LSC LIGO Scientific Collaboration

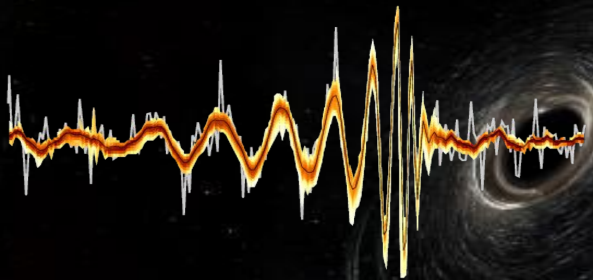
Georgia Tech Center for Relativistic Astrophysics College of Sciences

Containing BayesWave: Deploying LIGO analyses on the OSG

James Alexander Clark

james.clark@ligo.org

[GW150914 numerical relativity simulation (SXS collab.)]



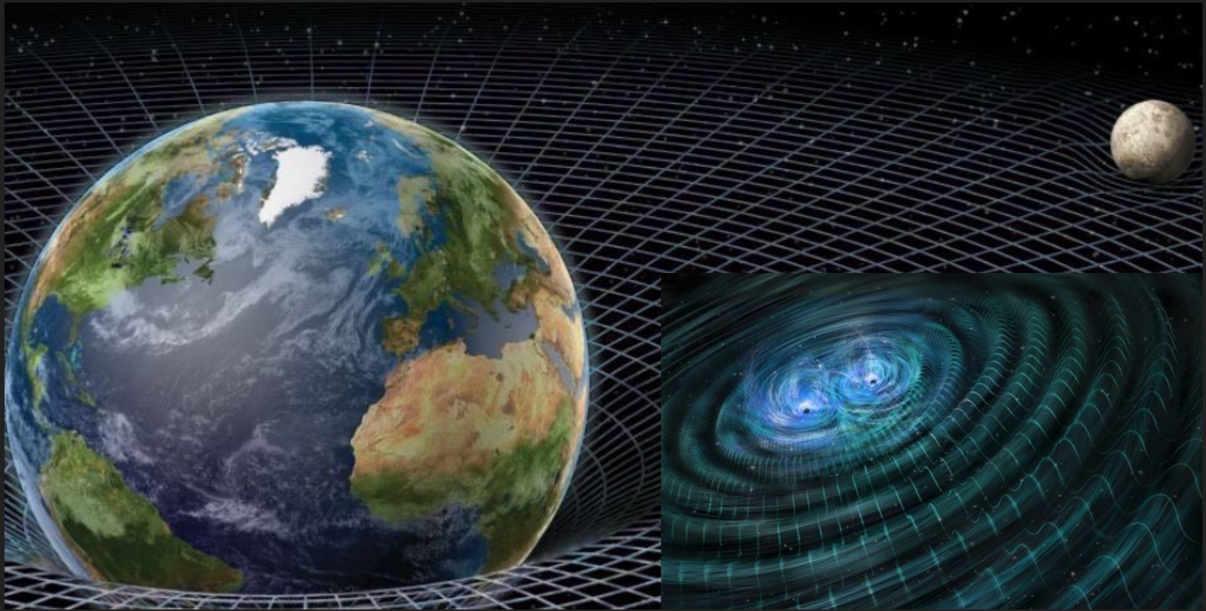
Un-modeled GW150914 reconstruction

NITRD MAGIC, March 7th 2018

LIGO-DCC-G1800346

Gravity: curvature of spacetime due to mass

Gravitational waves: perturbations in spacetime due to acceleration of mass



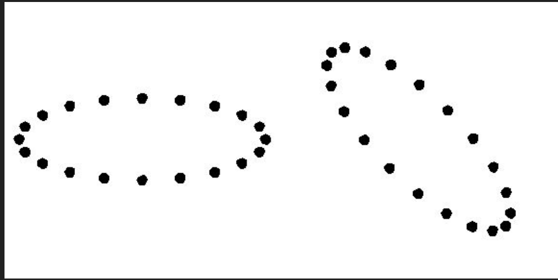
Earth-Moon orbit as motion in curved spacetime

Inset: gravitational wave emission from neutron star coalescence

Effect: tidal deformation h=dL/L

2 independent polarisations

Wave propagating out of page through freely-hanging particles:



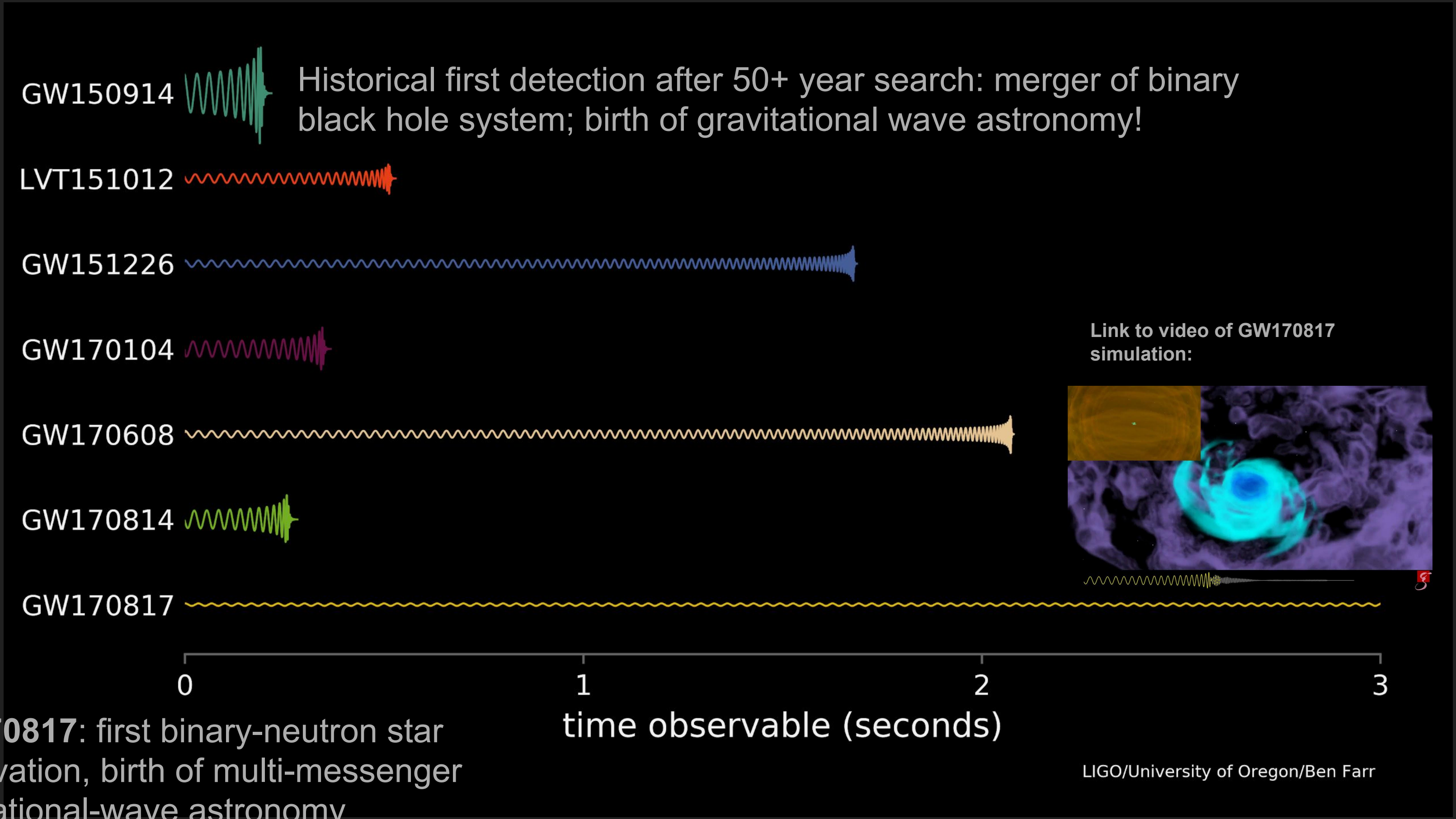
Magnitude: dL = h x L

Tidal strain magnitude h is tiny: h ~ 10^{-21} for a *loud* source

dL ~ 10^{-18} m for km-scale detector

Gravitational Wave Detections To Date

Historical first detection after 50+ year search: merger of binary black hole system; birth of gravitational wave astronomy!



Link to video of GW170817 simulation:

GW170817: first binary-neutron star observation, birth of multi-messenger gravitational-wave astronomy

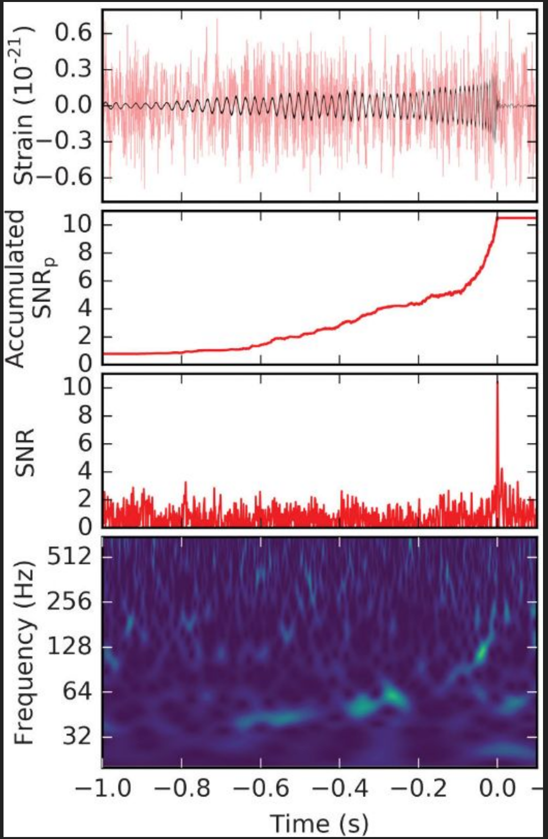
Gravitational Wave Data Analysis

- Observational data: noisy strain time series / detector
- Gravitational wave signals: O(10) ms -- O(1000) sec transients, continuous & stochastic signals
- Transient detection problem:
    - Known morphology: matched-filtering
    - Unknown morphology: statistical excess in time-frequency maps

Challenging computational problems in transient analyses:

1. *Detection*: Monte-Carlo simulations to reach statistical significance >5-sigma
2. *Bayesian source characterisation*:
    a. Analytic models for compact binary mergers with 15-dimensional parameter space
    b. Waveform reconstructions for un-modeled phenomenology

Main processing & validation generally embarrassingly parallel



GW151226: strain observations, matched-filtering & time-frequency decomposition , B. P. Abbott et al, Phys. Rev. Lett. 116, 241103 (2016)

LIGO Compute Resources

Many types of supply: dedicated, allocated, opportunistic, Many providers in the US and abroad:

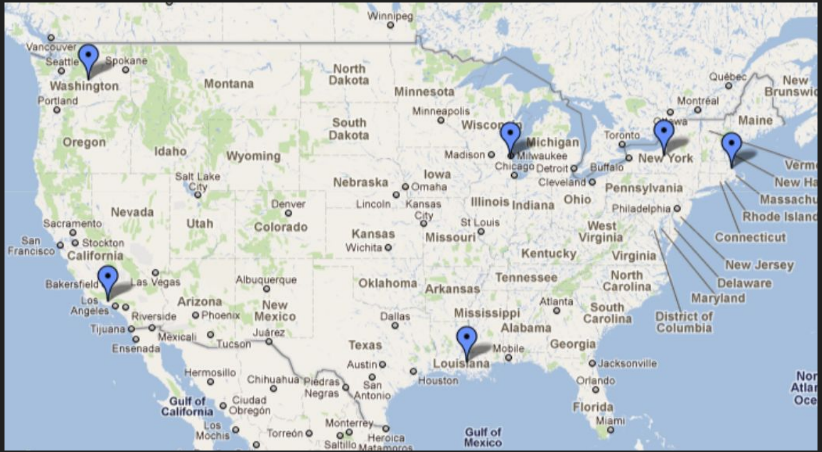
- Dedicated LIGO Lab clusters (HTC)
- Dedicated LSC clusters (HTC)
- Virgo clusters (mostly allocated on shared resources, HTC)
- PI clusters (shared, HTC and HPC)
- Campus/regional shared clusters (allocated, HTC and HPC) e.g., OrangeGrid, PACE, SciNet
- National shared supercomputers (allocated, HTC and HPC) e.g., XSEDE, Blue Waters
- Opportunistic cycles (campus clusters, DOE labs, HEP clusters, etc.)

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

1. O1 LDG usage: ~83%
2. O1 OSG usage: ~17%

LIGO Analysis And The LIGO Data Grid (LDG)

LIGO Scientific Collaboration operates a collection of Tier 1,2 compute resource: the LDG



LDG sites: computational and data storage resources with grid computing middleware to create a coherent and uniform LIGO data analysis environment

LDG users manually partition workflows across sites, and expect dependencies, shared filesystems, persistent storage. HTCondor/DAGMan & Pegasus for workflow management.

LIGO Analysis And The Open Science Grid (OSG)

As of ~2015, LIGO production analyses began to access the OSG

Challenges to LIGO OSG deployment:

- Data access: LDG provides compute *and* data storage. Users/workflows expect POSIX-access to input strain data
  - Use StashCache & CVMFS extensions to publish data to CVMFS repo: **`ligo.osgstorage.org`** when data is not hosted locally (see arXiv:1705.06202)
- Software availability: LDG sites provide reference OS & standard LSC software stack
  - Initial OSG deployments: deploy LSC software stack & specific analysis codes to CVMFS
  - **Now: Docker images deployed automatically to CVMFS repo: `containers.ligo.org` via gitlab CI**

CVMFS: global filesystem used to distribute publicly-available software.

Case Study: BayesWave

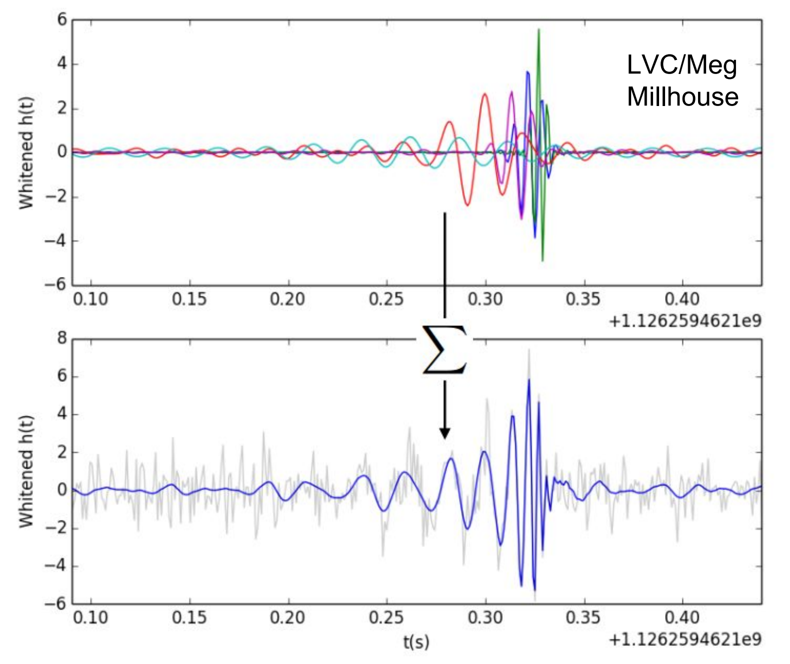
BayesWave [arXiv:1410.3835]: evaluates Bayesian evidence & produces posterior samples for a) gravitational wave (GW) signal model of arbitrary morphology, b) instrumental glitch model, c) Gaussian noise

Signals modeled as superposition of arbitrary number of Gaussian wave-packets (wavelets)

Number & parameters of wavelets determined by RJMCMC

$$h_+(t) = \sum_{i=0}^{N_s} \Psi(t; A_i, f_{0,i}, Q_i, t_{0,i}, \phi_i)$$

$$h_\times(t) = \epsilon h_+(t) e^{i\pi/2},$$



Containing BayesWave

- BayesWave: combination of C executables & python plotting / web-page generation scripts
- Builds & runs natively in LDG environment, main dependency: LIGO Algorithm Library (LALSuite)
- LIGO maintains a variety of Docker images, including a nightly build of LALSuite: https://hub.docker.com/u/ligo/
- Create BayesWave image from LALSuite:nightly base + BayesWave source in ligo gitlab
- Dockefile from: https://git.ligo.org/james-clark/bayeswave/tree/O2_online-branch

```
FROM containers.ligo.org/lscsoft/lalsuite:nightly
ARG version
RUN echo "Building bayeswave"
MAINTAINER James Alexander Clark <james.clark@ligo.org>

# Dependencies
USER root
RUN apt-get update && apt-get install --assume-yes build-essential

# Copy and build BayesWave
WORKDIR /
COPY install.sh /
COPY src /src
COPY postprocess /postprocess
COPY test /test
COPY utils /utils
RUN ls -lth
RUN sh install.sh / /opt/bayeswave
RUN rm -rf install.sh /src /postprocess /test /utils
RUN mkdir -p /cvmfs/oasis.opensciencegrid.org/ligo/frames

# BayesWave env
ENV PATH /opt/bayeswave/bin:${PATH} \
        PATH /opt/bayeswave/postprocess:${PATH} \
        PATH /opt/bayeswave/postprocess/dist:${PATH} \
        PATH /opt/bayeswave/postprocess/skymap:${PATH} \
        PATH /opt/bayeswave/postprocess/skymap/dist:${PATH} \
        PYTHONPATH /opt/bayeswave/postprocess:${PYTHONPATH} \
        PYTHONPATH /opt/bayeswave/postprocess/skymap:${PYTHONPATH} \
        PYTHONPATH /opt/bayeswave/utils:${PYTHONPATH}

ENTRYPOINT ["/bin/bash"]
```

Build & Deploy

Use gitlab CI script to build, deploy & test container

```
stages:
        - build
        - test

build-o2_online:

        only:
        - O2_online-branch
        stage: build
        script:
        - docker login -u gitlab-ci-token -p $CI_BUILD_TOKEN containers.ligo.org
        - docker build --rm --no-cache -t containers.ligo.org/james-clark/bayeswave:o2_online  ./
        - docker push containers.ligo.org/james-clark/bayeswave:o2_online

test-o2_online:

        only:
        - O2_online-branch
        stage: test
        script:
        - docker login -u gitlab-ci-token -p $CI_BUILD_TOKEN containers.ligo.org
        - docker run -u $(id -u):$(id -g) containers.ligo.org/james-clark/bayeswave:o2_online "/opt/bayeswave/test/test-bayeswave.sh"
```

BayesWave & Singularity

Run BayesWave singularity container via wrapper script:

```
#!/bin/bash
curl ipinfo.io # Geolocation for OSG testing
singularity exec \
      --bind /cvmfs/oasis.opensciencegrid.org/ligo/frames/ \
      --home ${PWD} \
      --contain \
      --writable \
      /cvmfs/ligo-containers.opensciencegrid.org/james-clark/bayeswave:o2_online \
      /opt/bayeswave/bin/bayeswave "$@"
```

Bind CVMFS data repository (ligo/frames) to directory inside the container for data access

Wrapper script: full control over container & seamless operation on LIGO resources, OSG resources and even local workstations

Workflow managed via HTCondor

Summary

Gravitational wave astronomy is here & increasingly resource hungry as sensitivity improves

LIGO production analyses historically served by dedicated resources managed by full-time LIGO personnel: standard software stacks & local data

Desire to exploit shared/allocated/opportunistic resources motivates containerized solutions

Distributed LIGO analyses spearheaded by pycbc search for binary mergers

Relatively new developments: seamless deployment to traditional & non-traditional resources

Containerisation & CI deployment to CVMFS will be a game-changer

BayesWave containers & OSG: potential expansion from expensive follow-up to full-blown search; experience now informing other, similarly expensive Bayesian inference algorithms

Thanks!

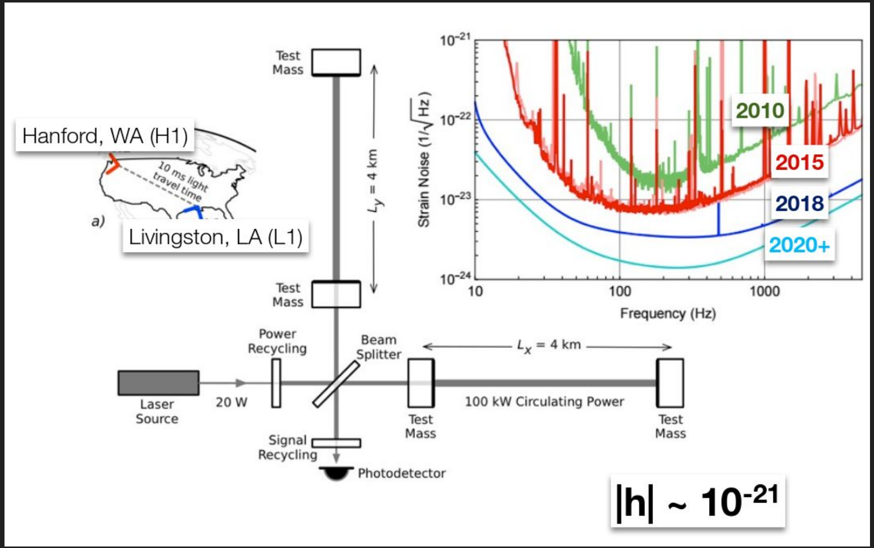
Special thanks to the Open Science Grid, and particularly to Thomas Downes and Adam Mercer of UW Milwaukee for their tireless work to provide LIGO containers and to integrate LIGO gitlab with CVMFS, as well as countless others in the LIGO Scientific Collaboration and Virgo Collaboration who make the science possible.

Supplementary

Gravitational Wave Detection

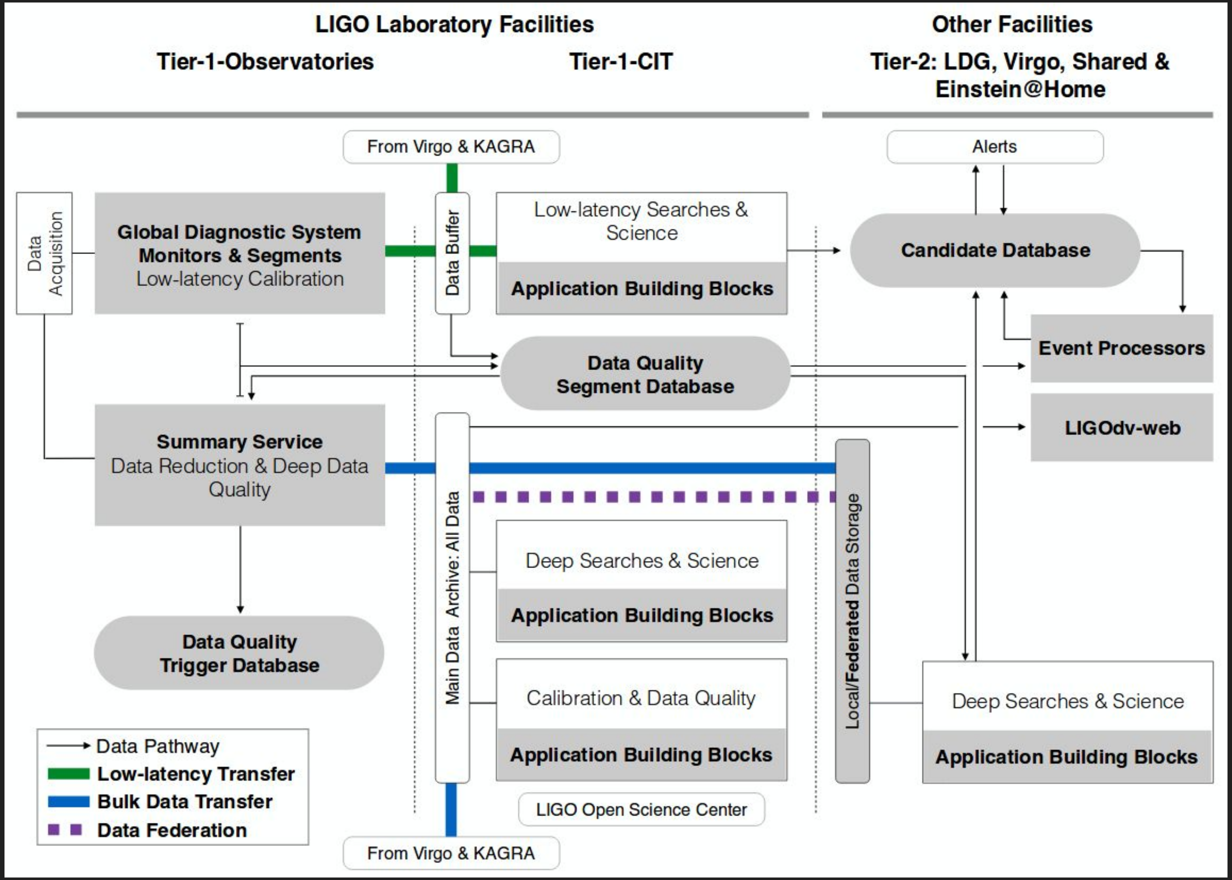
Modern detection via Michelson interferometers: measure interference pattern from grav-wave propagation

LIGO: 2x 4km-long interferometers in WA + LA



Worldwide network of ground-based detectors in operation, with more planned

LIGO Data Grid Operations



The Networking and Information Technology Research and Development (NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674, Fax: 202-459-9673, Email: nco@nitrd.gov, Website: https://www.nitrd.gov