Convergence of Large Scale Computing and the Big Data Revolution for Multi-Messenger Astrophysics

Eliu Huerta
Gravity Group
gravity.ncsa.illinois.edu
National Center for Supercomputing Applications
Computational Science and Engineering Faculty Fellow
Department of Astronomy
University of Illinois at Urbana-Champaign

AstroData2020 @ Caltech

December 5th 2018

Outline

Gravitational Wave Astrophysics

Multi-Messenger Astrophysics

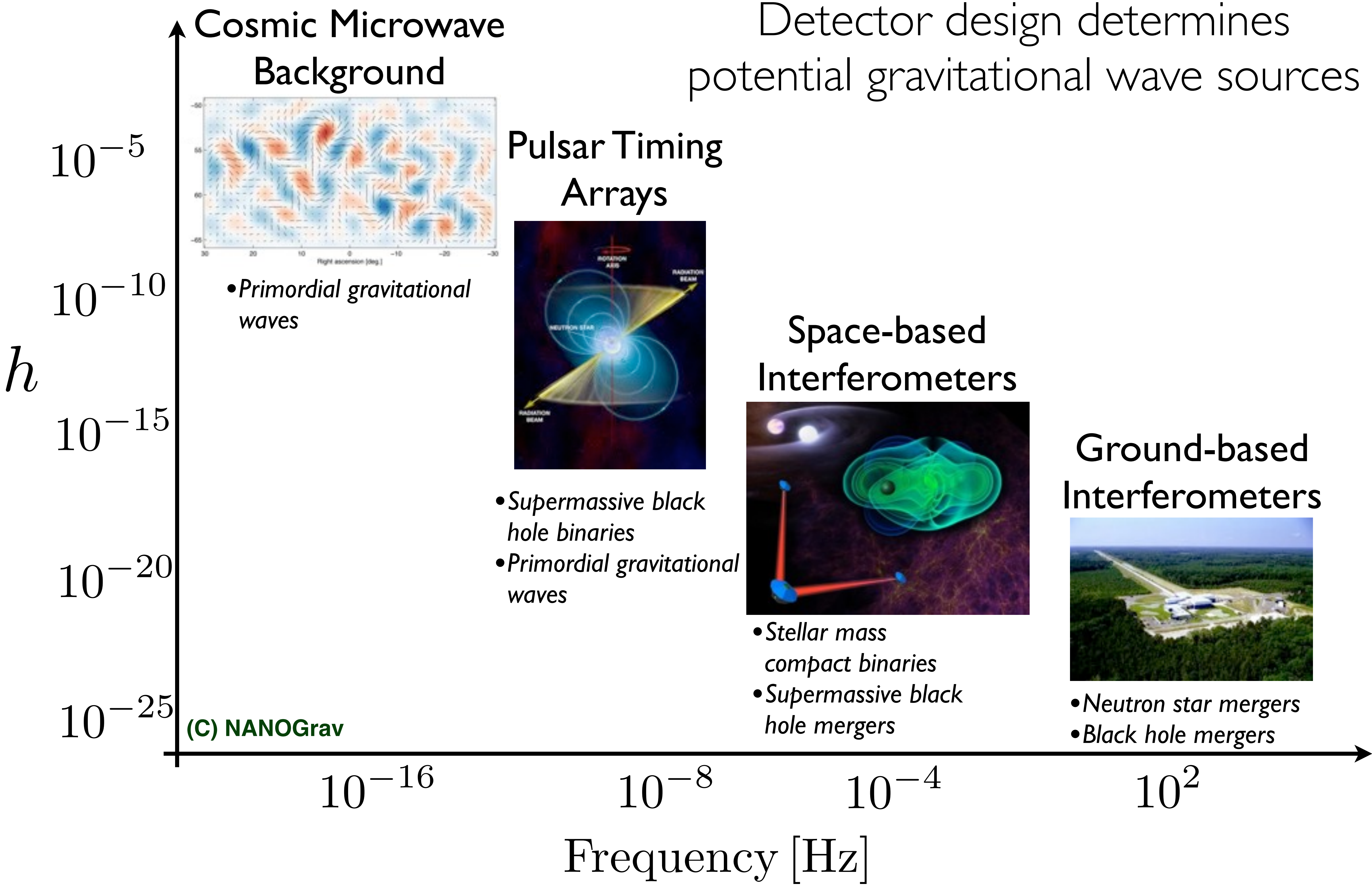
Outstanding theoretical and computational challenges in Multi-Messenger Astrophysics

The rise of the Data Revolution

Convergence of the Data Revolution with Multi-Messenger Astrophysics

Consider the source

Detector design determines potential gravitational wave sources



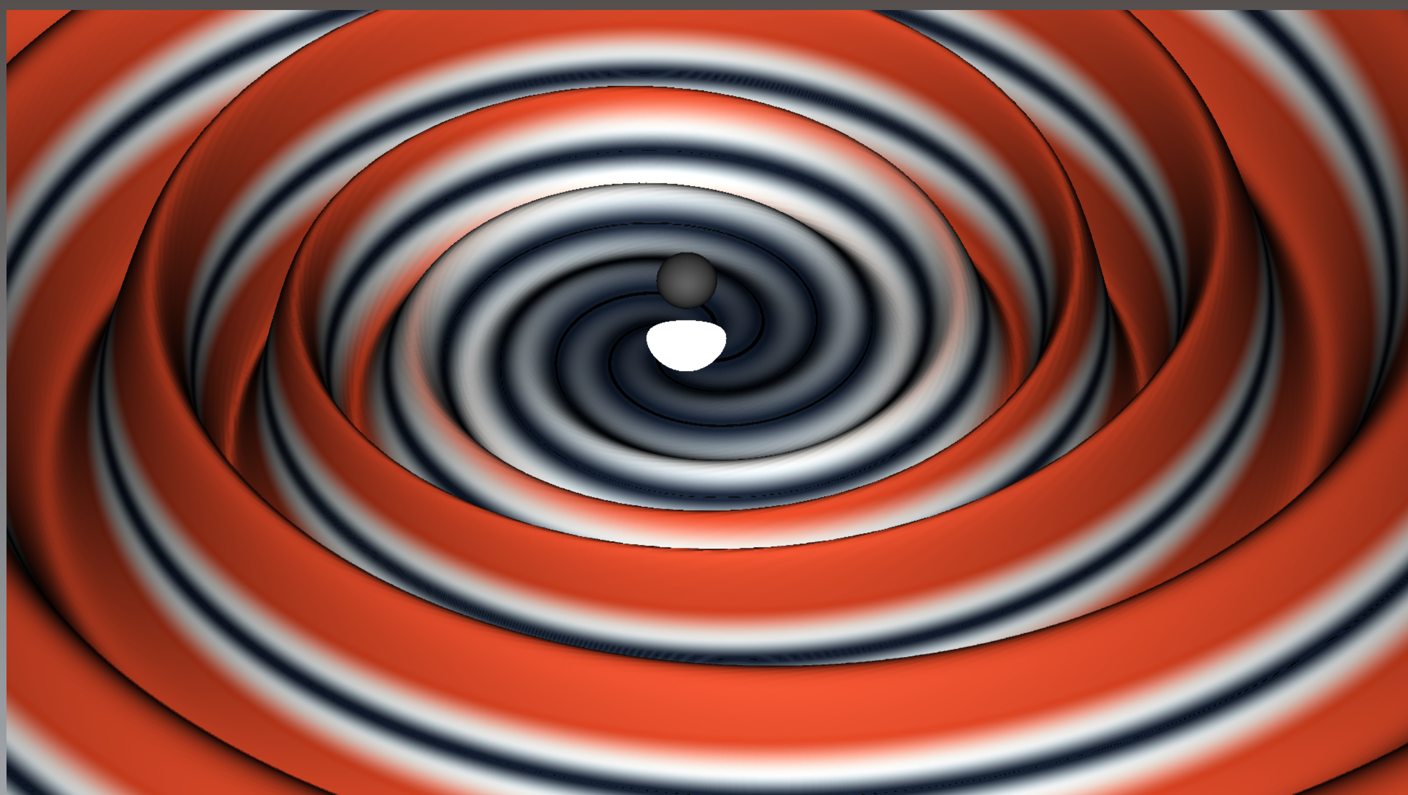
Gravitational waves exist

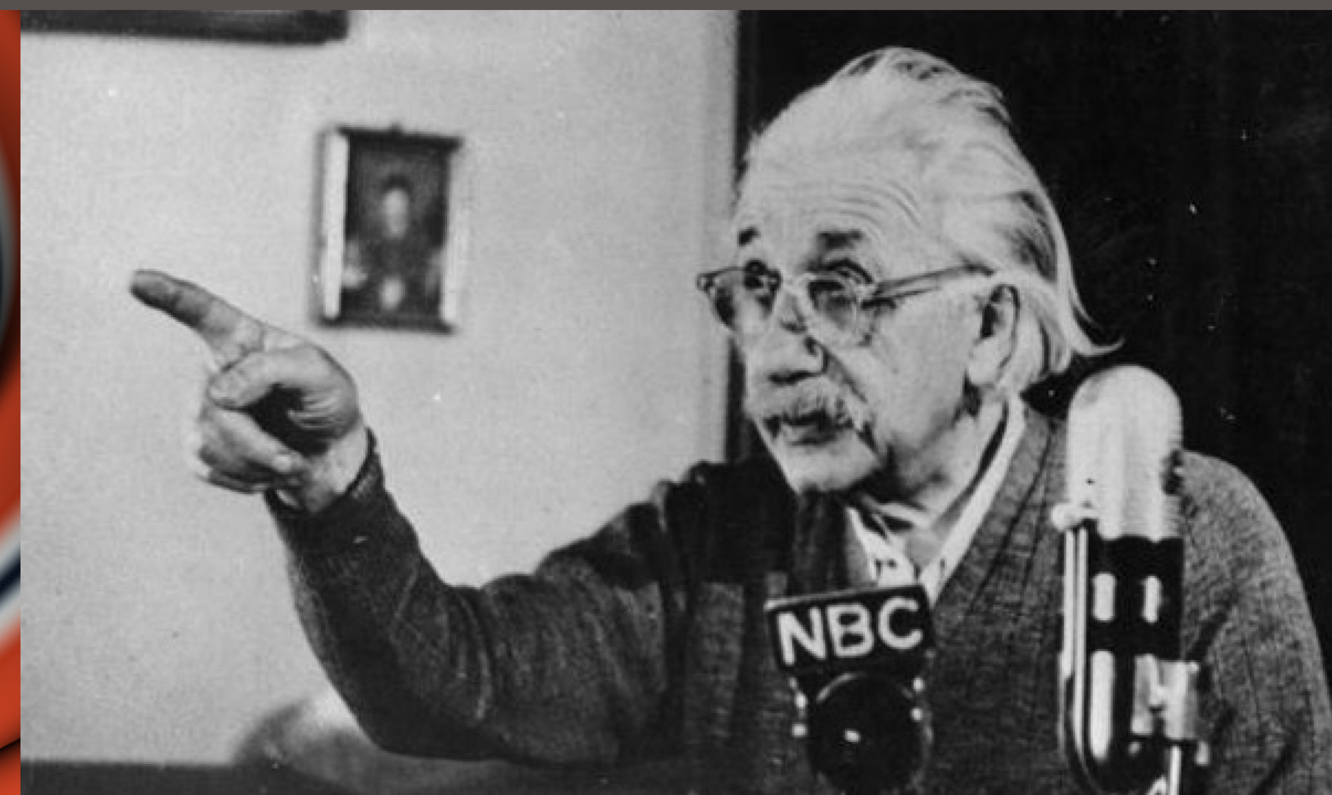
Systems of two black holes form and collide within the age of the Universe

Einstein’s theory of general relativity is correct in the most extreme astrophysical environments

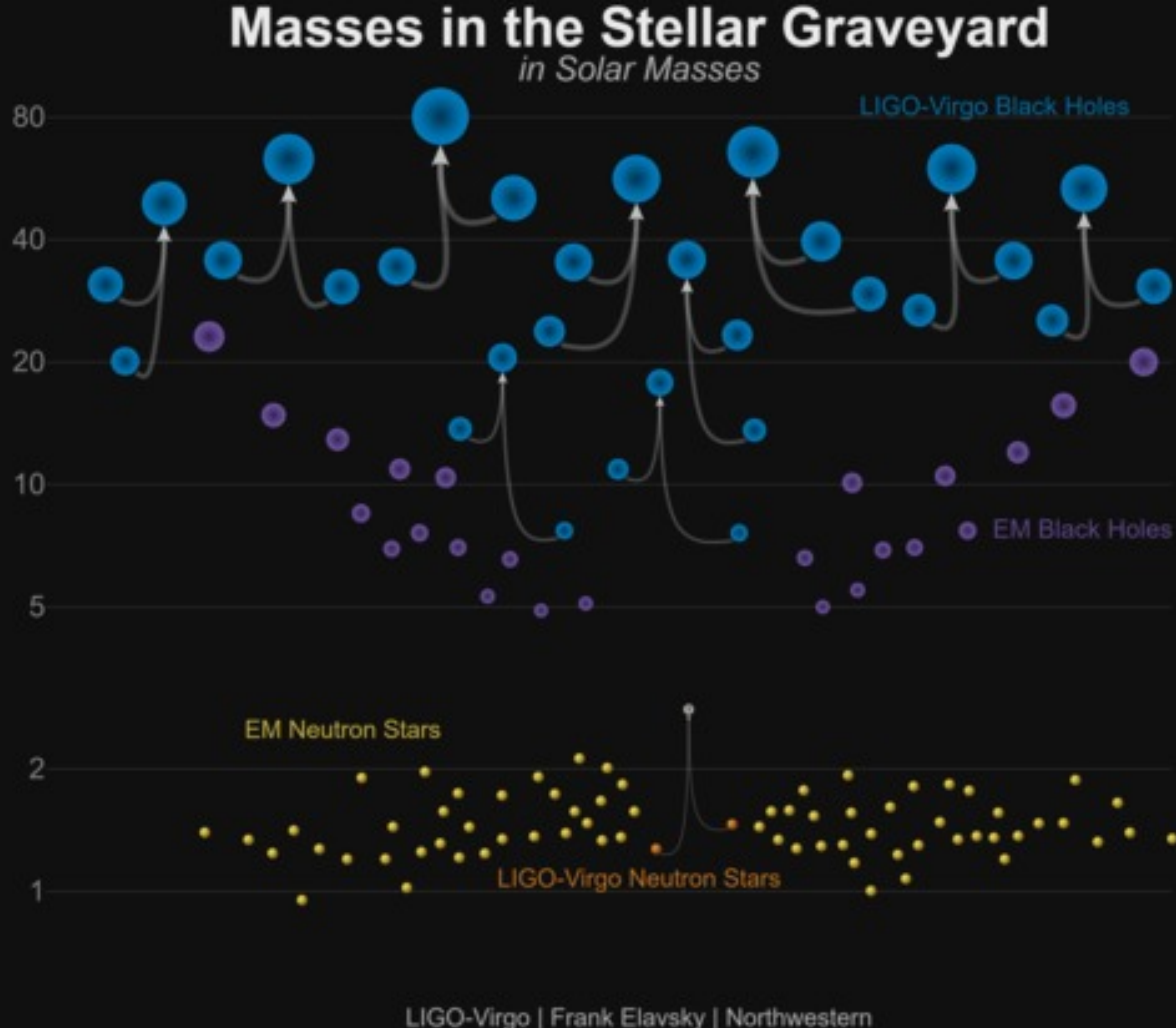
https://www.youtube.com/watch?v=RYWK26jklDg

Don't look at the Dark Universe

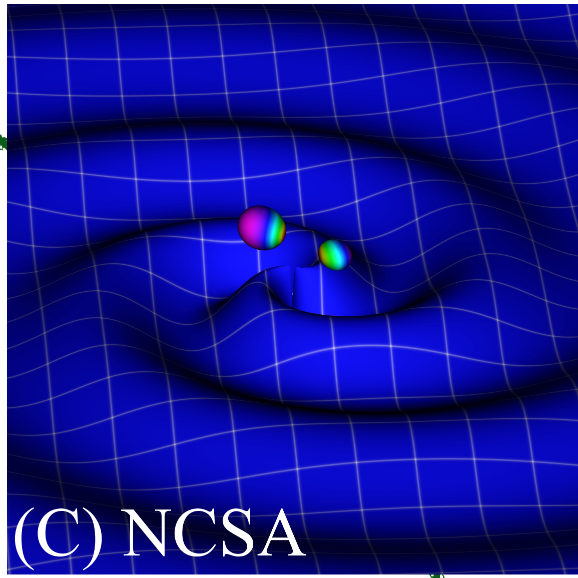




listen to it!

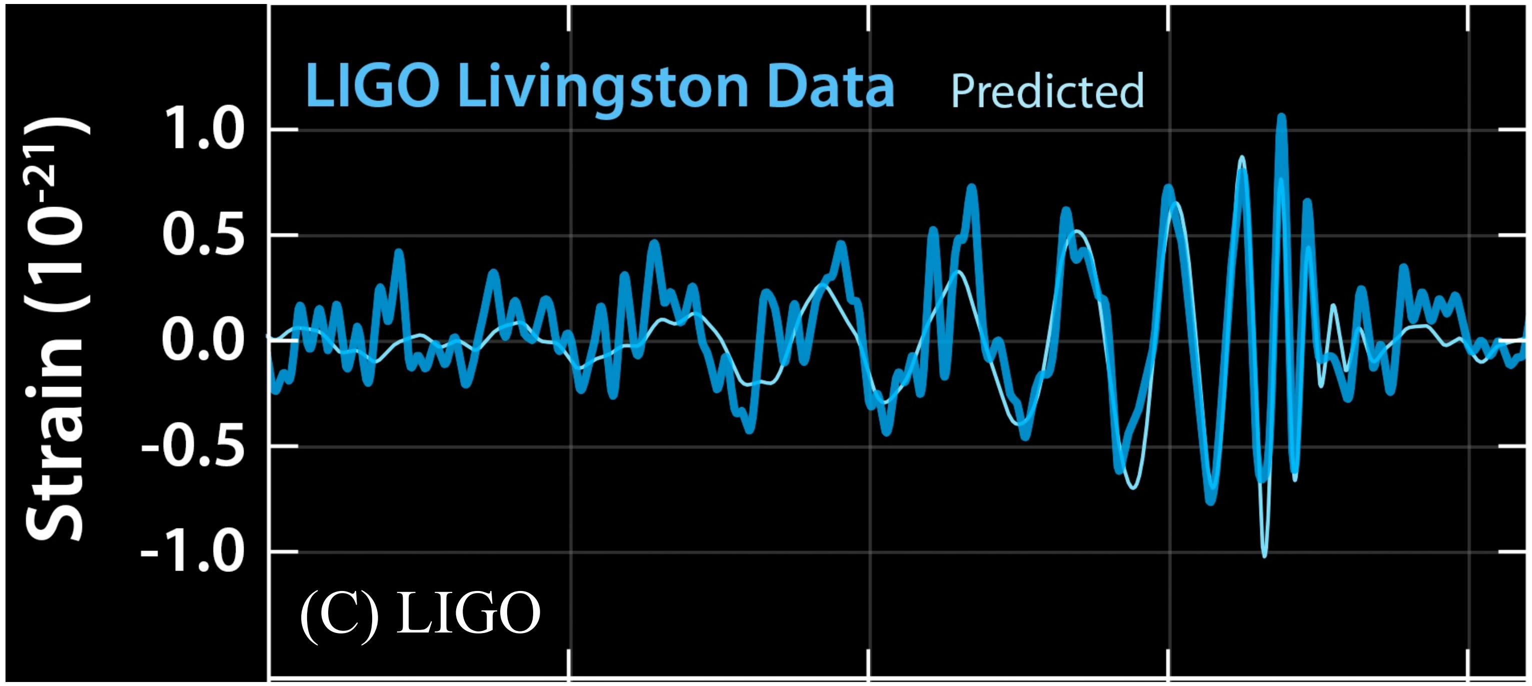

Masses in the Stellar Graveyard
in Solar Masses
LIGO-Virgo Black Holes
EM Black Holes
EM Neutron Stars
LIGO-Virgo Neutron Stars
80
40
20
10
5
2
1
LIGO-Virgo | Frank Elavsky | Northwestern

Models and simulations



Scientific Discovery

Observations



Theory

$G_{\mu\nu} = 8\pi T_{\mu\nu}$

Routine: black hole and neutron star collisions

Future: supernovae, oscillating neutron stars….

"For the greatest benefit to mankind"

Alfred Nobel

The Royal Swedish Academy of Sciences has decided to award the

2017 NOBEL PRIZE IN PHYSICS

to

Rainer Weiss
Barry C. Barish
Kip S. Thorne

"for decisive contributions to the LIGO detector and the observation of gravitational waves"

CURRENT STATUS

Detection of several gravitational wave sources consistent with stellar mass binary black hole systems

We will continue to detect black hole binaries. Is there anything else to wish for?

The future looks bright until you realize that…

Current algorithms target 4D parameter space

LIGO and Virgo will continue to increase their sensitivity

… $1b+ facility, with ever increasing sensitivity, and we can only cover a 4D parameter space…

5D more to cover

New detectors will be added to the existing network

Poorly scalable algorithms

Adding more cycles will not cover all the physical parameter pace

Detection to publication is a 3 month cycle

New gravitational wave sources to be detected

At design sensitivity, we will see events every 15 mins… is this a blessing or a curse?

Current algorithms target 4D parameter space

LIGO a
to inc

5D more to cover

New de
to th

Poorly scalable algorithms

Dete
is

Adding more cycles will not cover all the physical parameter pace

New gravitational wave sources to be detected

At des
ever
is this a

Future directions

Change the existing paradigm of gravitational wave astrophysics

Key ingredients

Numerical and analytical relativity to understand and model sources

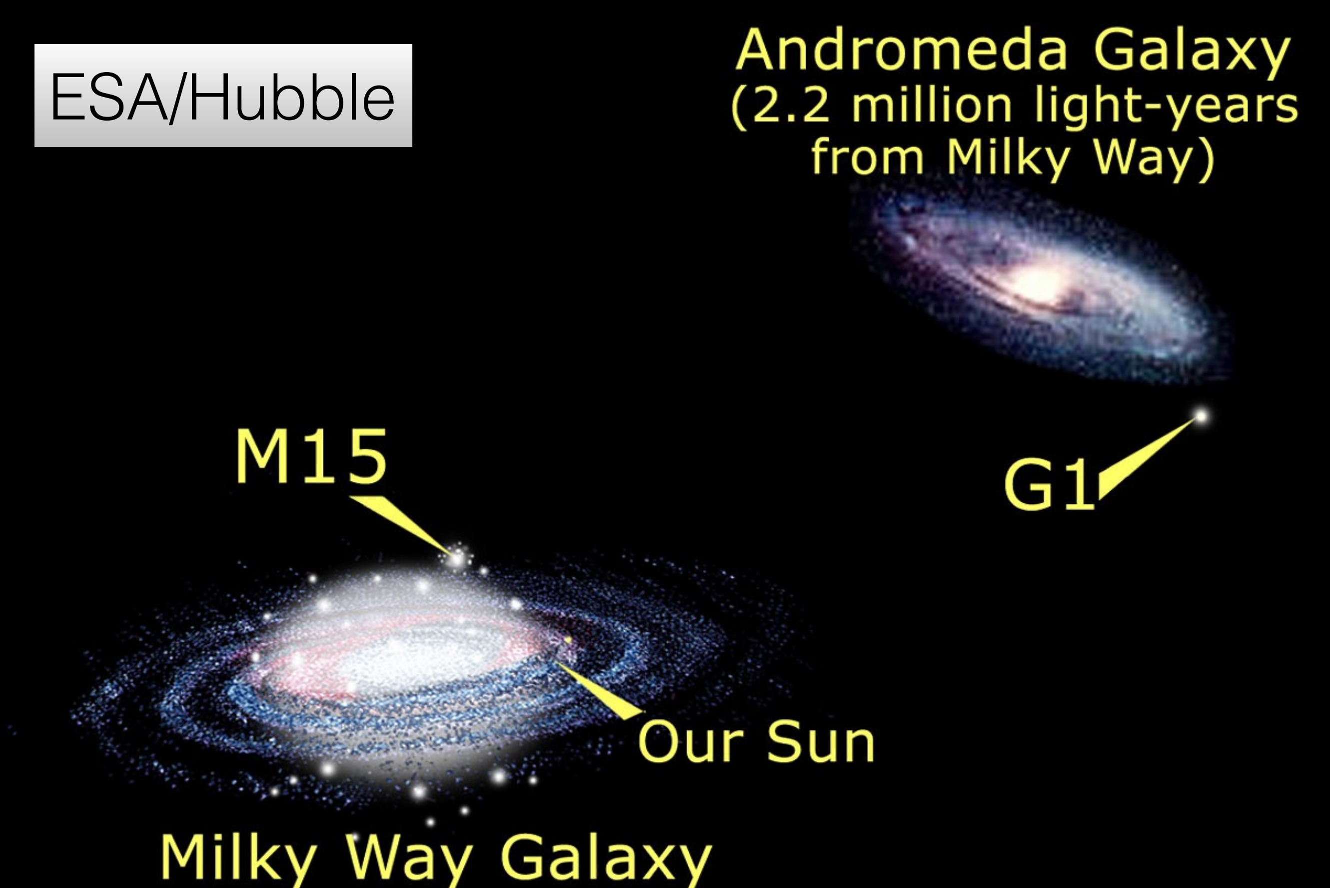
New scalable algorithms to cover the entire 9D parameter space

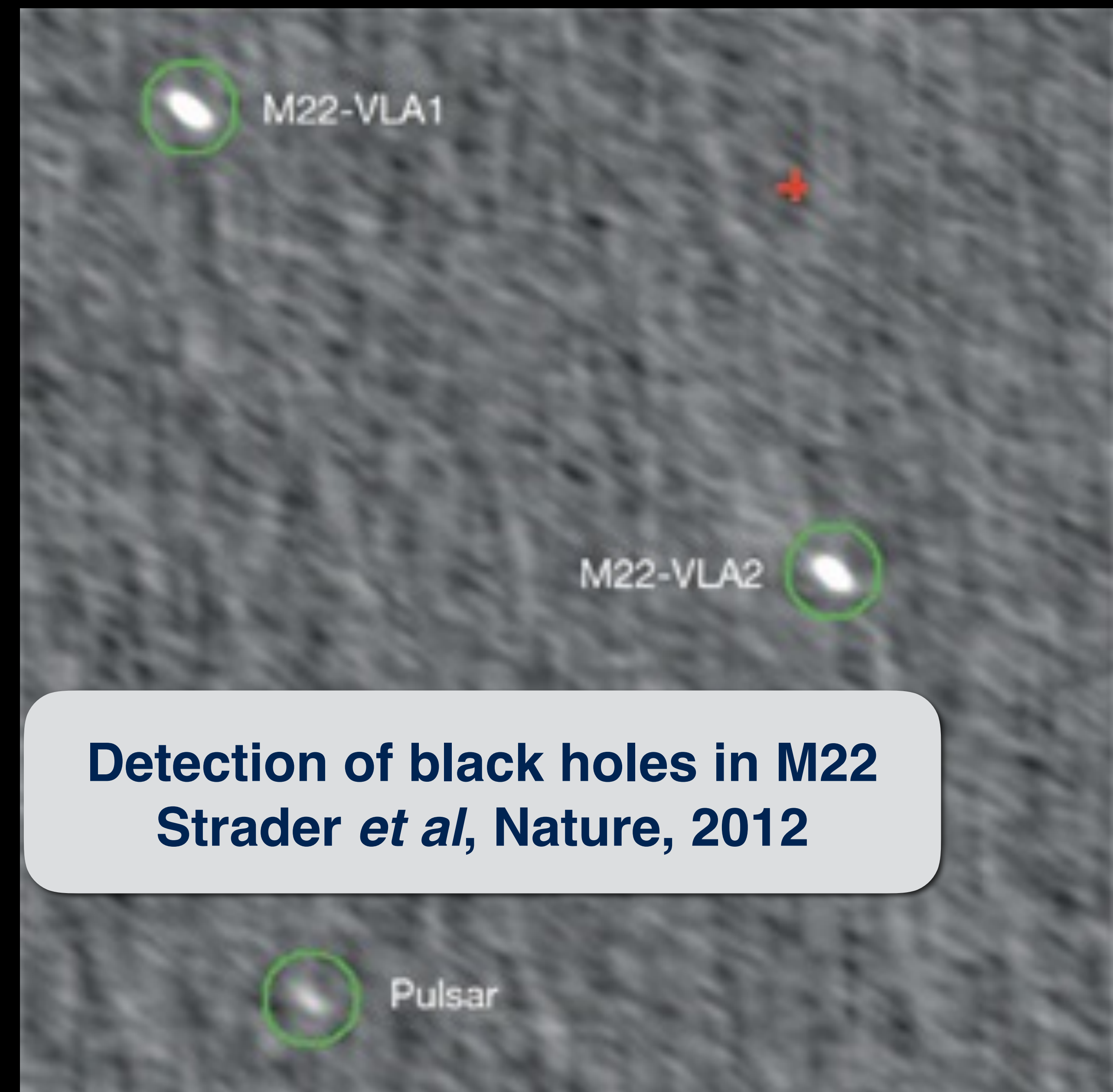
Detect and characterize sources in real time

Compact binary populations in dense stellar environments

Globular clusters known to have black holes

ESA/Hubble



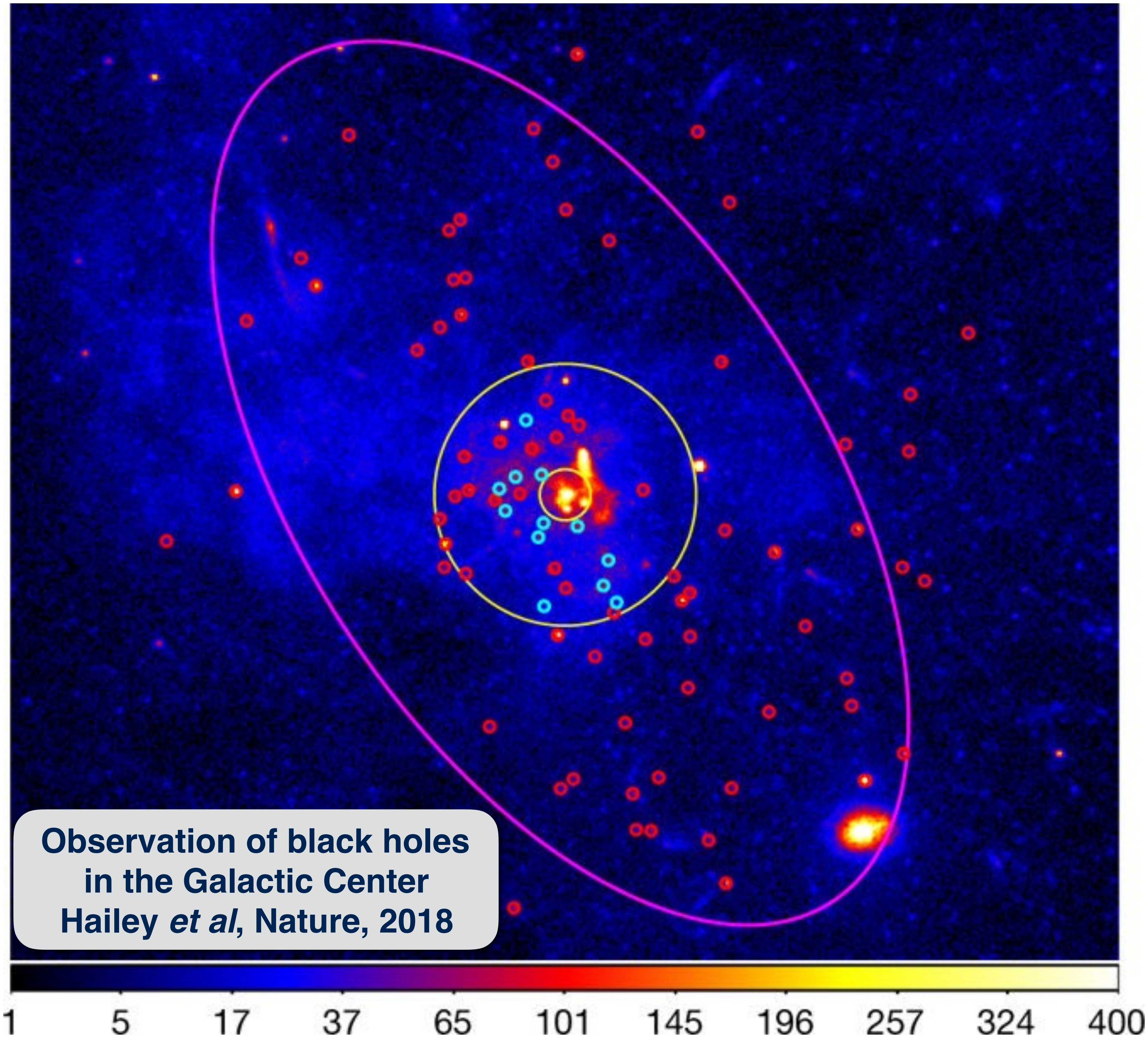


Detection of black holes in M22
Strader *et al*, Nature, 2012

N-body algorithms underpredict compact binary population in clusters
Antonini, ApJ, 2013

Eccentricity "cleanest signature" of black hole mergers in clusters
Samsing, ApJ, 2014

Compact binary populations in dense stellar environments



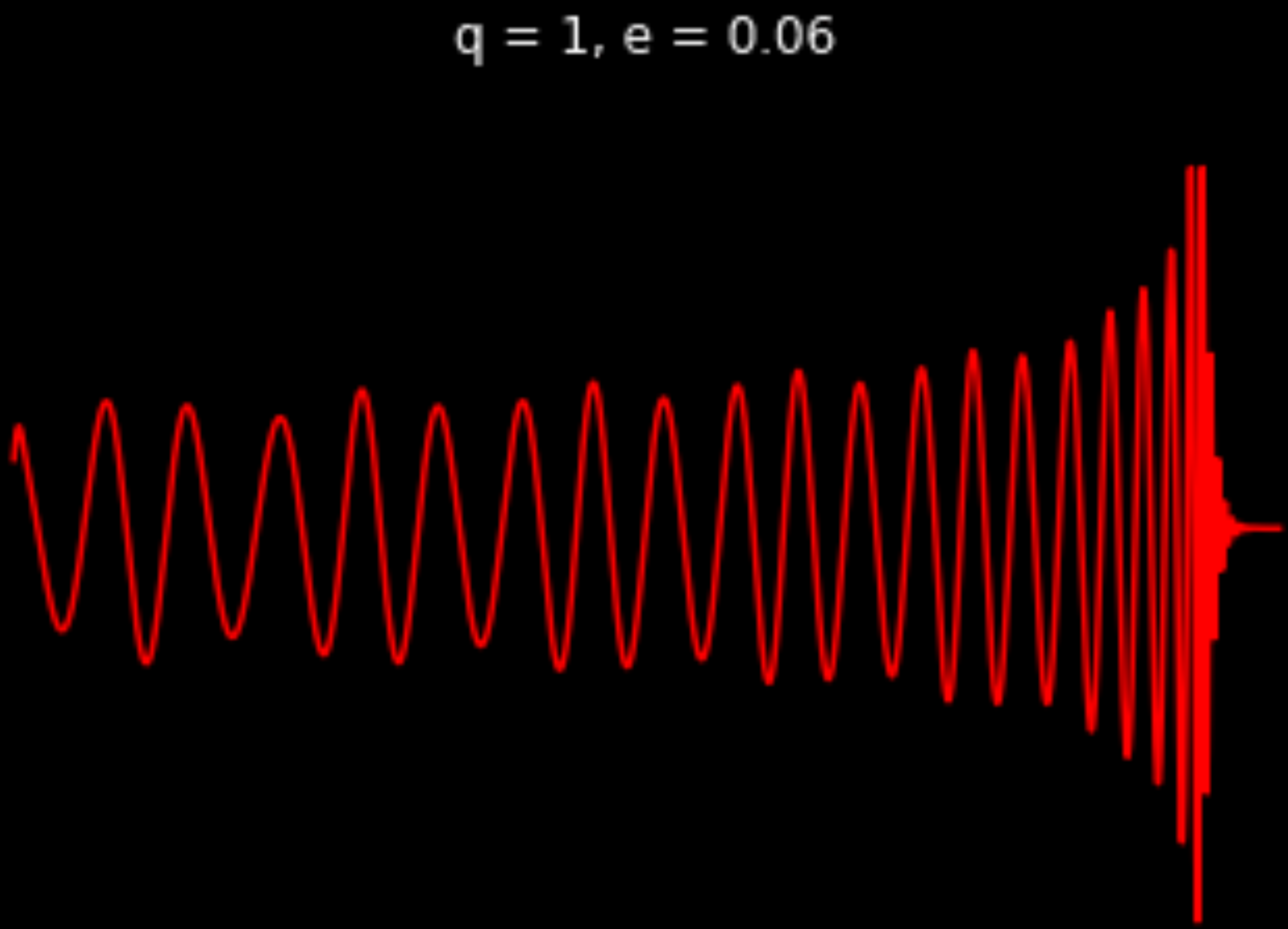
Evidence of compact source populations both in Galactic Clusters and the Galactic Center

Search for compact binary populations in these environments is warranted!

Model these sources
Loutrel, Pretorius, Yunes

100+ numerical relativity waveforms to characterize eccentric binary black hole mergers

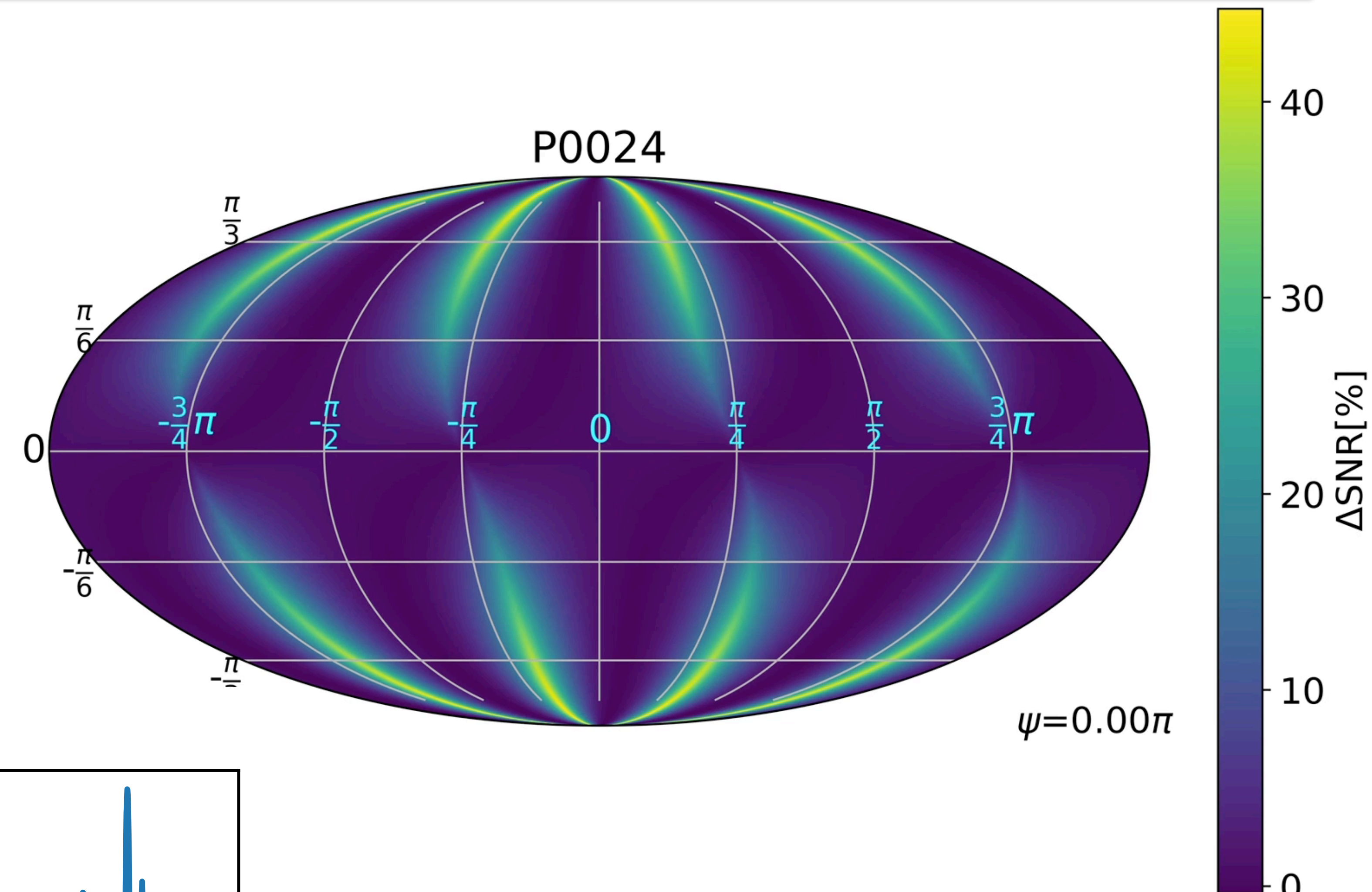
Huerta et al, in preparation

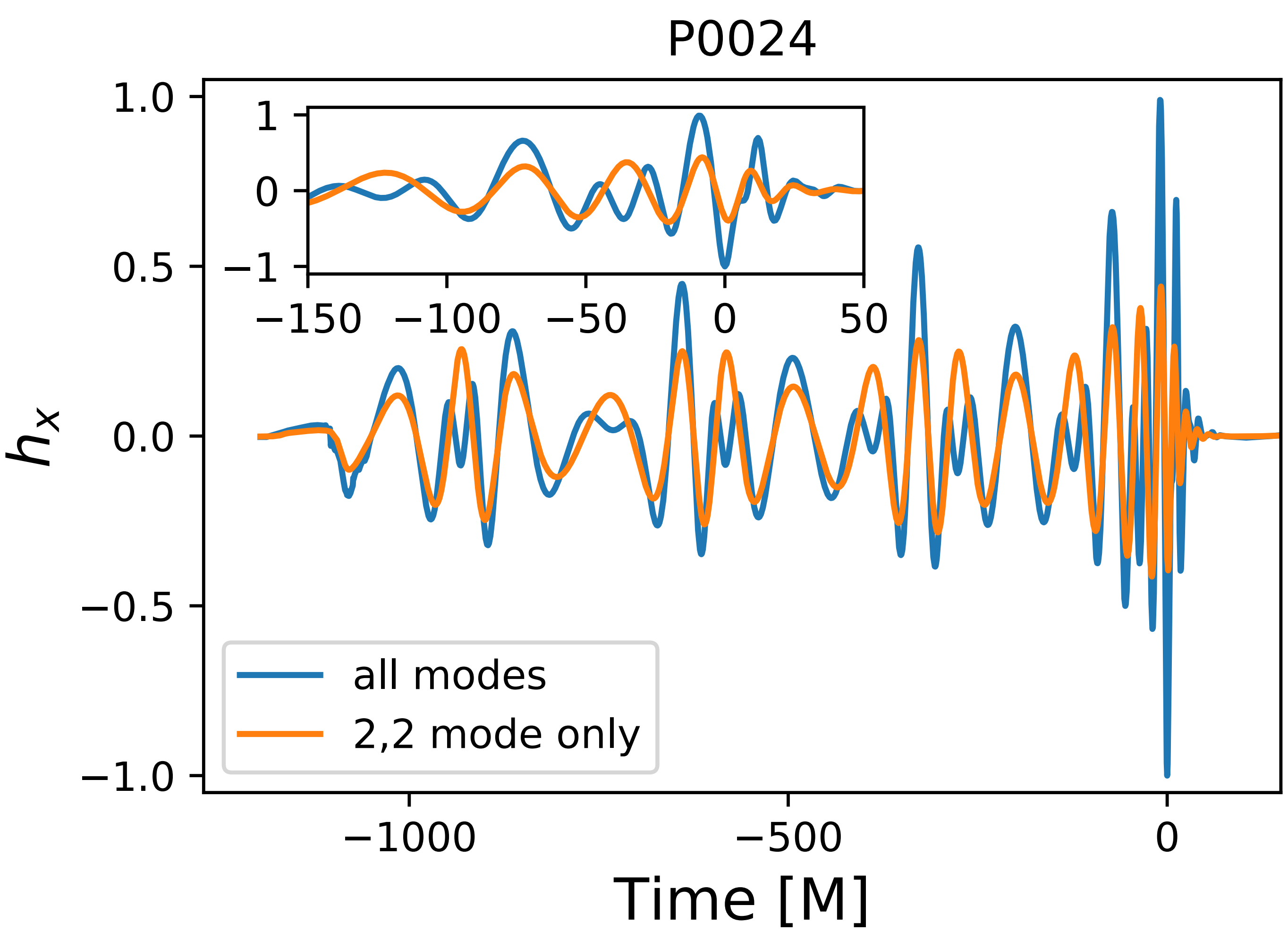


What about higher order modes?

Rebei, Huerta, et al, arXiv:1807.09787

Adam Rebei

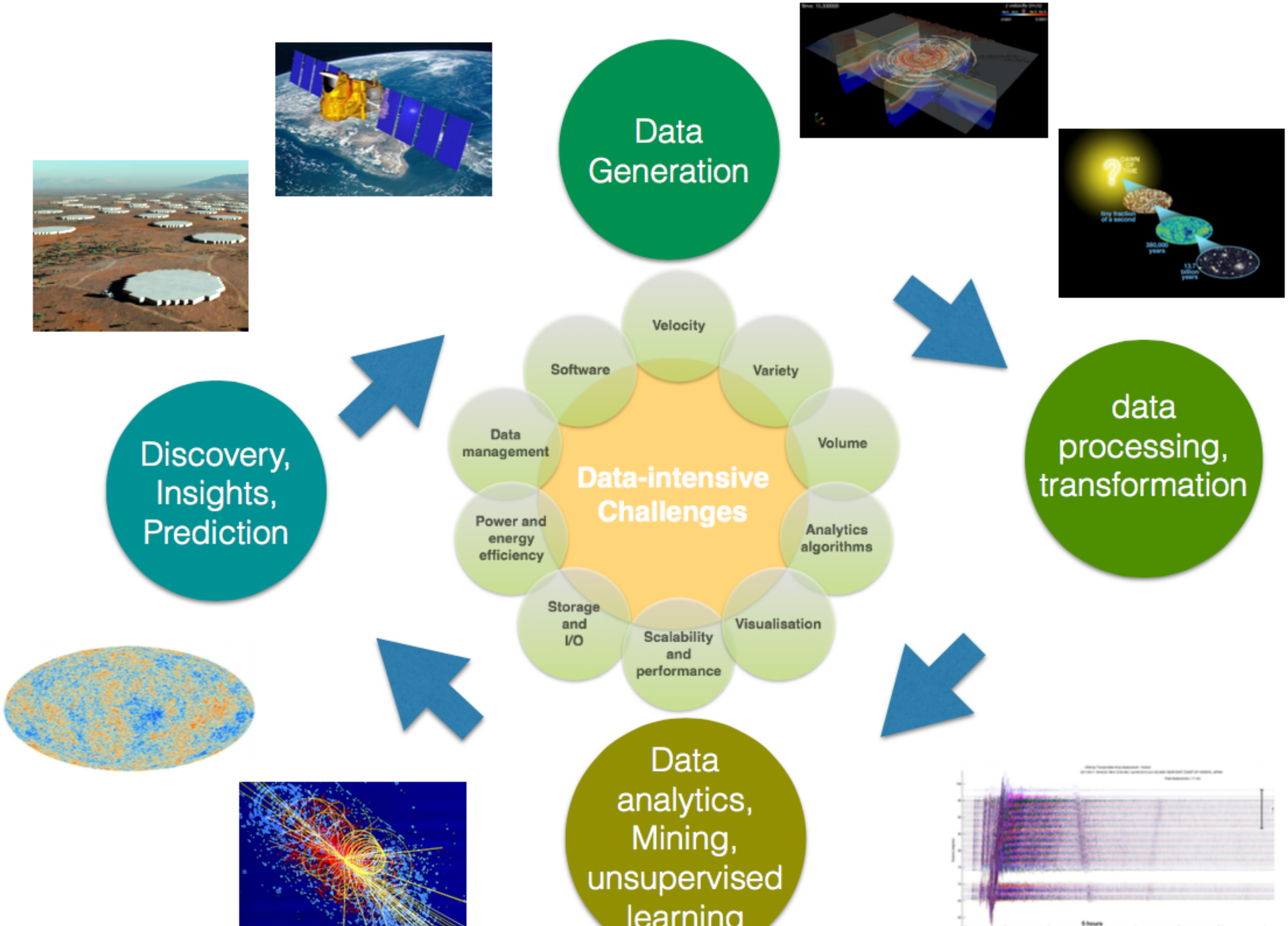




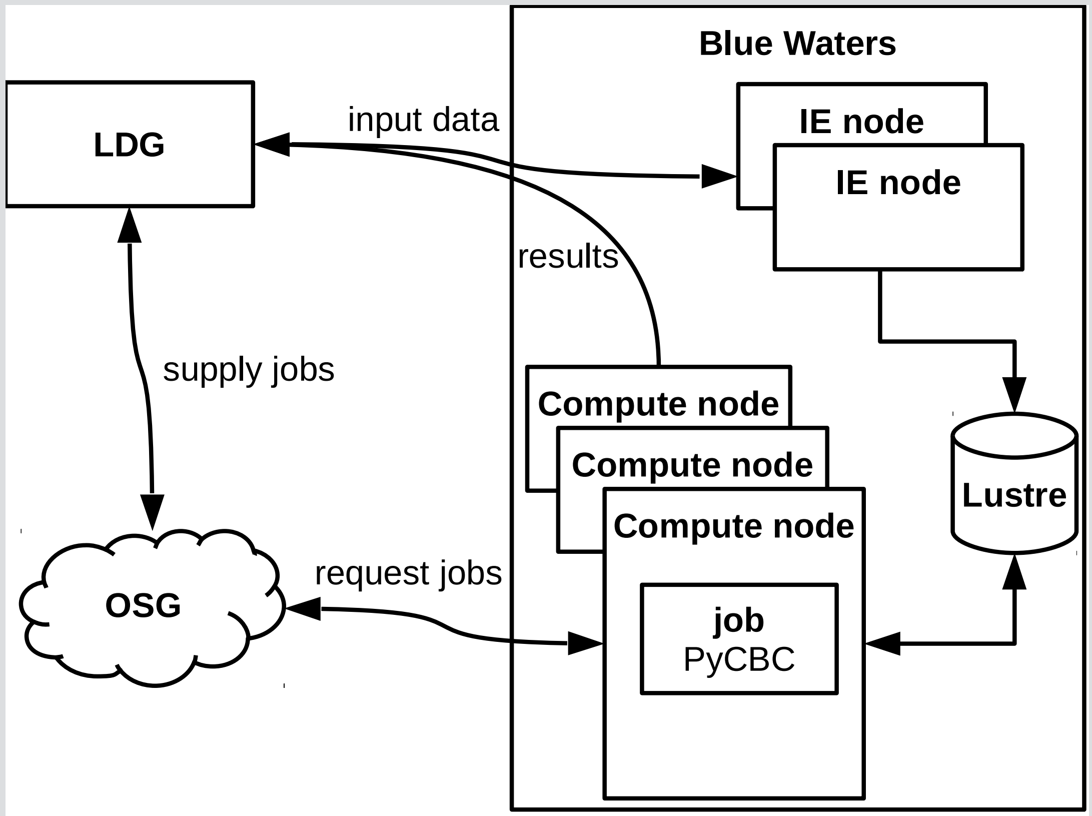
Essential to search for asymmetric mass-ratio binary black hole mergers

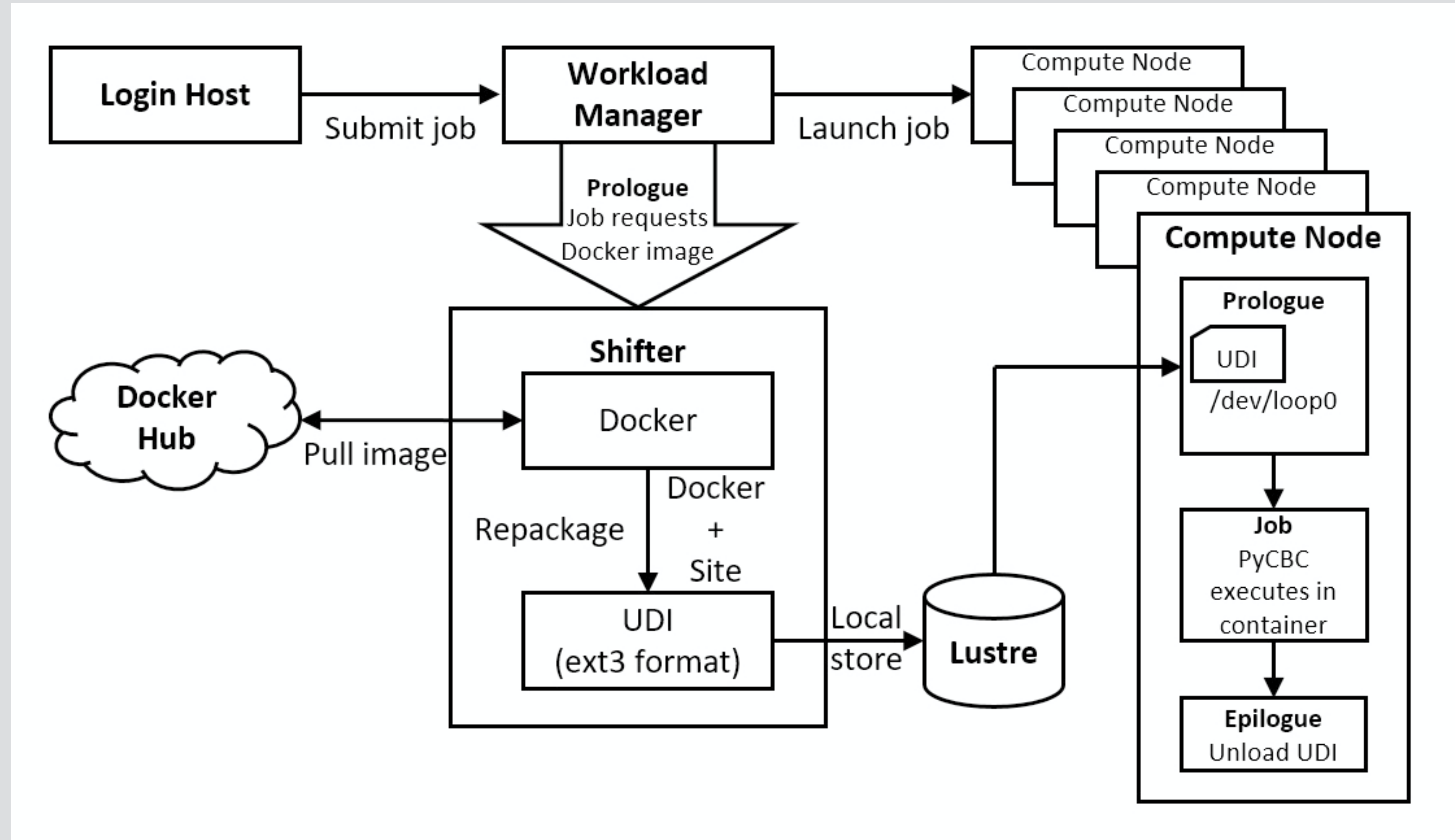
Increase up to 45% in signal-to-noise ratio for mass-ratio 10 black hole mergers

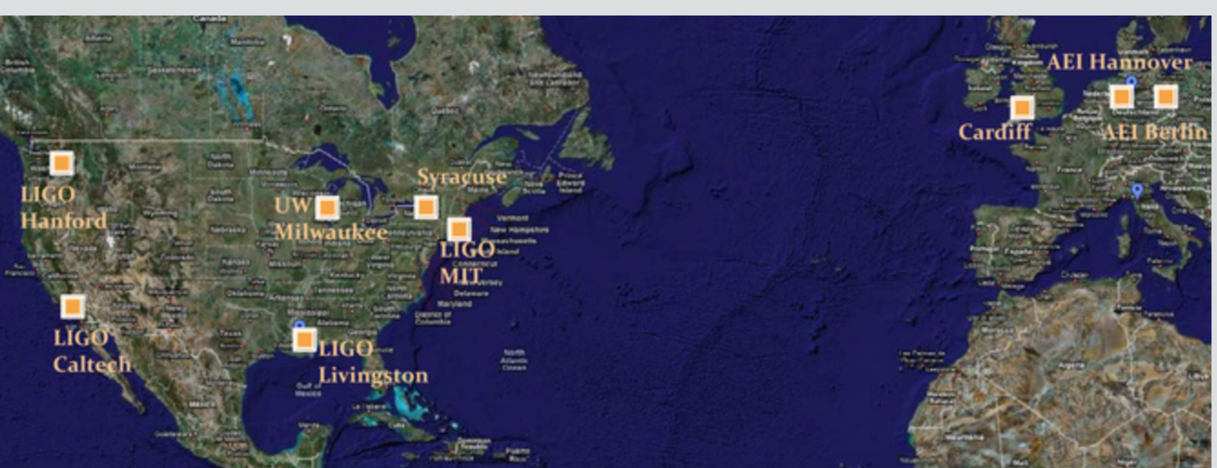
Distribution of needs in simulation and data-driven science in the science community

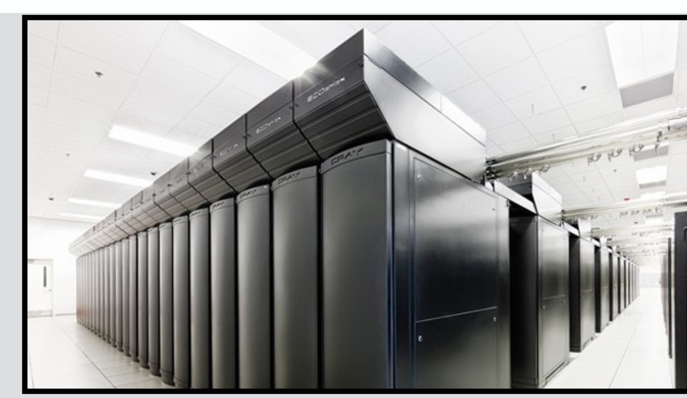


National Strategic Computing Initiative









LIGO Data Grid (LDG): 9 HTC dedicated clusters, 17k+cores
Stakeholder of Open Science Grid (OSG)
Huerta et al, eScience, 47, 2017

Containerized LIGO workflows can seamlessly use Blue Waters compute resources

SHIFTER

BOSS-LDG: A Novel Computational Framework that Brings Together Blue Waters, Open Science Grid, Shifter and the LIGO Data Grid to Accelerate Gravitational Wave Discovery

E. A. Huerta[1], Roland Haas[1], Edgar Fajardo[2], Daniel S. Katz[1], Stuart Anderson[3], Peter Couvares[3], Josh Willis[4], Timothy Bouvet[1] Jeremy Enos[1], William T. C. Kramer[1], Hon Wai Leong[1] and David Wheeler[1]

[1]NCSA, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
{elihu, rhaas, dskatz, tbouvet, jenos, wtkramer, hwleong, dwheeler}@illinois.edu
[2]University of California, San Diego, La Jolla, California 92093, USA
emfajard@ucsd.edu
[3]LIGO, California Institute of Technology, Pasadena, California 91125, USA
{anderson, peter.couvares}@ligo.caltech.edu
[4]Abilene Christian University, Abilene, Texas 79699, USA
josh.willis@acu.edu

Accepted to eScience, 13th IEEE International Conference on eScience

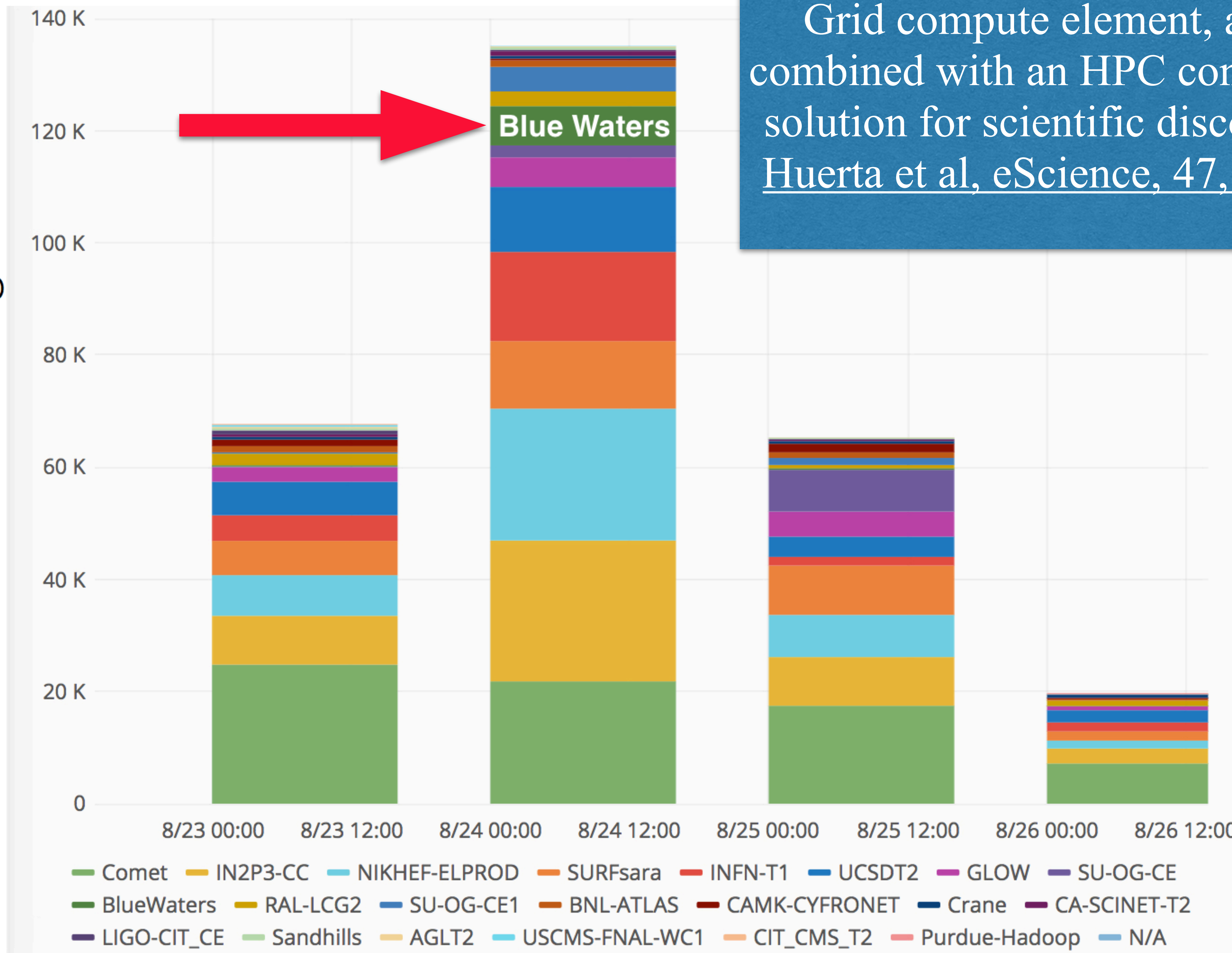
New framework used during the last several weeks of aLIGO-VIRGO second discovery campaign (O2)
Blue Waters was the largest contributor for gravitational wave searches at several points by the end of O2

National Strategic Computing Initiative

[eahuerta@ldas-osg ~]$ condor_status -pool osg-ligo-1.t2.ucsd.edu -af GLIDEIN_Site | sort | uniq -c

```
1126 BlueWaters
  29 Caltech
 399 CCIN2P3
 200 CNAF
 610 Comet
   9 FNAL
  51 LIGO-CIT
  13 Nebraska
 899 NIKHEF
 127 Omaha
  24 PL_CAMK-CYFRO
  51 Purdue
 201 RAL
 200 SURFsara
 638 UCSD
  17 Wisconsin
```

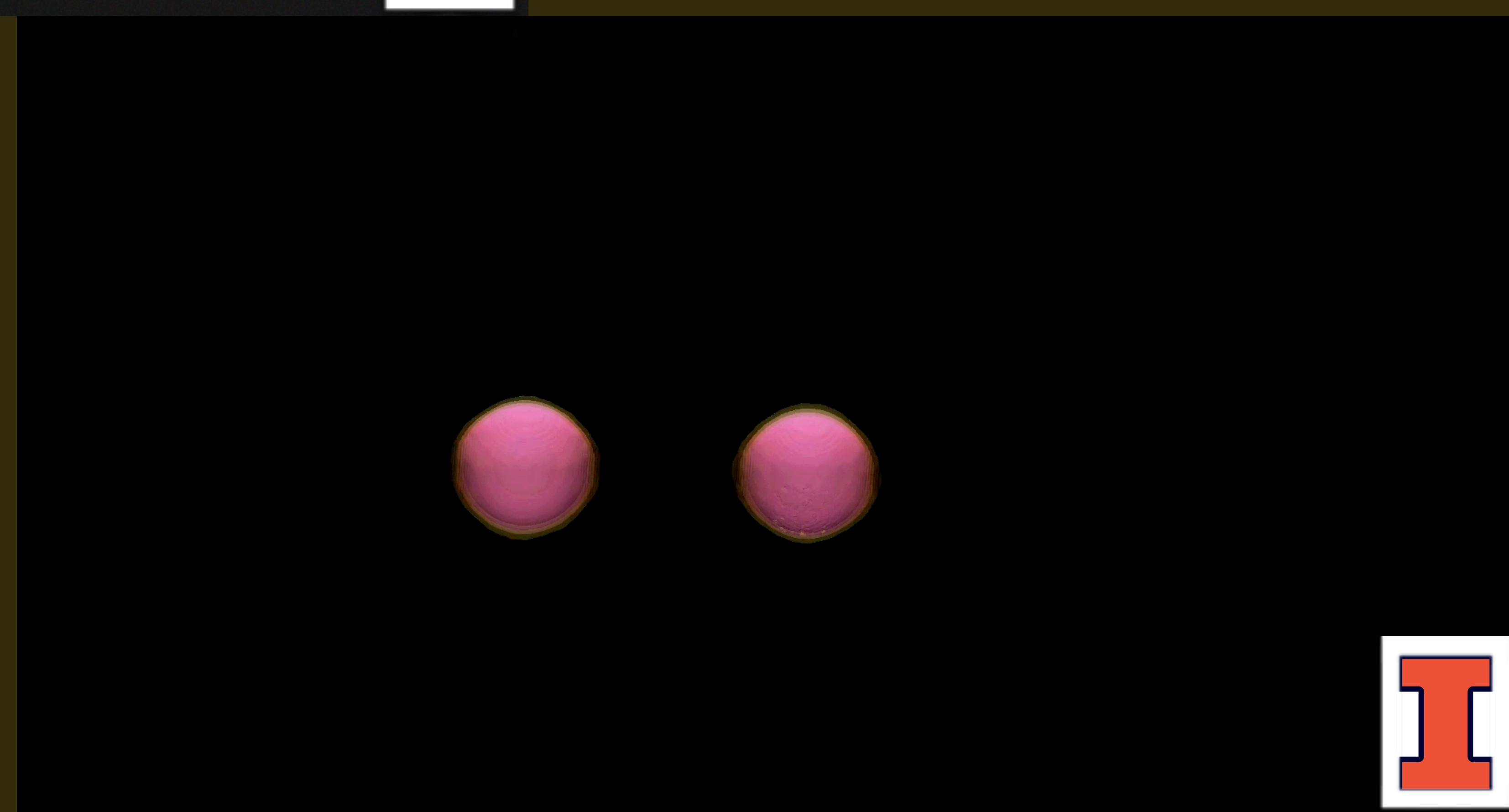
First time Blue Waters is configured an as Open Science Grid compute element, and combined with an HPC container solution for scientific discovery
Huerta et al, eScience, 47, 2017



Multi-Messenger sources: combination of Einstein's general relativity with magnetohydrodynamics and microphysics

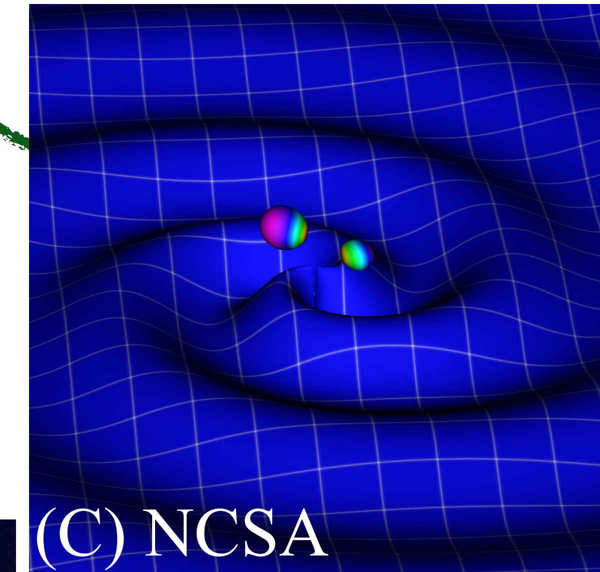
Equation of state leads to a long-lived hyper-massive neutron star

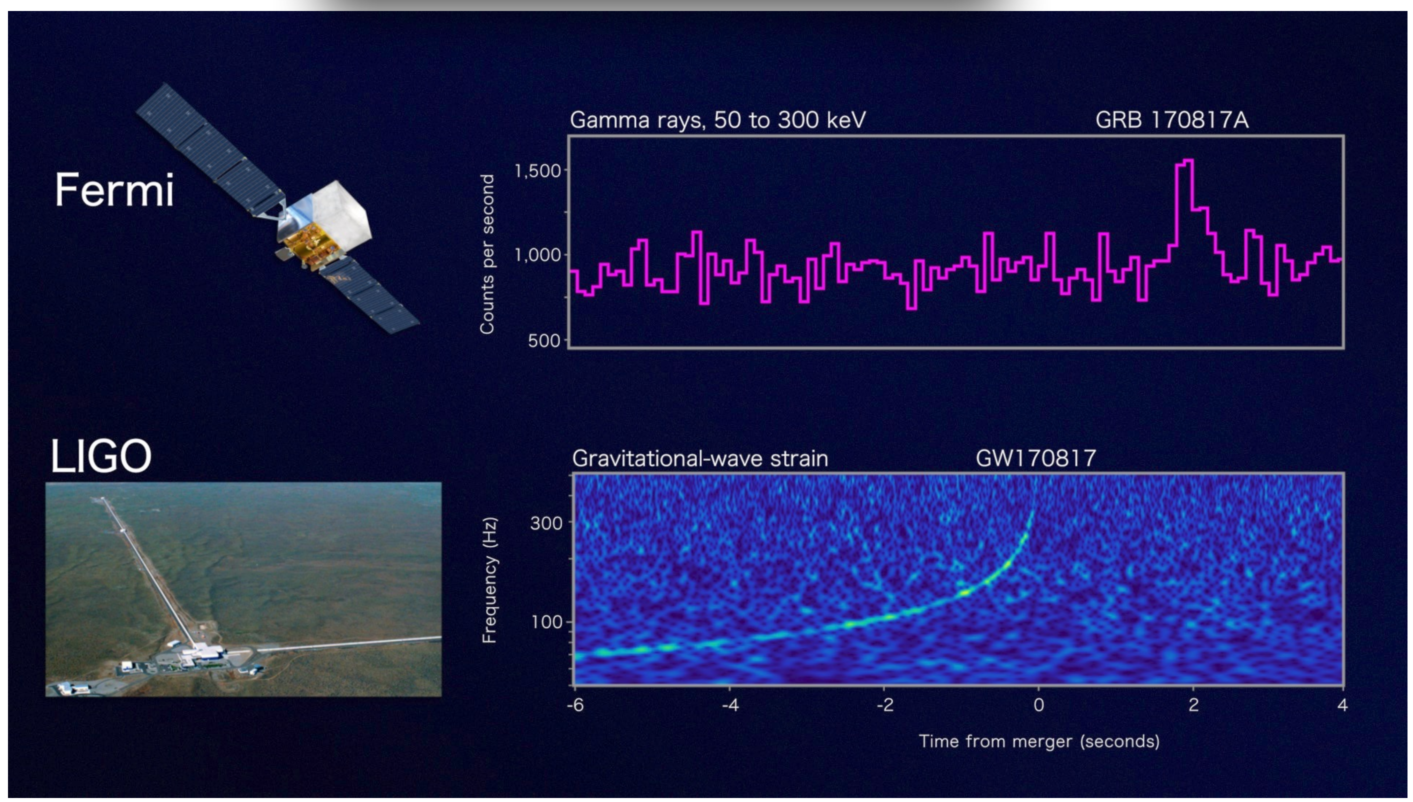
Equation of state leads to a hyper-massive neutron star that promptly collapses into a black hole



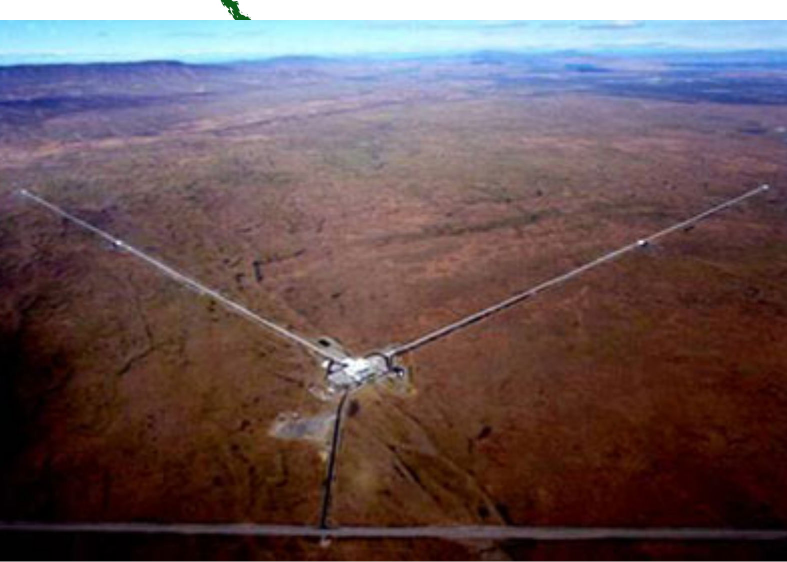
Models and simulations

Scientific Discovery





Observations



Theory

$$G_{\mu\nu} = 8\pi T_{\mu\nu}$$

Routine: black hole and neutron star collisions

Future: supernovae, oscillating neutron stars….

Gravitational Wave Discovery

Existing algorithms are computationally expensive and poorly scalable

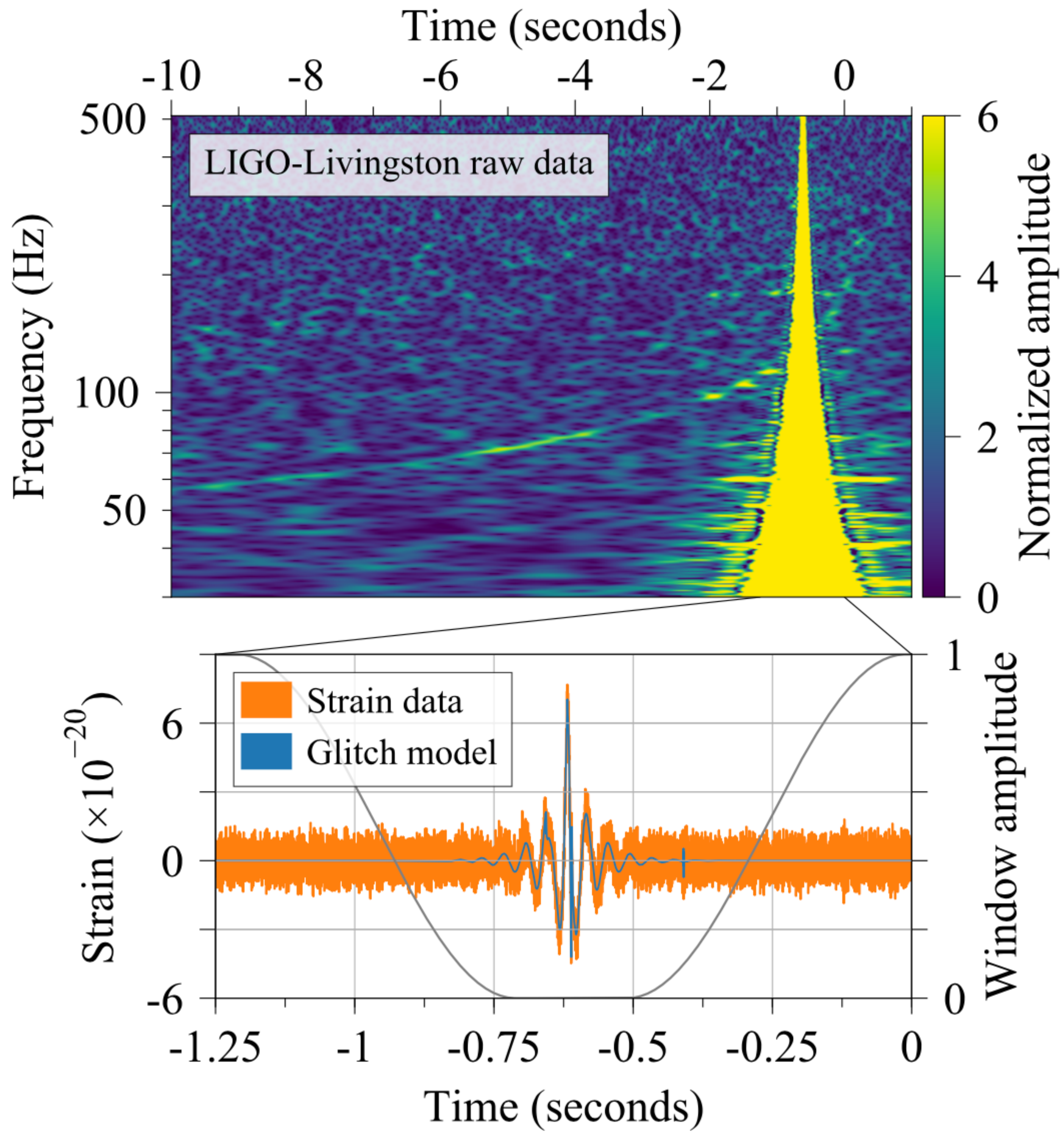
Extension to explore a deeper parameter space is computationally prohibitive

We only probe a 4-dimensional manifold out of the 9-dimensional signal manifold available to LIGO

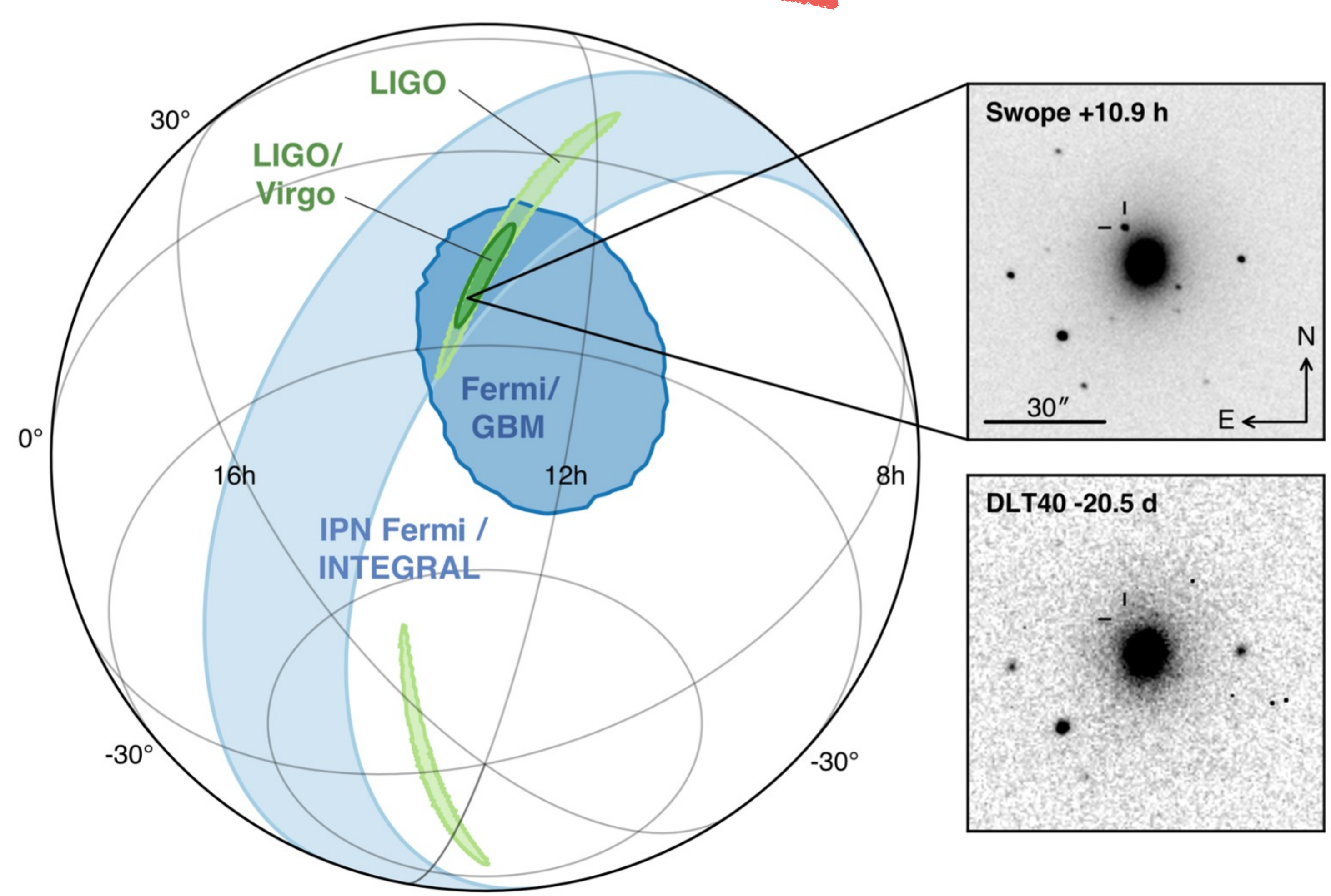
Are we missing astrophysically motivated sources in LIGO data

KAGRA and LIGO-India will eventually come on-line…

Do we go and seize all HPC and HTC resources to detect and characterize new GW sources in a timely manner?



Significant time-lag between gravitational wave detection and production of sky map

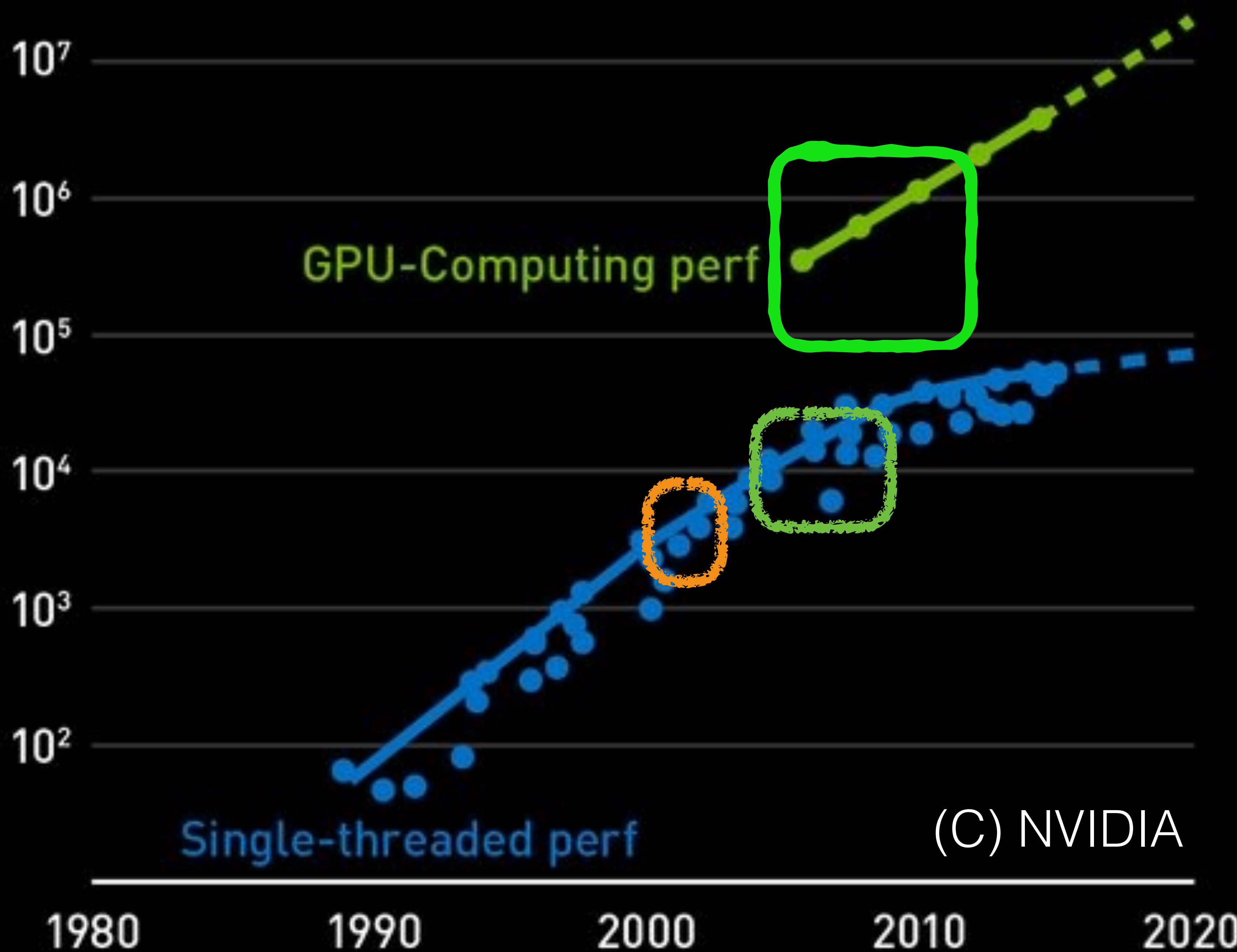


What if we could do this in real-time?

What if we could handle noise anomalies with no human intervention?

On disruptive changes and data revolutions

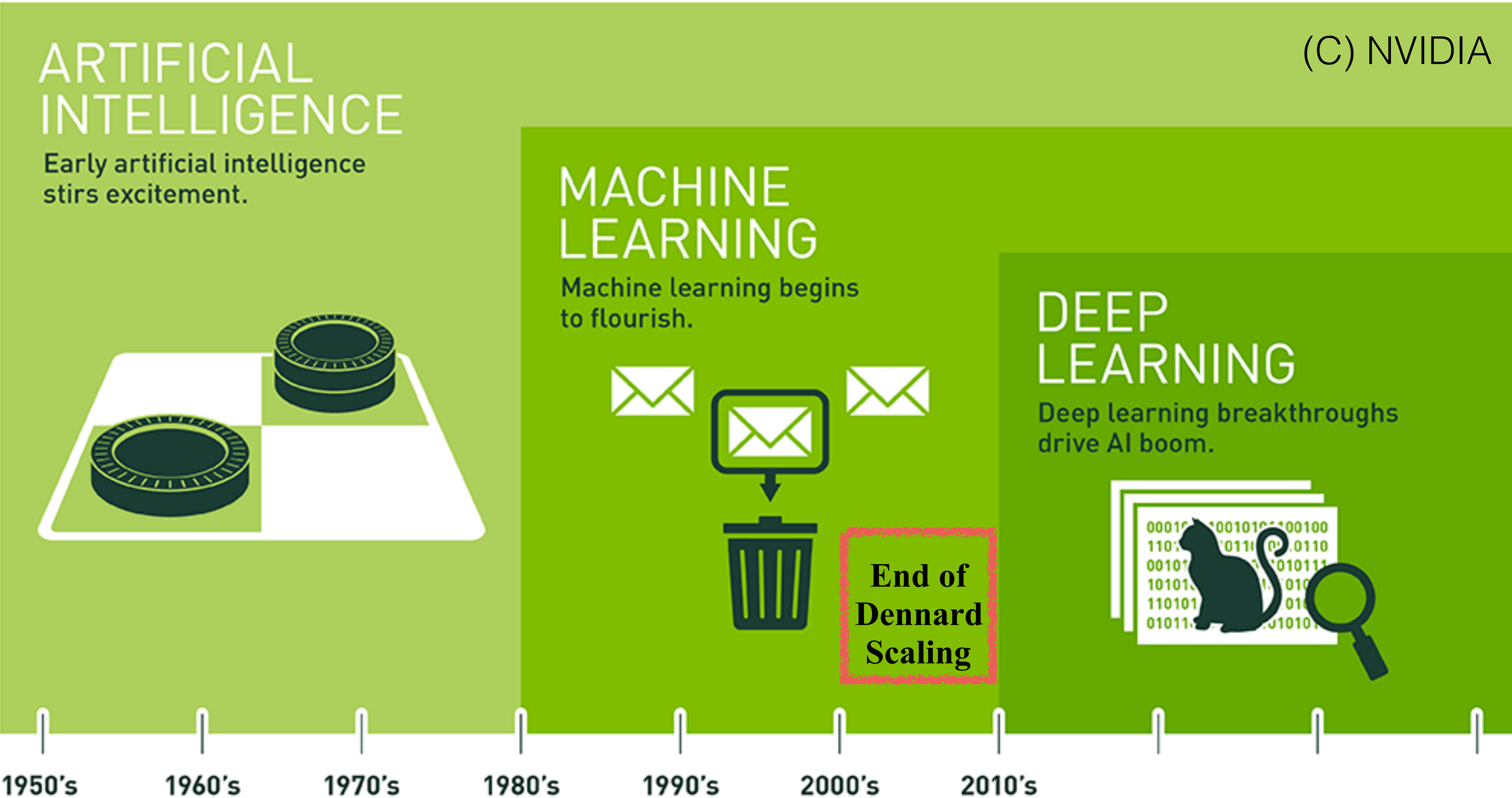
HPC and Big Data Revolution Coexist
Roadmap for Convergence



2012
Boom of interest in infrastructure and tools for big data analytics in cloud computing environments

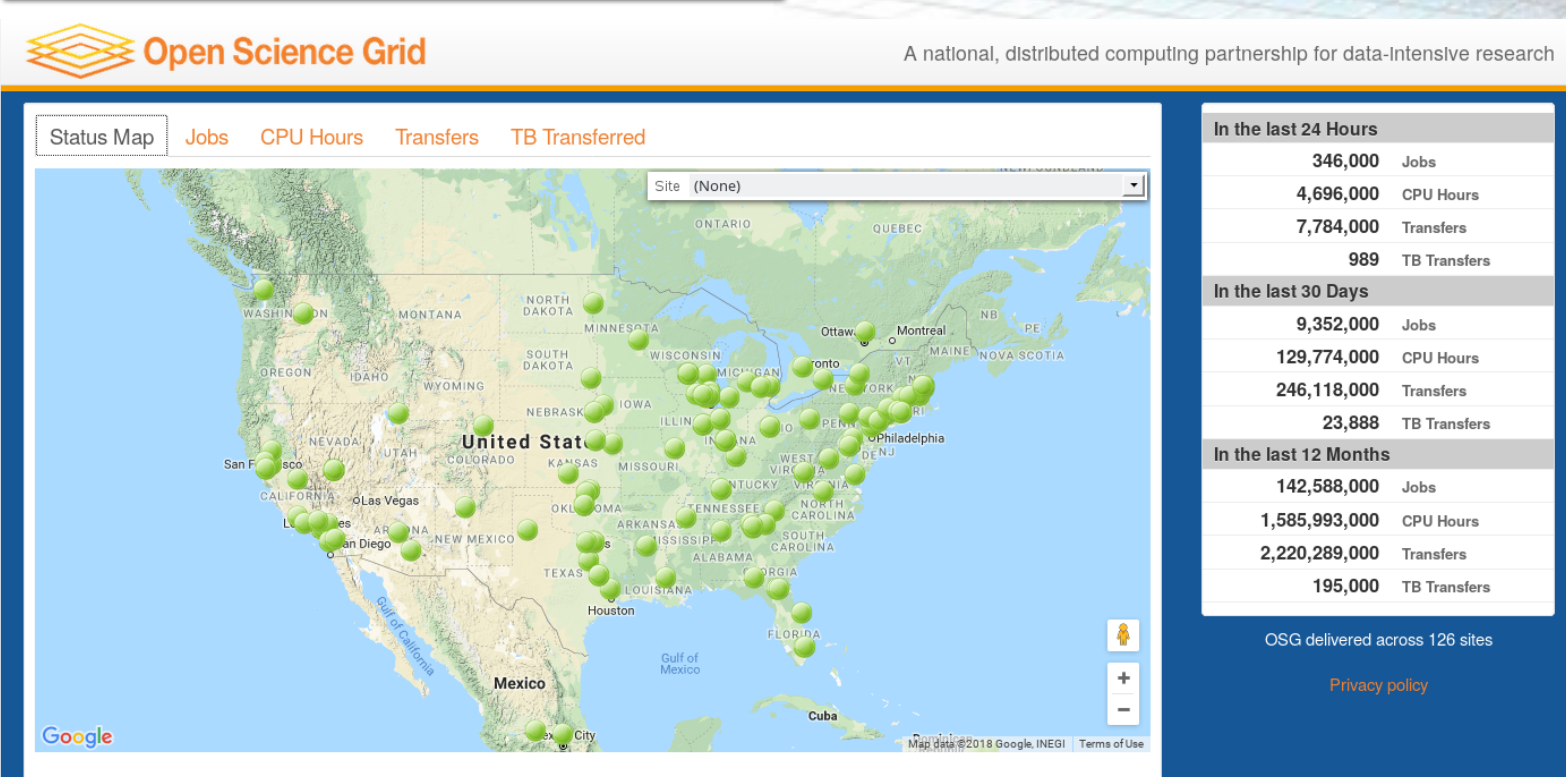
2015
US Presidential Strategic Initiative: convergence of big data and HPC ecosystem

Deep Learning
From optimism to breakthroughs in technology and science



Trends in **simulation** and data driven science

Interoperability of cyberinfrastructure resources



Open Science Grid as a universal adapter for disparate compute resources and science communities

Emergent trends for simulation and data driven science

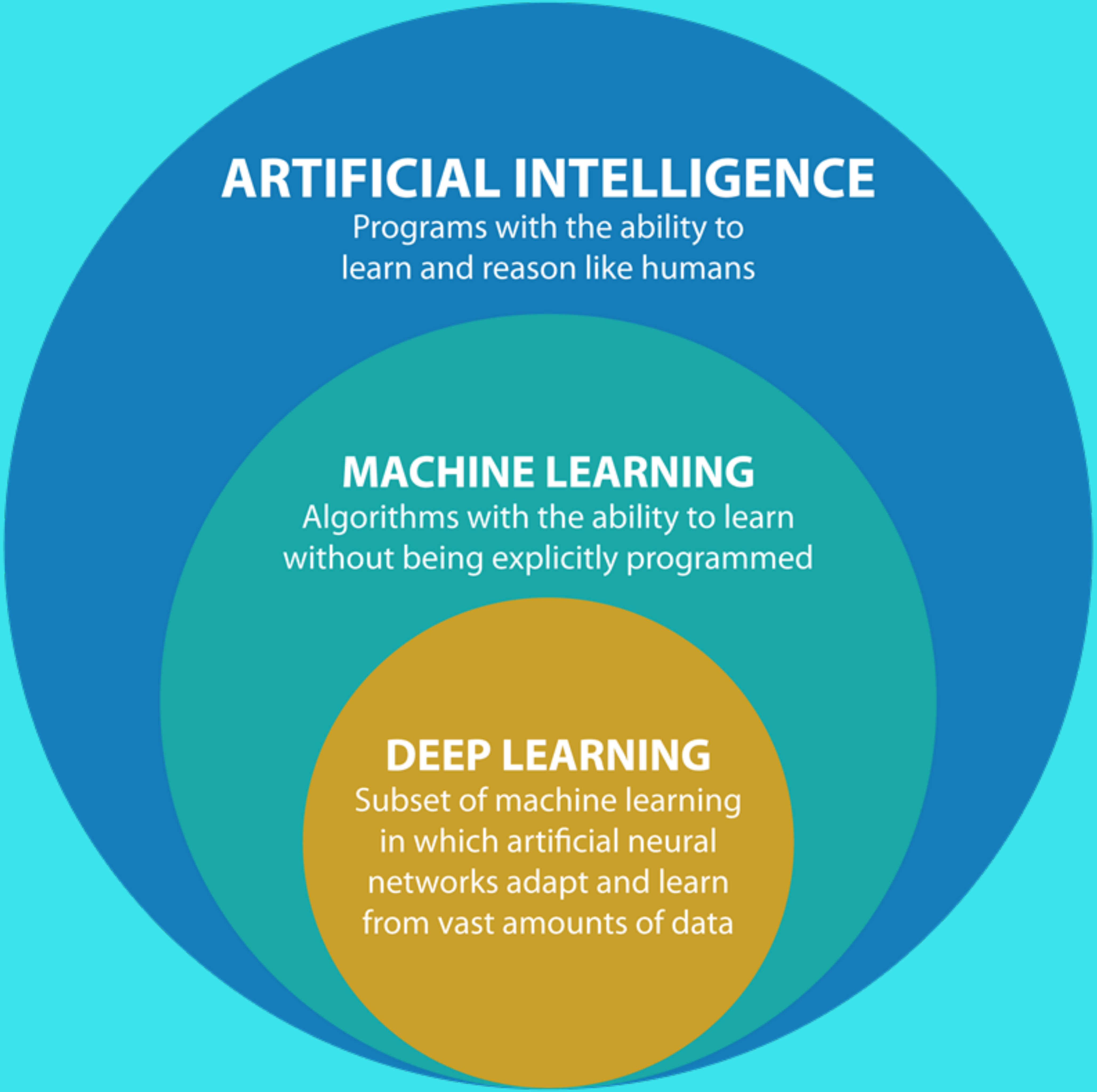
- US Presidential Strategic Initiative: convergence of big data and HPC ecosystem

- European Data Infrastructure and European Open Science Cloud: HPC is absorbed into a global system

- Japan and China: HPC combined with Artificial Intelligence (AI)

    - Japan: $1billion over the next decade for big data analytics, machine learning and the internet of things (IoT)

    - China: 5-yr plan raises big data analytics as a major application category of exascale systems


ARTIFICIAL INTELLIGENCE
Programs with the ability to learn and reason like humans
MACHINE LEARNING
Algorithms with the ability to learn without being explicitly programmed
DEEP LEARNING
Subset of machine learning in which artificial neural networks adapt and learn from vast amounts of data

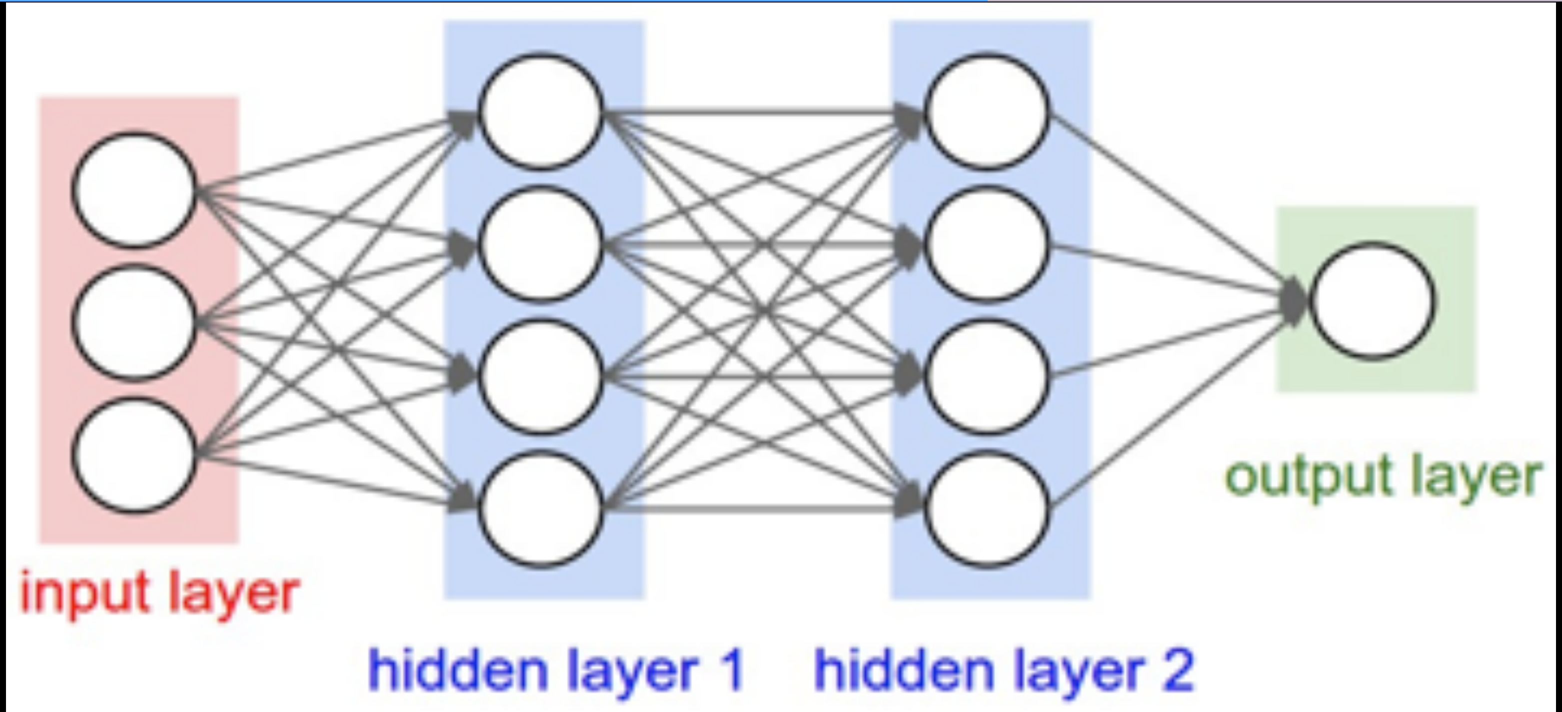
Deep Learning
Transforming how we do science

Overview

- Very long networks of artificial neurons (dozens of layers)
- State-of-the-art algorithms for face recognition, object identification, natural language understanding, speech recognition and synthesis, web search engines, self-driving cars, games…

Representation learning

- Does not require hand-crafted features to be extracted first
- Automatic end-to-end learning
- Deeper layers can learn highly abstract functions

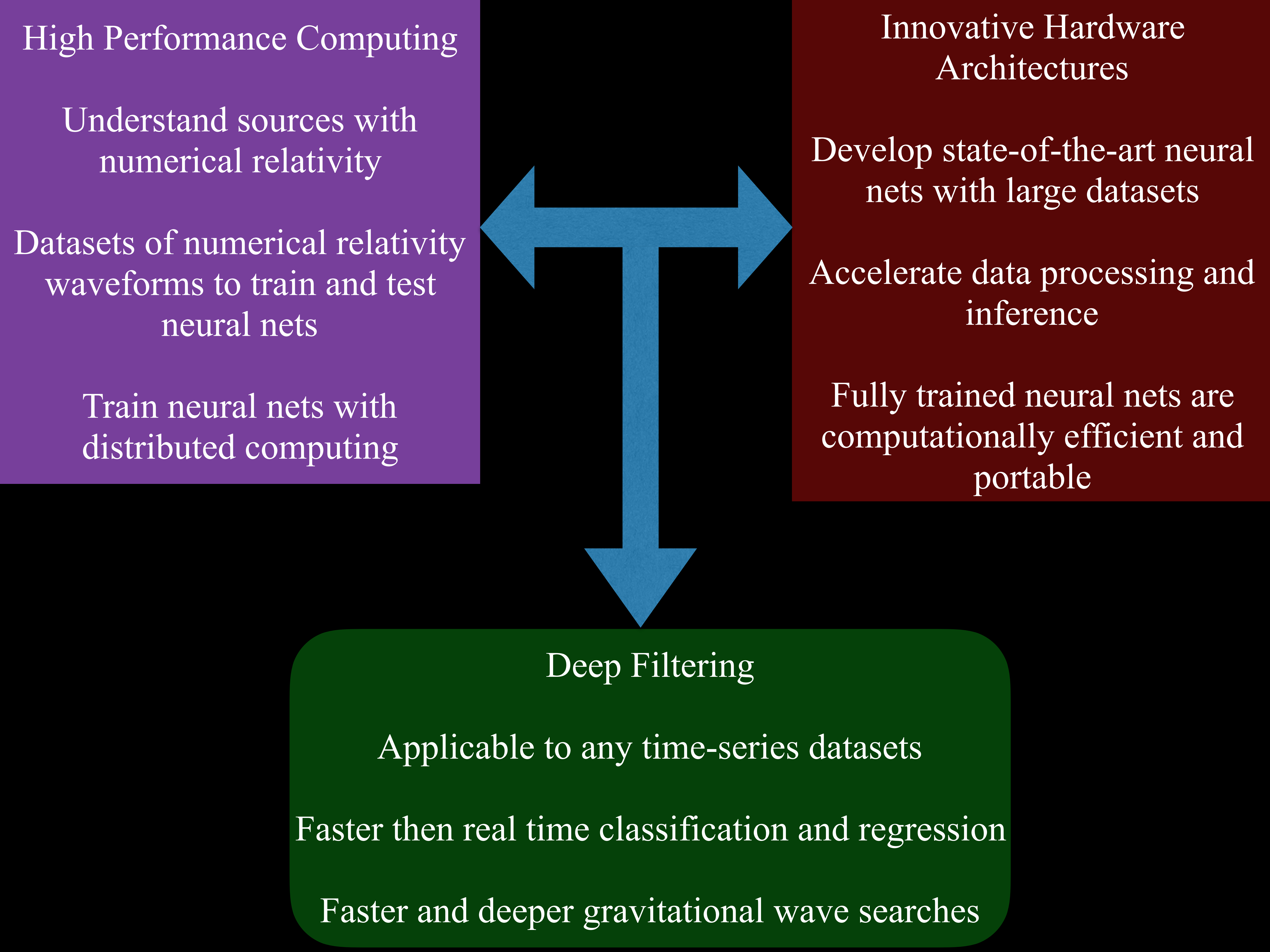


Innovate@NCSA

Adapt existing deep learning paradigm to do classification and regression of time-series data

Replace pixels in images by time-series vectors; pixel represents amplitude of waveform signals

Combine HPC to construct catalogs of numerical relativity waveforms with new deep learning training algorithms to find weak gravitational wave signals in non-Gaussian and non-stationary gravitational wave data



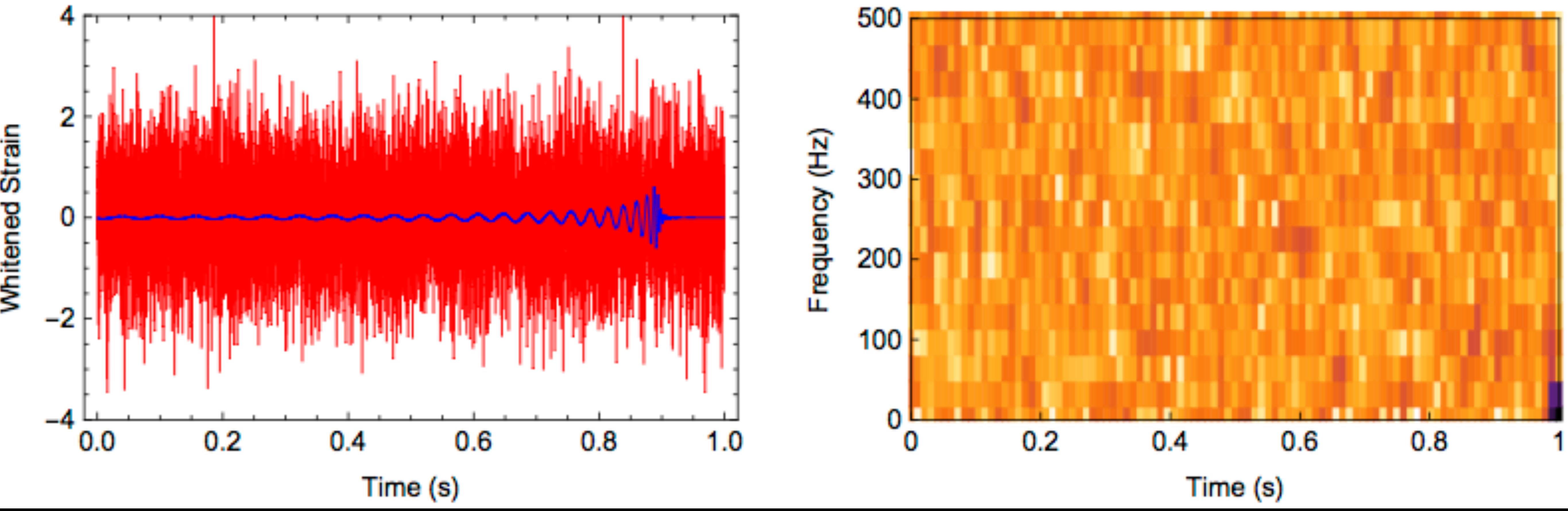
High Performance Computing
Understand sources with numerical relativity
Datasets of numerical relativity waveforms to train and test neural nets
Train neural nets with distributed computing
Innovative Hardware Architectures
Develop state-of-the-art neural nets with large datasets
Accelerate data processing and inference
Fully trained neural nets are computationally efficient and portable
Deep Filtering
Applicable to any time-series datasets
Faster then real time classification and regression
Faster and deeper gravitational wave searches

Deep Filtering

D George & E. A. Huerta, *Physical Review* D 97, 044039 (2018)

First scientific application for processing highly noisy time data series

Using spectrograms is sub-optimal for gravitational wave data analysis

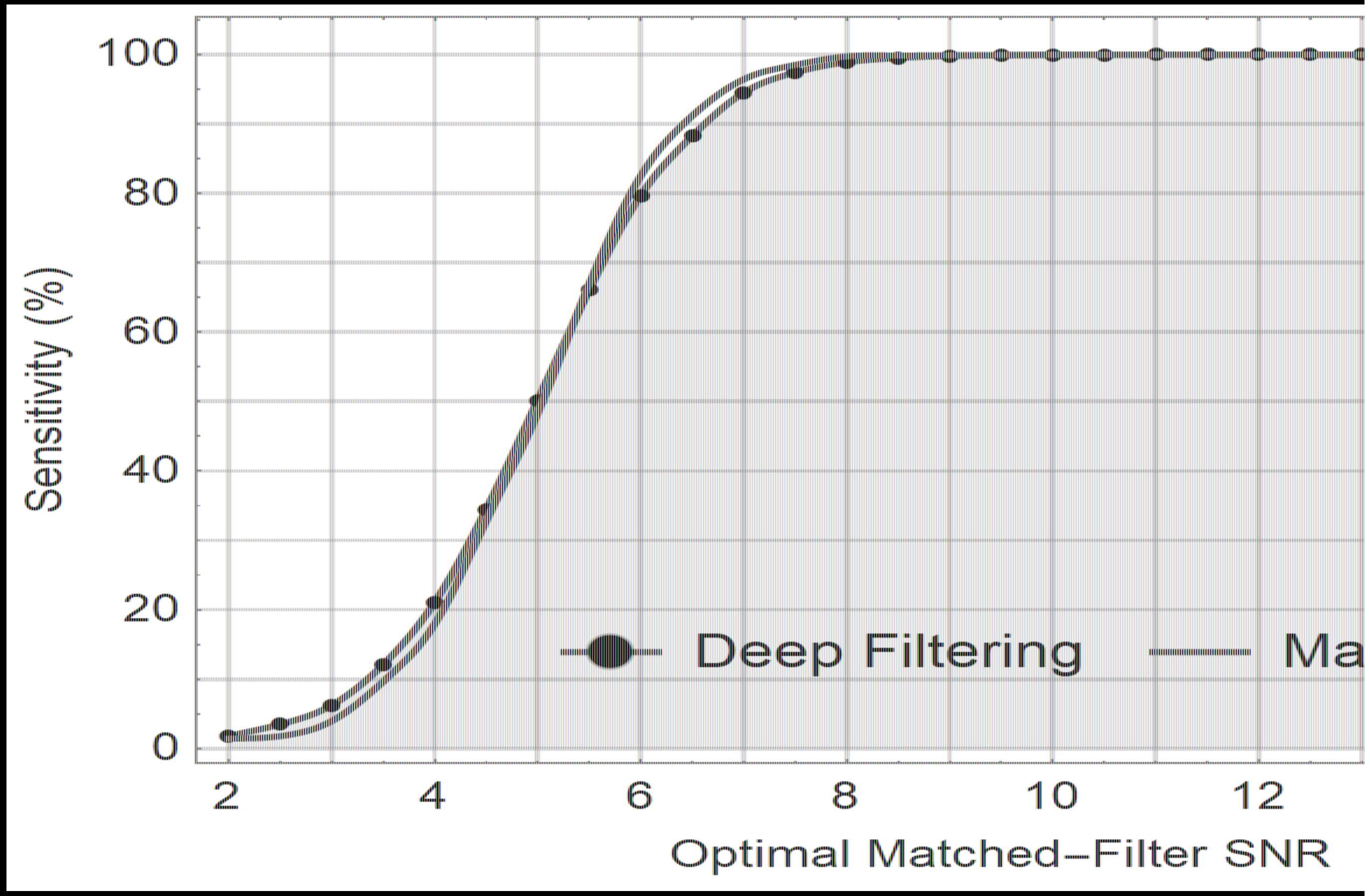


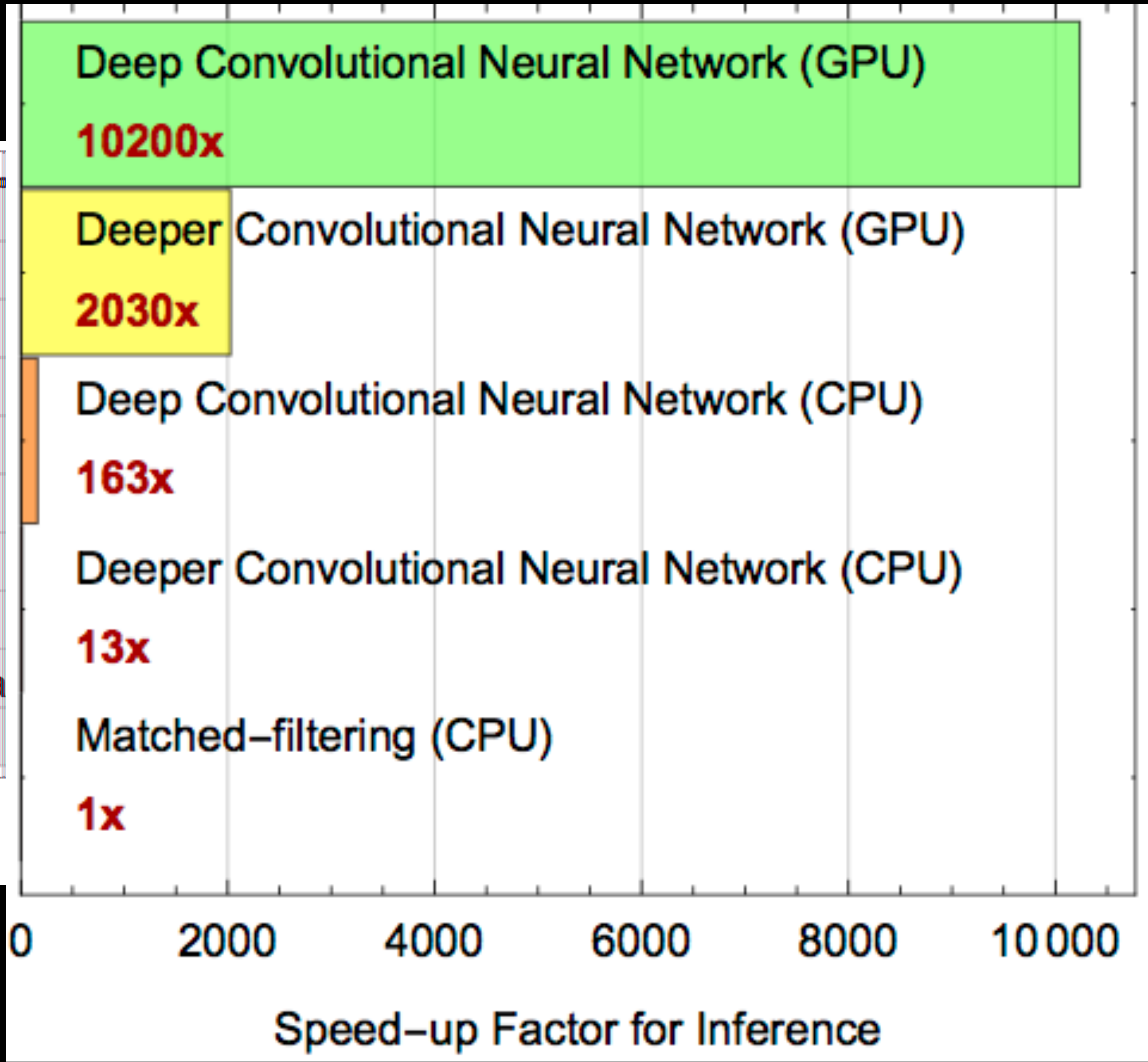
Deep Filtering

D George & E. A. Huerta, *Physical Review* D 97, 044039 (2018)

First scientific application for processing highly noisy time data series

**Sensitivity for detection is similar to a matched filter in Gaussian noise…
but orders of magnitude faster…**



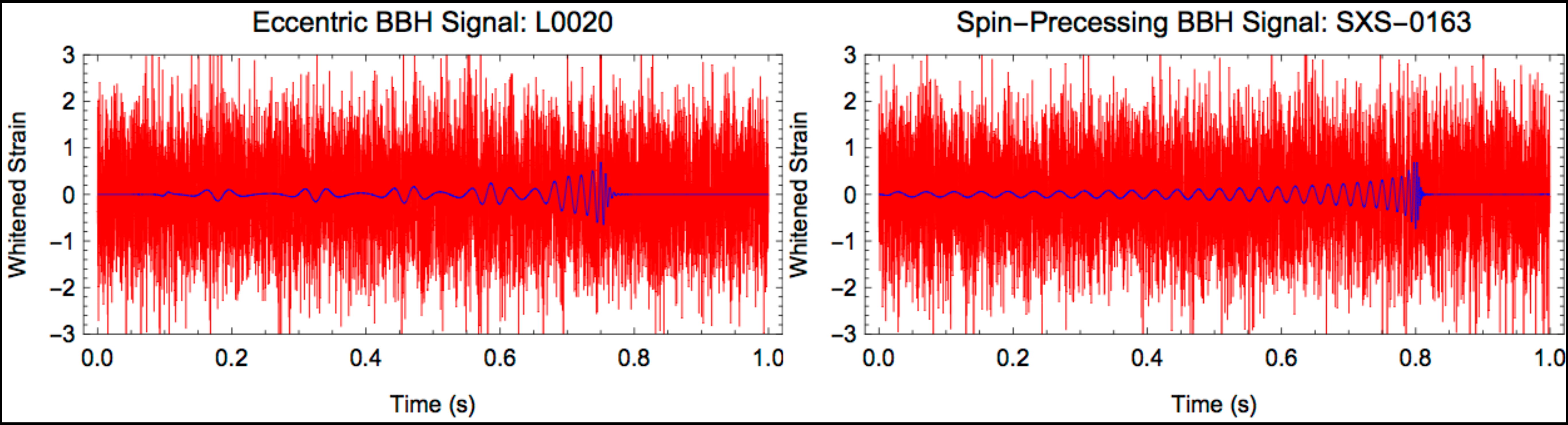


Deep Filtering

D George & E. A. Huerta, *Physical Review* D 97, 044039 (2018)

First scientific application for processing highly noisy time data series

Sensitivity for detection is similar to a matched filter in Gaussian noise…
but orders of magnitude faster…
and enables the detection of new types of gravitational wave sources

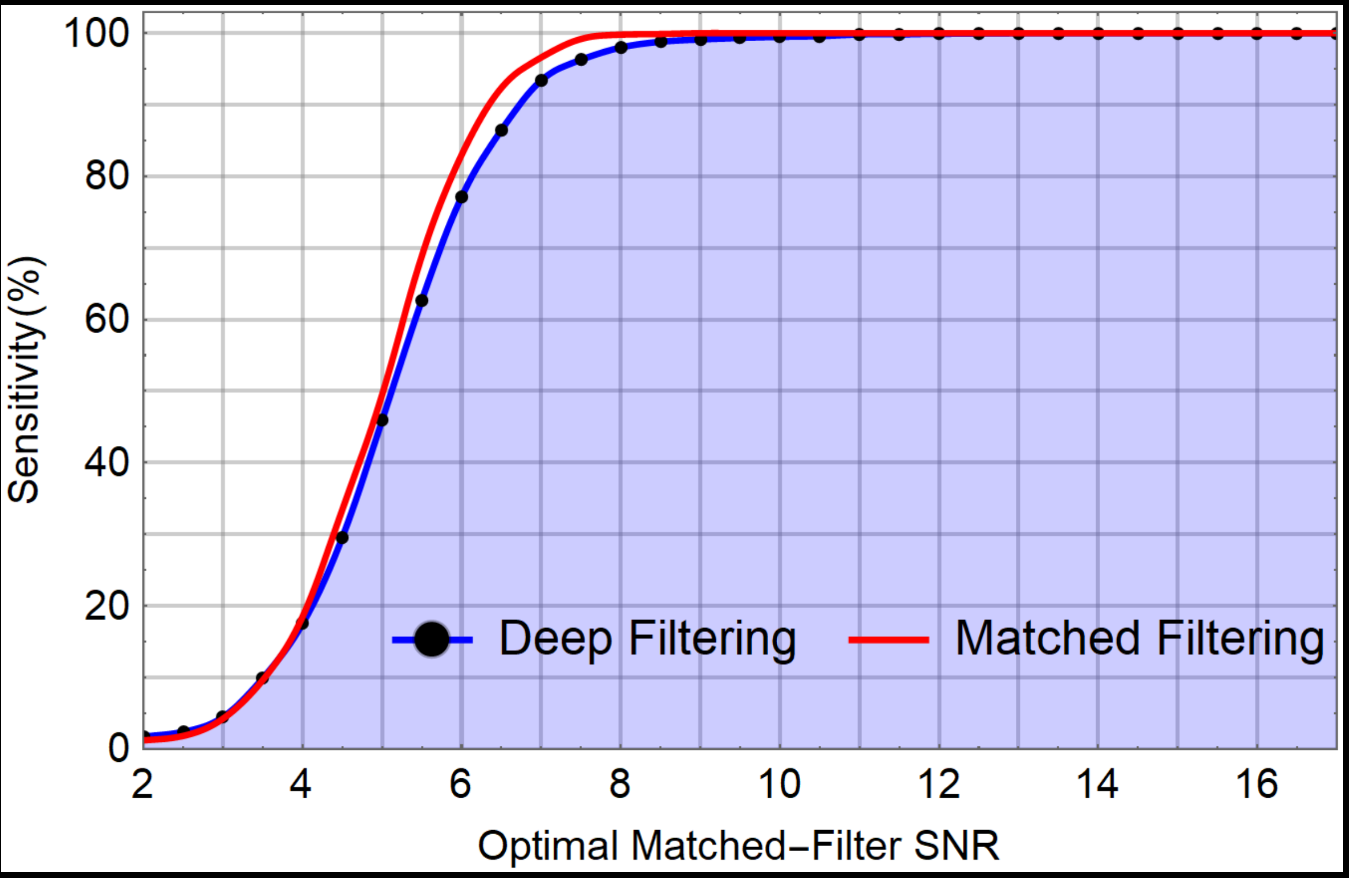


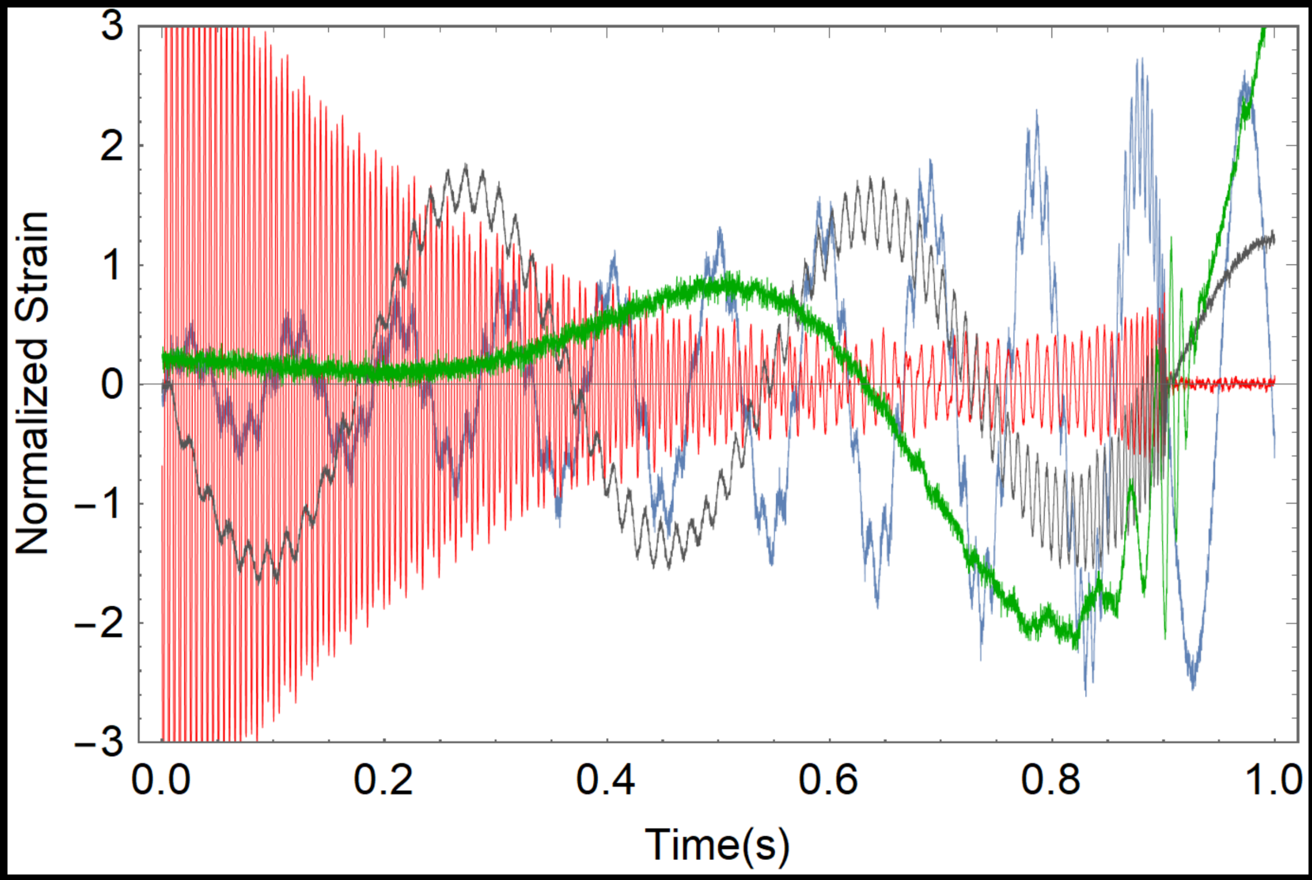
Deep Filtering

D George & E. A. Huerta *Physics Letters* B 778 (2018) 64-70

First scientific application for processing highly noisy time data series

As sensitive as matched-filtering
More resilient to glitches
Enables new physics
Deeper gravitational wave searches faster than real-time

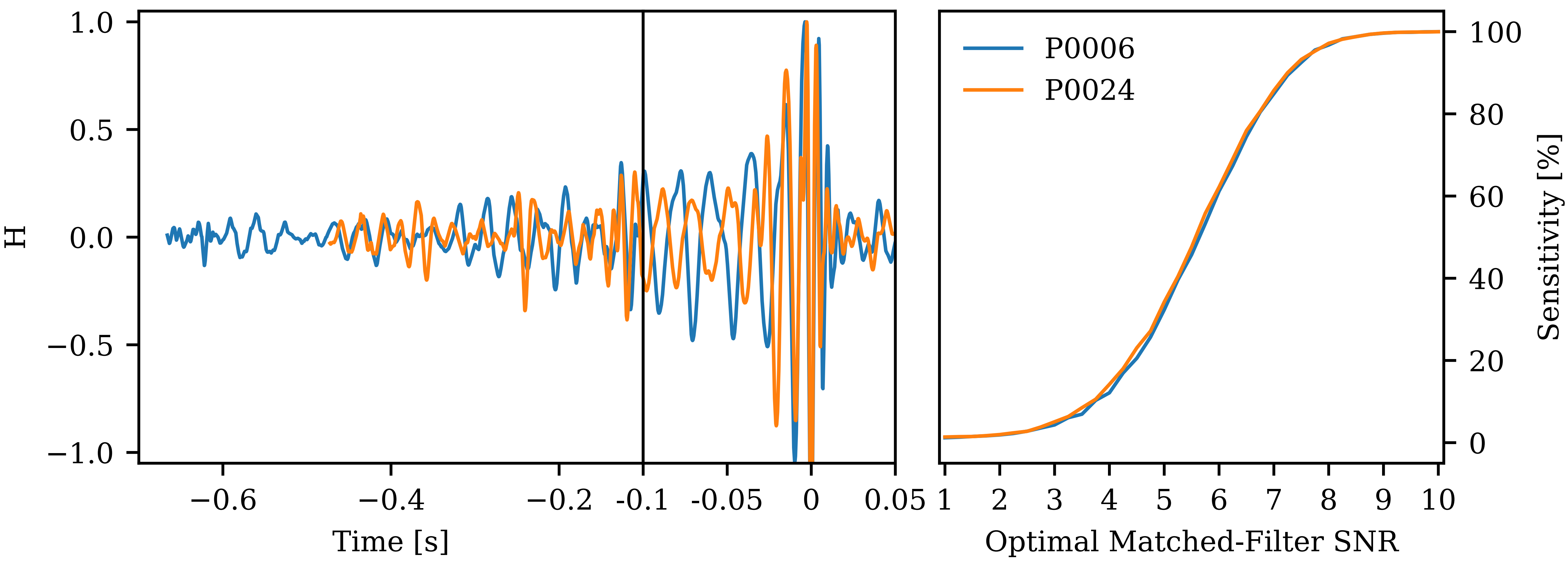




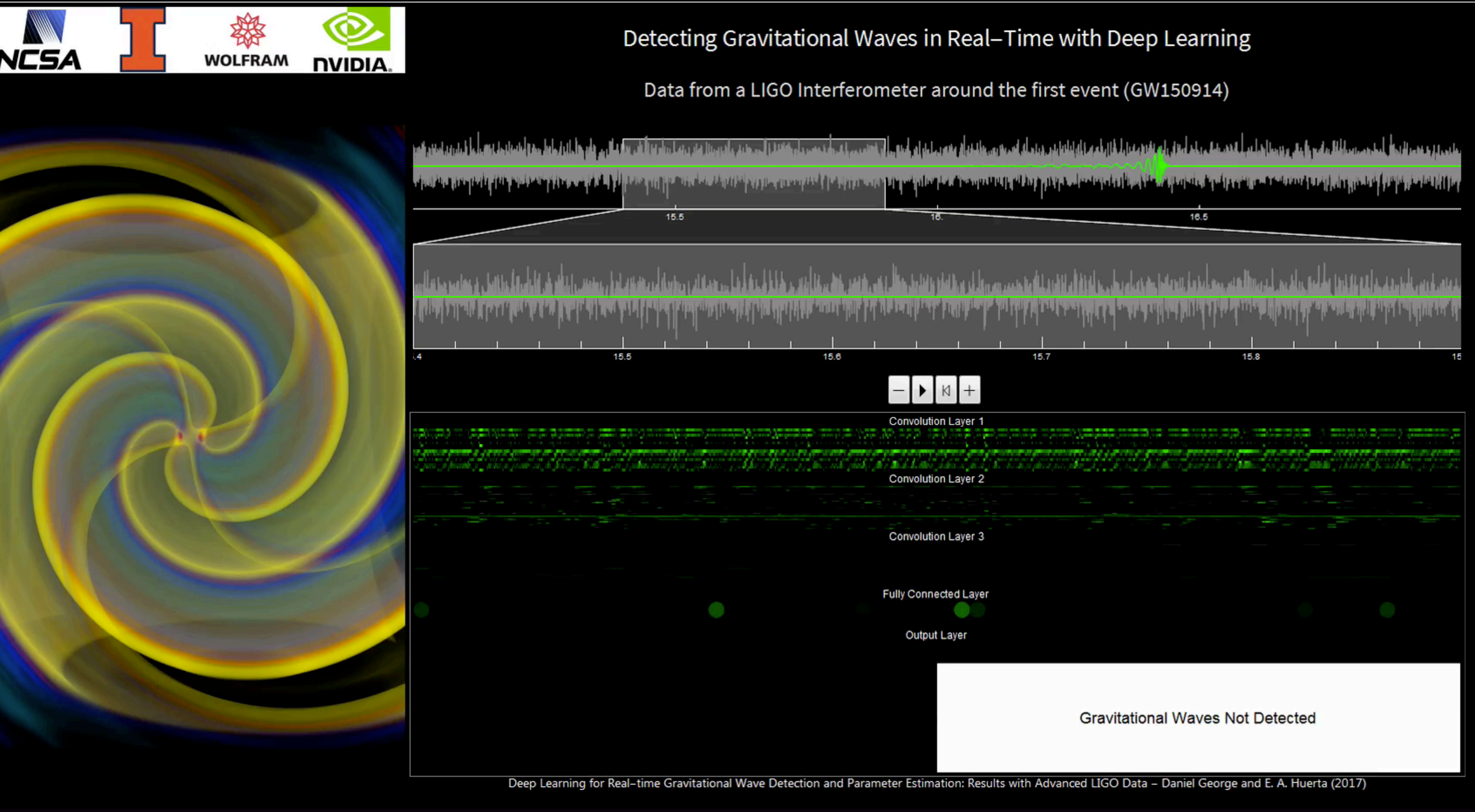
Deep Filtering

Rebei, Huerta, et al, arXiv:1807.09787
First scientific application for processing higher-order multipoles waveforms in highly noisy time data series

A new class of gravitational wave sources can be seamlessly detected by deep learning with the same accuracy we can identify quasi-circular waveform signals



https://www.youtube.com/watch?v=87zEIl_hkBE



FUSION OF AI & HPC & SCIENTIFIC VISUALIZATION

REAL-TIME DETECTION AND REGRESSION OF REAL EVENTS IN RAW LIGO DATA

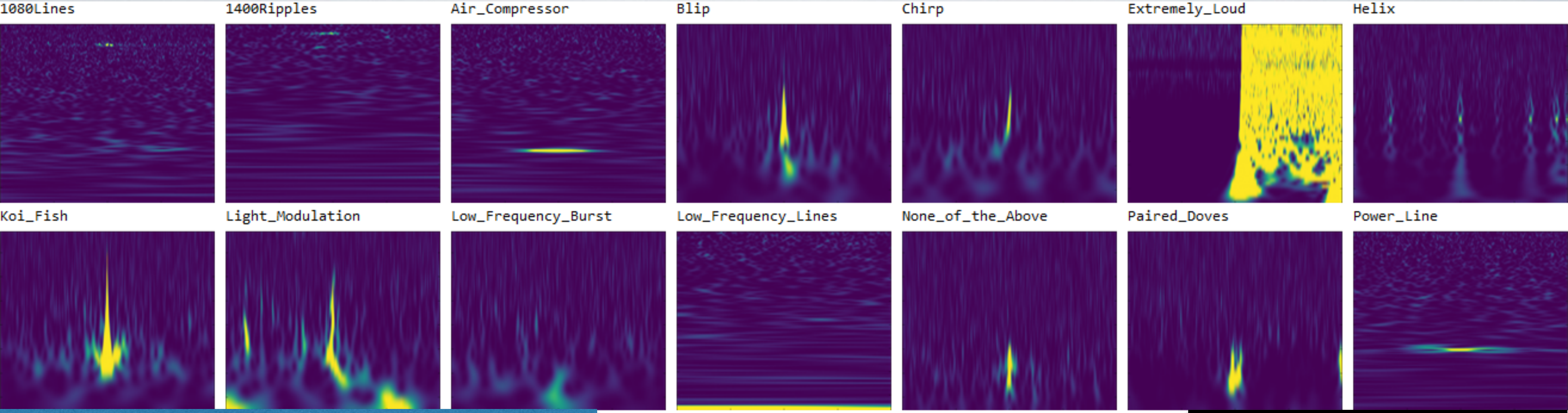
Deep Learning for Observational Astronomy

Post-process images to classify and cluster noise anomalies in real-time

Goal: enable real-time discovery with the Large Synoptic Survey Telescope (LSST); 15TB of data per night, thousands of triggers per second

Case study: LIGO data from first observing run

Noise anomalies in LIGO data

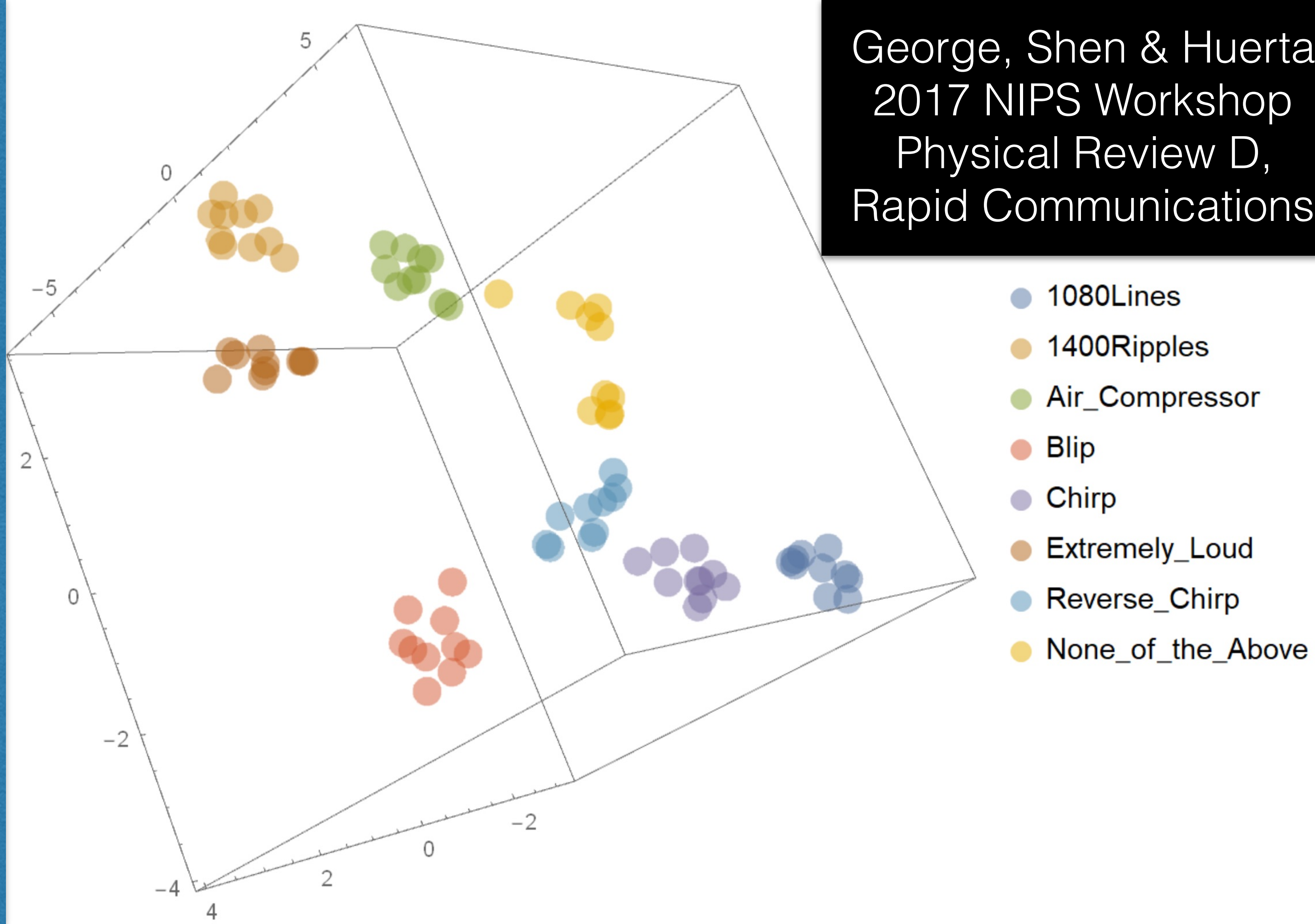


Classify and cluster anomalies according to morphology. First application to LIGO Science

Perform the same task fort unlabeled datasets

Classify and cluster new classes of glitches in real-time

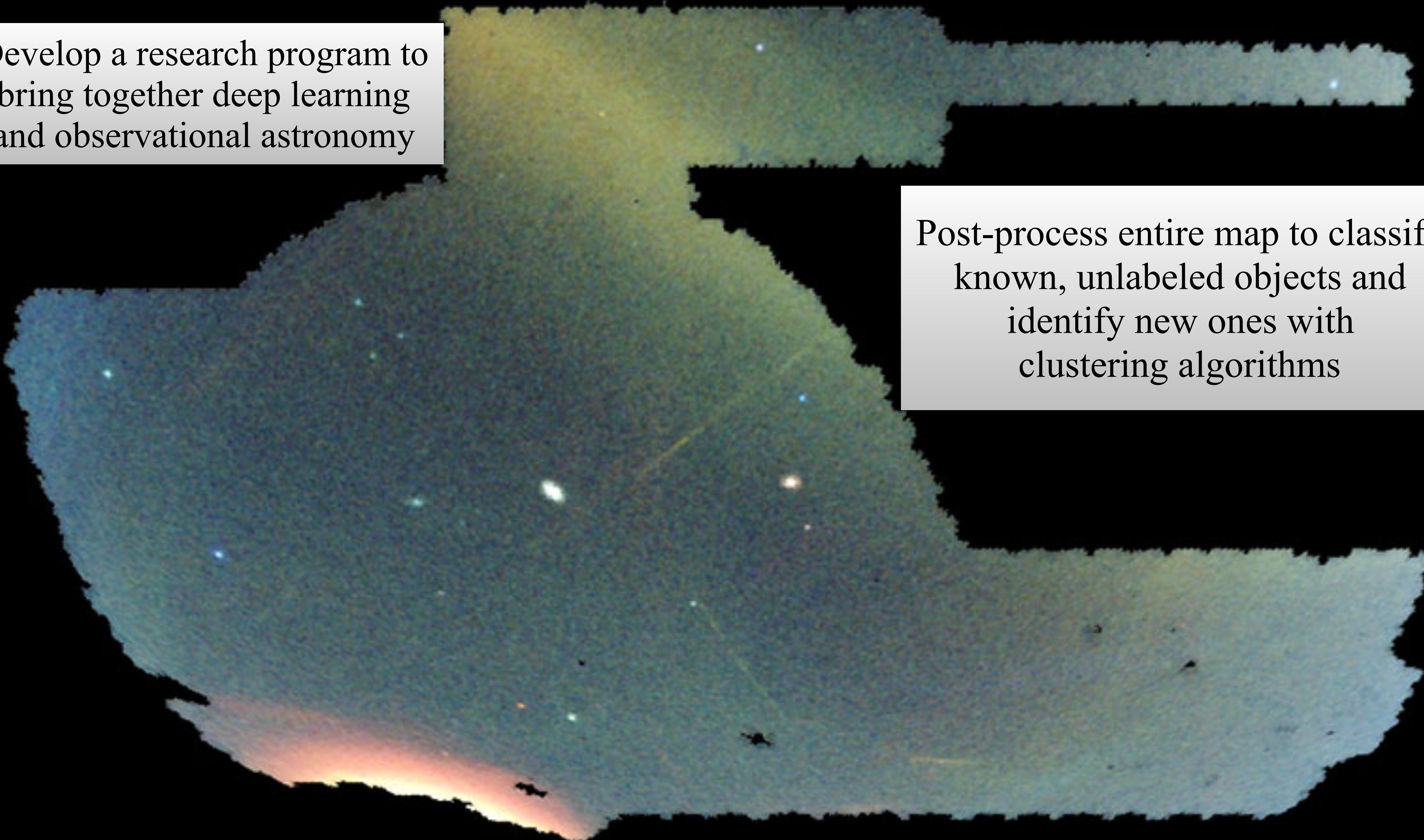
Add this new capability to Deep Filtering



George, Shen & Huerta
2017 NIPS Workshop
Physical Review D,
Rapid Communications

Now consider these anomalies

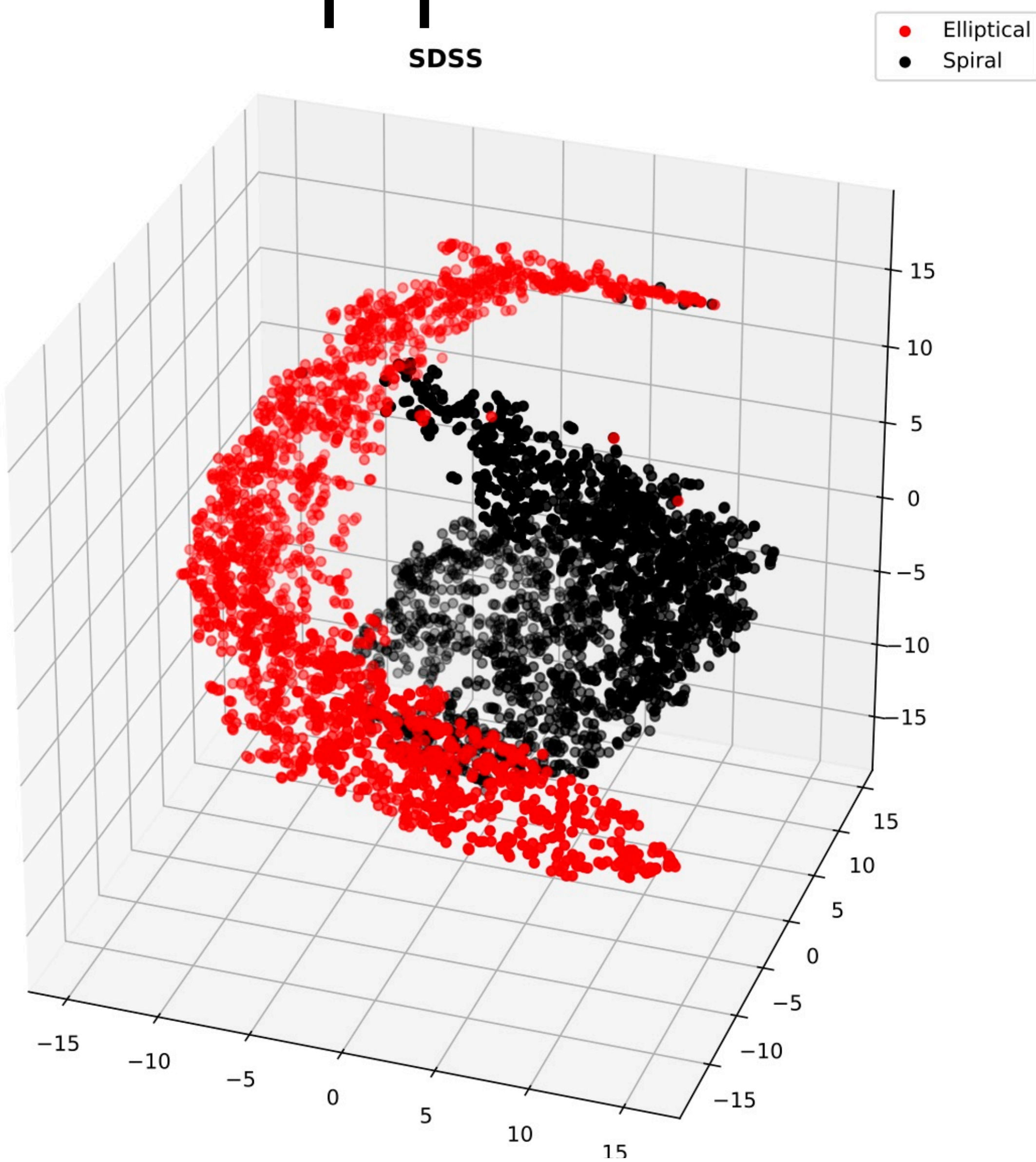
Develop a research program to bring together deep learning and observational astronomy



Post-process entire map to classify known, unlabeled objects and identify new ones with clustering algorithms

Dark Energy Survey (DES)

Application to galaxy images



Khan, Huerta, Wang and Gruendl, arXiv:1812.02183

36k+ raw galaxy images from the Sloan Digital Sky Survey clustered according to morphology and 3 filters

0.9999943

0.99999285

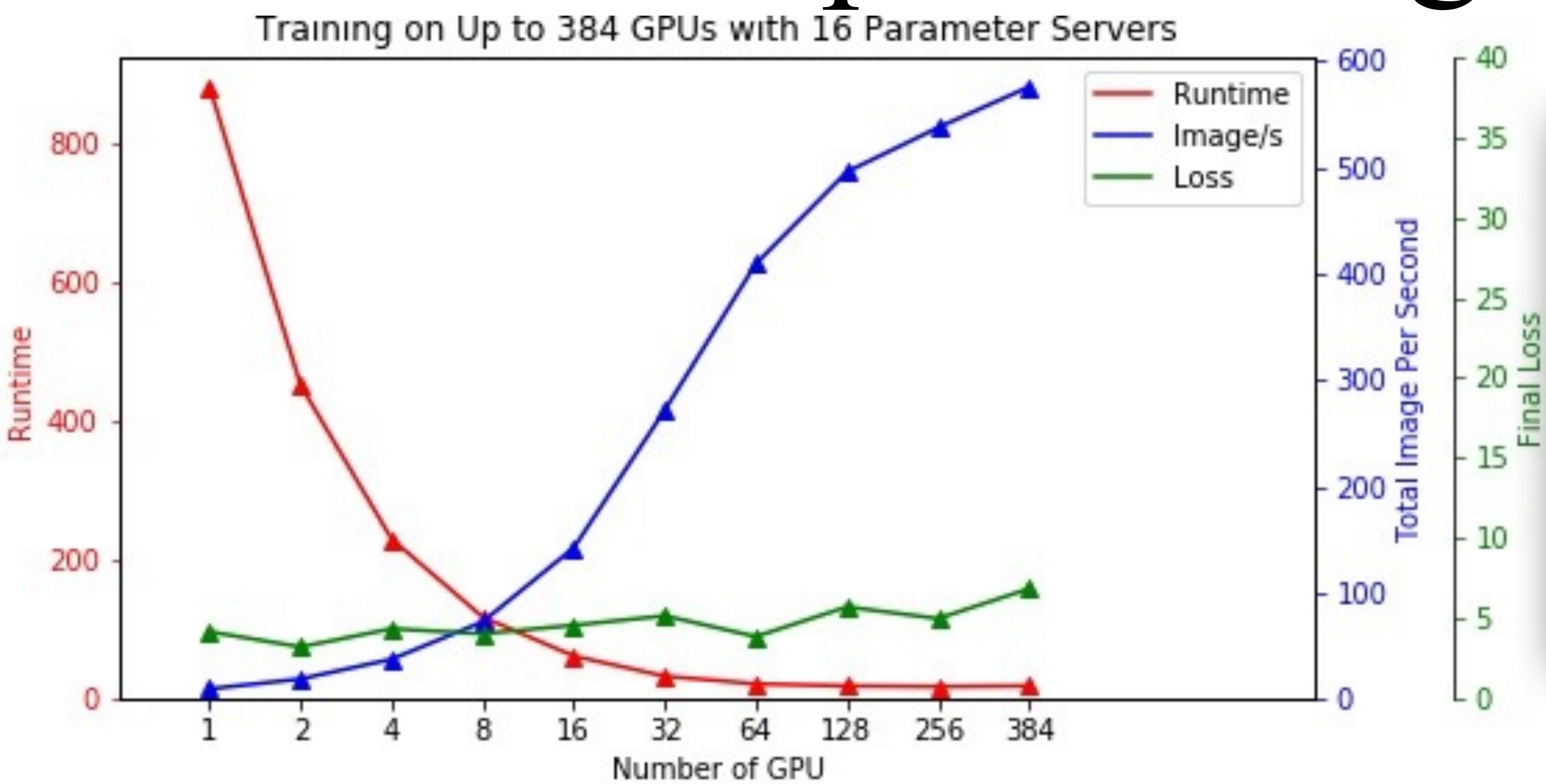
https://www.youtube.com/watch?v=n5rl573i6ws

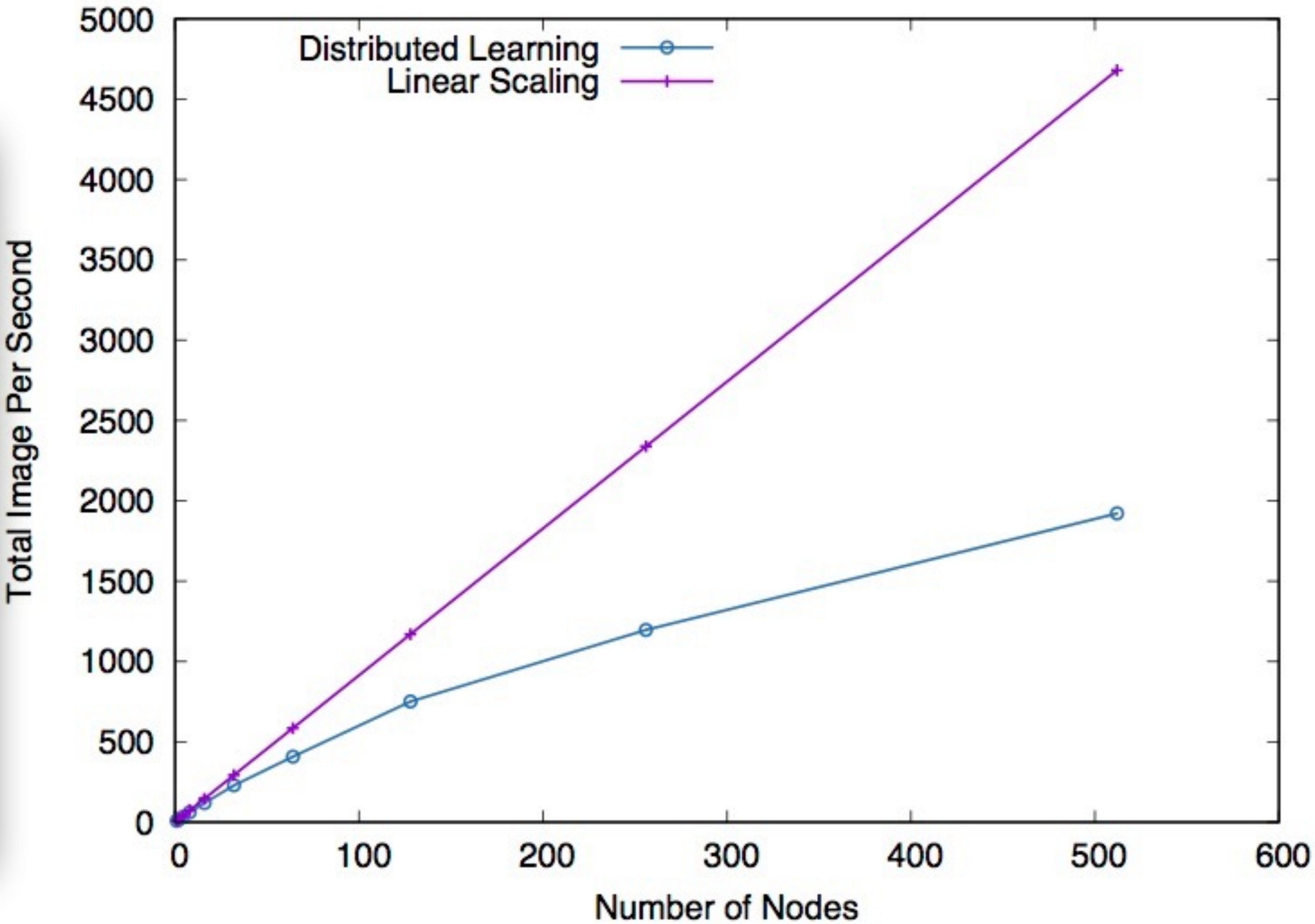
Deep Learning at scale

What is the optimal neural network architecture to enable discovery in higher dimensional signal manifold?

Is it possible to design deeper and more accurate neural net models using larger training datasets while also reducing the length of the training stage?

Deep Learning at scale

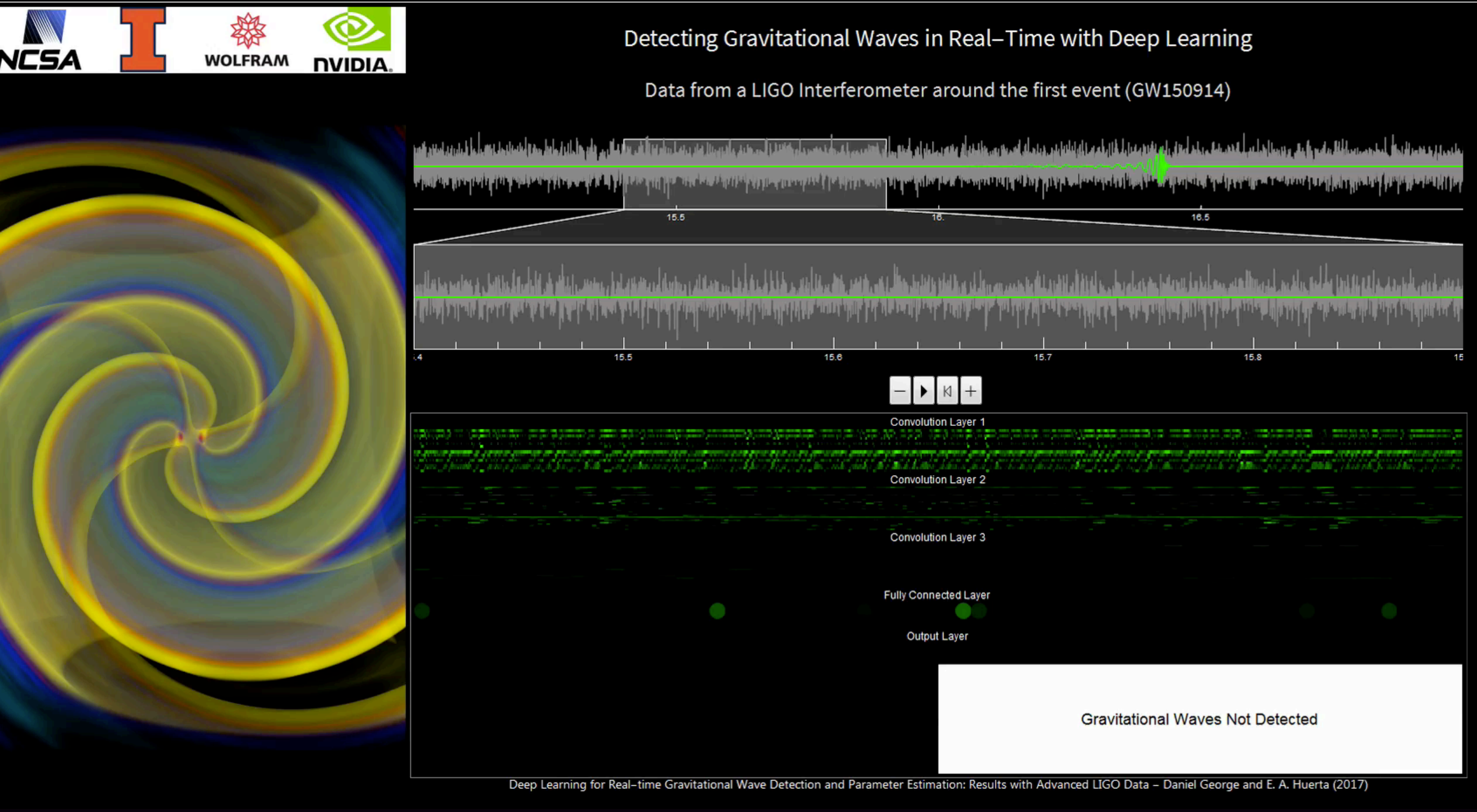




Scientific Machine Learning

- What do neural nets learn?
- Reproducible training methods
- How do we interpret their results?
- What is the cost of failure?
- Where is AI heading?

https://www.youtube.com/watch?v=87zEII_hkBE



FUSION OF AI & HPC & SCIENTIFIC VISUALIZATION

REAL-TIME DETECTION AND REGRESSION OF REAL EVENTS IN RAW LIGO DATA

NCSA

XSEDE
Extreme Science and Engineering
Discovery Environment

NSF

I

NVIDIA®

Capital One

DEPARTMENT OF ENERGY
UNITED STATES OF AMERICA

Argonne
NATIONAL LABORATORY