LIGO-G1900989

Updates from LIGO

james.clark@ligo.org

Condor Week 2019

What's LIGO (and who's this guy?)

LIGO: large (4km) experiment to detect cosmic gravitational waves

LIGO Scientific Collaboration (LSC): international consortium of ~1300 scientists (plus ~500 in Virgo) engaged in hardware operation & development, data analysis and astrophysics

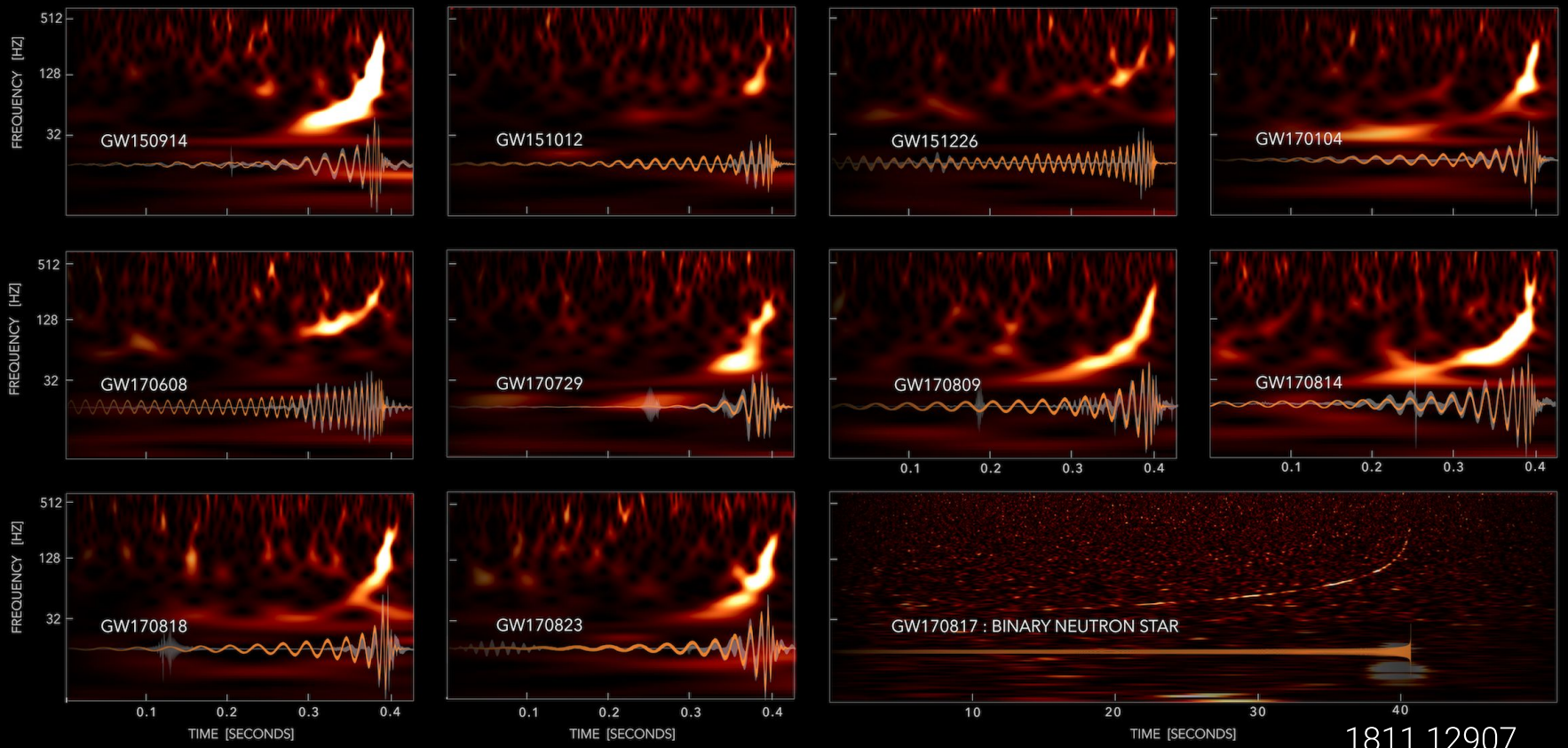
Sept 2015 - Jan 2016: historical first detections of gravitational waves from binary black hole (BBH) mergers (see e.g. 2017 Nobel prize in physics)

Nov 2016 - Aug 2017: multiple BBHs, first binary neutron star, inc. unprecedented electromagnetic follow-up campaign

Me: research scientist @ GATech

- 2005 - 2017: LIGO data analysis background (transient signals of uncertain morphology)
- ~2017 - now: LIGO DevOps / OSG & bulk data management (and a bit of science)

GRAVITATIONAL-WAVE TRANSIENT CATALOG-1



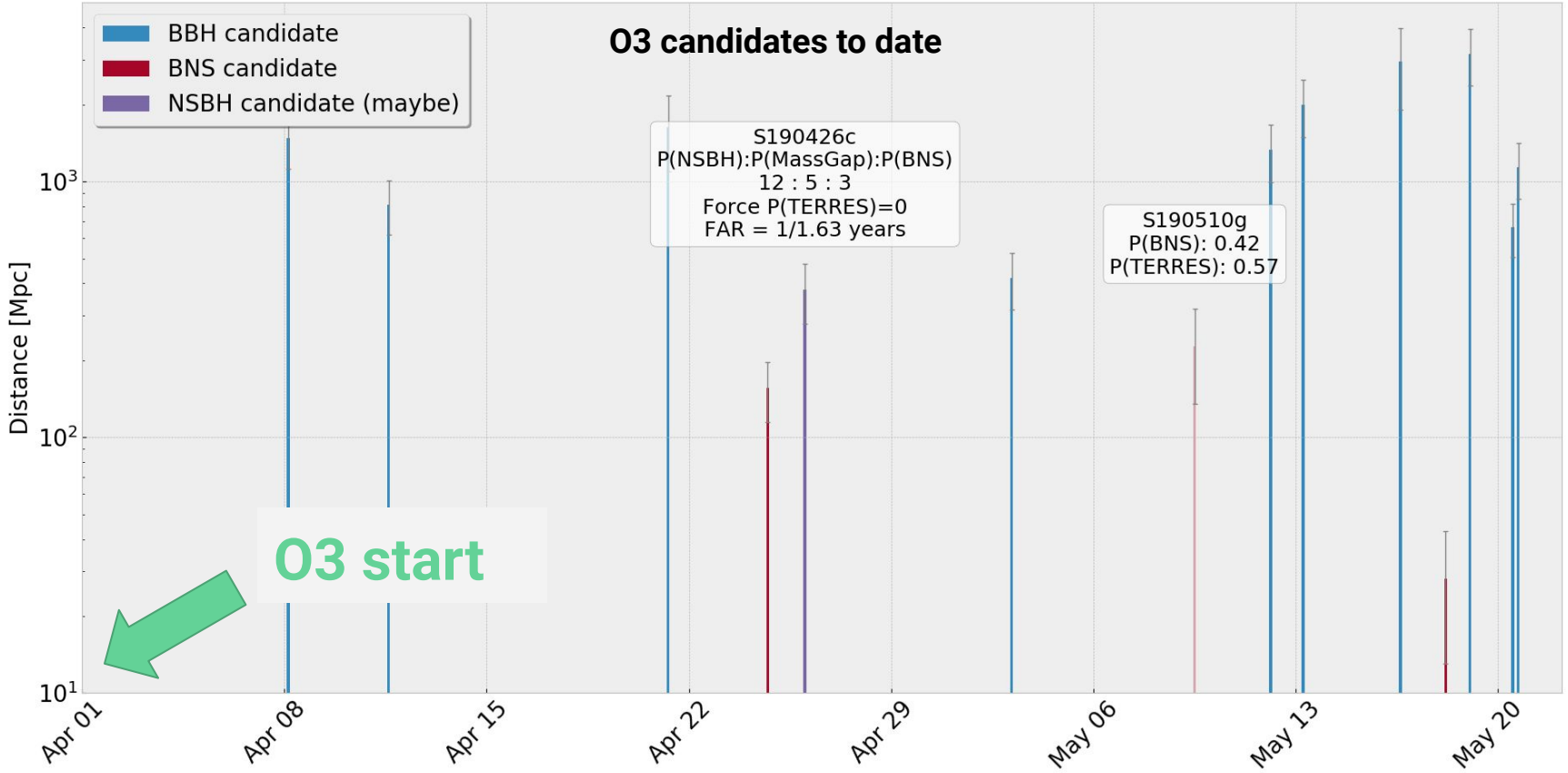
1811.12907

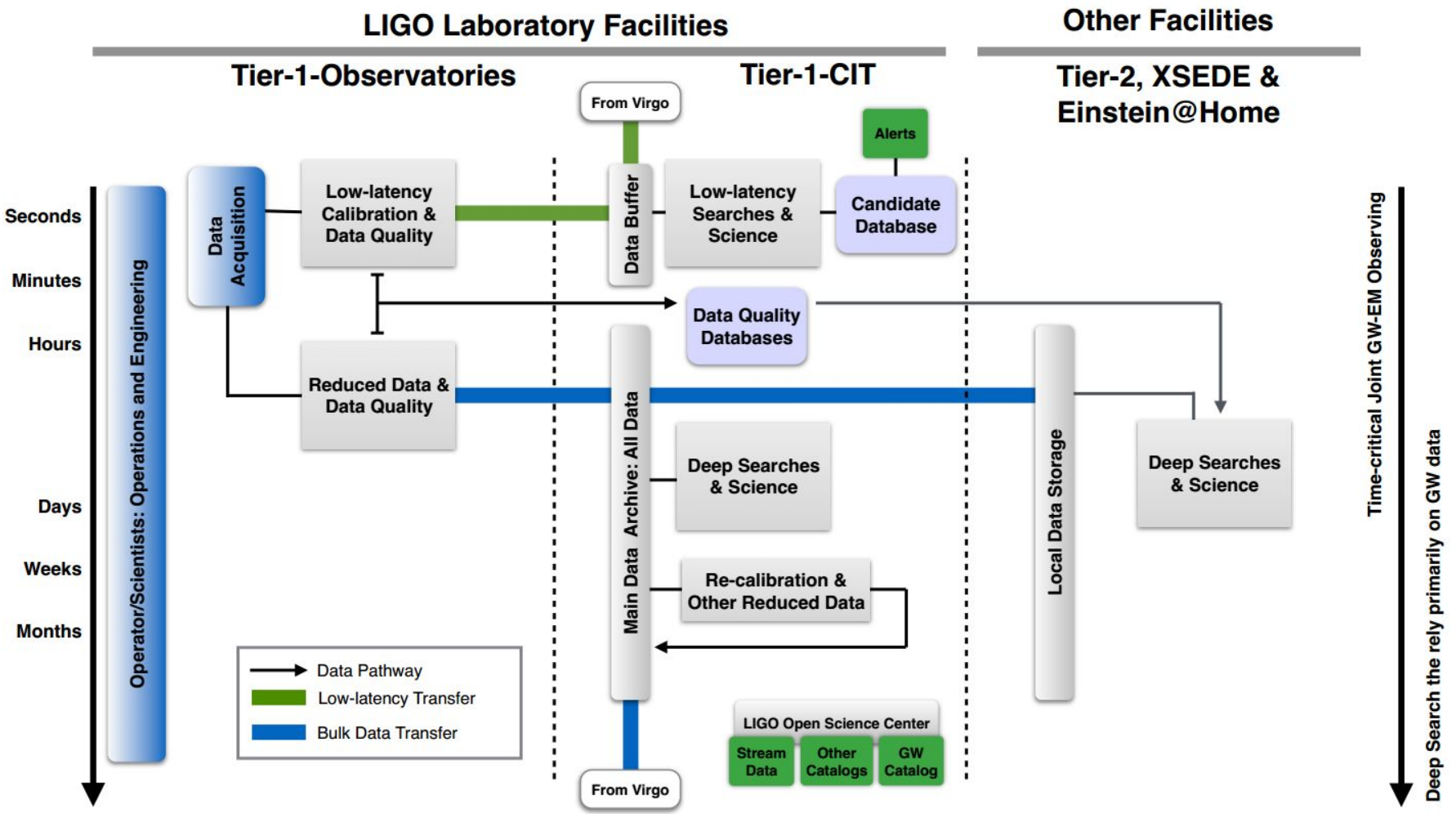
LIGO-VIRGO DATA: HTTPS://DOI.ORG/10.7935/82H3-HH23

WAVELET (UNMODELED) EINSTEIN'S THEORY

S. GHONGE, K. JANI | GEORGIA TECH

3rd LIGO/Virgo observing run (O3): 1 year




LIGO Laboratory Facilities
Other Facilities
Tier-1-Observatories
Tier-1-CIT
Tier-2, XSEDE & Einstein@Home
From Virgo
Alerts
Seconds
Minutes
Hours
Days
Weeks
Months
Operator/Scientists: Operations and Engineering
Data Acquisition
Low-latency Calibration & Data Quality
Data Buffer
Low-latency Searches & Science
Candidate Database
Data Quality Databases
Reduced Data & Data Quality
Main Data Archive: All Data
Deep Searches & Science
Re-calibration & Other Reduced Data
Local Data Storage
Deep Searches & Science
Data Pathway
Low-latency Transfer
Bulk Data Transfer
LIGO Open Science Center
Stream Data
Other Catalogs
GW Catalog
From Virgo
Time-critical Joint GW-EM Observing
Deep Search the rely primarily on GW data

Computing Scale

Data rates

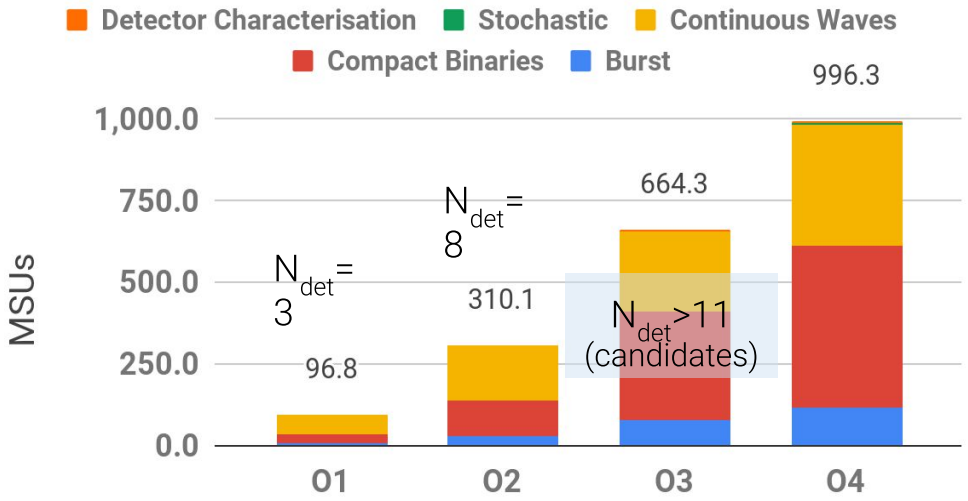
- Channels per site: ~200,000
- Strain per IFO: 0.12 MB/s
- Raw and Reduced data: 0.85 PB/yr
  - ~3 TB of strain data
- User data: 2.1 PB/yr

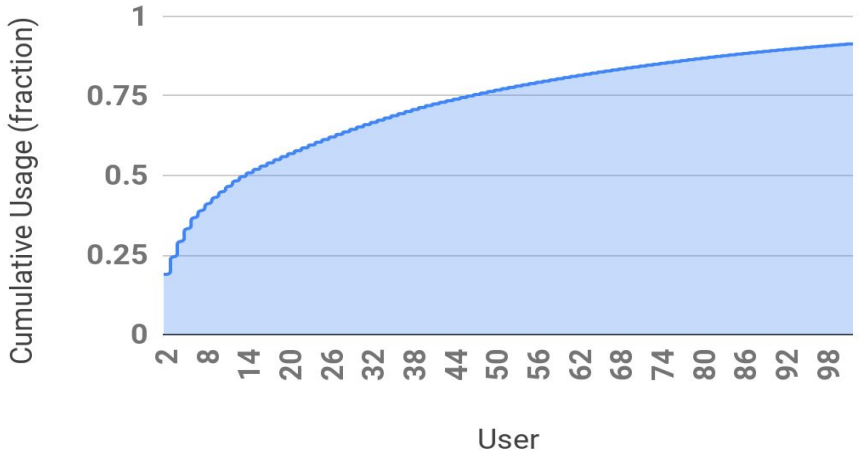
Total computing requirements

- MSU=10^6 E5-2670 core hours
- O3 projected usage: ~600 MSU

LIGO-Virgo Users:

- ~700 total, ~400 active in last year
- Top 50 users drive ~70% of demand



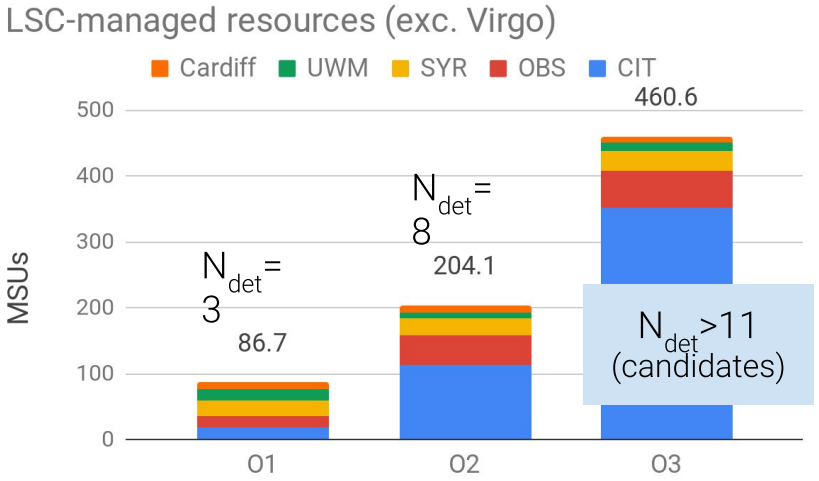


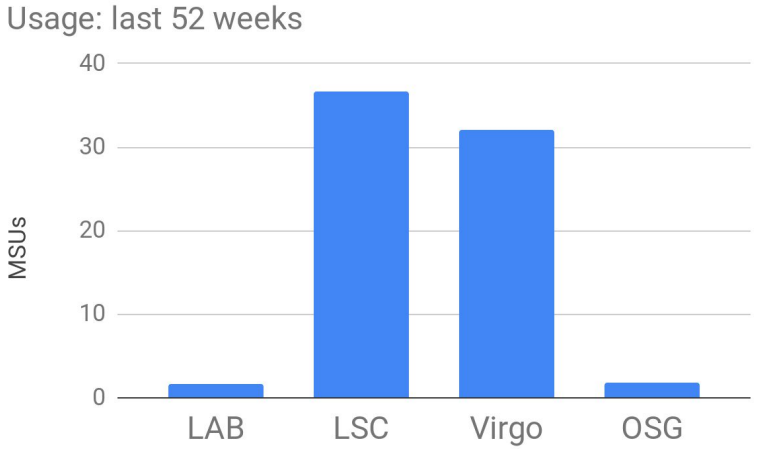
Supply

Diverse set of resources: dedicated, allocated & opportunistic

- LIGO lab
- LSC-institutions
- Virgo (mostly allocated)
- PI clusters (shared)
- Campus / regional shared clusters
- National supercomputers (XSEDE, Blue Waters)
- Opportunistic cycles
- Future: OzGrav, commercial cloud

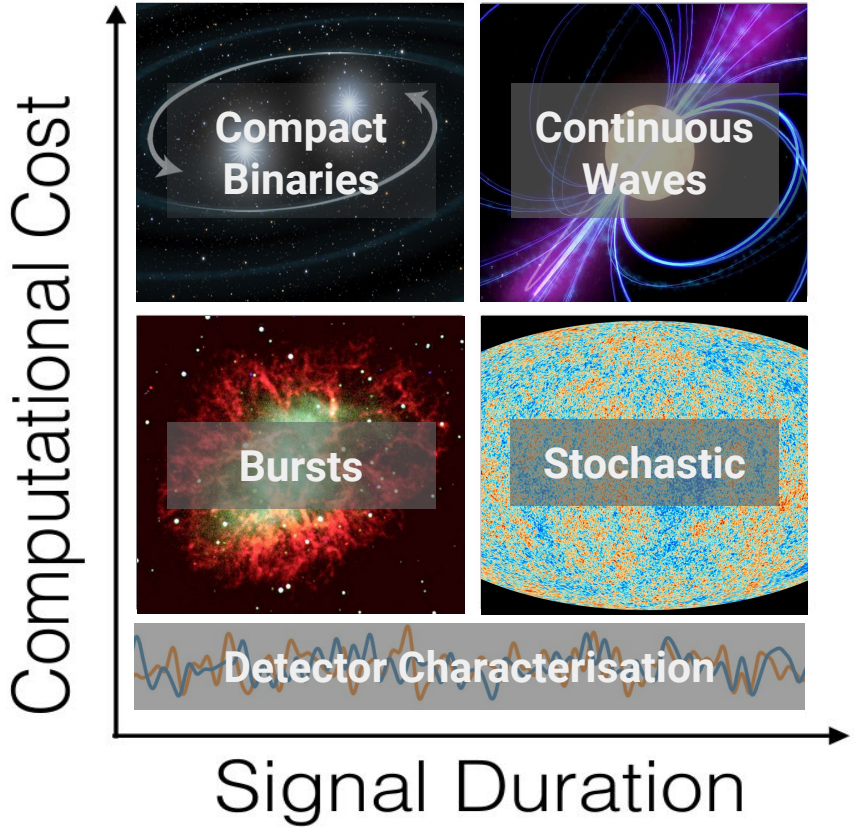
Runtime environments: LDG & OSG





LIGO Data Analysis

- Astrophysics groups:
  - Propose science goals
  - Determine analysis algorithms
  - Write and run applications
- Detector characterization: supports the commissioning teams, astrophysics groups
- Diverse algorithms and methods → heterogeneous demands on computing infrastructure
- Bulk of our analyses are ~~embarrassingly~~ pleasantly parallel → HTCondor & Pegasus



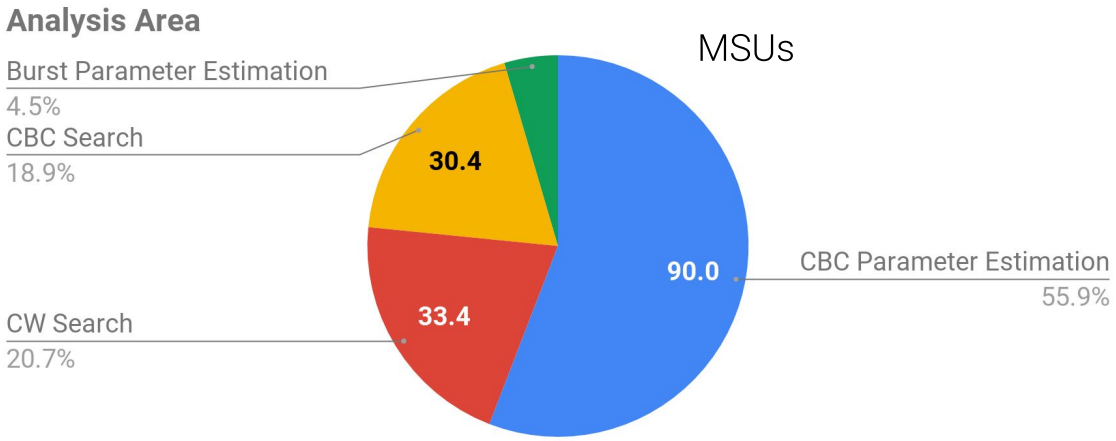
Main consumers

Demand dominated by:

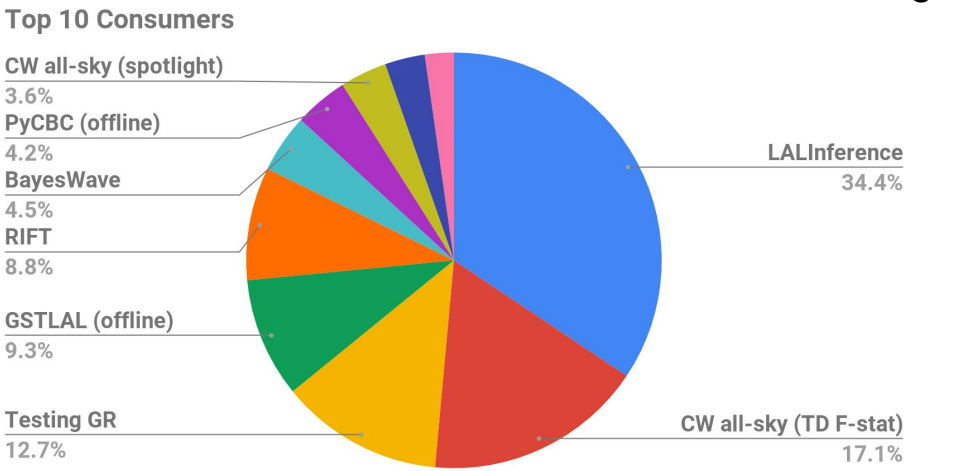
Parameter estimation - follow-up short data segments with densely sampled templates

CW searches - downsampled data, *large* numbers of templates

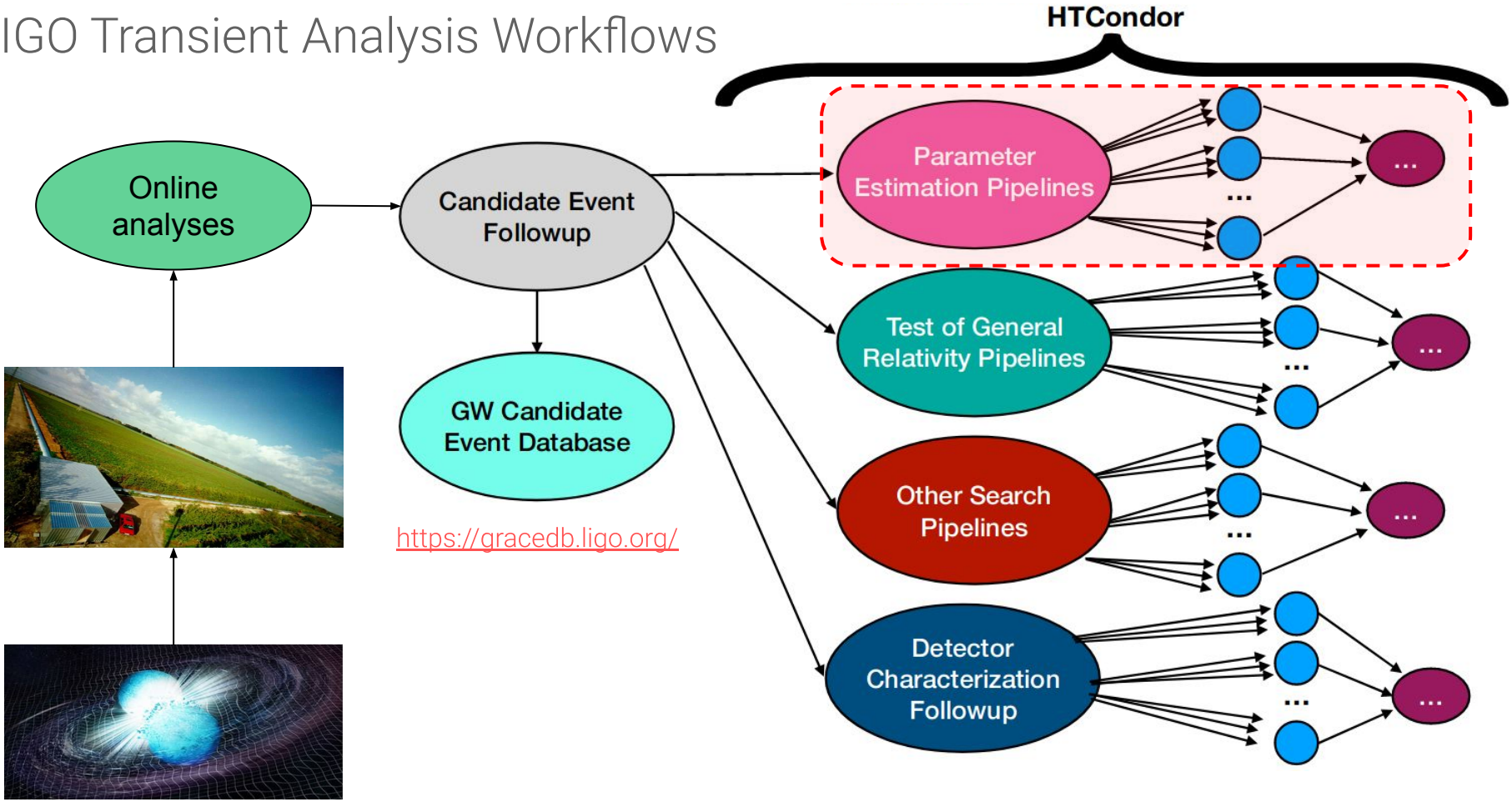
Offline CBC searches - compare large volumes of data, relatively sparse template banks



Last 52 weeks' usage



LIGO Transient Analysis Workflows



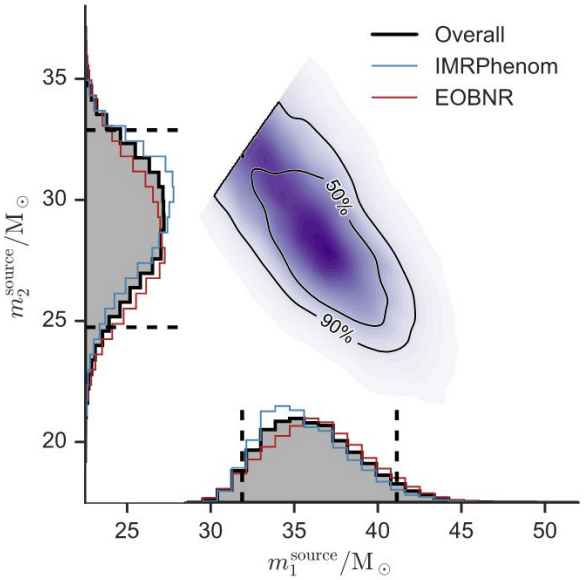
Determine source properties with Bayesian inference

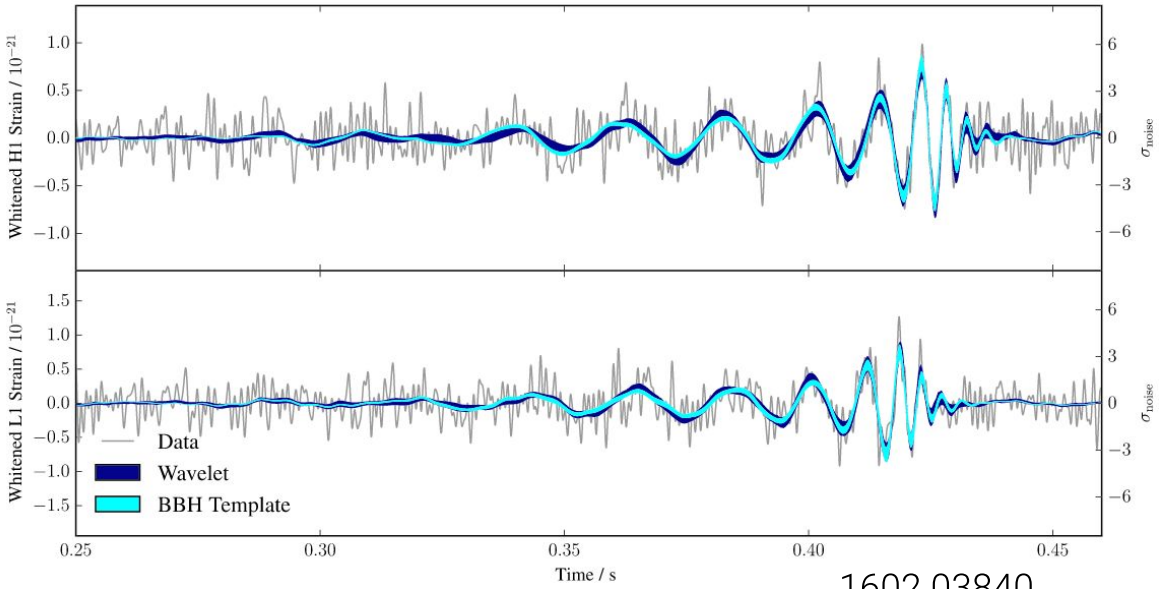
$$p(\vec{\theta}|\{d\},H) = \frac{p(\vec{\theta}|H)p(\{d\}|\vec{\theta},H)}{P(\{d\}|H)}$$

Generally:

- small data volumes
- ~simple workflows

- Templates: **θ** = sky-location, spins, masses, distance, …
- Wavelets: **θ** = sky-location, number of wavelets, wavelet frequency, bandwidth, …



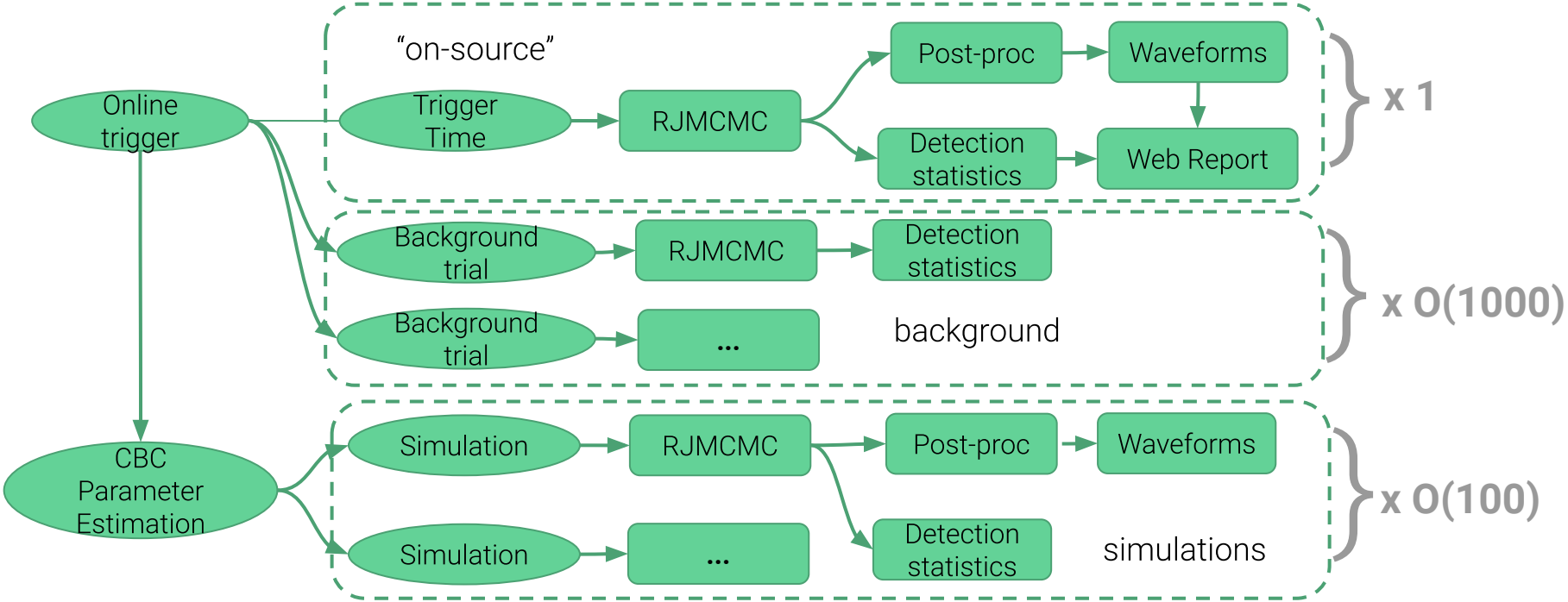


1602.03840

Focus: BayesWave (1410.3835)

Morphology-independent ('burst') algorithm (wavelets) for O(1 s) transients

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



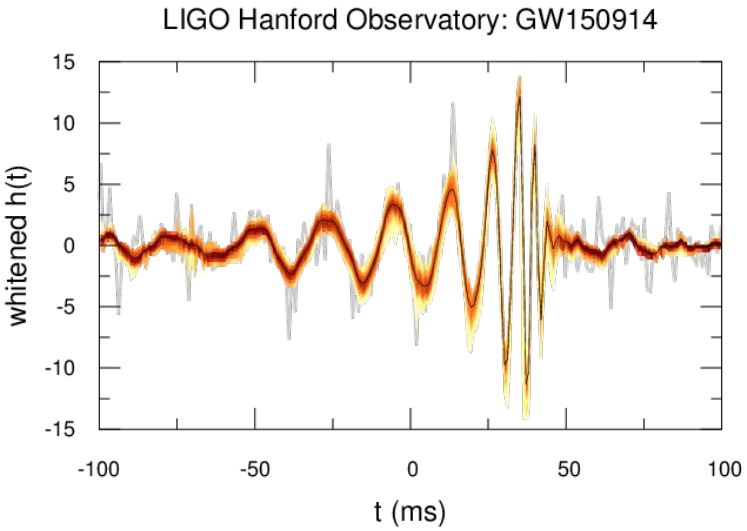
Focus: BayesWave

Individual BayesWave jobs: ~12-48 hours

- Huge parameter space
  - $D_{signal} = 5\,N_{wavelets} + 4$
  - $D_{glitch} = 5\,N_{detectors}\,N_{wavelets}$
- Trans-dimensional RJMCMC: non-parallelisable
- Checkpointing critical!

Recent developments:

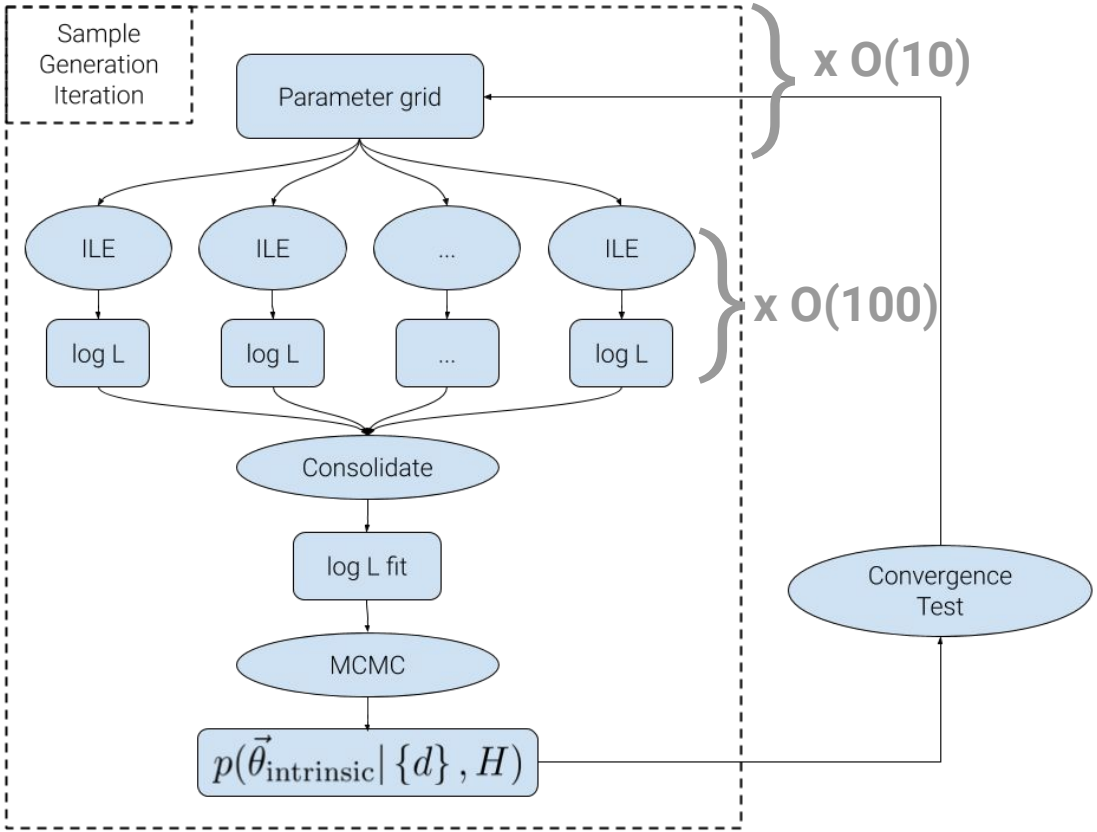
- Periodic checkpointing & file transfers: ✅
- Using CVMFS for:
  - Singularity containers ✅
  - LIGO Conda environment ✅
  - Proprietary LIGO data ✅
- OSG-deployed: ✅



User concerns re checkpointing:

- Cannot access intermediate results 😞
- Lose all progress after `condor_rm`'ing DAGs 😟

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

Focus: RIFT

LIkelihood eval: timeseries manipulations & matrix operations: numpy → cupy

- CPU likelihood eval: 407s
- GPU likelihood eval: 21s

Significant implications:

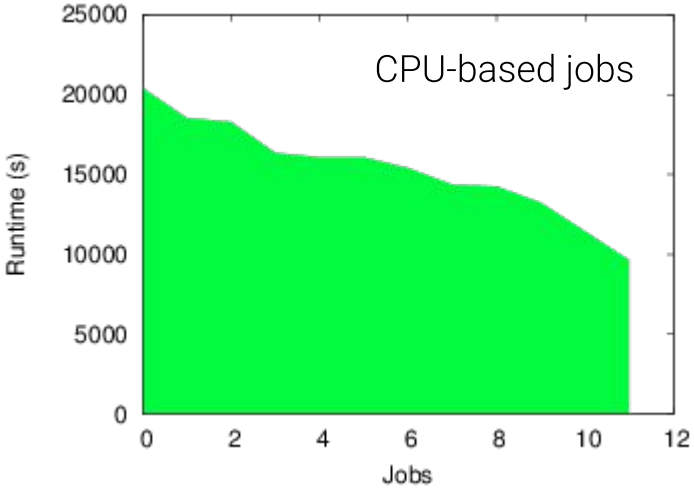
- rapid EM-followup
- simulations and probability coverage tests

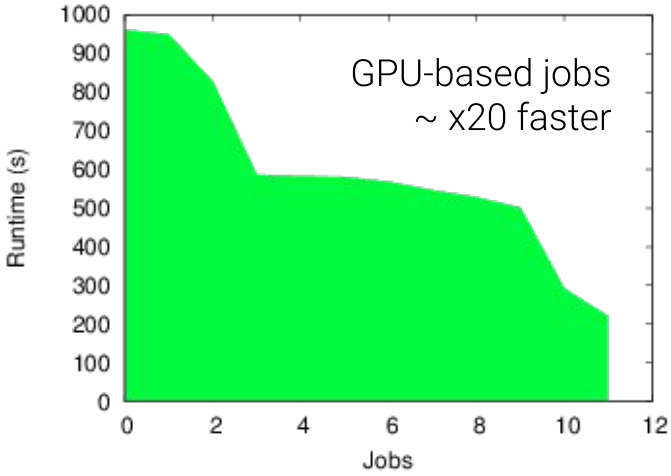
In the last month

- Containerised in LIGO CVMFS & running on OSG CPUs
- Deployed to PRP GPU cluster via docker

CVMFS:

- Usual LIGO strain data requirement
- WIP: distribute template data via CVMFS





Concluding remarks

O3: multiple candidates / week

“Strain” on in-house resources increasing → OSG etc growing in importance & opening up more resources (GATech, LSU, PRP, ...)

3 (out of >> 10) pipelines OSG-friendly

CVMFS for bulk data & software

Modern DevOps tech now playing a huge role (e.g., continuous integration, containerisation)

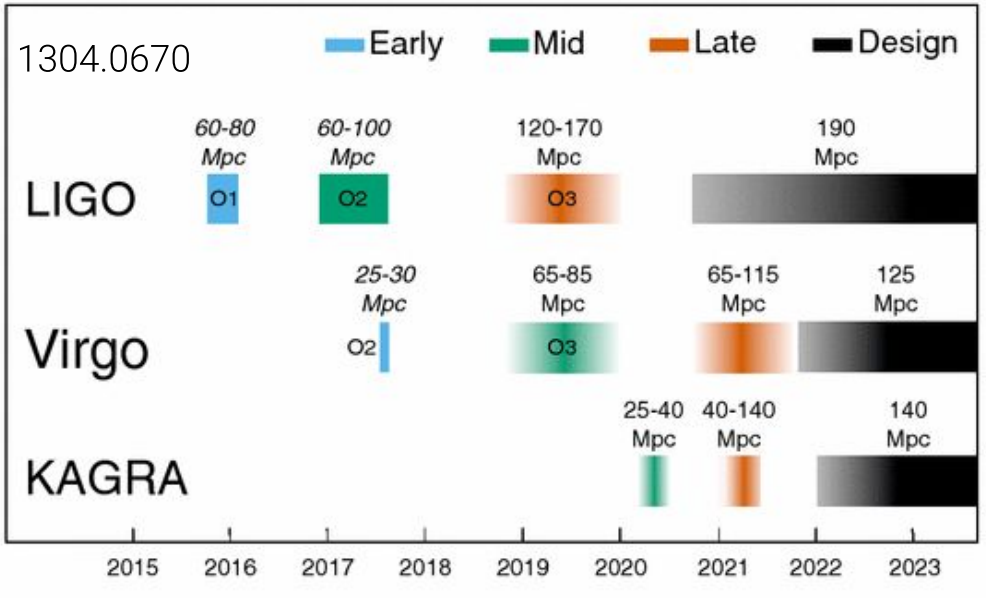
Synergy in LIGO-HTCondor-OSG community paying dividends

end

GW Forecast

O3 now running for 7 weeks:

- 9 BBH (cf. ~few / week)
- 1 - 3 BNS (cf.1/month - 1/year)
- O3 ends ~April 2020
- KAGRA (Japan) ~late O3



2020+ runs:

- aLIGO/AdVirgo nominal
- 4 - 80 BNS / year

2024+ runs:

- + LIGO India
- 11 - 180 BNS / year

RIFT → CVMFS

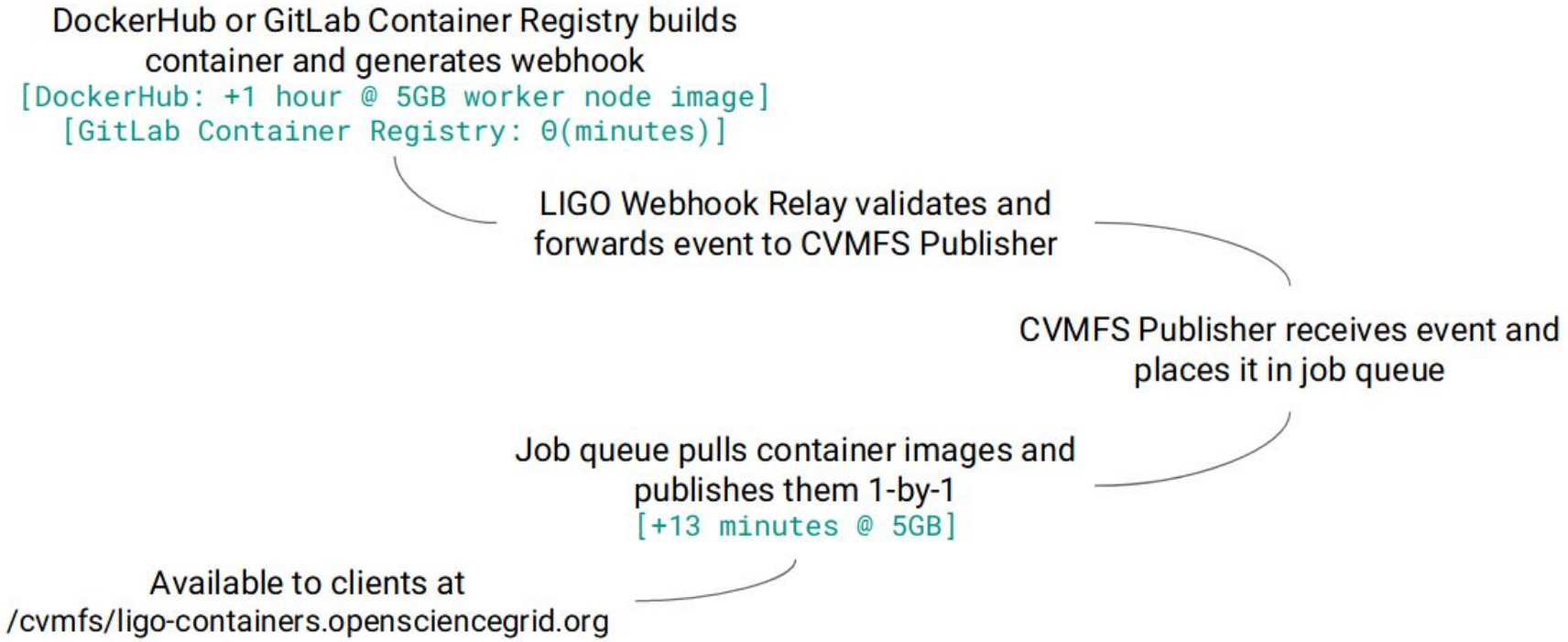
A summary of the data itself:

- updated ~once/year
- public — no authentication or access control needed
- Numerical relativity data sets look like:
- CIT data is O(100GB) total, consisting of O(100MB) HDF5 files
- GTech/RIT data are each O(100MB) total, consisting of O(1MB files)
- New, additional datasets will be released soon which may ~double the dataset size (i.e., probably not order of magnitude increases)
- Also helpful to host "surrogate" data (models for numerical relativity waveforms)
- O(10) waveform files ~10 MB - 10 GB

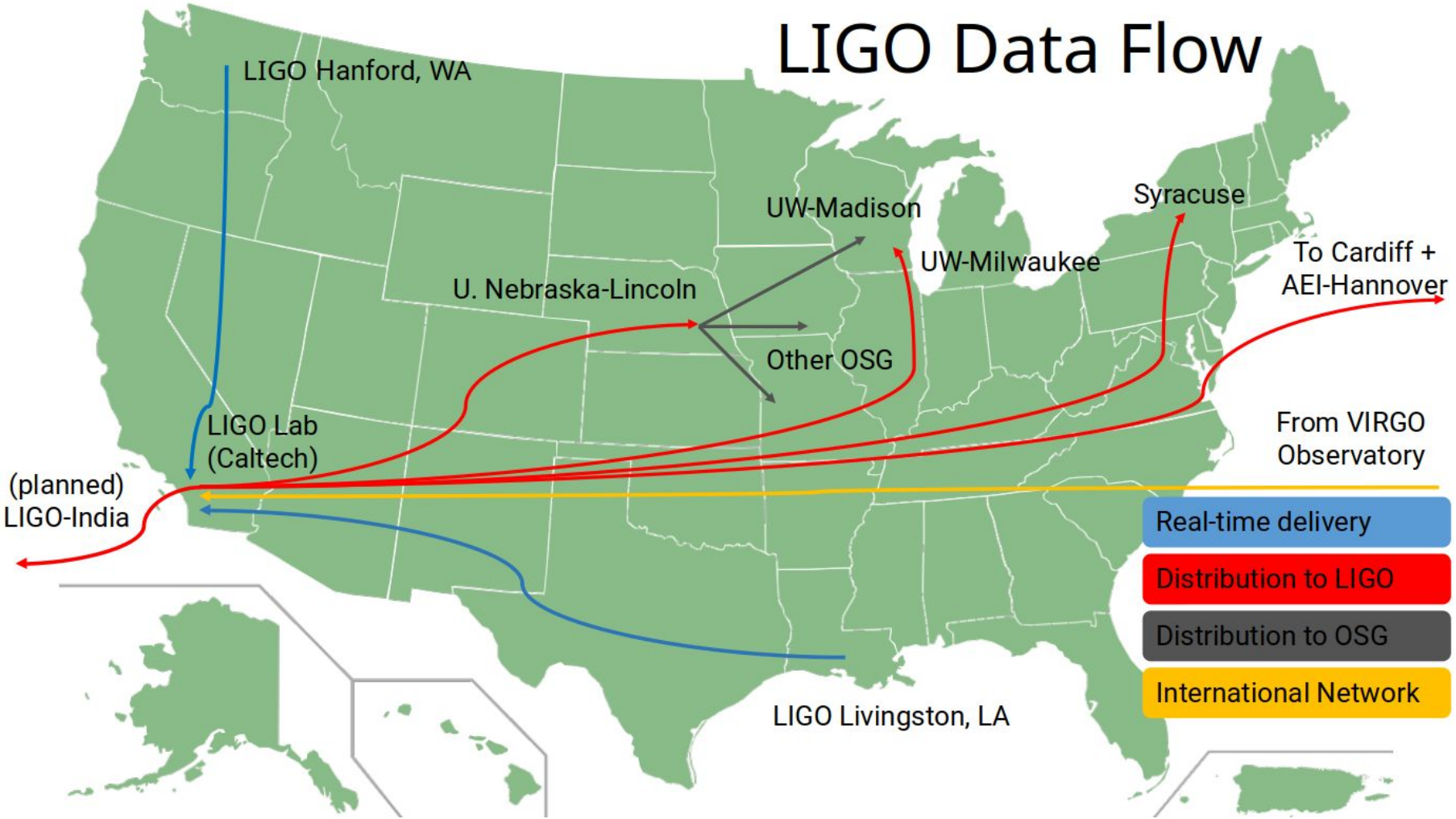
Usage pattern:

- At any one time, in full production, expect up to O(10-20k) RIFT jobs running on all LDG & OSG resources from all users
- 1 job runtime ~ 1 hour
- Each numerical relativity job performs a single read operation of up to 20-30 waveform files.
- A surrogate job reads a single surrogate file once

Containers in CVMFS



Within hour, a developer can test changes via Docker or on Open Science Grid using Singularity and CVMFS!


LIGO Data Flow
LIGO Hanford, WA
UW-Madison
Syracuse
UW-Milwaukee
To Cardiff +
AEI-Hannover
U. Nebraska-Lincoln
Other OSG
LIGO Lab
(Caltech)
From VIRGO
Observatory
(planned)
LIGO-India
Real-time delivery
Distribution to LIGO
Distribution to OSG
International Network
LIGO Livingston, LA