Future challenges for the VIRGO experiment

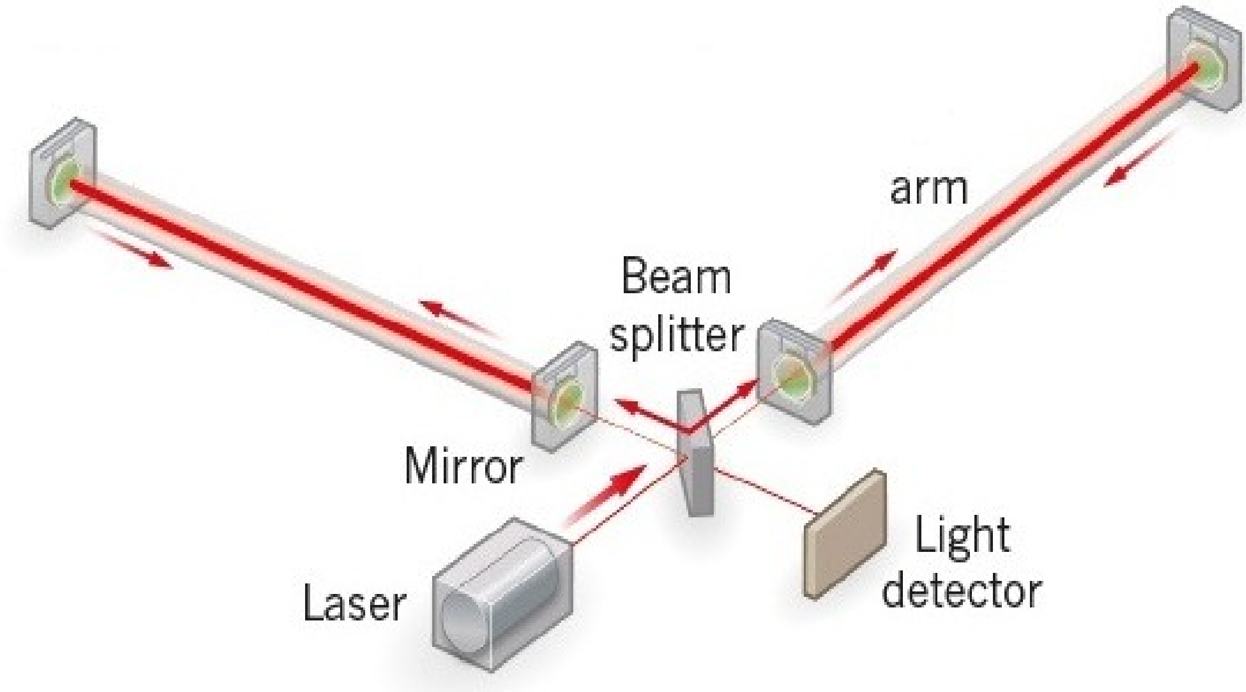
Lazzaro Claudia - INFN PD

LVC Collaboration

LNGS 22 May 2017

Outline

- Virgo and LIGO Collaborations
- Data analysis of interferometers network
- Analysis algorithms and computing resource requirements/usage
- Future plans

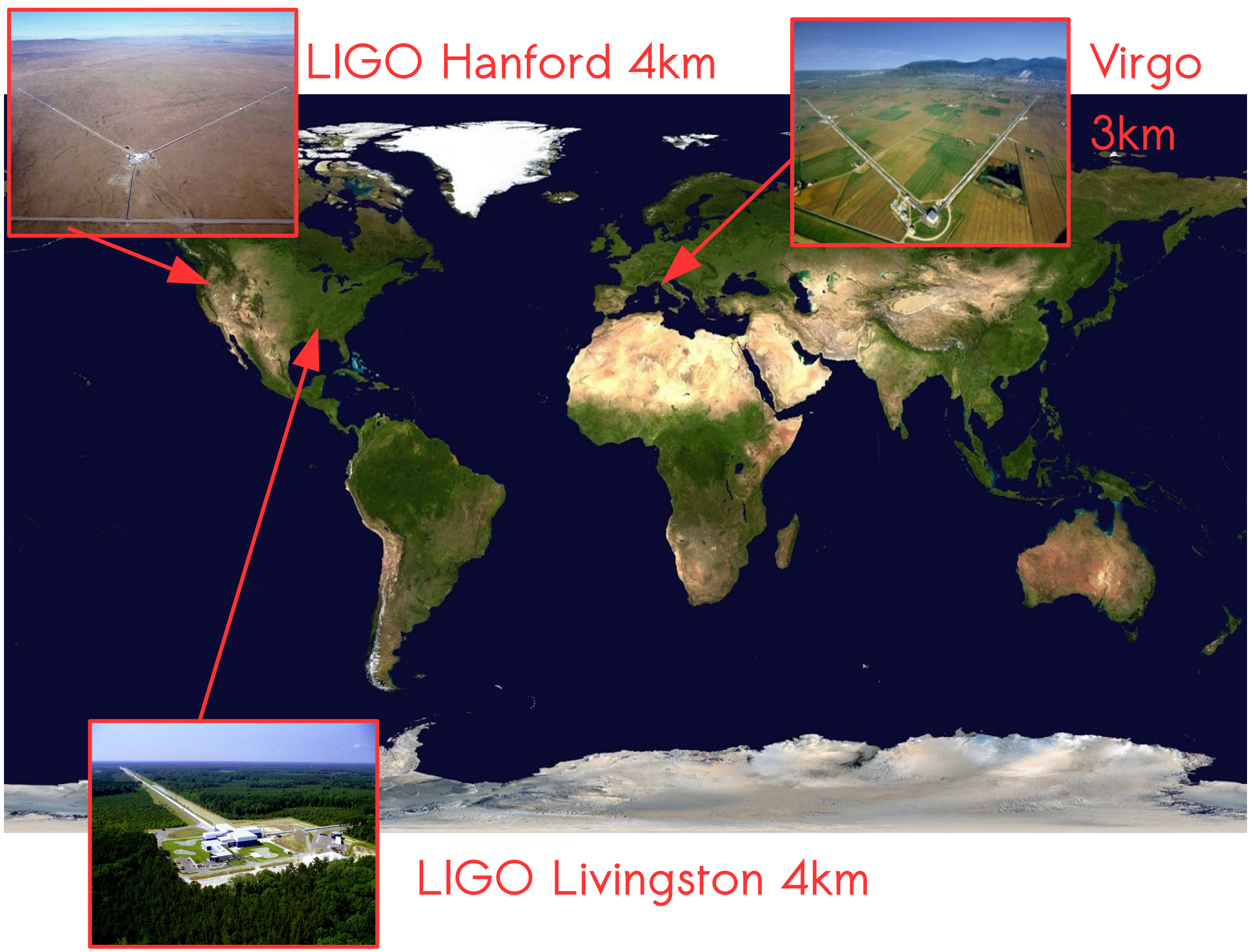


LVC - Computing centers

Agreement signed between Virgo, LIGO and GEO600 detectors to share data

Analyse simultaneously data from a network of detectors allows to enhance scientific output and reduce background

Data and software shared in the LVC collaboration



Virgo:

- EGO - Cascina detector site
- CNAF - Bologna, Italy (INFN)
- CCIN2P3 - Lyon, France (CNRS)
- Nikhef
- Poland

LIGO:

- LLO, LHO: Detector sites
- ATLAS AEI Hannover
- CIT - Caltech
- SU - Syracuse
- NEMO
- ...others

Data analysis workflow: from detectors to physical results

Data taking, monitoring, commissioning and calibration

- Gravitational channel h(t): time series signal, sampling rate 20 kHz
- Thousand of auxiliary channels saved with different sampling rates (range: >Hz to few kHz)

Virgo:

- Continuous stream of data (data flow ~3TB/day), to be transferred (CNAF, CCIN2P3)
- The computing center at Cascina is dedicated to data production, commissioning, detector characterization

Data analysis workflow

Data taking, monitoring, commissioning and calibration

Detector characterization and data quality →

Analysis of the auxiliary channels performed:

- Both low latency (minutes latency) and offline
- Both single interferometer data analysis, and analysis of all network data

Data analysis workflow

Data taking, monitoring, commissioning and calibration

Detector characterization and data quality

GW searches & low latency GW searches

- Simultaneously analysis of the network data (h(t) channel)
- Different analysis (searches) have been developed to address different sources and signals

- Pipelines are built on several algorithms, therefore they require different computing resources and input/output data management
- Low latency searches have been implemented (since few minutes to hours depending on searches and pipelines) to promptly identify GW candidates and send GW alert to EM partners

Virgo data Transfer

Stream of data (raw data) from Virgo to CNAF/CCIN2P3 ~ 3TB/day

Nominal transfer requirements: 50 MB/s (2 X the nominal data acquisition)
Typical throughput reach during tests about 65MB/s toward CNAF and 25MB/s toward CCIN2P3

(LIGO requirements is: 15MB/s in science mode, 26MB/s in commissioning)

~ 1/10 of raw data (h(t) channel +selection of auxiliary chan.) are transferred to LIGO:
The procedure to transfer the Virgo data to CNAF and CCIN2P3 has been updated; it uses the Ligo Data Replicator (LDR) data transfer framework all over in the LIGO-Virgo collaboration and performs reliable data transfer. (LDR is based on Globus and GridFTP):

Last week tests: data transferred to CNAF ~66MB/s and 28MB/s to Lion, analysis on the performance and complexity are ongoing

Low latency data transfer Virgo↔ LIGO: it includes h(t) channel only, peer-to-peer transfer handled by MBTA pipeline

Astrophysical searches & pipeline

CBC search

Transient signals

Coalescing Compact Binary Systems (Neutron Star-NS, Black Hole-NS, BH-BH): Strong emitters, well modeled

Well known signals
Template search

Asymmetric Core Collapse Supernovae weak emitters, not well-modeled ('bursts'), transient
Cosmic strings, soft gamma repeaters,
pulsar glitches

Not known signals
burst search

Burst search

Continuous signals

Cosmological stochastic background (residue of the Big Bang, cosmic GW background, long duration)
Astrophysical stochastic background

Stocastic search

Spinning neutron stars (known waveform, long/continuous duration)

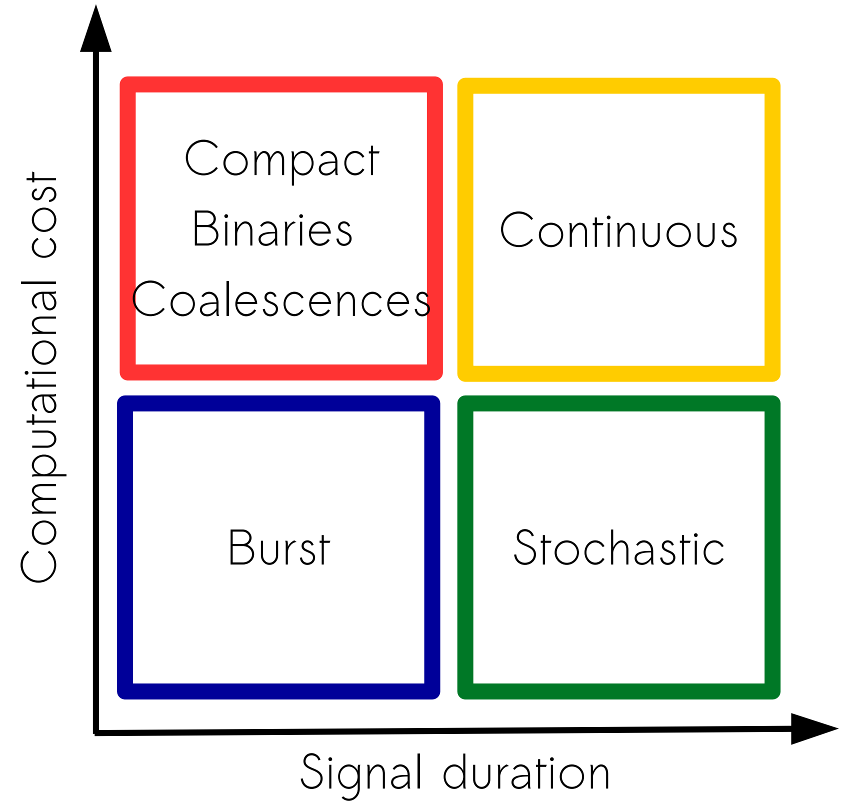
known signals

Continuous search

Pipeline computational cost

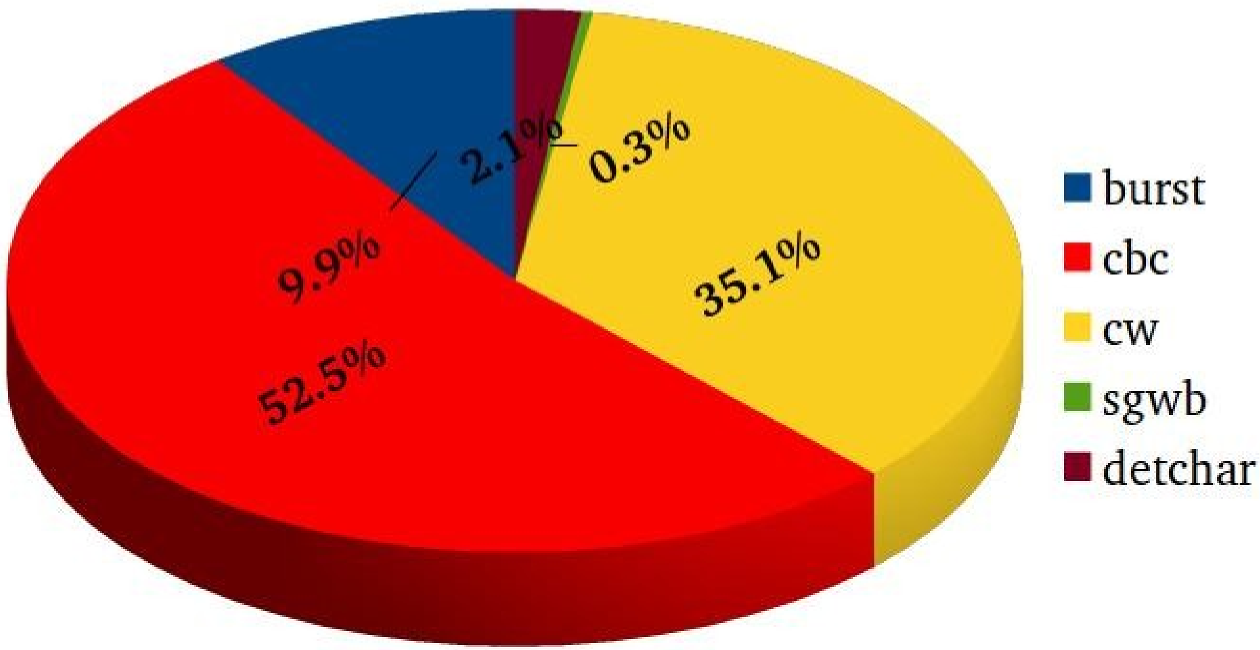
Different algorithm pipelines are implemented for each group/astrophysics search

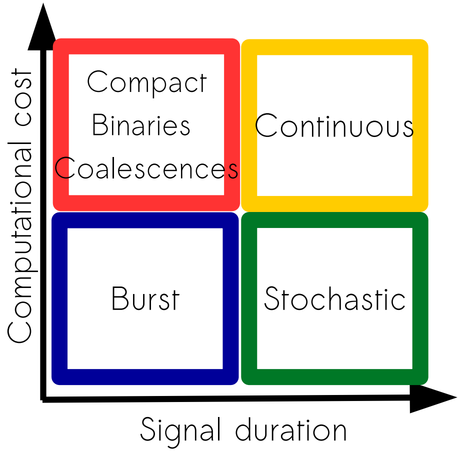
- Continuous Waves (CW): Fourier transform/matched filtering, frequency Hough transform
  all-sky searches aim to explore source parameter space as large as possible
  → high computational resources needed
- CBC: template searches based on match filtering algorithm (big bank of simulated signals needed).
  Fourier transforms, matched filtering, χ^2 calculation, parameters estimation
  → high computational costing pipelines
- Burst: wavelets, Fourier transformation, energy evaluation, likelihood calculation, parameters estimation searches are planned
  → medium/high computational cost
- Stochastic → Low computational costing pipelines
- Detector characterization



Pipeline computational cost

Amount of CPU core hours last 52 weeks in LVC





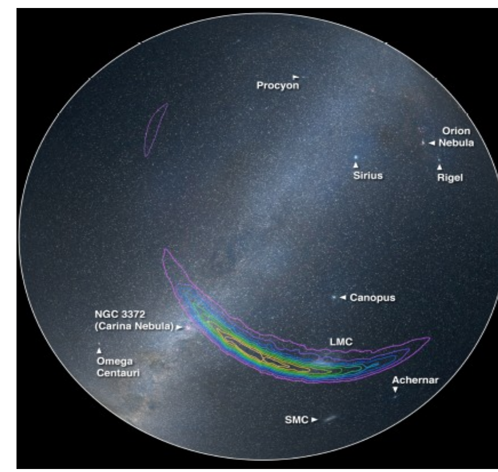
Searches	CPU core hours
CBC	65980786
Burst	12613653
CW	43317276
Detchar	2678844
Sgwb	426713
Total	125017272

Computational issues can limit the scientific goal (space parameters of analysis is limited)

Low latency GW data analysis

Online analysis: triggers to e.m. telescopes (and neutrino detectors.
82 MOU signed, 60 telescopes are ready for triggers in O2
Requirements: event trigger alerts and data quality information within few minutes; rapid source sky localization and parameter estimation

Many pipeline developed for different searches have been configured for low latency analysis (CBC, Burst, GRB)



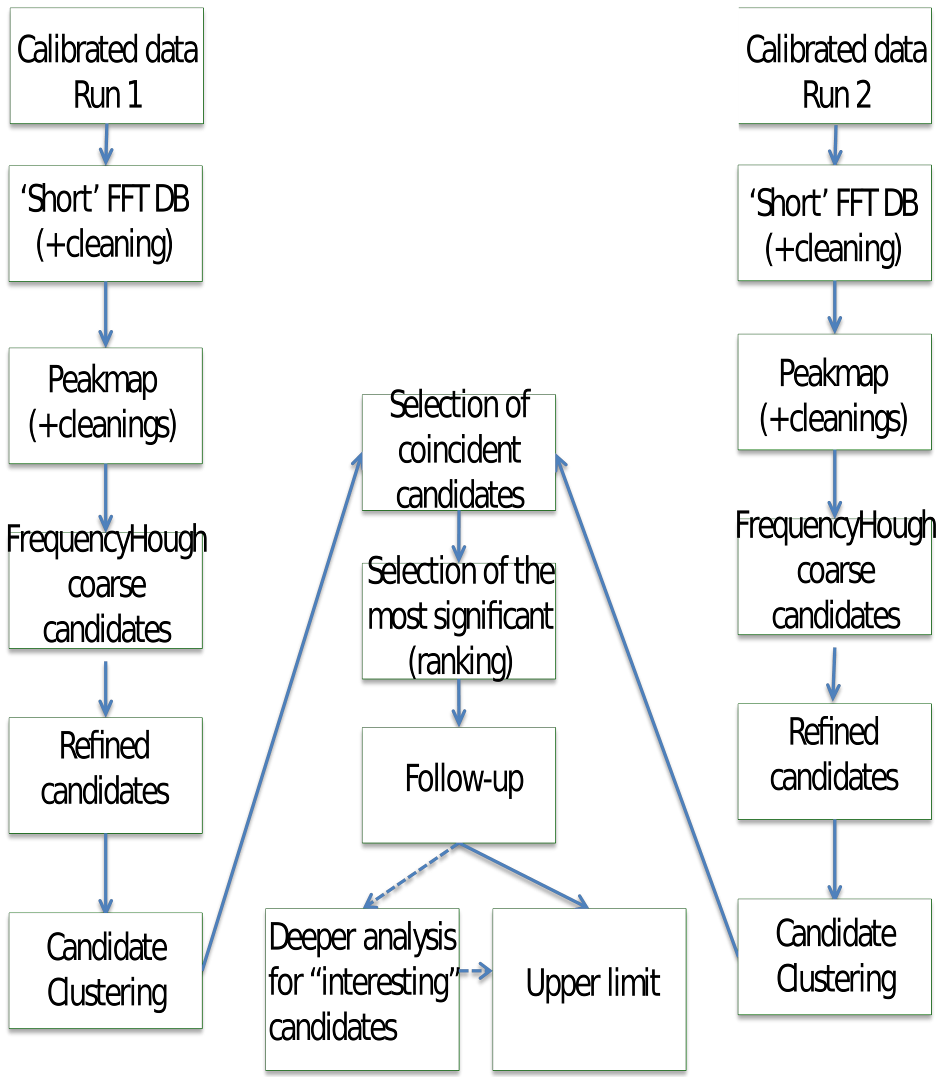
Advanced Virgo h(t) are transferred from Cascina to LIGO Clusters and advanced LIGO h(t) are transferred from LIGO to Cascina to guarantee the low-latency workflows (few seconds).
~ 8 GB/day from Cascina to aLIGO;~16 GB/day from aLIGO to CNAF and Lyon

Low latency Virgo pipeline MBTA (Multi Band Template Analysis), will run in Cascina (not high computational cost)

Virgo

CW - Rome

Continuous waves algorithms generally require relatively small amount of pre-processed data but many CPUs



Need to consider/analyse months of data to increase signal to noise ratio

The signal depends on source position in the sky, source frequency and frequency derivative → huge amount of point to be analyzed in parameter space

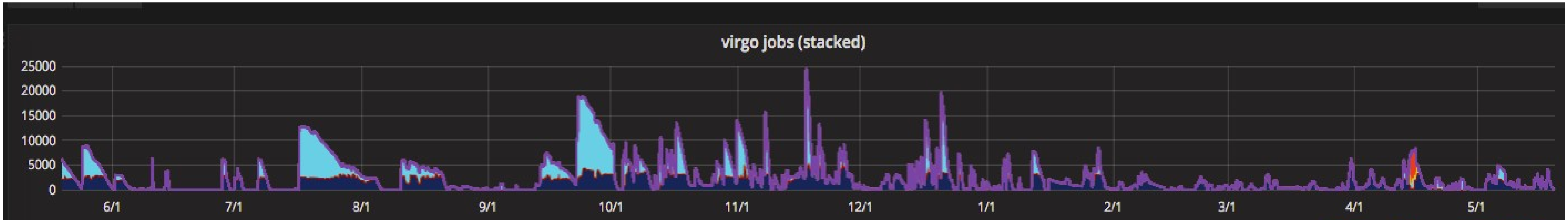
Extending the parameter space to be covered increases the computational cost of the analyses

CW algorithms discussed here are based on well-known methodologies, new other methods have been developed https://arxiv.org/abs/1703.03493

CW (all-sky) - Rome

- the frequency band to be searched for (0-2kHz) is split in several sub-bands, each analyzed in a job completely independent → embarrassingly parallel analysis
- Data of each interferometer are analysed separately. This is followed by candidates coincidence and follow-up of the most significant coincidences
- The core of the analysis is based on the Hough Transform. Computationally heavy due to the large explored parameter space

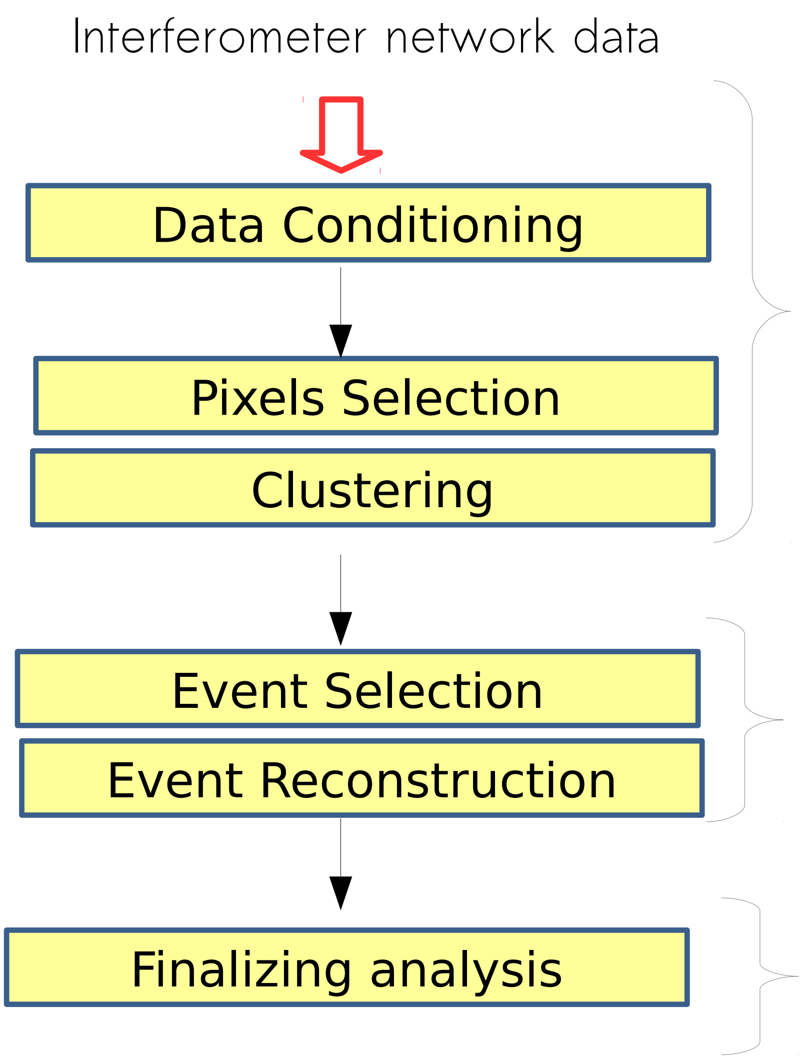
CW searches (Rome group) run mainly at CNAF. During last year:
-average number of jobs running each day 417
-energy used 15.9 kHSE06



	min	max	avg	current
grid.running	0	6129	1417	88
local.running	0	98	0	0
grid.pending	0	20620	1219	0
local.pending	0	8	0	0
grid.suspended	0	5880	20	0

CW searches are being successfully run also at Nikhef (via grid) and successful tests have been done to the Taiwan ASGC (Academia Sinica Grid Computing Center).

Coherent Wave Burst (cWB)



Data conditioning (removing frequency lines and whitening procedure) → FFT of the data data (single interferometer)

Pixels selection is performed on time frequency map of data (TF map is calculated at different resolution level) → wavelet transformation

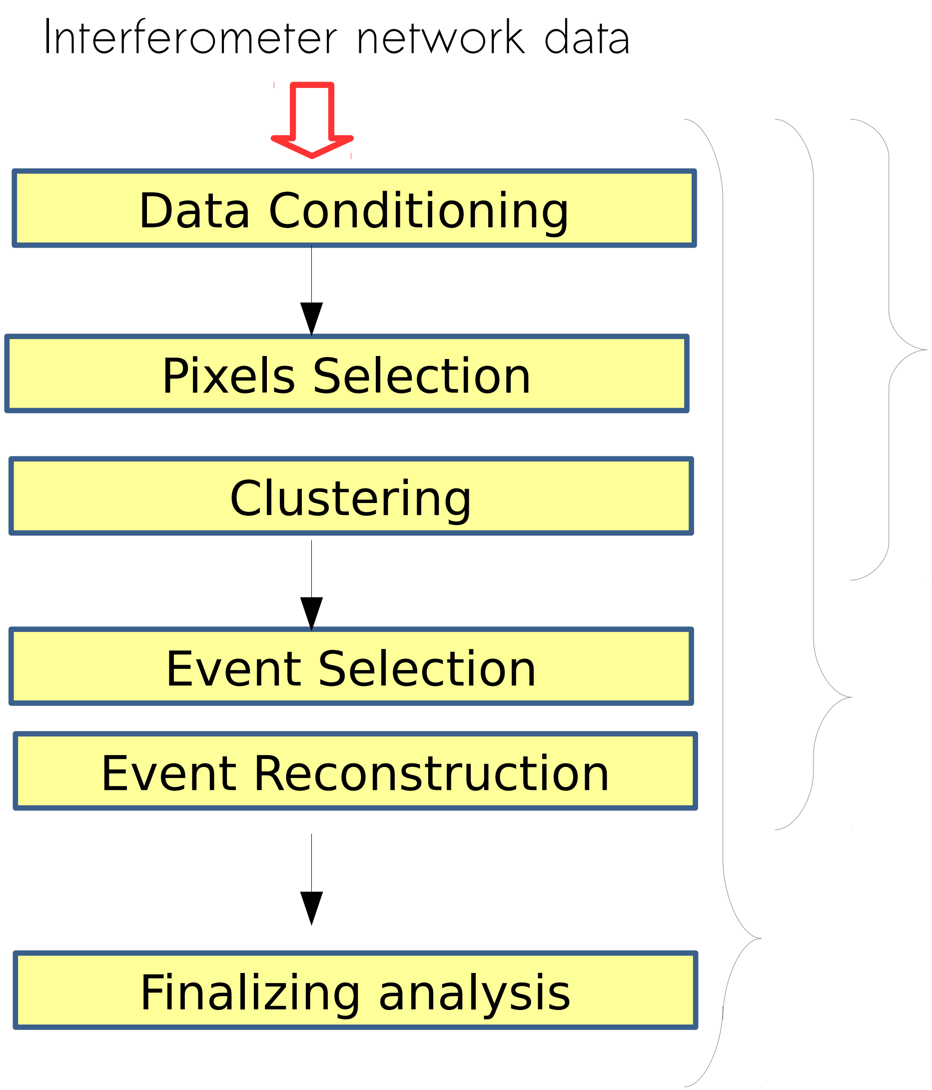
The analysis is performed coherently and between detectors since the pixel selection step.

Likelihood is calculated to select and reconstruct the event; the likelihood is maximized over a loop on the possible sky position of the source

Post production cut and selection, significance estimation of candidate

The time to be analysed is divided in segment (length defined by user, generally 600s), → Each segment analysis is defined in a job

Coherent Wave Burst (cWB)



Distribution of the jobs elapsed time at intermediate stage for an analysis of background estimation, (O1, LH, short all sky search)

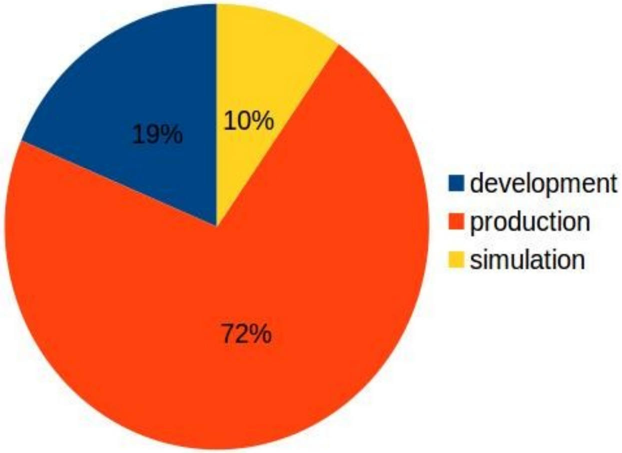
Mean time: 0.12h

Mean time: 0.21h

Mean time: 0.21h

Coherent Wave Burst (cWB)

CPU hours used by cWB pipeline (offline) to anlyse O1 data:



To develop the pipeline:

Allsky search	1722228
Imbh	301746
supernovae	16106

Production:

allsky	7351548
imbh	279465
Supernovae	188839

Simulations:

Allsky	68707
Imbh	827617

Estimation of candidate significance (or upper limits) is performed analysing data years of time equivalent (production). Runnig cWB version for O2 analysis

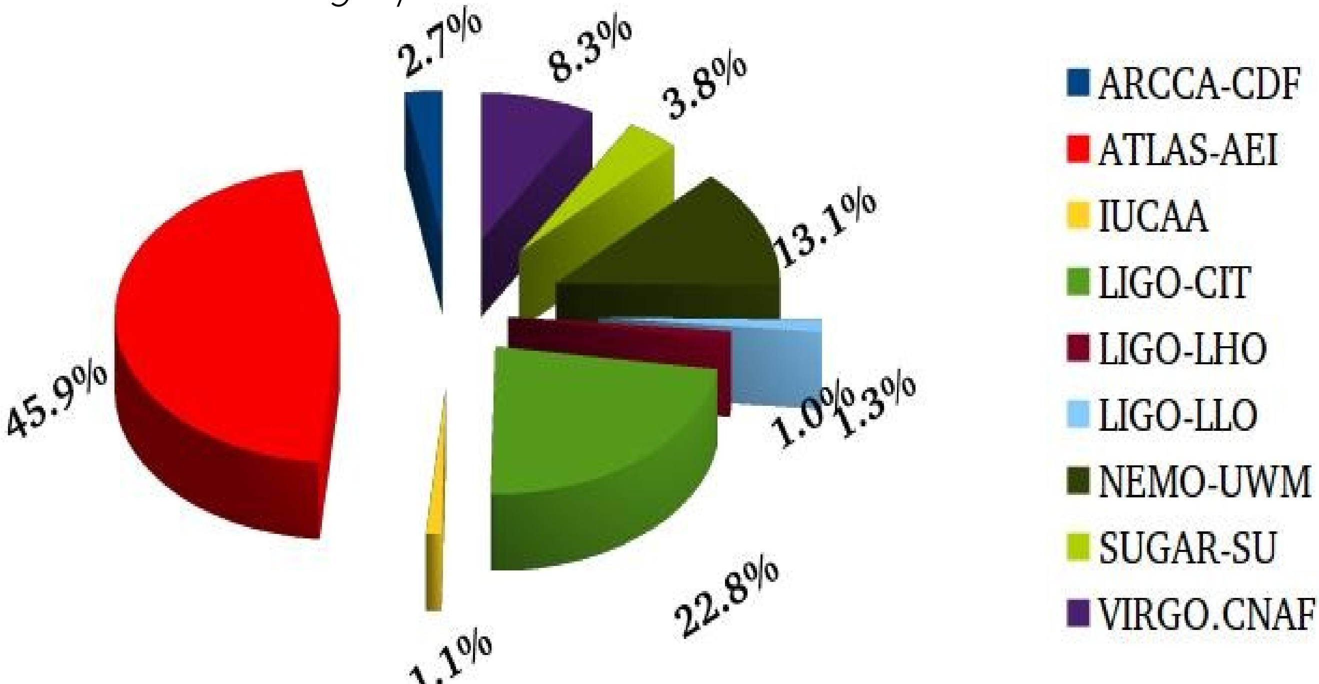
	h/job estimated	running job time	cpu time/year
All sky low frequency LH:	98h	0.21h	18h/year
All sky high frequency LH:	98h	0.82h	73h/year
All-sky long duration LH:	98h	0.29h	26h/year

Running job time depends on threshold applied. (to be defined)

For the first detection 68000 years of background time equivalent have been estimated

Custer resources used

LDG accounting system



LIGO Data Analyses:

- >60 software pipelines implemented.
- computing demand dependence by pipeline: top 10 pipelines cost 90% of computing demand.

LDG accounting system doesn't account jobs submitted via OSG, therefore CNAF is underestimated in this plot in fact almost 30% of the jobs are submitted to CNAF via OSG

Punturo's talk

Two runtime software environments: LIGO Data Grid (dedicated clusters), Open Science Grid (OSG)

LCS: 8 main clusters with shared file-system, job submission and work-flow through Condor

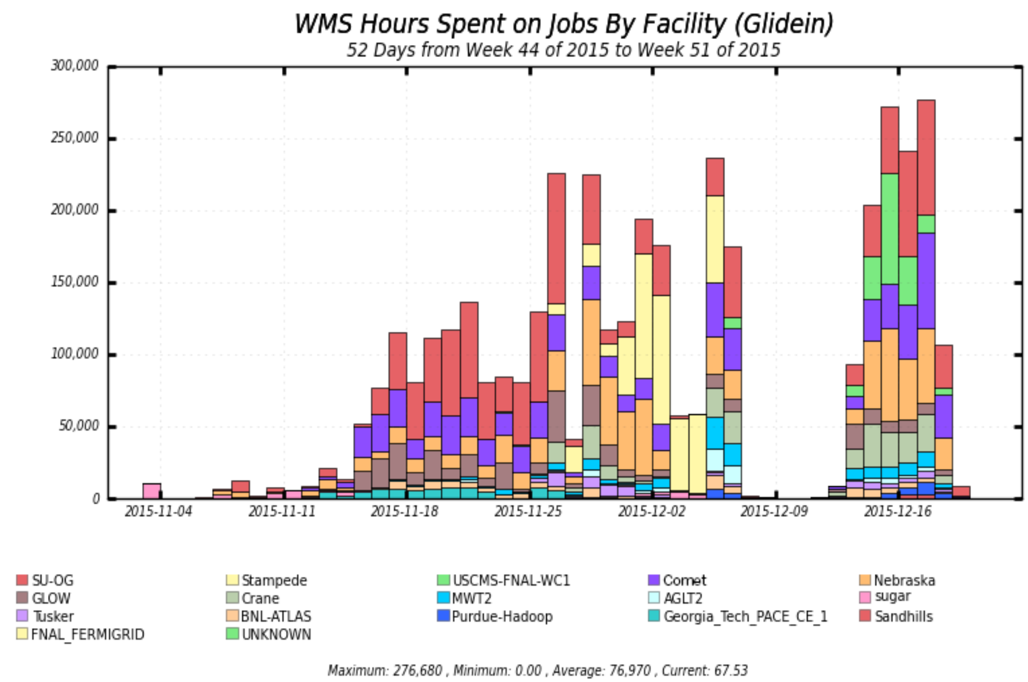
Dedicated LIGO Lab and LSC clusters cover 83% of required resources for O1 analysis

Open Science Grid - LIGO

Computing resources requirements is not uniformly distributed in time, computing request significantly peaks at certain times -> many clusters can handle this easily than isolated ones.

Open Science Grid can be considered an "adapter"
to allow LSC data analysts to submit their searc
pipelines via a familiar Condor interface at a lo
LDG site, running on different external resourc

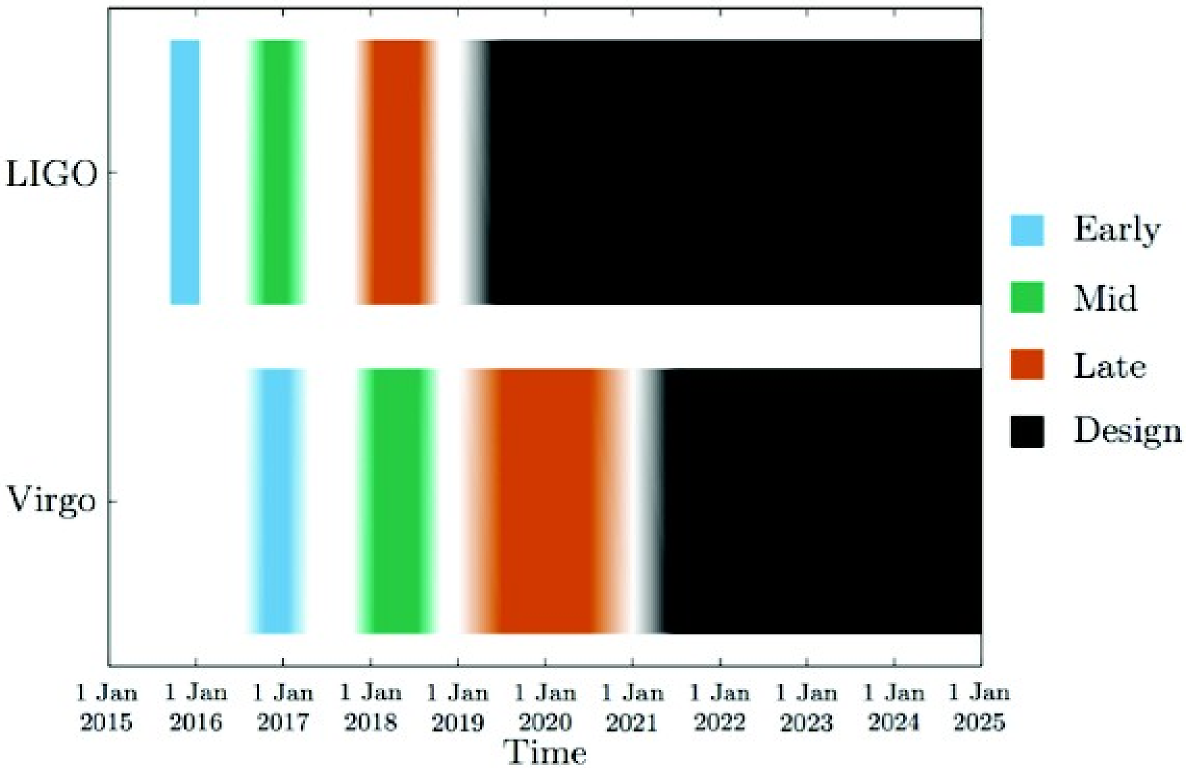
The LIGO analyses running across >20 differer
OSG resources, >20 million OSG CPU-hours f
O1.



Production offline CBC analysis (pyCBC pipeline) utilized OSG in O1.

https://www.opensciencegrid.org/osg-helps-ligo-scientists-confirm-einsteins-last-unproven-theory/

Near future



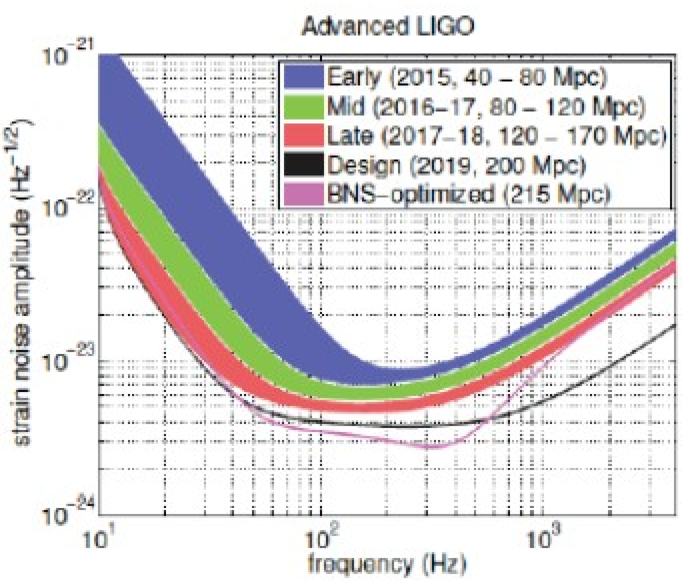
Performance upgrades by steps, interleaved by scientific observation runs: O2, O3, design

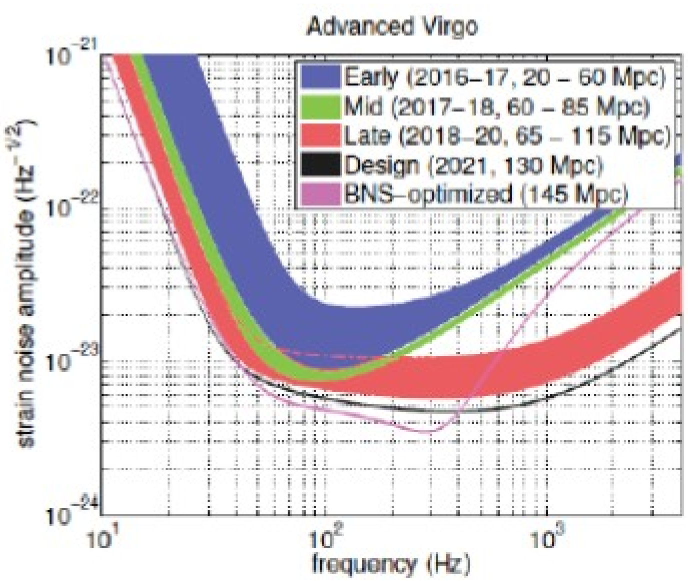
Data taking O2: ongoing

Design sensitivity:

Binary BH detection rate: 10-500 /year

Binary NS detection rate: 0.2-200 /year





More interferometers, more computing resources needed

During O2 Virgo that will join the network→ cpu requirements will increase

The rise of cpu needed is due to
adding new interferometers (depending on different algorithms):

- For the algorithms which analyzed separately the interferometers data, and performed finally a coincident analysis, the cpu requirements increase linearly with the number of detectors
- Algorithm performing a coherent/coincident analysis can have different behavior. Ex: comparison for cWB similar analysis LH and LHV network

	h/job estimated	running time job	cpu time/year
All sky low frequency LH:	98h	0.21h	18h/year
All sky low frequency LHV:	50h	0.491h	85h/year

improved sensitivity (and the scientific goals):

- Better sensitivity (ex: for template search, it will requires longer templates) and larger band of frequency available for the search

Could GPUs be a solution to parallelize and optimize the code?
GPU version of CBC search algorithm (pyCBC) is under development, and tests are ongoing

Conclusions

- O2 is ongoing:
  - test of data taking and data transfer ongoing (last week ~66MB/s and 28MB/s to Lion)
  - Different pipelines can run at CNAF, CCIN2P3, Nikhef
  - Virgo/CNAF/CCIN2P3/Nikhef are providing computer resources, total amount of computing resources provided is a now a accountable amount of the total LSC requirements

- Short/middle term (O2-O3): O1 analysis is almost completed, O2 is now ongoing, the three interferometers analysis data will improve the cpu requirements with respect O1

- Middle / Long term (>2020): detector network is expected to include more interferometers (LIGO India, KAGRA)

Back up

Virgo data Storage

Amount of data is a few hundred TB/yr.

Data is stored temporarily in a circular buffer at Cascina

(Backup selected data for crash recovery)

Copy of the data will be storage and back up at CNAF and CCIN2P3

CNAF storage

	2015	2016	2017	>2018
Disk (TB)	445	592	656	720

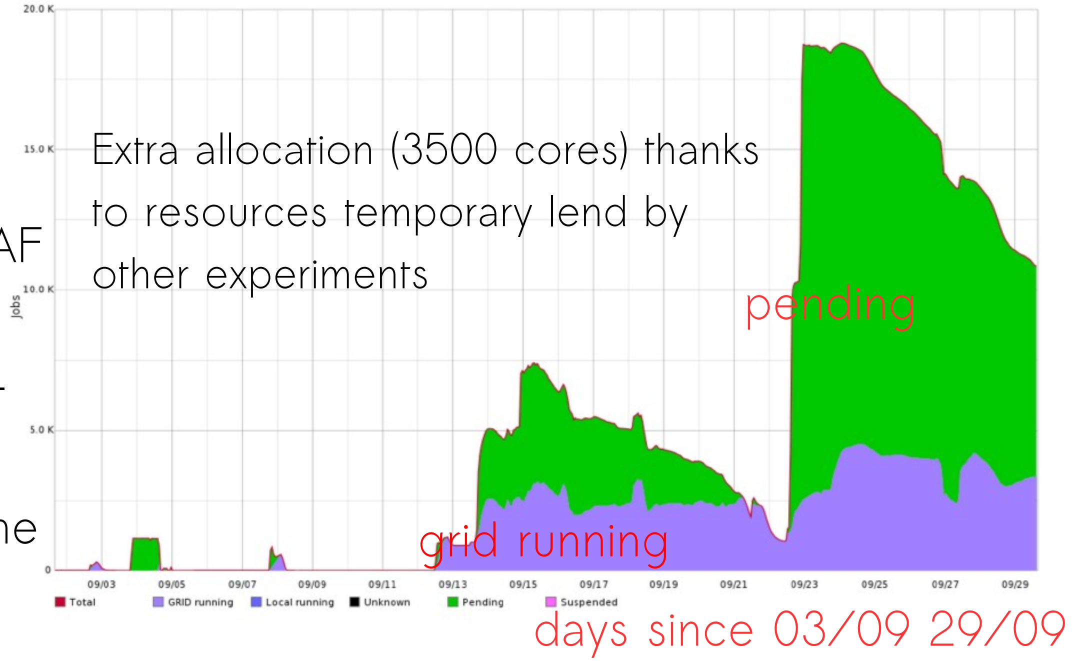
The quantity of data storage is no overwhelming, but the data have to be accessible to many users (hundreds of person involved in data analysis).

Virgo computing resources used

CNAF resources requested

	2015	2016	2017	>2018
CPU (HS06)	10000	25000	36000	78000

CW searches (Rome group) runs mainly at CNAF
Mainly all-sky CW searches are better handled by parallel computing: frequency band (few Hz-2kHz) can be split in several sub-bands each analyzed in a job completely independent on the others.
Other pipelines are working to be ready to use CNAF resources (cWB -burst pipeline)



CNAF 2016 (till 28/09): 14.3kHSE06 (Virgo pledge 25kHSE06, about 2500cores/days)

CCPIN2P3: 5k HS06 hours requested (0.7 kHS06 hours has been used), searches regularly use CCIN2P3 CPU: the EM-follow up Virgo team, the cosmic string search team (Virgo), the long transient search (Virgo)

Nikhef: 25kHS06 in the Dutch National GRID infrastructure