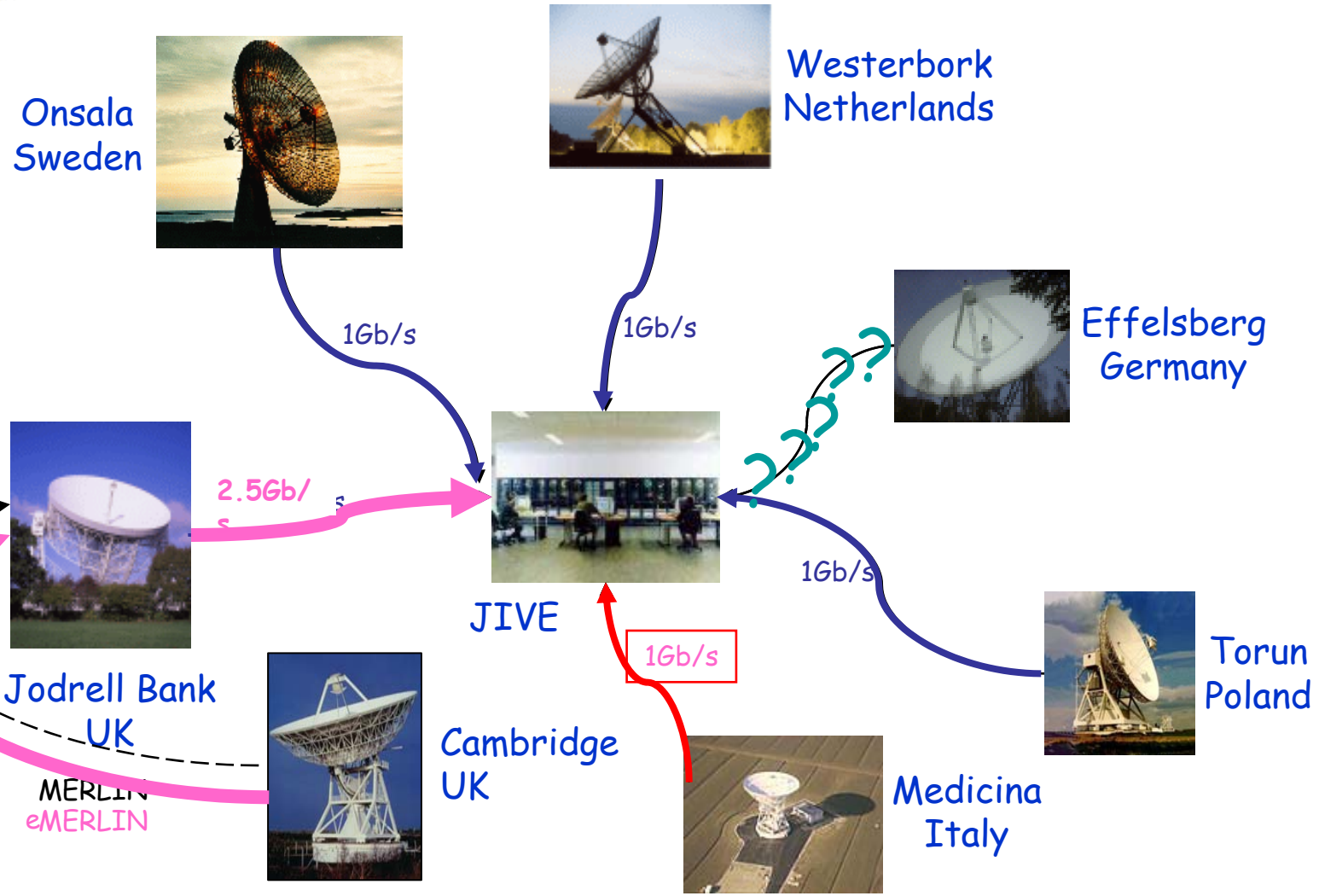


2006 IDEA Awards
idea.internet2.edu

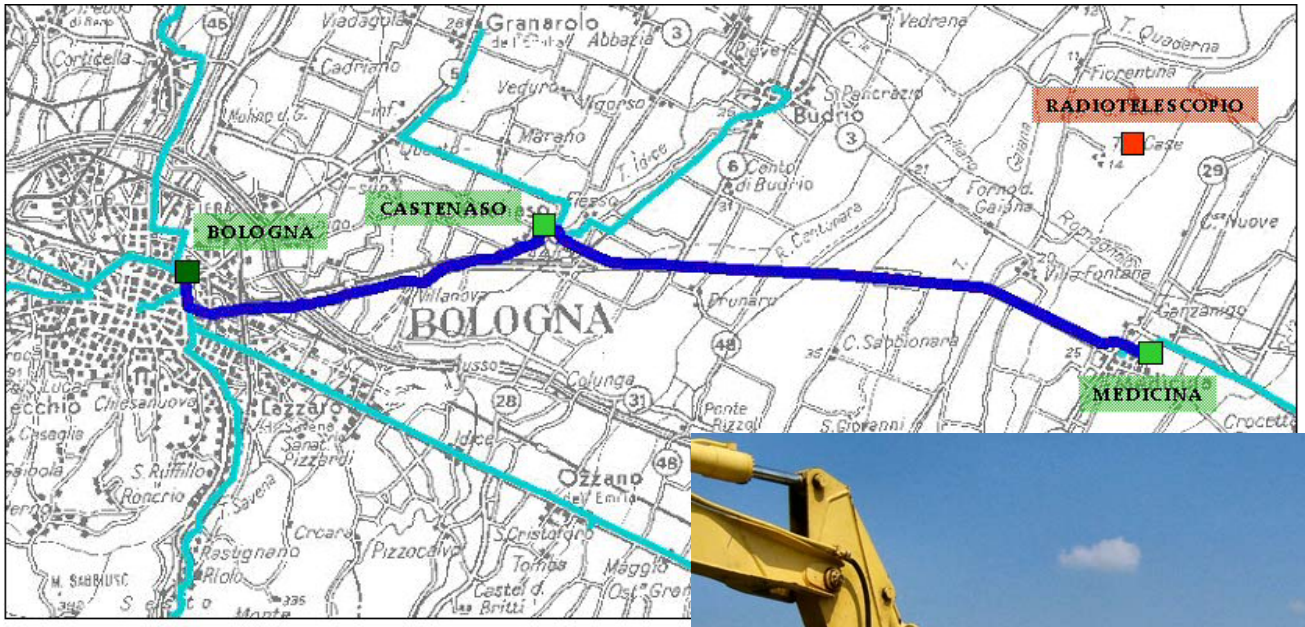
e-VLBI project wins IDEA award:

The e-VLBI project, led at JIVE, has received a prestigious award for its progress in high speed data transfer at the inaugural IDEA awards ceremony held on April 21st 2006.

RadioTelescopes connections

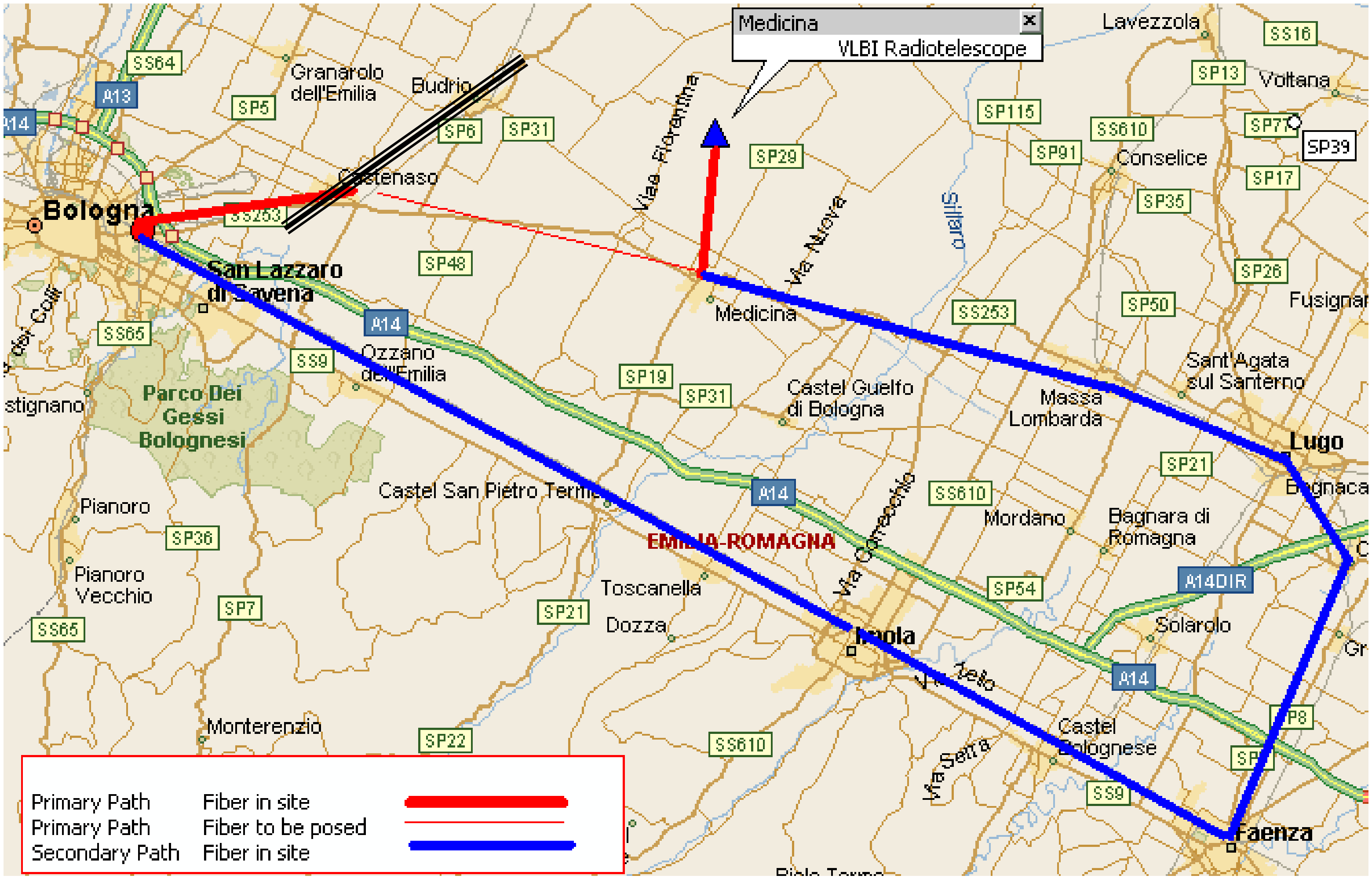


Il Collegamento di Medicina



Tra Medicina e il Radiotelescopio (6Km) e' stata posata una fibra nel Maggio 2005

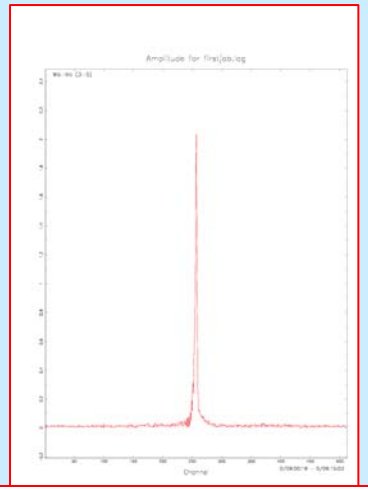
Le FS regionali dirottano l'e-Vlbi



Il collegamento con l'Olanda (senza Jumbo frames !!)

Gennaio –Febbraio

Prove tecniche e tuning



Marzo: FRANGE

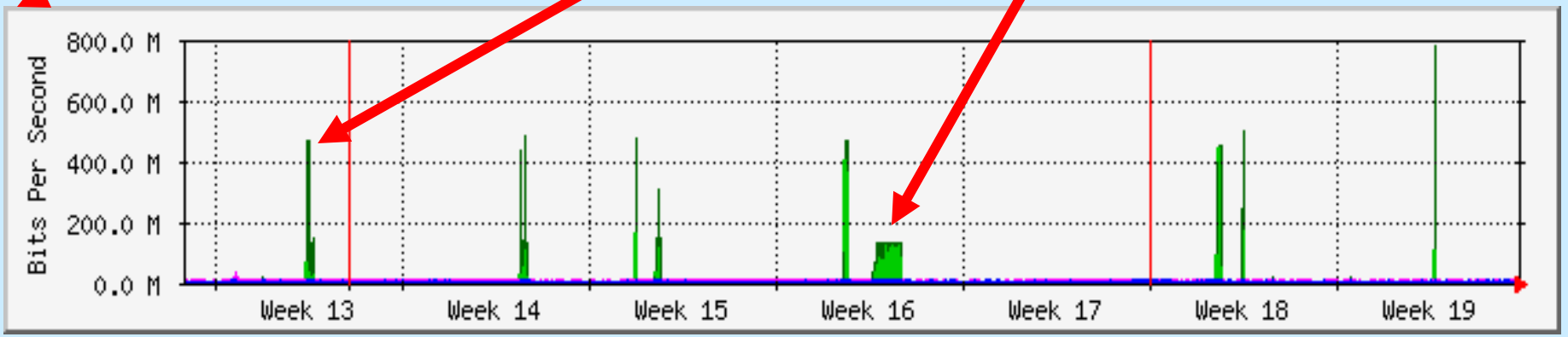
20 – 21 Aprile

Run Osservativo

128 Mbit

Test

750 Mbit



e-Vlbi in Italia

Medicina (BO) 32m : Operativa a 128-256 Mbit/sec. Sul percorso definitivo avra' una dark-fiber. (1 Giga e oltre)

San Basilio (CA) 64m:
In costruzione. Accordi con la regione Sardegna per il collegamento.

e-Vlbi in Italia II

Noto (SR) 32m : Non collegata

Si sente la mancanza di una infrastruttura di rete nel Sud della Sicilia (e-vlbi, Nemo, Universita')

La correlazione dei dati

Per correlare i dati che provengono da piu' antenne

- **Ieri**: Un supercomputer con harware dedicato
- **Oggi**: tra 100 e 1000 CPU e "software furbi"

Si possono costruire, di volta in volta, "radiotelescopi ad hoc" con le diverse antenne

Si possono realizzare i correlatori in tecnologia Grid ???

E-VLBI e Grid

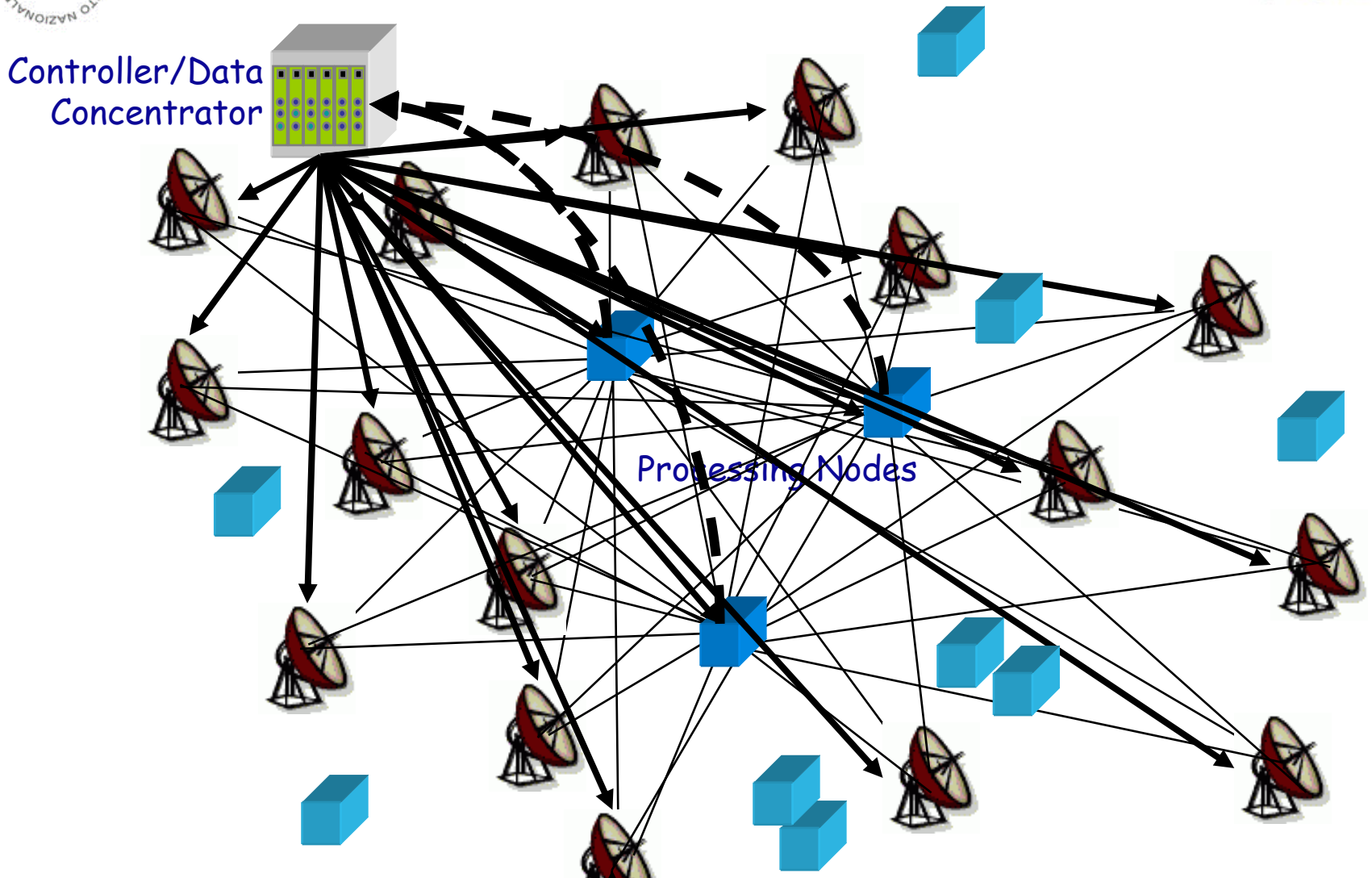
argomenti contro la "GRID correlation":

- Un sistema dedicato e' generalmente piu' efficiente di un array di processori general purpouse
- Il grid e' sostanzialmente un "batch", la correlazione e' un processo fortemente "real time".
- Processi distribuiti richiedono duplicazione dei dati ai singoli nodi.

argomenti a favore della "GRID correlation":

- Processori dedicati sono MOLTO costosi mentre i general purpose processors sono MOLTO economici
- Il middleware richiesto e' gia' in parte sviluppato per il GRID
- Se i dati sono "time-sliced", i dati possono essere gestiti come nel caso di un sistema centrale anche se i calcoli possono essere distribuiti sulla rete.

Correlation: GRID Computation task



E comunque .....

Occorre tanta tanta tanta banda disponibile . . . .

. . . . durante i giorni in cui vengono condotte le osservazioni (3-5 giorni/mese … per ora :-})