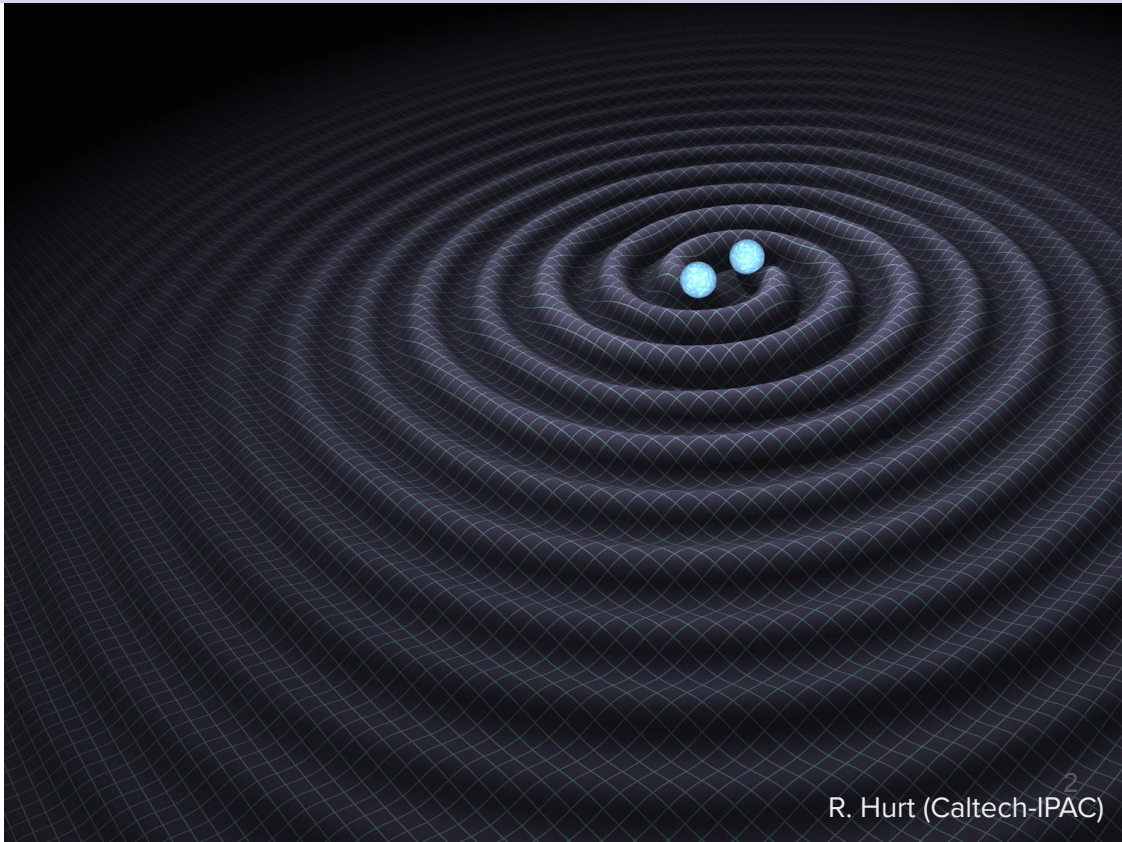
Finding Compact Binary Mergers with GstLAL

Cort Posnansky

Throughput Computing 2024
July 8th, 2024

An overview

- I help search for gravitational waves from distant astrophysical collisions.
- I do this with a data analysis pipeline called GstLAL.
- This work requires high throughput computing clusters.
- My goal is to improve scalability so we can do more science, and learn more about the universe.



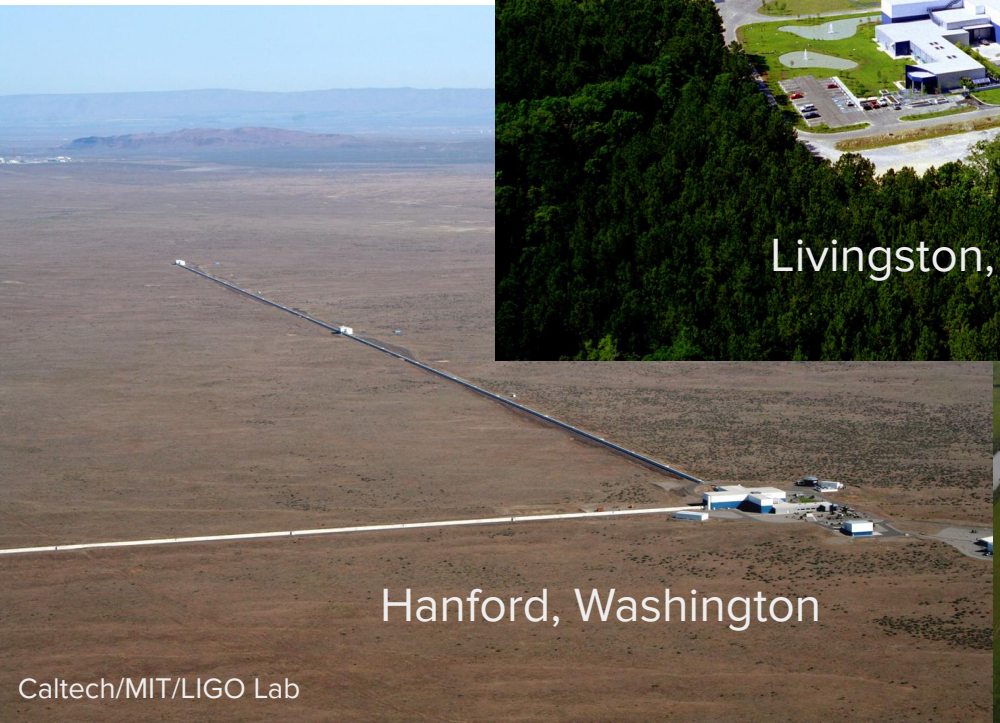
R. Hurt (Caltech-IPAC)

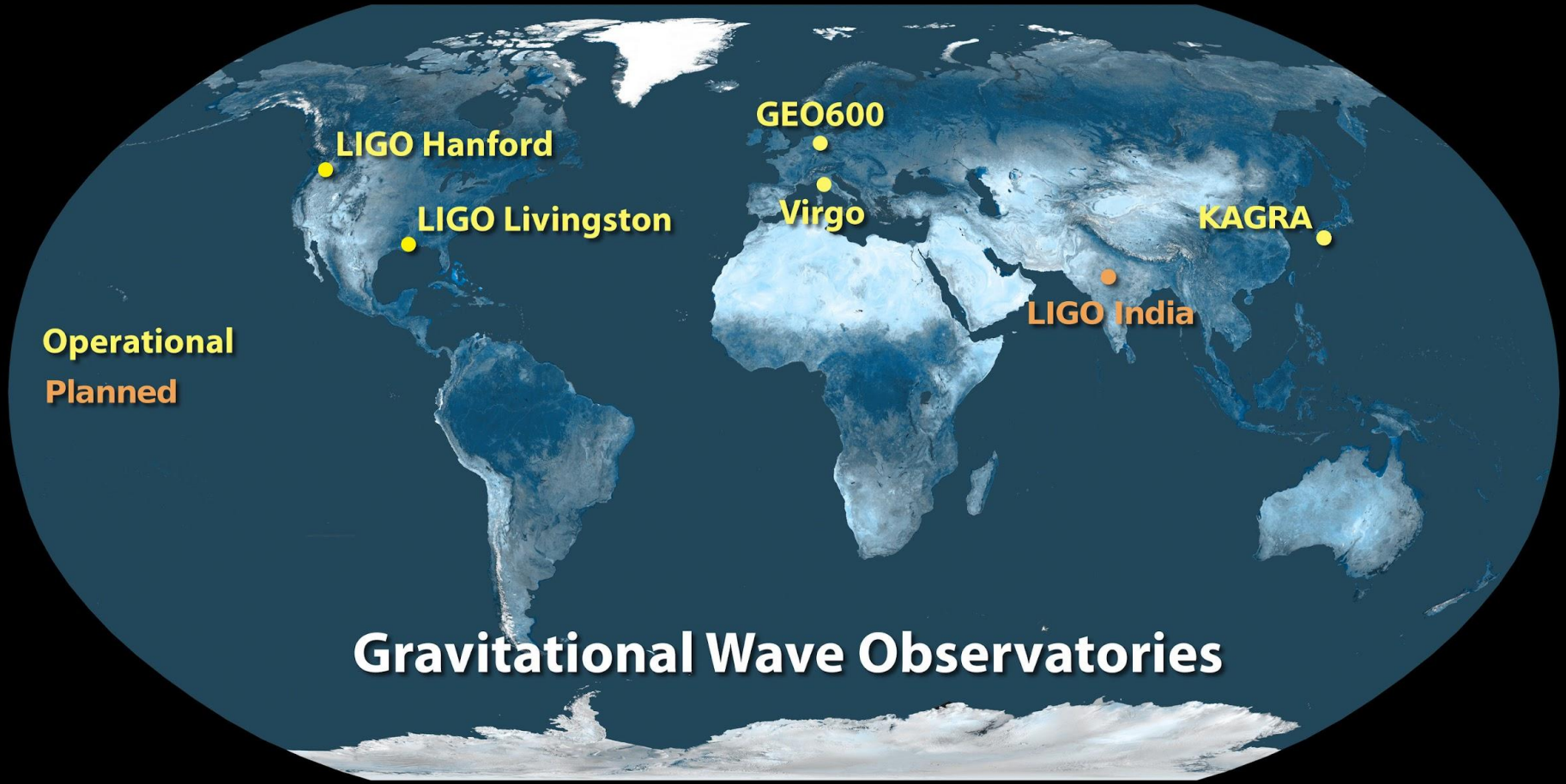
Gravitational Wave Observatories

Kilometers long!

LIGO
VIRGO
KAGRA

Seismic isolation!
Tunnels in vacuum!
Cryogenic cooling!





Caltech/MIT/LIGO Lab

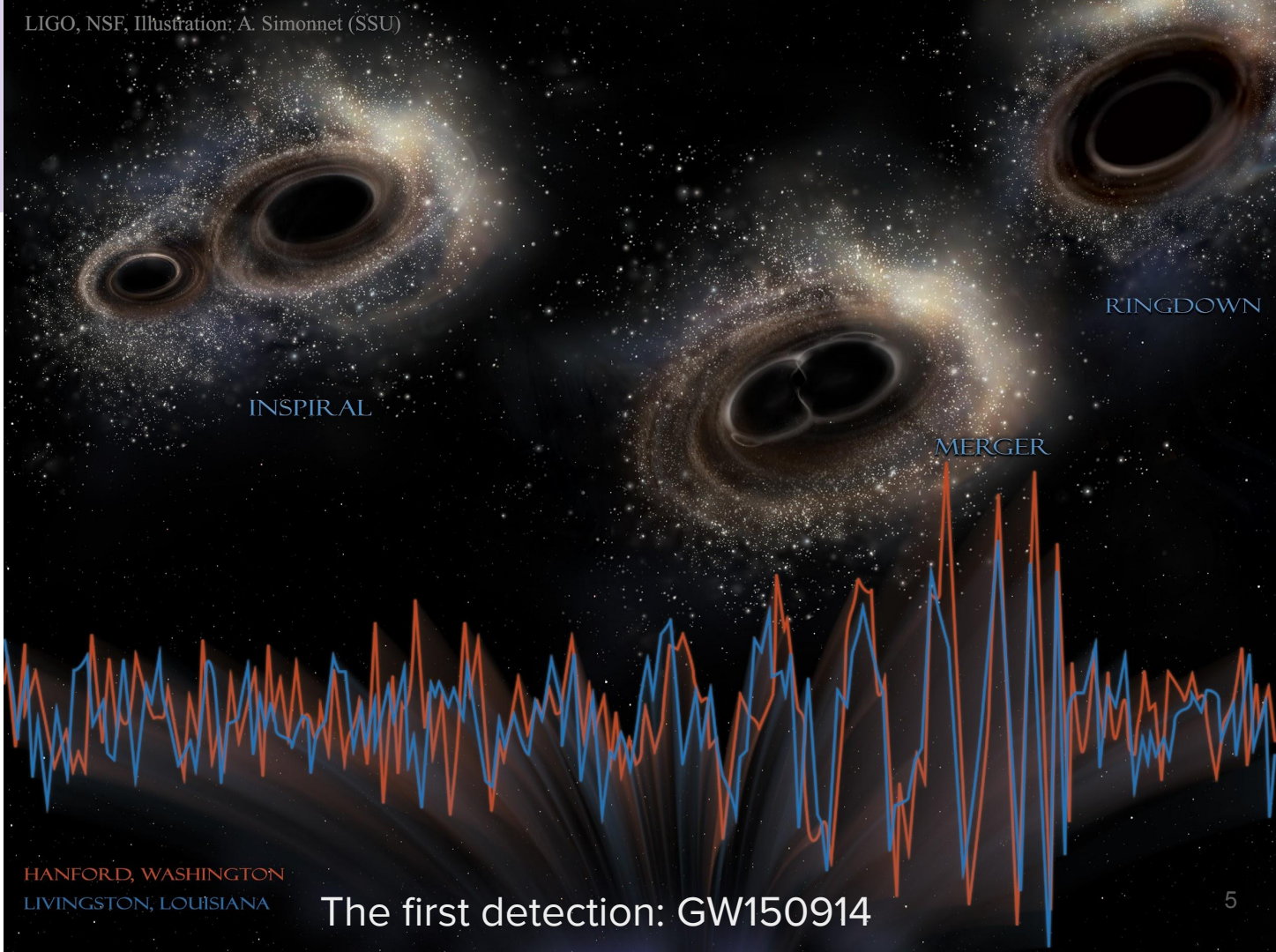
An example signal

Amplitude and frequency increase until the merger.

This was a loud event, but noise is still significant.

GstLAL compares pre-computed waveforms to data.

Simulation of this event: https://www.youtube.com/watch?v=flvFpFUzEXY



The first detection: GW150914

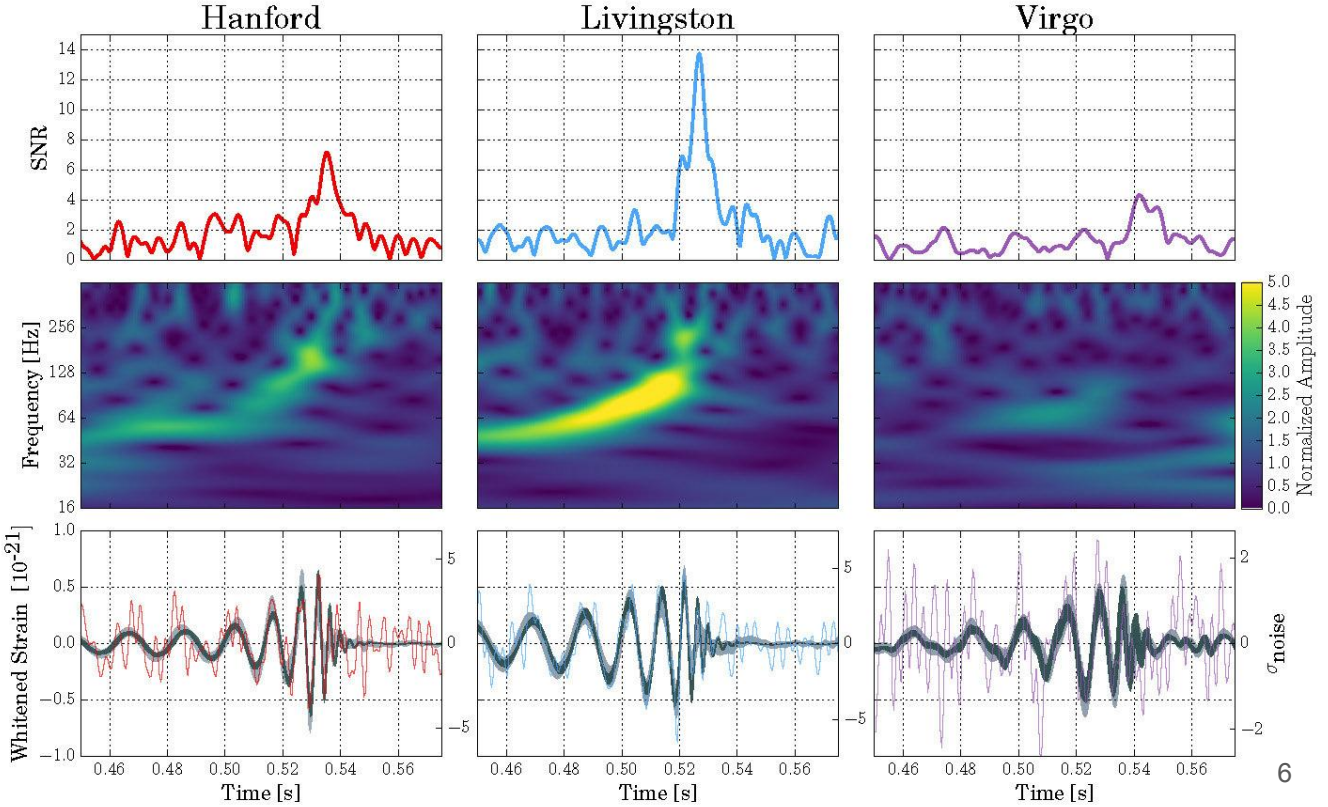
GstLAL computes Signal to Noise Ratio (SNR)

This is GW170814

LIGO/Caltech/MIT/LSC

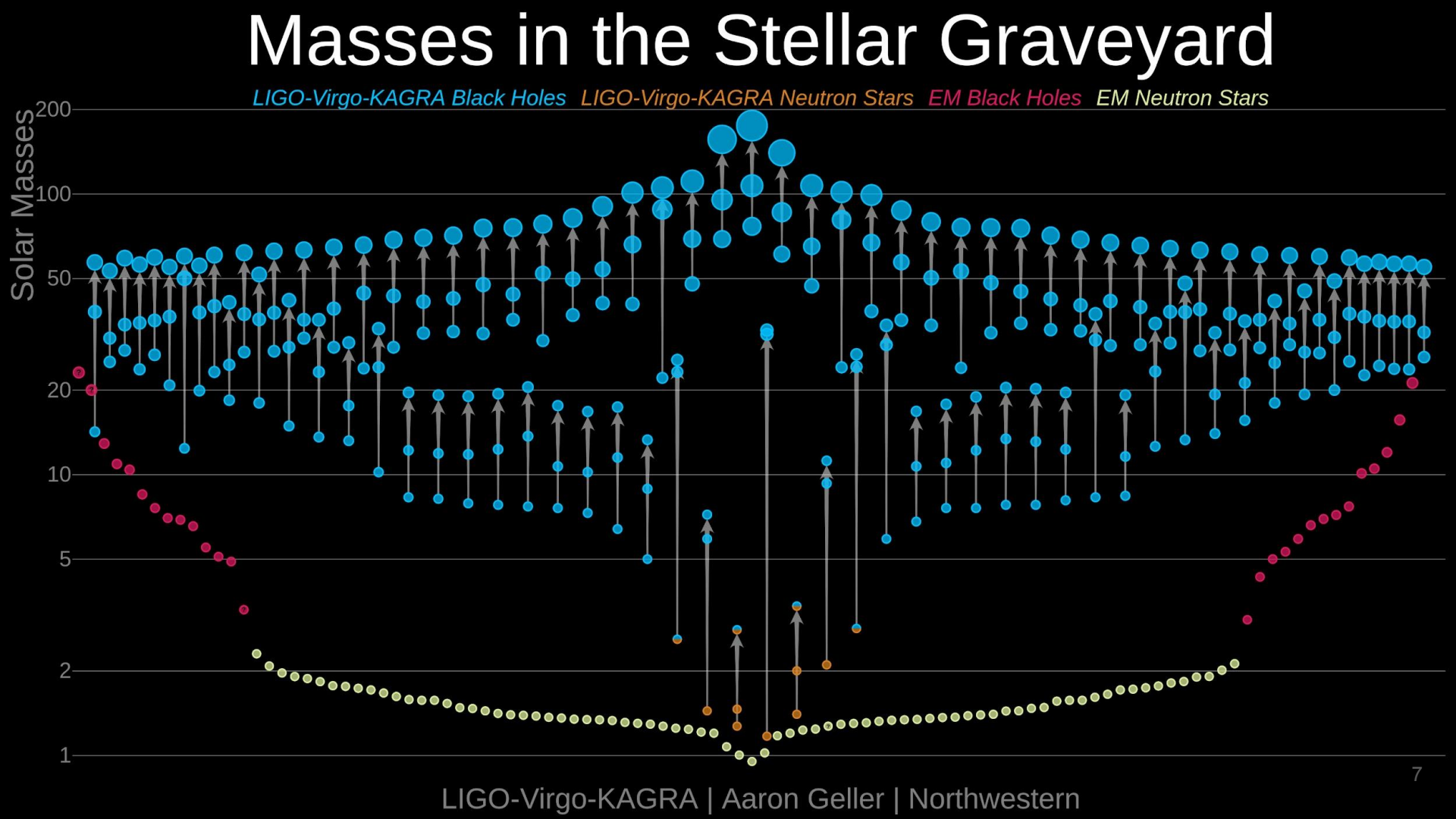
A better match with a pre-computed waveform gives a larger SNR.

We search for millions of these waveforms!



LIGO/Caltech/MIT/LSC

Masses in the Stellar Graveyard



LIGO-Virgo-KAGRA | Aaron Geller | Northwestern

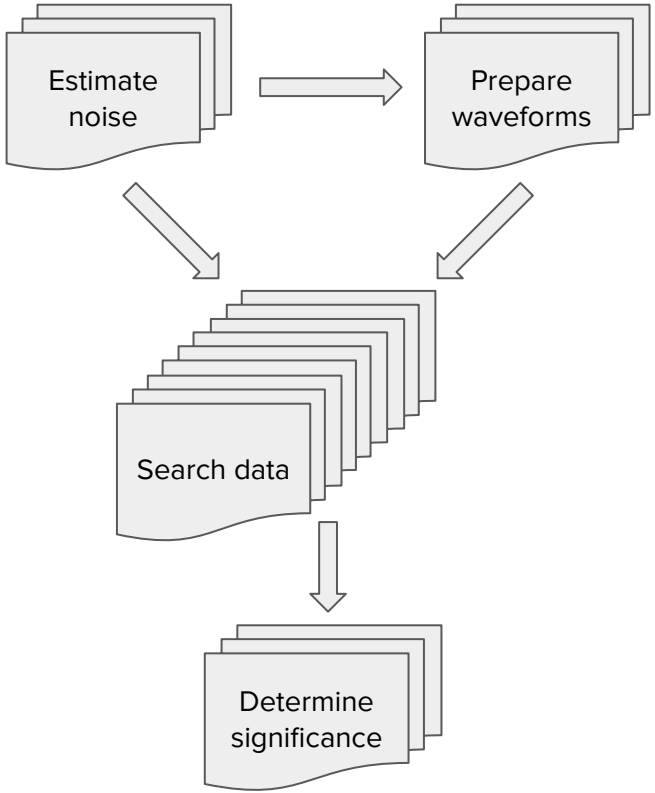
An analysis is a Directed Acyclic Graph (DAG)

GstLAL programs are organized into HTCondor DAGs.

The size of a DAG depends heavily on the amount of data being searched, and the types of waveforms included.

A typical job uses 1 cpu, a few GB of memory and disk, and runs for around 20 minutes.

Searching a week of data for two million waveforms involves O(100K) jobs.



My current project

My goal is to increase our throughput so we can search for more types of waveforms, and make more discoveries.

I intend to submit all our archival search DAGs from one LVK computing site, and have jobs run on our other sites through the OSG.

The biggest challenge is the number and size of network transfers. OSDF solves this problem for us, and I'm working to adapt our pipeline to use these caches.

Although this workflow is not finished, initial tests from an OSG access point (ap40) have worked well, and the throughput is great!

Lessons learned

Although I am still new to the world of high throughput computing, here are a few tips for new users:

- Try never to wait on computers. The bottleneck should be humans.
- Containerize your software.
- Moving files can be complicated. Plan early for how to distribute your data.
- Bookmark the `condor_submit` documentation. Trust me :)

Thank you!

Questions?