The Search for Gravitational Waves

Alexander Nitz
MPI for Gravitational Physics
(Albert Einstein Institute)

PASC18, July 4th, 2018
Gravitational Wave Polarizations

- Gravitational waves have two physical degrees of freedom. These manifest as two polarizations, called the “plus” and the “cross” polarizations.
Gravitational Wave Polarizations

- Gravitational waves have two physical degrees of freedom. These manifest as two polarizations, called the “plus” and the “cross” polarizations.
First Detection

~36 and 29 solar mass black holes merging

~400 Mpc away
GW1701817 and GRB170817A

GRB170817A occurs \((1.74 \pm 0.05)\) seconds after GW170817

It was autonomously detected in-orbit by Fermi-GBM (GCN was issued 14s after GRB) and in the routine untargeted search for short transients by INTEGRAL SPI-ACS
Types of Sources

- Compact Binary Mergers
- Continuous waves
- Unmodeled Coherent Bursts
- Stochastic background
Matched Filtering to extract the signal

Signal embedded in strain data $s(t)$

Model waveform filters: $h(t)$

$\rho = \frac{(s|h)}{\sqrt{(h|h)}}$

$(a|b) = 4 \int_{0}^{\infty} \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} df$

SNR
Bank of Templates

- For the purposes of detection we simplify our waveforms to only require 4 parameters.
  - Two Component Masses
  - Spin of each object aligned with the orbital angular momentum.
  - Sky location / Distance / Phase analytically maximized over
- We are not including all the physics!
  - Precession / higher order modes / matter effects

\[ \sim 400,000 \] templates needed for O2 analyses which may grow to several million for future analysis.
Detecting a Compact Binary Merger

- Model the signal
  - post-Newtonian, analytical / numerical relativity
    - See Sascha Husa’s Talk
  - Phenomenological models and reduced order models
- Use matched filtering to generate SNR time series
- Check the signal consistency
  - Does it appear in multiple observatories?
  - Is the morphology correct?
  - Are there other confounding factors?
    - See Andy Lundren’s Talk
- Estimate statistical significance
Offline Detection

- Deep Analysis with our best understanding of the data
- Detailed Simulation campaigns to understand sensitivities and biases.

Low Latency Detection

- Critical to alert other astronomers to perform follow-up observations
Offline Detection

- *Embarrassingly* Parallel
  - Each template can be analyzed separately
  - Data can be broken into independent blocks of time
- High Throughput Computing
- $10^5 \rightarrow 10^6$ compute tasks
- $\sim 100$ million weighted core-hours for O2 analyses, projecting 200 million for O3.

Pegasus Workflow Management System
- Abstract Representation of Workflow using Pegasus
- Handles input / output file movement
- Batch submission of work (HTCondor, Slurm, etc).
- Enables running on the LIGO Data Grid, Xsede, and Open Science Grid
Available Resources

LIGO Data Grid (~80% of computing in O2)
- Dedicated LIGO Lab clusters
- Dedicated LSC clusters
- ~40,000 cores available

Available Through the Open Science Grid (~20% of computing in O2)
- Virgo Clusters (Shared)
- Individual PI resources (Shared)
- Campus/regional shared clusters (Allocated) e.g. OrangeGrid, PACE, SciNet
- National Supercomputers (Allocated) e.g. XSEDE, Blue Waters

+ Volunteer Einstein@Home ~ 5 petaflops ~ 40,000 active users
OSG Usage

- Many different clusters
- Various software and hardware environments
- Data Access
  - ~5TB of LIGO data is stored under CVMFS at the Univ. of Nebraska
  - > 1PB of data transfer
  - ~10 Gbps sustained transfer rates from Nebraska storage
- Software environments are containerized
Optimizations

- Costs are limited to a few kernels
  - Template generation
    - Polynomial expression or spline interpolated reduced order model
  - Correlation
  - Fourier Transform
    - Well optimized existing libraries (FFTW, MKL)
  - Peak finding
Porting to the GPU

- Decision to focus on CUDA due to the availability of hardware and software support.
- Heavy use of consumer GPUs due to their high single precision performance and cost
- cuFFT
- Kernels are largely ported .... with one exception
Template Generation

- Waveform models for templates frequently evolve, incorporating new physics. Not practical to have a GPU implementation of each.
- Leverage *fixed template banks*: provided we can efficiently store and reconstruct the generated waveforms.
- Waveforms themselves are highly oscillatory, but can write as a product of a smooth amplitude function, and complex exponential of a smooth phase. Feasible to store interpolation data for these.
Future GPU Development

• Test runs just referenced showed a roughly 5x speedup relative to current O2 CPU runs

• The cards are capable of more, and we are investigating several avenues to improve this performance:
  • NVIDIA’s Multi-Process Service (MPS): allows running multiple CUDA processes on one GPU (hardware Tesla or newer)
  • Hide latency using multiple streams
  • Batch processing and reordering of some tasks, similar to optimizations currently used in low-latency pipeline (PyCBC Live)
Low Latency Detection of Gravitational Waves

- We expect electromagnetic counterparts to be possible if the merger contains a neutron star.
- We need to quickly tell other astronomers where to look!
- Automated Alerts from Several Analyses
  - *PyCBC Live* / GstLAL / MBTA looking for compact binary mergers
  - cWB / oLIB looking for excess power
Analysis Segment N

Data Conditioning

0.5s 3.5s

Template Duration

Variable Length (16 - 256+ s)

Valid Results

Overwhitening

8s 3.5s .5s

Data Conditioning

Overwhitening

Maximum Latency = 20 seconds

Analysis Segment Duration = 32+ seconds

Signal Present

Ready for Segment N

Submit Candidate to GraceDB

Ready for Segment N+1

Strain Data

How Data is Used for Filtering
PyCBC Live: Basic Architecture

- Direct matched-filtering in the frequency domain using the full frequency domain waveforms
- Uses MPI for parallelization across many nodes and OpenMPI for parallelization among the cores on each node.
- Many “worker” nodes do matched filtering fully independently
- A single “control” node collects triggers, estimates the background, and finds candidates
- 160 cores needed to analyze bank of 360,000 templates.
- ~30s from GW passing and detection
Summary

- Offline Analyses are embarrassingly parallel and can take advantage of high throughput computing resources
  - Pegasus WMS allows us to map our tasks to various computing resources
  - Throughput can be improved by taking further advantage of the Open Science Grid
  - Port to GPUs is underway but we only benefit if all the core kernels are ported to avoid the data transfer

- Online Analysis requires *dedicated resources* to achieve the lowest latency
Gravitational Wave Open Science

- LIGO Data is released through LIGO Open Science Center (LOSC)
  - Losc.ligo.org
  - (Soon to become GWOSC with the joining of Virgo and future instruments)

- Majority of LIGO analysis software is public
  - LALSuite / PyCBC / gstLAL / others
  - losc.ligo.org/software

- PyCBC: FOSS Python toolkit for gravitational-wave data analysis
  - pycbc.org/tutorials

github.com/gwastrot/pycbc