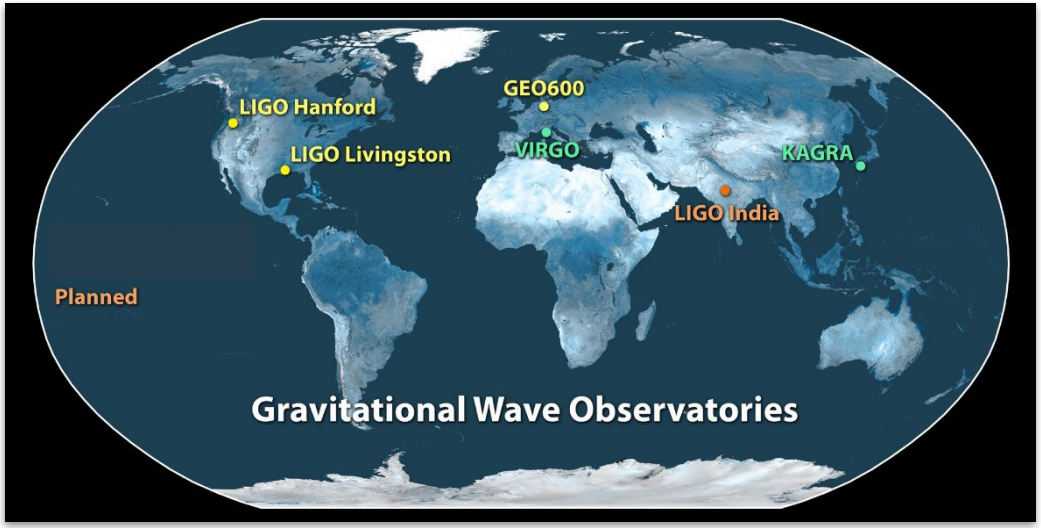
G2001437

IGWN on the grid

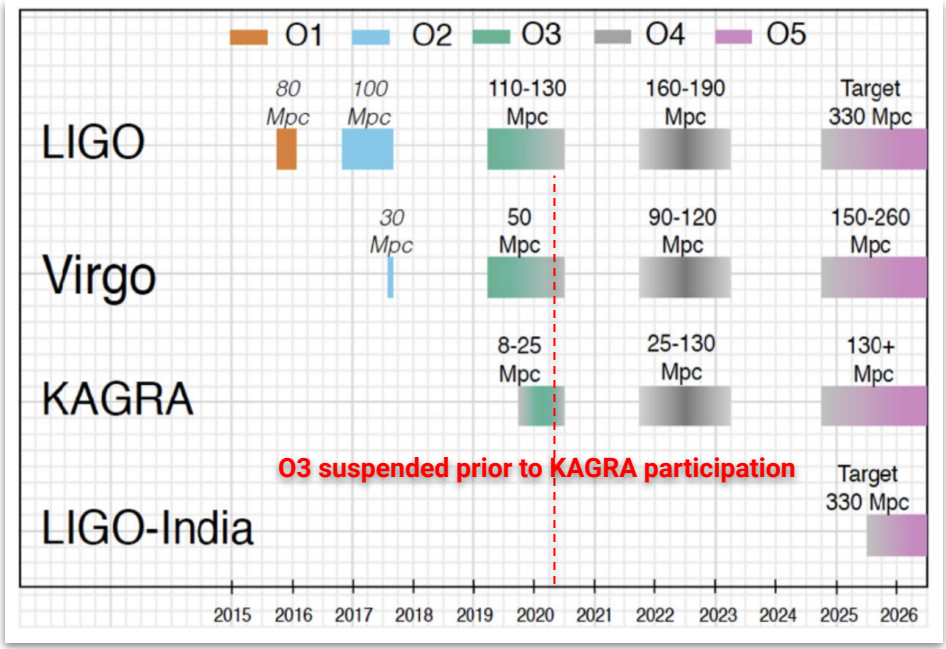
More gravitational waves and more resources

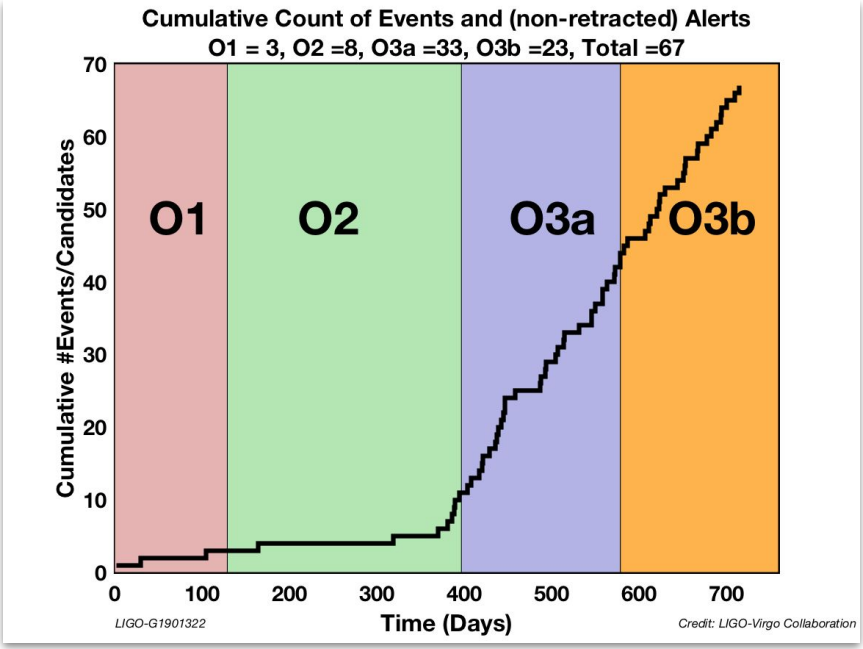
james.clark@ligo.org, OSG All-Hands Meeting 2020

IGWN: International Gravitational-Wave Network



Recent History / Near-future





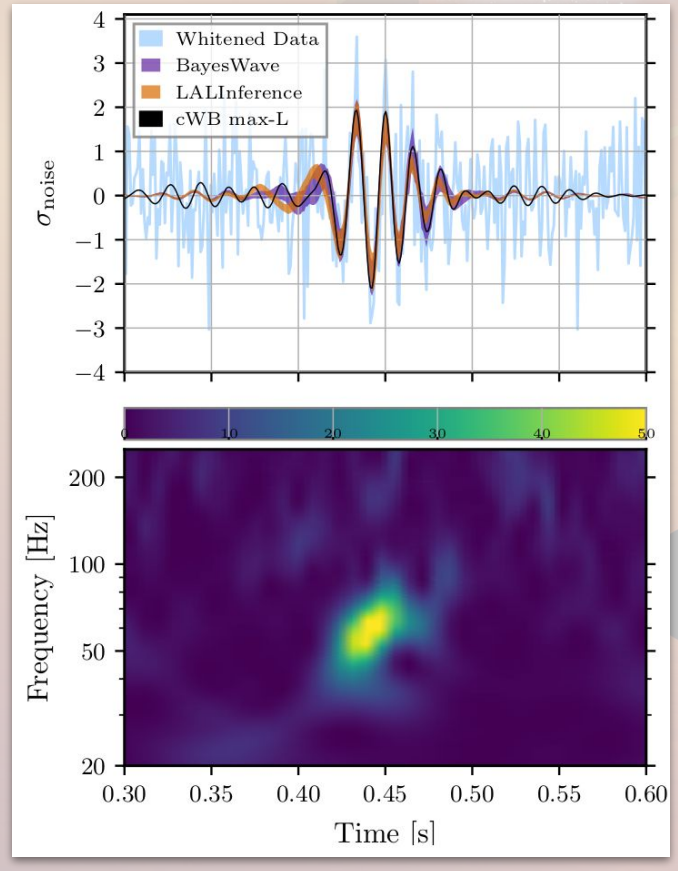
Living Reviews in Relativity **21**, 3 (2018)

"Advanced era" - wildly successful: 67 public alerts [14 published BBH + 1 BNS]

O3: originally scheduled 1 calendar year, joint LIGO+Virgo(+GEO600), KAGRA to join spring 2020...

Big Announcement: GW190521

Discovery: 2009.01075
Implications: 2009.01190



Detected May 21st 2019:

- Heaviest merger yet seen!
- Masses: 85+66 M_{sun}, remnant: 142 M_{sun}
  - ~8 M_{sun} radiated in GWs
- Redshift 0.48 -- 1.1
  - Universe ~half current age
- Potential **new** population of black holes

Major challenge to understand formation:

- Primary mass lies in pair-instability gap → stars shouldn't get this big on their own...
- Stellar merger? dynamical capture scenarios?

IGWN Grid: IGWN sites + OSG infrastructure → blend of opportunistic & dedicated / allocated resources

LIGO Data Grid

LIGO Hanford & Livingston

CalTech

UW Milwaukee

Other submit sites coming online (e.g. Nikhef)

- IGWN-OSG submit node
- Local HTCondor pool
- Local networked storage (large, persistent, backed-up /home)
- HTCondor IGWN Central Manager

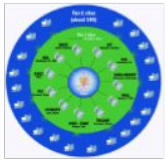
UCSD

Factory

Nautilus

IGWN front-end

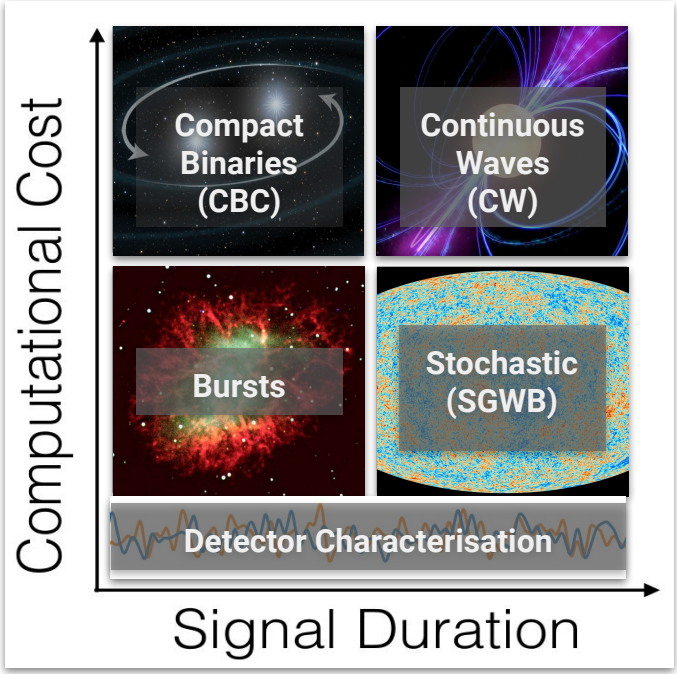
Open Science Grid



WLCG/Virgo resources

HPC Centers

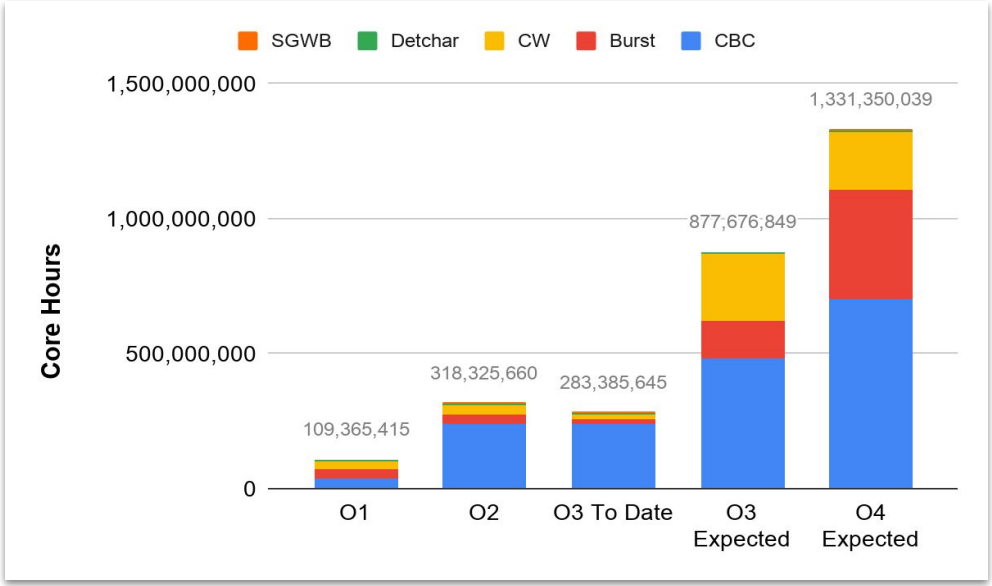
LIGO-affiliated sites (e.g. GATech)



Astrophysics groups determine & deliver science goals/priorities

Many analyses: segment data || split parameter space → naturally amenable to HTC approach

Long, rich relationship with HTCondor & OSG



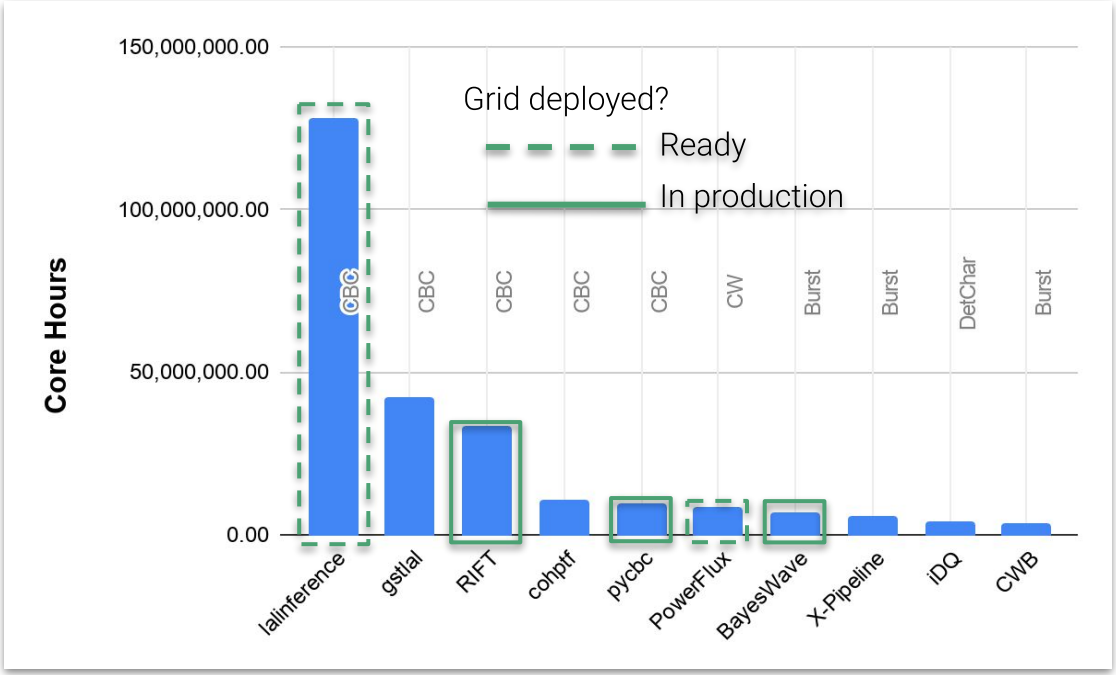
~10% of total consumption accessed thru OSG

Current capacity should meet O4 demand thanks to:

- provisioning new sites, including thru OSG
- rigorous prioritization exercise
- assume no overlap between O3/O4 analyses

Opportunistic resources → absorb **surprises**, spur development & novelty

Getting on the grid



Out of ~**70** total analysis pipelines, **10** consume **90%** of O3 hours *to date**

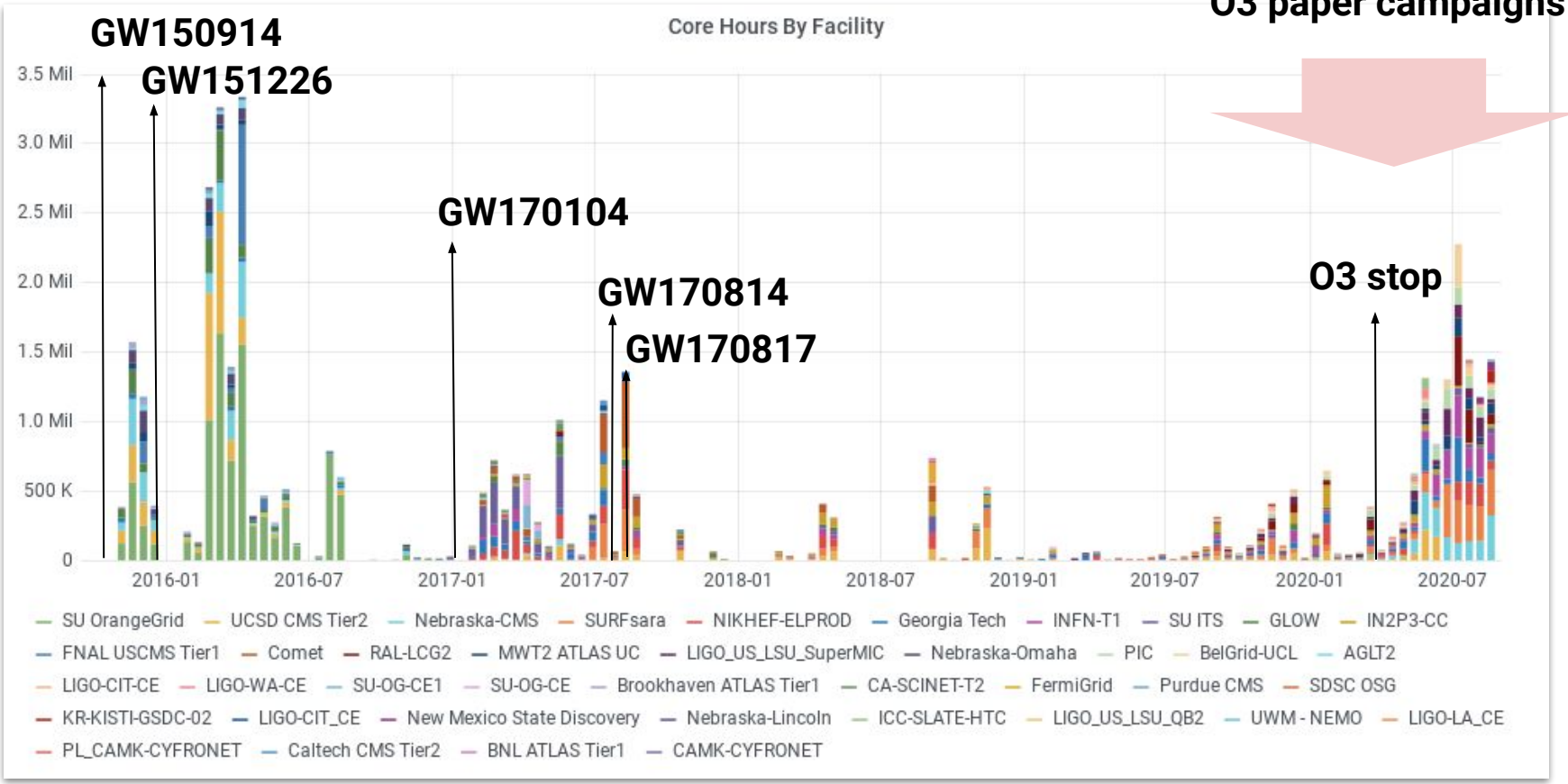
5 in/near production on the grid:

- **PyCBC**: searches archival for black hole &/or neutron star mergers
- **RIFT**: BH &/or NS merger parameter estimation
- **BayesWave**: model-independent signal characterization
- **LALInference** (and successor **Bilby**): BH &/or NS merger parameter estimation
- **Powerflux**: search for continuous signals from e.g. unknown pulsars

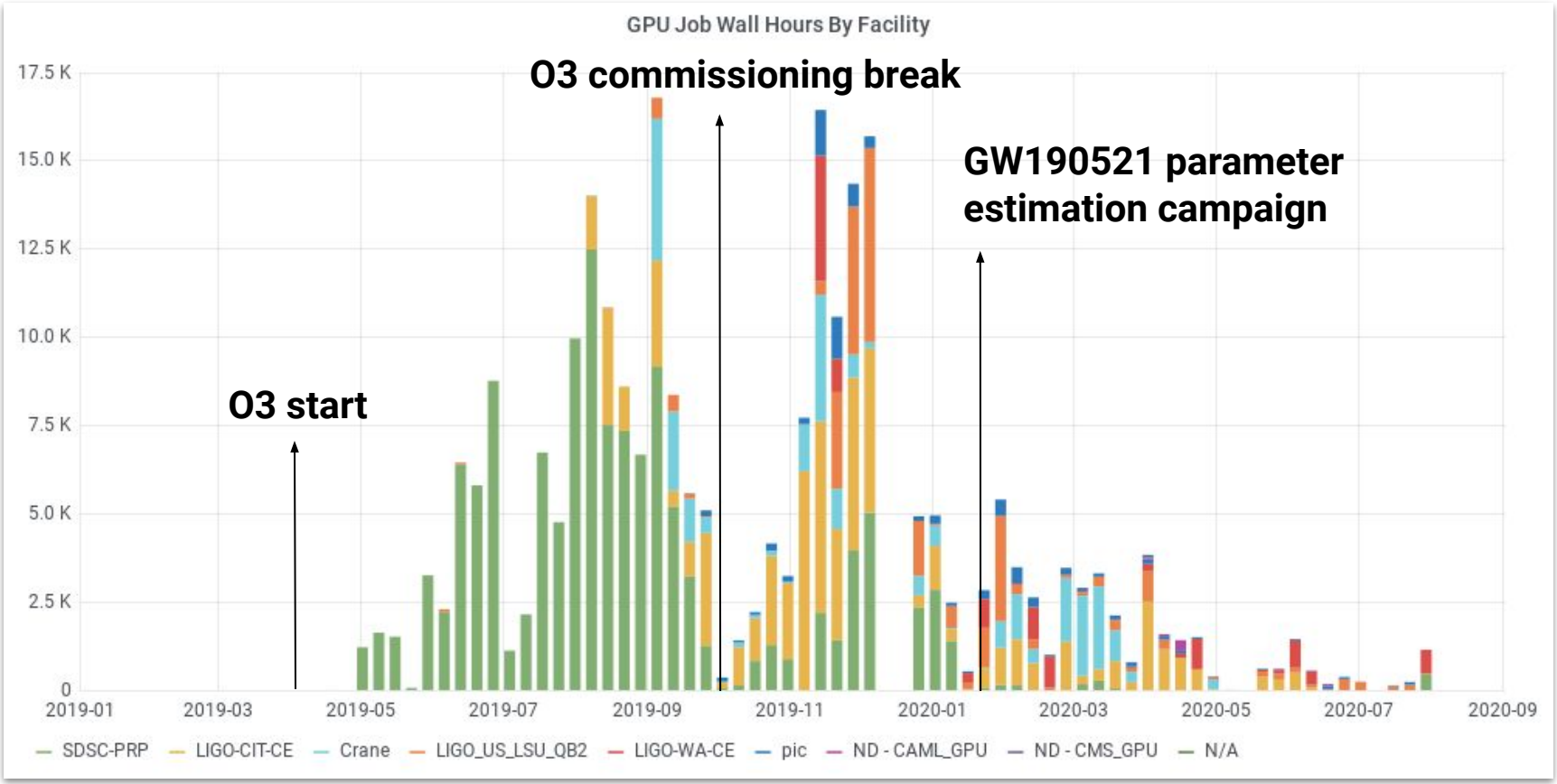
Onboarding new pipelines to IGWN-OSG grid, driven by (e.g.):

- Motivated (e.g. OSG school!), rapid development cycle
- Suitability / adaptability; classical HTC workflow? Latency/data requirements? Job lifetimes?
- Scientific prioritisation / hardware access

IGWN-OSG CPU History



IGWN-OSG GPU History



Software On The Grid

Legacy approach:

- Reference software environment installed on all LDG resources from OS-native packages
- developers build from source, iterate/petition for dependencies

Recent move to conda:

- Reference **conda** environments
- Pre-built envs available from IGWN CVMFS (`/cvmfs/oasis.opensciencegrid.org/ligo`)
- Developers create conda envs and/or build inside the reference environment
- Pipeline code releases → trigger conda feedstock → gitlab-CI rsync's OASIS, reference env updates

.. and containerization:

- LIGO gitlab instance → docker registry → sync's with OSG CVMFS singularity repository
- Scientists have distributed access to pre-conda'fied code & pre-release developmental code

Data On The Grid

Primary data product: gravitational wave “frame” files

- Multi-channel storage format for GW instrumental time-series, similar to but pre-dates HDF5
- Various levels of data-reduction & calibration versions, typically contain O(100 - 1000)s of gravitational wave strain from a given detector [O(100 - 1000)MB per file]

Data access patterns vary but *typical* IGWN-OSG job reads ~up to 3 files per execution attempt

- Workflows with many 1000s of jobs, developed with expectation of frame replicas on local storage
- Explicit file transfers → increased development & bandwidth overheads

Enter IGWN frame data in CVMFS:

- Proprietary: `/cvmfs/oasis.opensciencegrid.org/ligo/frames`
- Public: `/cvmfs/gwosc.osgstorage.org/gwdata`

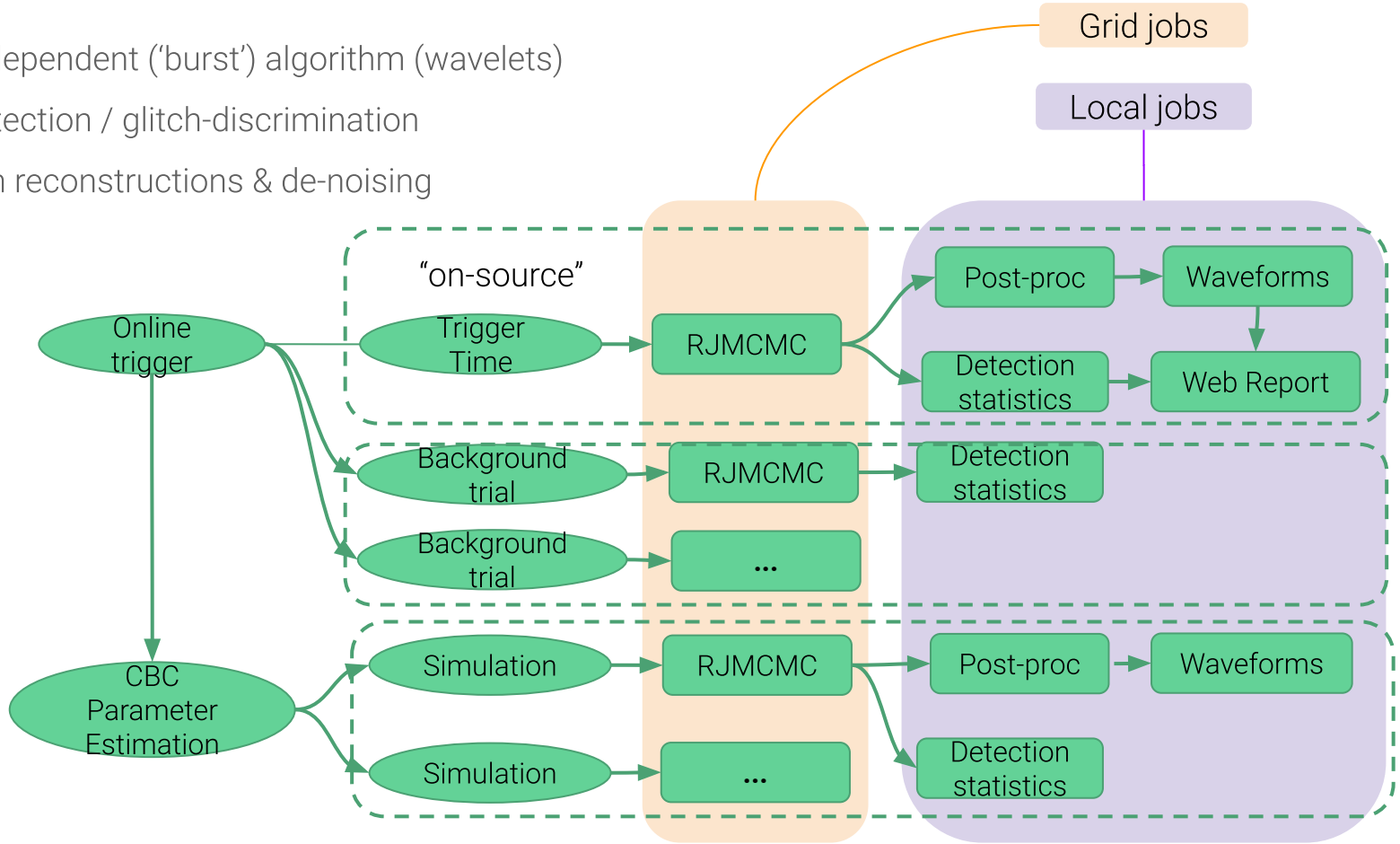
Ongoing/new projects:

- HTCondor plugin to permit file transfer with IGWN data discovery utils
- Formalizing distribution of intermediate / auxiliary analysis data through CVMFS

BayesWave https://docs.ligo.org/lscsoft/bayeswave/

Morphology-independent ('burst') algorithm (wavelets)

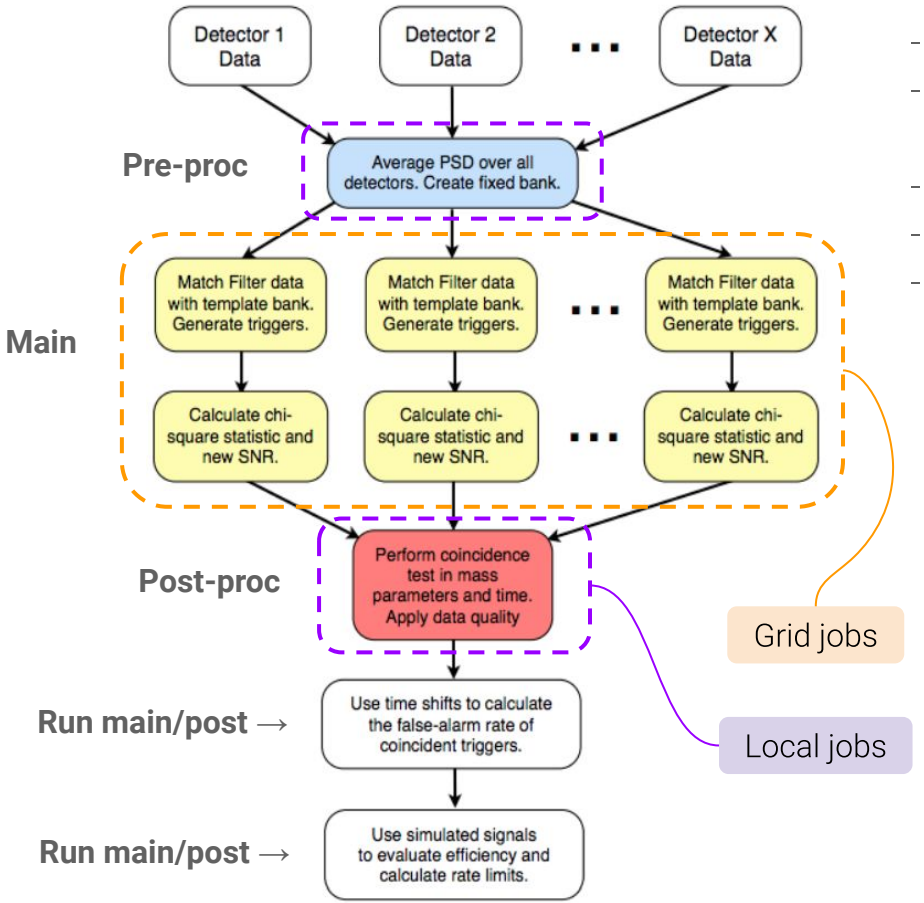
- signal-detection / glitch-discrimination
- waveform reconstructions & de-noising

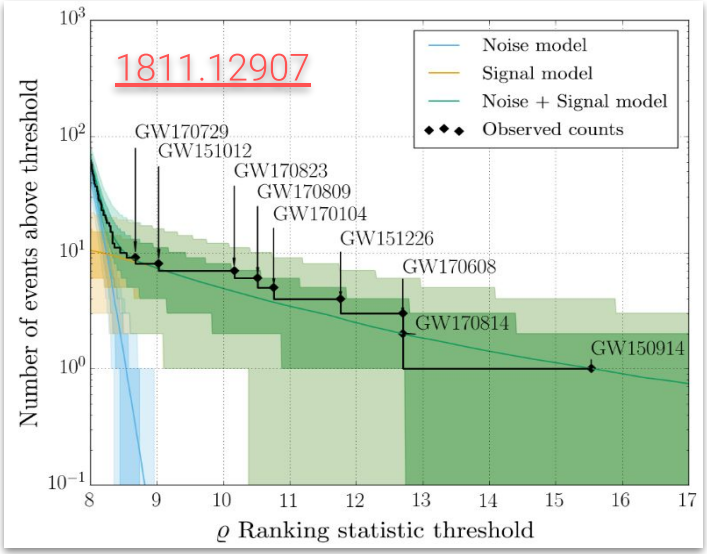


PyCBC (“off-line”) https://pycbc.org/

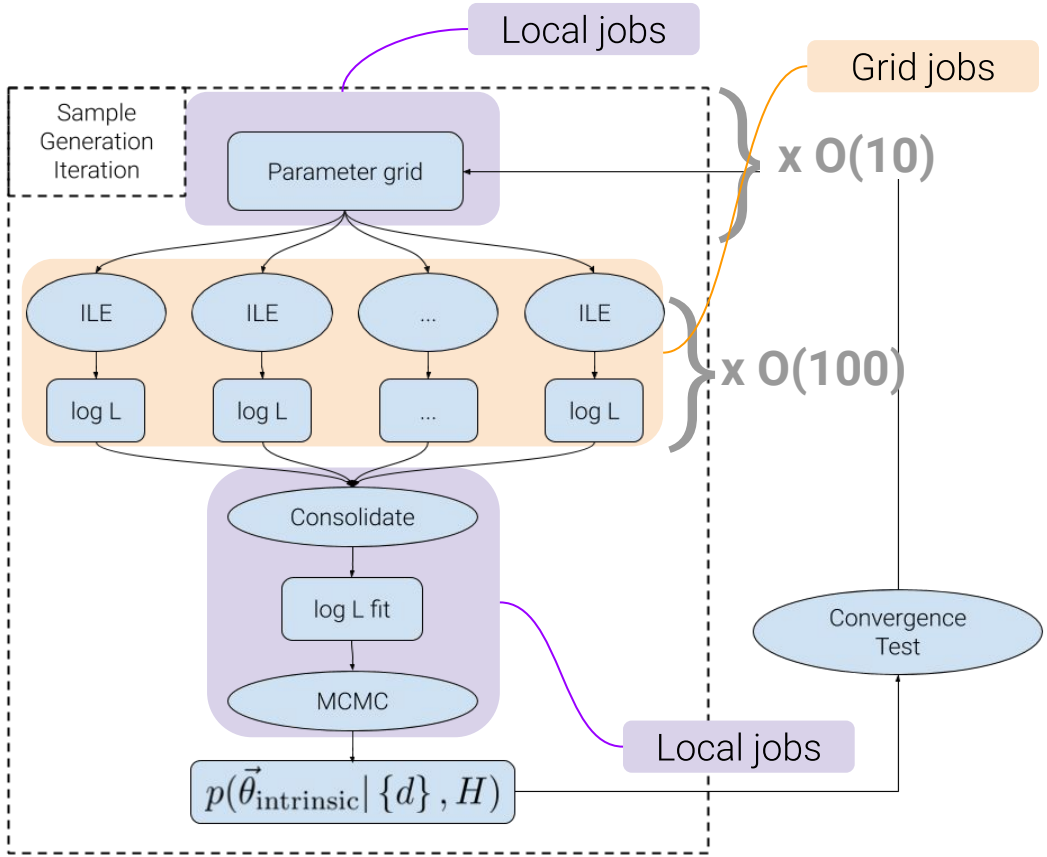
Matched filter search for BH &/or NS mergers

- Pre-, post-process stages → local HTCondor pools
- Main matched-filter jobs → IGWN-OSG grid
- Parallelism: 1 job ~ 2000s, repeat on background synthesised from time-shifts & simulations
- O(100k) jobs per day of data
- Early adopter for Pegasus WMS
- GPU version available for O3 analysis





Focus: RIFT (1805.10457)



Rapid parameter inference on gravitational wave sources via Iterative FiTting

Standard inference (e.g., `LALInference`):

- Approximate analytic models
- MCMC-based sampling in 15-D

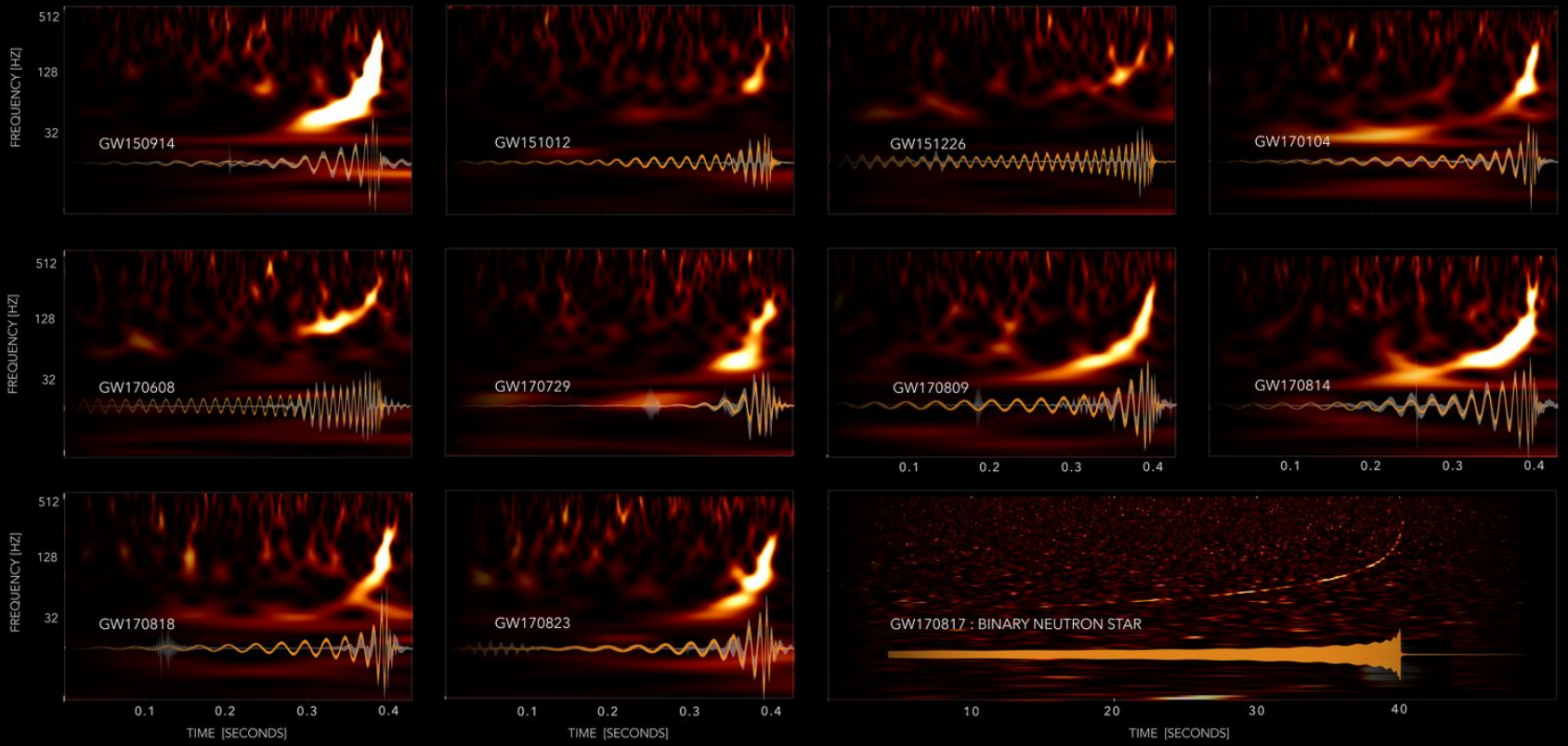
RIFT:

1. Start with discrete grid of waveforms
2. **I**ntegrate **L**ikelihood over **E**xtrinsic parameters
3. Fit marginal likelihood to grid
4. MCMC samples using likelihood fit
5. Refine grid, repeat

Developed for rapid evaluation of waveform approximants

Extended to direct use of numerical relativity waveforms from HPC simulations

GRAVITATIONAL-WAVE TRANSIENT CATALOG-1



LIGO-VIRGO DATA: HTTPS://DOI.ORG/10.7935/82H3-HH23

WAVELET (UNMODELED)

EINSTEIN'S THEORY

IMAGE CREDIT: S. GHONGE, K. JANI | GEORGIA TECH

Summary

Tremendous progress:

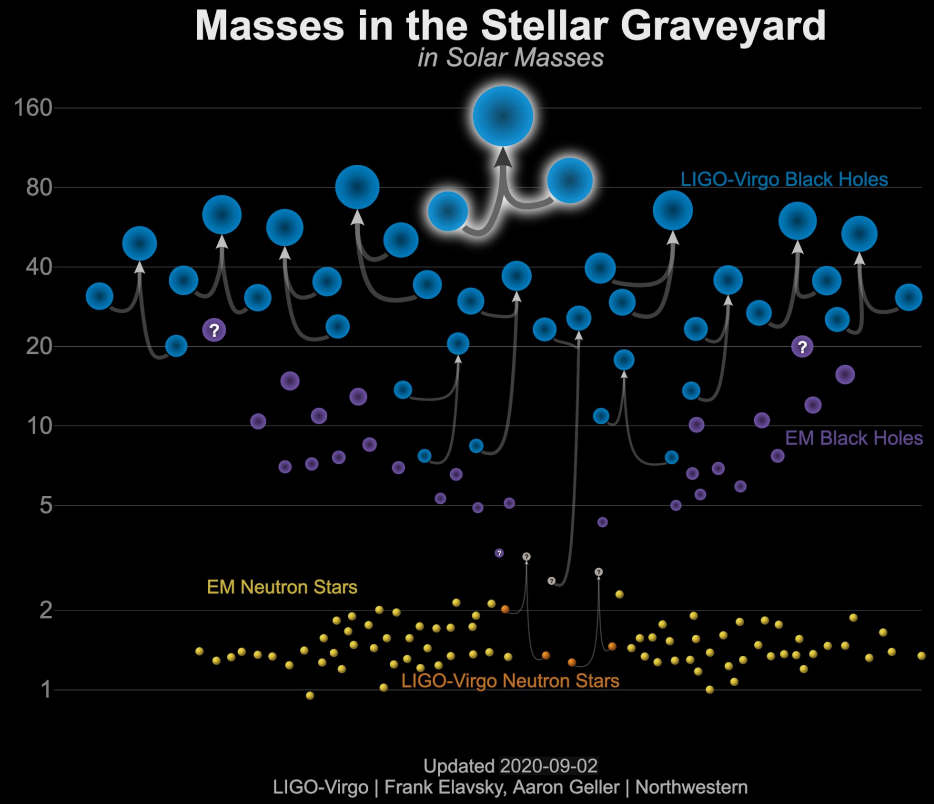
- Easy data & software distribution
- Many of our resource-hungry analyses now running in / near production

Coming soon:

- O3 publications…
- O4: ~late 2021/early 2022 +/- COVID
- Expect: ~10 BNS + ~80 BBH detections / year

OSG is central to long-term IGWN strategy

- Services & education/support, as well as resources!



THANK YOU