Journées Advanced Virgo+

L’analyse des données,
les partenaires,
et le calcul.

Marie Anne Bizouard, LAL
31 mai 2018, LKB ENS Paris

Data analysis

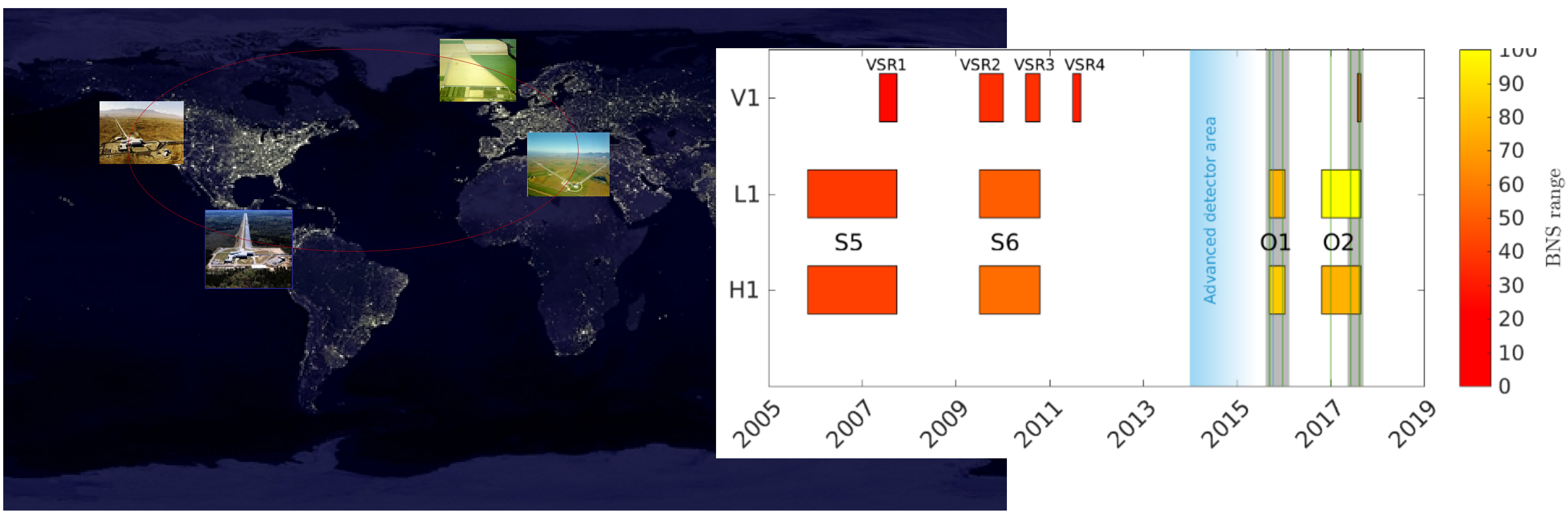
- The organisation: « the single machine »
  - Since 2007, LIGO, GEO and Virgo data analysis is a joint activity. Rules spelled out in the LIGO/Virgo MOU renewed each ~3 years.

    → Exchange of data, joint running planning, commissiong/shutdown.

    → Common data analysis groups + calibration groups and detector characterization (detchar) groups for each collaboration. LSC & Virgo co-chairs.

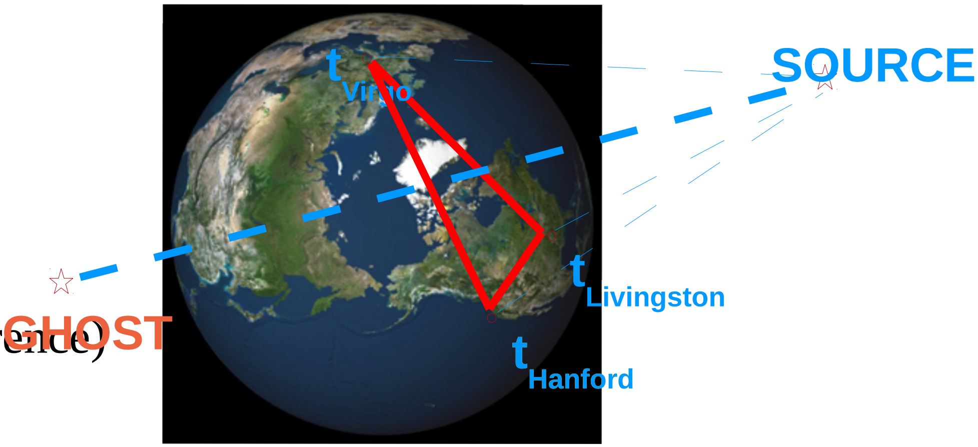
    → Common data analysis council (DAC).

    → Common publications (joint editorial boards).



Data analysis

- The single machine :
  - Increase the sky coverage
  - Source reconstruction (localization)
  - Reduce false alarm (coincidence/coherence)
  - Assess discovery confidence



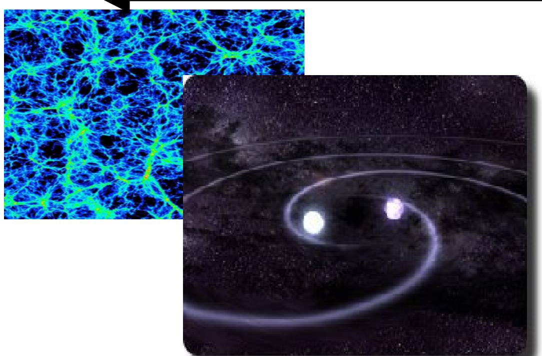
- Missions:
  1. Find all GW sources in LIGO/Virgo/GEO detectors data.
  2. Extract all possible physics results : Fundamental physics tests and measurements : H_0, graviton celerity, test of equivalence principle, constrain the nuclear matter EOS, …
  3. Provide alerts to the outside world and especially to « observers » (multi-messenger analysis).

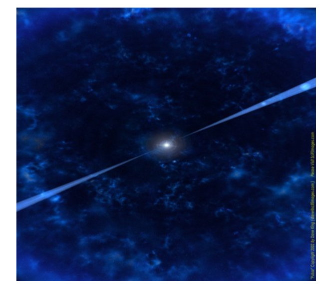
Data analysis

- Physics groups:

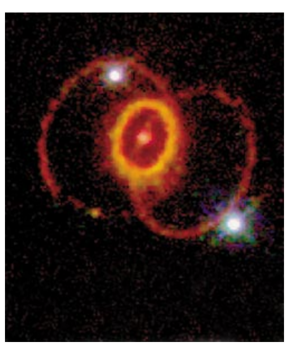
Short duration (~1s) — Long duration (∞)

Waveform known





Waveform unknown

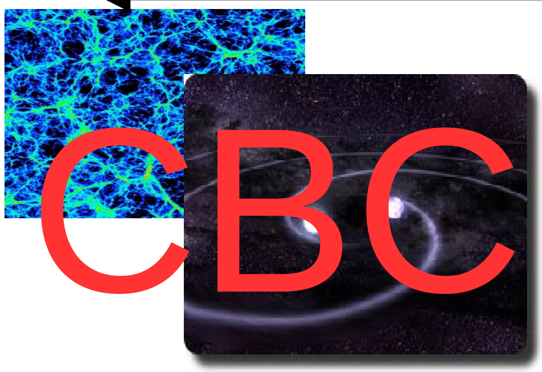


Data analysis

- Physics groups:

Short duration (~1s) — Long duration (∞)

Waveform known



CBC

CW

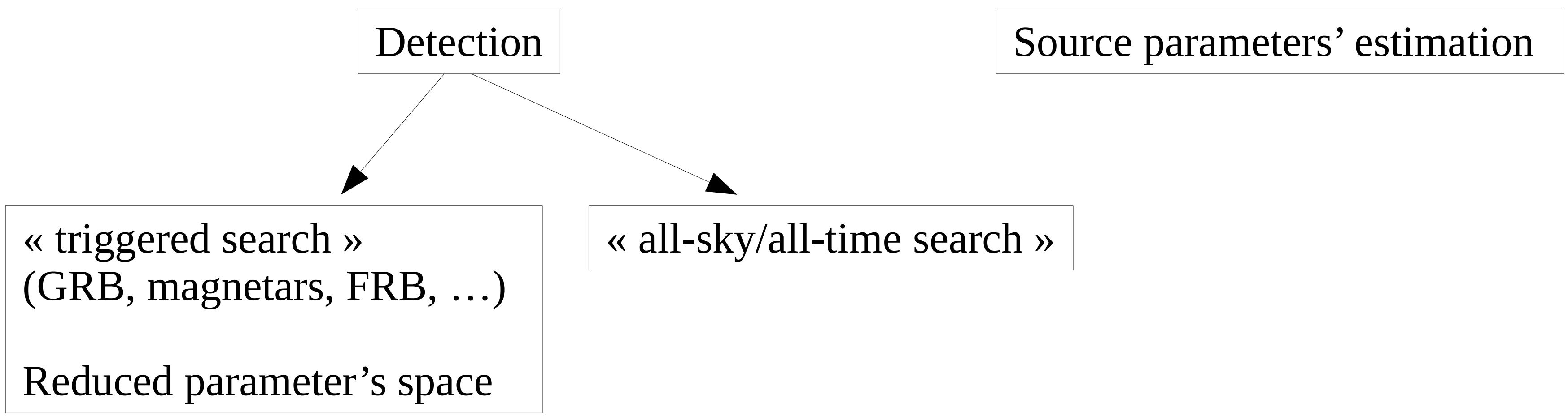
Waveform unknown

burst

stochastic

Data analysis

- Signal processing for all sources :



Data analysis

- Signal processing for all sources :

Detection

Source parameters' estimation

« triggered search »
(GRB, magnetars, FRB, …)

Reduced parameter's space

« all-sky/all-time search »

Méthodes :

- Matched filtering quand la forme d'onde est paramétrisable.
- Cross-correlation entre 2 h(t)/detectors quand on ne connaît pas la forme d'onde.
- Excess power/time-frequency decomposition : un signal transitoire est modélisable par la somme de fonctions de sinus-gaussiennes.

- Multi-resolution clustering (sources transitoires non modelises) & peak lines (sources continues)

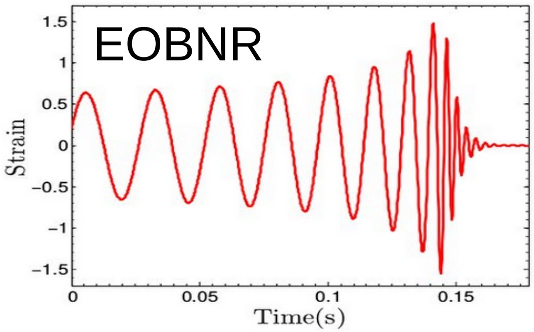
Data analysis : matched filtering

FFT of data

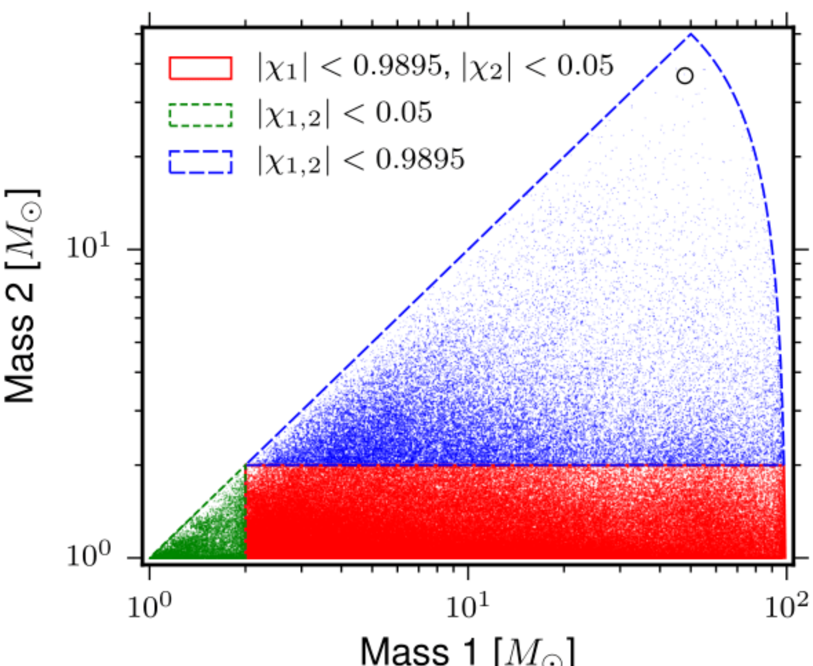
Template can be generated in frequency domain using stationary phase approximation

$$\mathcal{C}(t) = \int_{-\infty}^{\infty} \frac{\tilde{x}(f)\tilde{h}^*(f)}{S_n(f)} e^{2\pi i f t} df$$

Noise power spectral density (in this case this is the two-sided Power spectrum)



EOBNR-IHES waveform: m1=36Msun, m2=29Msun, nonspinning black holes



GW150914 search configuration:
Waveform templates: EOBNR with aligned spins
Online: low mass regime (<20 Msun)
Offline: 1-100 Msun

Data analysis

- Signal processing for all sources :

Detection

« triggered search »
(GRB, magnetars, FRB, …)

Reduced parameter's space

« all-sky/all-time search »

Source parameters' estimation

Methods :

- Bayesian inference when waveform is known.

$$p(\vec{\theta}|\mathrm{data},\mathrm{Model}) = \frac{p(\vec{\theta})p(\mathrm{data}|\vec{\theta},\mathrm{Model})}{Z}$$

$$Z = \int_{\Theta} p(\vec{\theta})p(\mathrm{data}|\vec{\theta},\mathrm{Model})d\vec{\theta}$$

- Hasting-metropolis MCMC to generate sample.
- Nested sampling to integrate posterior distributions.

Data analysis

- Signal processing for all sources :

Detection

« triggered search »
(GRB, magnetars, FRB, …)

Reduced parameter's space

« all-sky/all-time search »

Source parameters' estimation

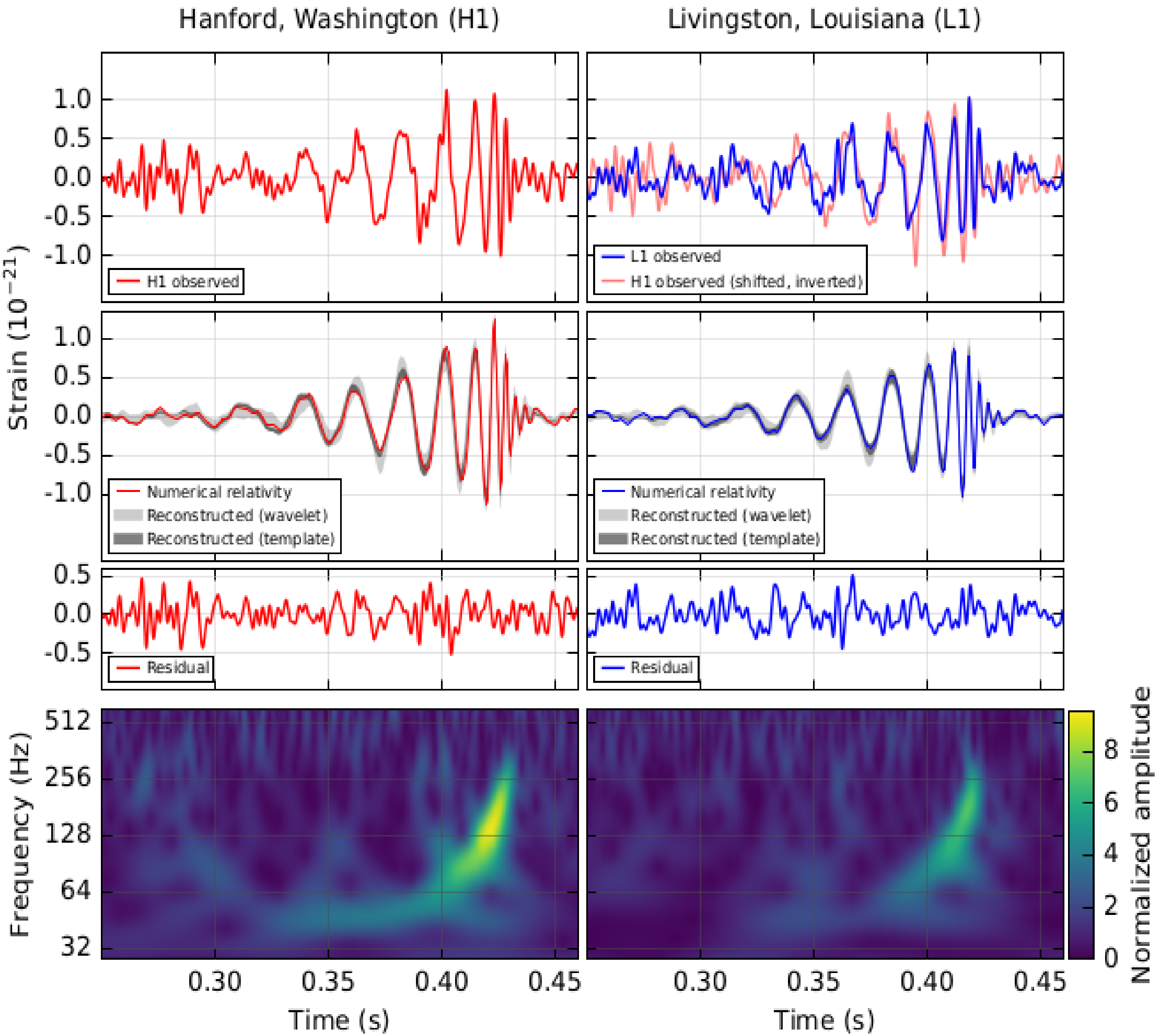
Methods :

- Bayesian inference when waveform is unknown.

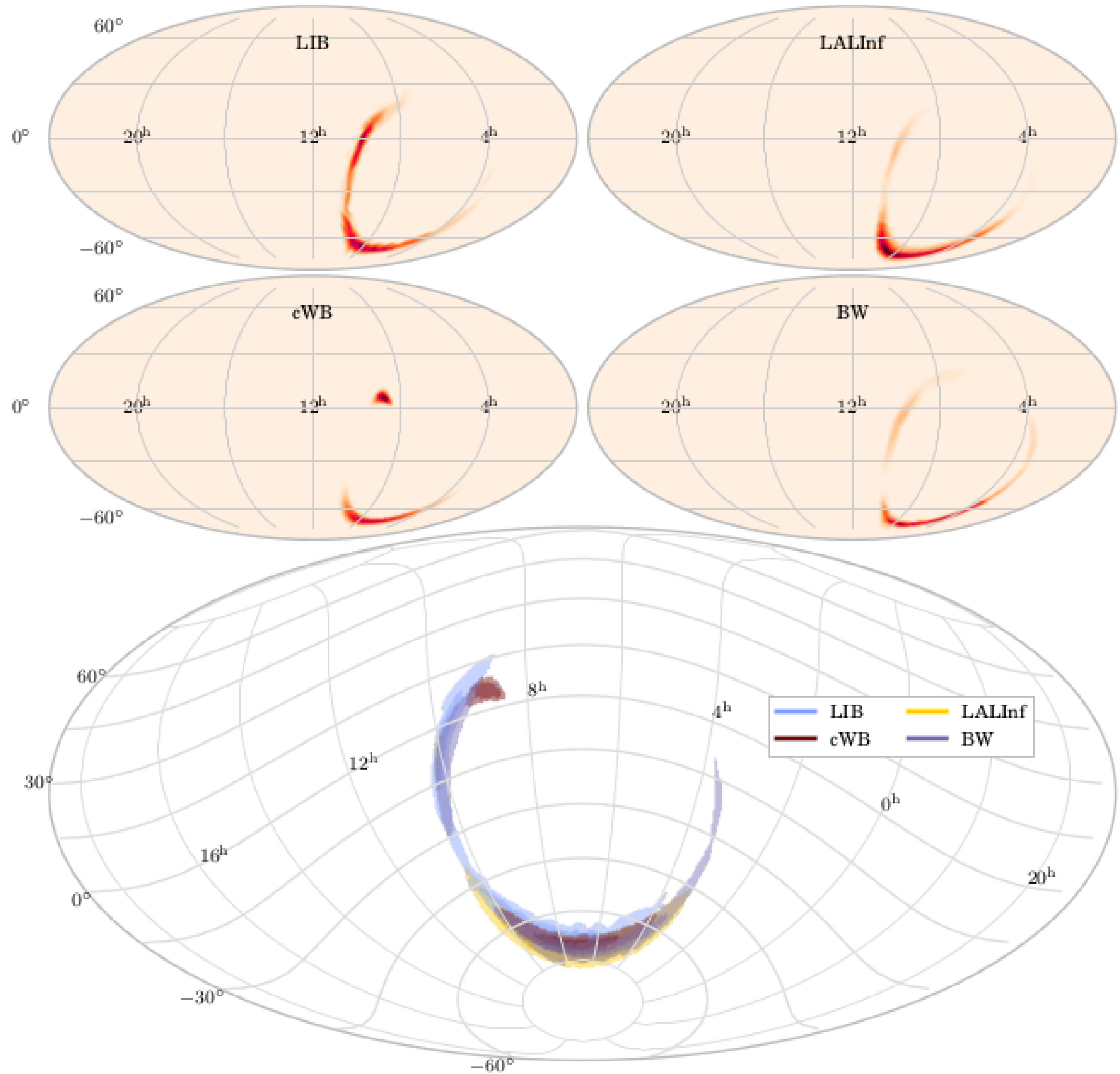
$$\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} F_i^+/\sigma_i \\ \vdots \\ F_N^+/\sigma_N \end{bmatrix} h_+ + \begin{bmatrix} F_i^x/\sigma_i \\ \vdots \\ F_N^x/\sigma_N \end{bmatrix} h_x + \begin{bmatrix} n_1 \\ \vdots \\ n_N \end{bmatrix} \rightarrow \boldsymbol{x} = \boldsymbol{F}_+^w h_+ + \boldsymbol{F}_x^w h_x + \boldsymbol{n}$$

Solution : $\hat{h} = (\mathbf{F}\mathbf{F}^{\mathbf{T}})^{-1}\mathbf{F}^{\mathbf{T}}\mathbf{x}$

Data analysis : parameters' estimation

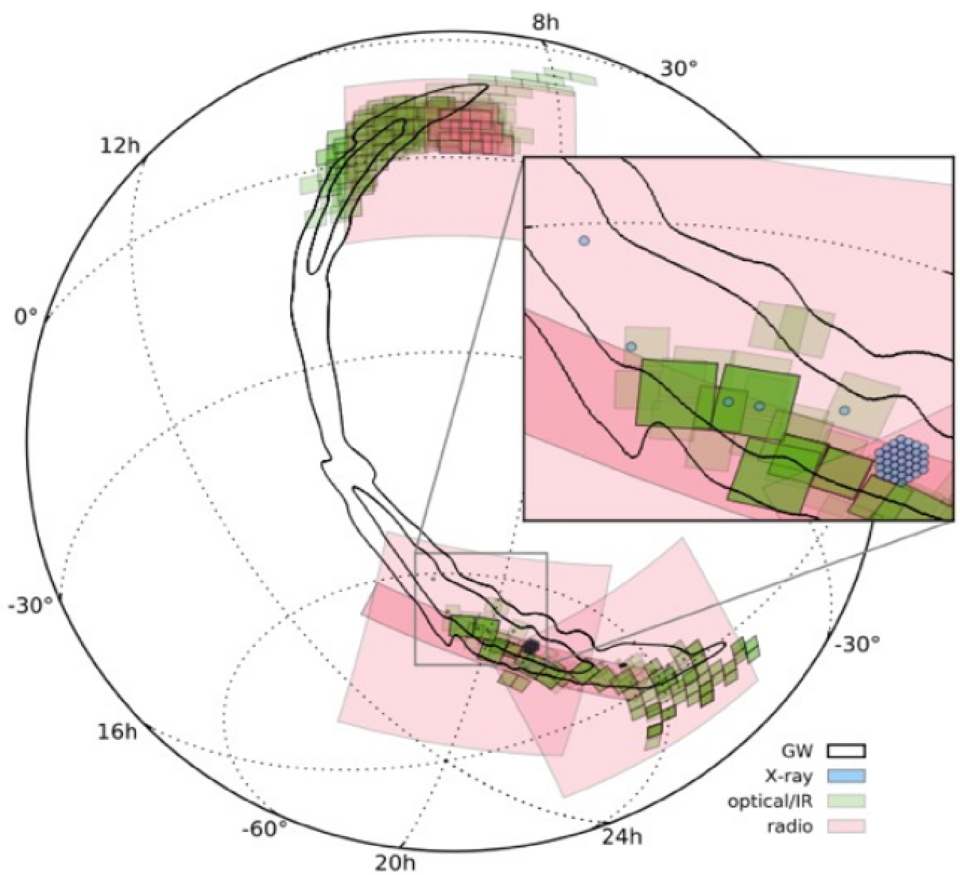


Data analysis : sky localization



O1-O2 electro-magnetic / neutrino follow-up

- ~95 MOUs (radio, optical, IR, X-ray and γ-ray) + neutrinos.

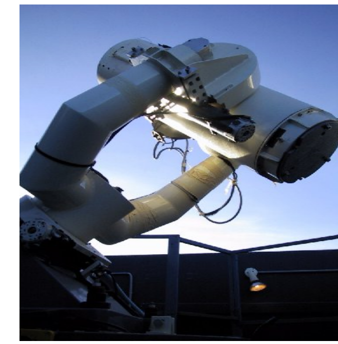


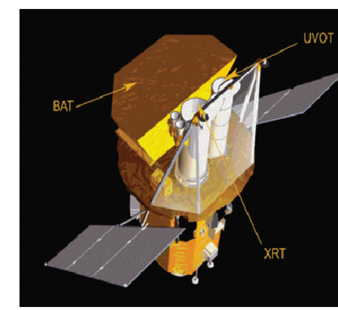
cWB sky map

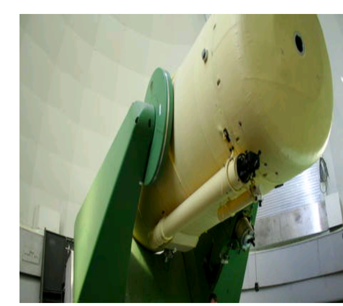
γ / X-ray observations

Optical observations

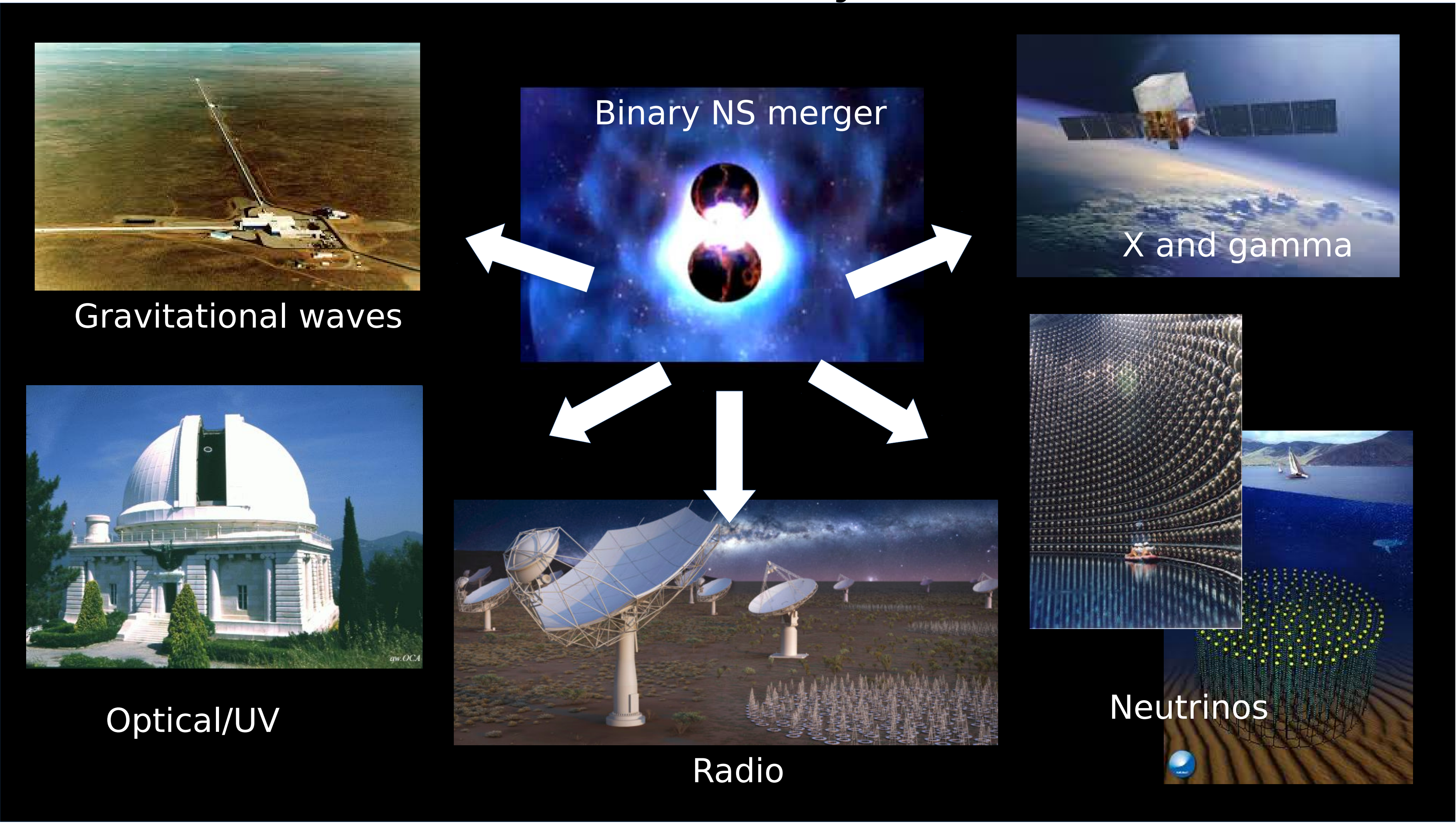
Radio Observations





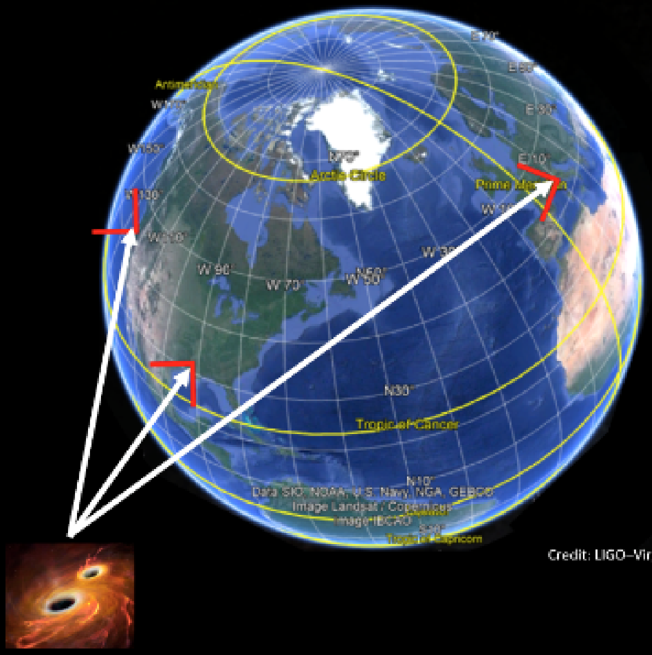


Multi-messenger analaysis : « triggered » & sub-threshold analyses

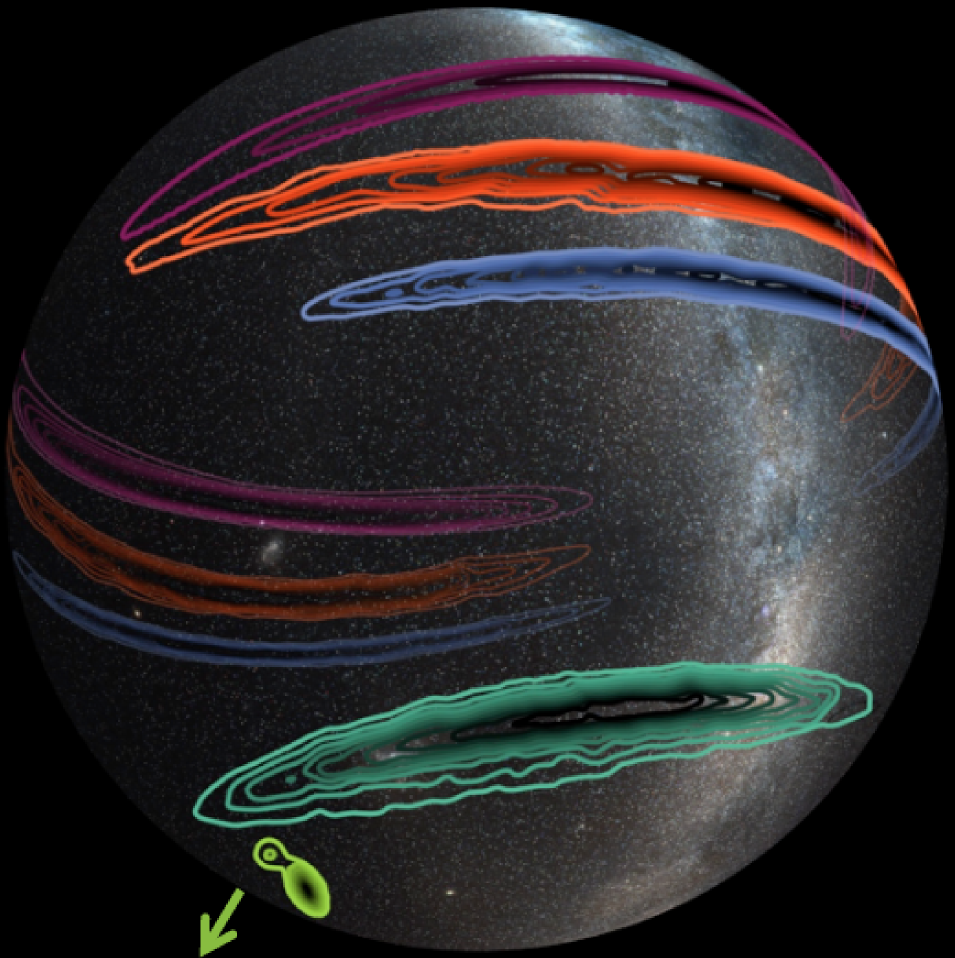


Data analysis : network role

2017 August 14



Credit: LIGO–Virgo



GW170814

Credit: LIGO/Virgo/NASA/Leo Singer
(Milky Way image: Axel Mellinger)

Transcient sources: low latency & GW alerts

All-sky / all-time transient searches during O1/O2

→ low latency (<5 min)
→ submit **private** GCN notices & circulars the "most significant" triggers
→ ~90 MOU currently active
→ all triggers kept in a private LVC data base
→ LIGO data public release 18 months after data taking.
Not yet any release fo Virgo data.

Targeted searches (external triggers)

→ GRB, SGR, AXP, magnetars
→ medium latency (GRB only)

Multi-messenger searches

→ neutrinos (HEN & LEN), FRB, GRB, ...
→ sub-threshold triggers

Latest — as of 20 June 2017 16:22 CDT

Query:

Get neighbors:

UID	Labels	Group	Pipeline	Search	Instruments
G250089		CBC	gstlal	HighMass	H1,L1
G250088		CBC	gstlal	HighMass	H1,L1
G250087		CBC	gstlal	HighMass	H1,L1
G250086		CBC	gstlal	HighMass	H1,L1
G250085		CBC	gstlal	HighMass	H1,L1
G250078		CBC	MBTAOnline		H1,V1
G250077		CBC	MBTAOnline		H1,V1
G250076		CBC	gstlal	HighMass	H1,L1
G250075		CBC	MBTAOnline		H1,V1

Low latency during O3 & GW alerts

- LIGO/Virgo will release public alerts for event candidats that we are confident about and can stand behind. They will look a lot like events in O1 and O2, except they will be instantly public.

O3 alerts == public alerts

  - For compact binary coalescences (CBCs), we aim at an overall astrophysical purity of 90% (e.g. at most 1 in 10 compact binary candidates on average will have instrumental or environmental rather than astrophysical origin). Threshold sets at FAR ~ 1/month – 1/year
  - More restrictive threshold for unmodelled burst sources : FAR ~ 1/year - 1/100 years.
- Which information will be in the public domain ?
  - Content :
    - Alerts should contain all of the information that is useful for searching for a counterpart (if there is one).
    - Details : significance, time, GW signal classification, 3D sky position and distance.
  - How to get them ?
    - Preliminary GCN notice within 5 minutes, without human vetting.
    - Initial GCN notice and circulars within ~4 hours with skymaps and classification.
    - Update GCN circulars, especially if better skymaps available.
- Collaborations with external groups : Science driven MOU

Science driven MOU : policy fundamentals

1. Objectives must be part of the science program of the LIGO-Virgo Collaborations.
2. Agreements/collaborations with non-LIGO-Virgo partners should not be “exclusive” for any of the science topics pursued.
3. Data/information/results privacy to be maintained at all times.
4. Joint publications of results upon mutual agreement and with the whole LIGO-Virgo author group.

Data flow models

- GW transient triggers below the detection standard that may improve a specific science/source search when analyzed jointly with the EM/neutrino sectors.
- Several MOUs with similar scope exercised in recent times/still in place:
  - High Energy Neutrinos (Antares, Icecube).
  - Gamma-Ray/X-ray transients sources (Fermi-GBM).
  - Core-collapse Supernova low energy neutrinos (Borexino, Icecube, KamLAND, LVD).
- Generally, not low-latency critical (until now) and with low opportunity cost.

Data flow models

- EM transient/neutrino triggers not in the public domain that may improve a specific science/source search when analyzed jointly in GWs.
- Several MOUs with similar scope exercised in recent times/still in place:
  - High Energy Neutrinos (Antares, Icecube).
  - Gamma-Ray/X-ray transients sources (Fermi-GBM).
  - Fast Radio Bursts (Green Bank Observatory, Parkes Radio telescopes).
  - Core-collapse Supernova low energy neutrinos (Borexino, Icecube, KamLAND, LVD).
- Generally, not low-latency critical (until now) and with low opportunity cost .

Data flow models

- EM transient information not in the public domain that may improve a specific GW search/detection potential.
- Several MOUs with similar scope exercised in recent times/still in place:
  - CCSN light curves, progenitor information (ASAS-SN, DLT40).
- Generally, not low-latency critical.

One specific case : the Hubble constant measurement
- Host galaxy redshift and peculiar velocity
- « Complete » galaxy catalogues until several hundred Mpc

→ Open calls for collaboration

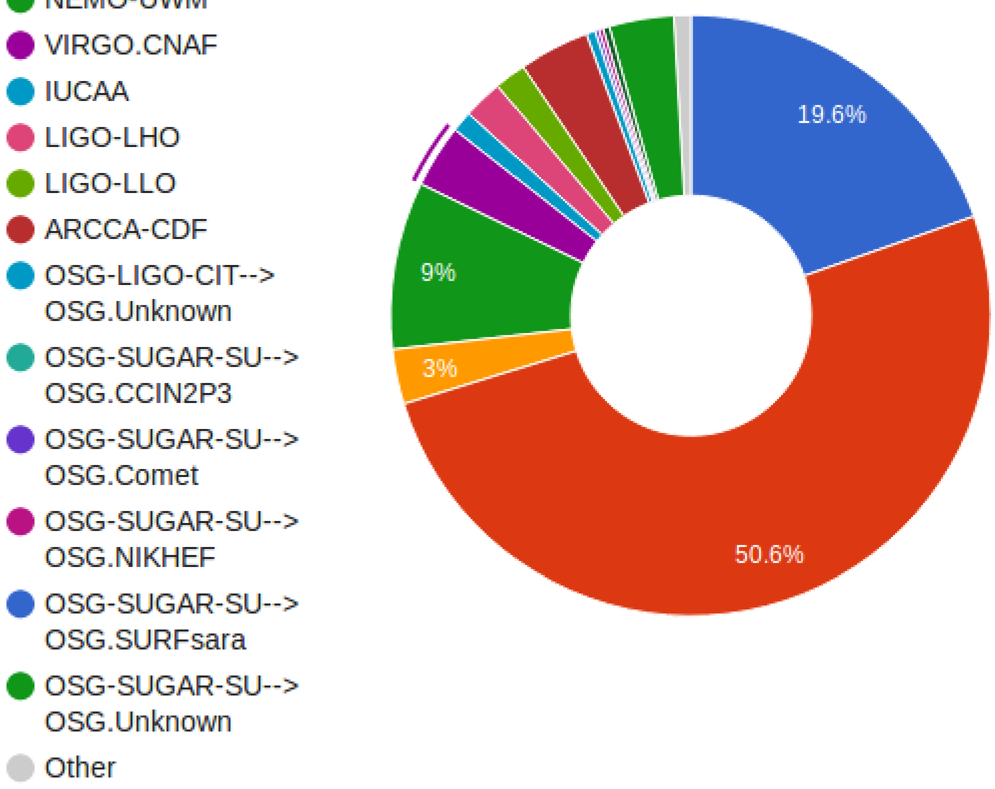
Data flow models

- Information on GW transient detection from LIGO-Virgo not in the public domain (OPA) that can be used in analyzing EM data jointly and for specific science targets:
  - Inclination, individual masses and spins, tidal parameters for binary mergers.
  - 3-D localization information including full error budget post-EM counterpart identification.
  - Waveform details on GW transient alert when not a binary merger.

Data analysis & computing means

- Each collaboration has its own organization :
  - Historically :
    - LSC : « owned centers ». homogenous workloads (condor)
    - Virgo : national computing centers (CC IN2P3, CNAF, NIKHEF, …). European grid (LCG). non homegenous workloads.
    - Virgo : no manpower/budget/organization for offline computing workload developments.
  - LSC is now using OSG to access different « external » centers including CNAF, CCIN2P3 and NIKHEF.
  - Virgo is lacking manpower for an ambitious DA computing model.
  - Current O2 computing usage :

LVC computing accounting :



Sept2016-Sept2017

Virgo 7.4 %

LVC computing accounting :

- O2 computing usage :

	DAC Estimate (SUs)	Actual Use (SUs)	% of Estimate
CW	145,113,600	1,710,793	1%
CBC	112,643,185	134,364,413	119%
Burst	27,276,670	20,278,027	74%
Detchar	3,048,903	3,781,855	124%
SGWB	894,520	2,348,540	263%
Grand Total	**288,976,878**	**162,483,628**	**56%**

in aLIGO Service Units (SUs)
1 SU = 1 Xeon E5-2670 CPU-core-hour

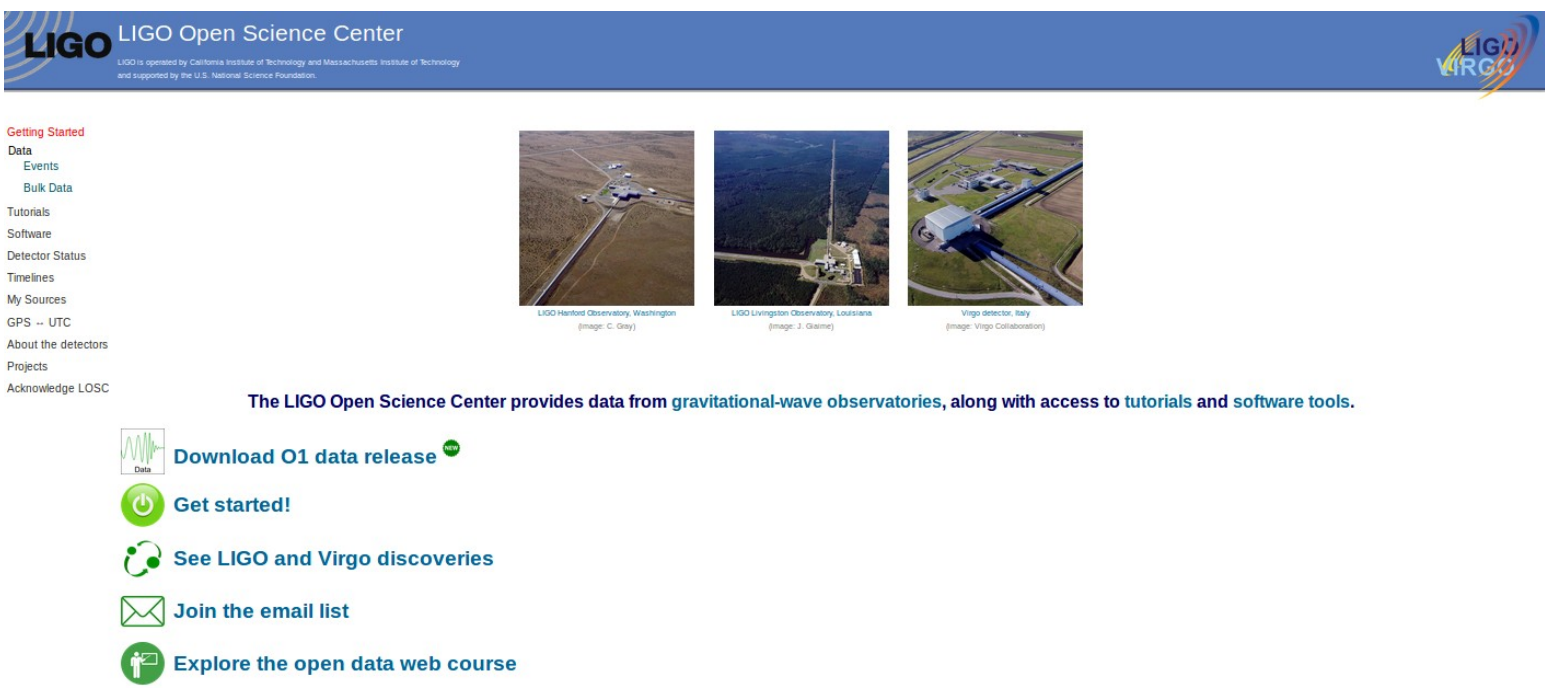
Virgo computing

- Several activities :
  - Cascina : federate logins, storage upgrade, low latency computing upgrade, ...
  - Data transfer, data access and file catalogues.
  - CCs : DIRAC for data transfer to Ccs and workload Virgo general infrastructure ?
  - /cvmfs to export software and

AdV+ in the open science era

- LIGO has adopted a LIGO data managament plan which specifies that LIGO data become public after XXX months of a run completion. The XXX months period getting reduced as time goes by (right now 18 months, then 12 months → 6 months after O4).
- LIGO has developed the LIGO Open Science Center where :
  - past run LIGO data set are available (S5 & S6 & O1).
  - GW events data snapshot (including Virgo data) are available when GW events are announced/published.
  - Software & tutorials to analyse the data + workshops (next one in Europe in 2019).

https://losc.ligo.org/about/



Conclusion

1. AdV+ data analysis : lots of good physics to do !
2. Multi-messenger analysis : open public alerts is our new model. But collaboration with partners on specific science subject will continue.
3. Computing for AdV+: still understaffed and lacking of a good organization.