Frontiers at the interface of High Performance Computing Deep Learning and Multimessenger Astrophysics

Roland Haas  
(PI: Eliu Huerta)  
Gravity Group  
gallery.ncsa.illinois.edu
National Center for Supercomputing Applications  
University of Illinois at Urbana-Champaign

Blue Waters Symposium  
Sunriver Oregon, June 4-7 2018
Outline

Trends in simulation and data-driven science

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

Existing facilities and services to address the needs of the science community

Emergent trends for simulation and data-driven science
Trends in **simulation** and data driven science

**Fusion of HPC and HTC**

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

**Interoperability of cyberinfrastructure resources**
Existing facilities to cover the needs of the science community

Two case studies:

• Use of Blue Waters for the discovery of two colliding neutron stars in gravitational waves and light

• Fusion of HPC & AI for gravitational wave astrophysics
Detecting gravitational waves

LIGO's raw data is noise dominated

Signal detection is computationally expensive
LIGO DATA GRID

9 clusters, 17k+cores
Connected to Open Science Grid and XSEDE since 2015

Wider detector network with ever increasing detection sensitivity demands more computational resources

NCSA-led team connected Blue Waters to the LIGO Data Grid, used it during O2
Huerta et al, eScience 47, 2017
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
HPC enables numerical simulations of neutron stars collisions: combination of Einstein’s general relativity with magnetohydrodynamics and microphysics.
First time Blue Waters is configured as an Open Science Grid compute element, and combined with Shifter for scientific discovery. Huerta et al., eScience, 47, 2017.
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?
On disruptive changes and data revolutions

End of Dennard Scaling

2004
HPC reaches inflection point

2009-2012
International Exascale Software Initiative
On disruptive changes and data revolutions

HPC and Big Data Revolution Coexist
Roadmap for Convergence

2012
Boom of 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

ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.

MACHINE LEARNING
Machine learning begins to flourish.

DEEP LEARNING
Deep learning breakthroughs drive AI boom.


End of Dennard Scaling
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
Deliverable: create skymap in real-time and estimate source’s parameters even if signals are contaminated by noise anomalies.

Wish list: handle noise anomalies in real-time and with no human intervention.

(C) LVC, Phys. Rev. Lett. 119, 161101 (2017)
Innovate

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

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

Fuse AI (deep learning algorithms) and HPC (catalogs of numerical relativity waveforms and distributed learning) to find weak gravitational wave signals in raw LIGO data
Deep Filtering

First scientific application for processing highly noisy time data series

Using spectrograms is sub-optimal for gravitational wave data analysis
Deep Filtering

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

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

---

Eccentric BBH Signal: L0020

Spin-Precessing BBH Signal: SXS-0163
Deep Filtering

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
Conclusions

• Blue Waters contributed to LIGO's detection of a neutron star binary

• Deep neural networks detect LIGO signals at least as efficiently as current methods

• Use Blue Waters to train networks for 7D LIGO parameter space
High Performance Computing

Understand sources with numerical relativity

Datasets of numerical relativity waveforms to train and test neural nets

Train neural nets with distributed learning

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 than real time classification and regression

Faster and deeper gravitational wave searches
Detecting Gravitational Waves in Real-Time with Deep Learning

Data from a LIGO Interferometer around the first event (GW150914)


Gravitational Waves Not Detected

FUSION OF AI & HPC & SCIENTIFIC VISUALIZATION
REAL-TIME DETECTION AND REGRESSION OF REAL EVENTS IN RAW LIGO DATA

https://www.youtube.com/watch?v=87zEll_hkBE
Multimessenger Astrophysics

LIGO Virgo
KAGRA...

Numerical and Analytical Relativity

Deep and machine

Fusion of HPC and AI to accelerate and maximize discovery

Raw Data

Raw Data

DES

LSST

NCSA Gravity Group vision for Multimessenger Astrophysics
A primary goal of the National Strategic Computing Initiative is to foster the convergence of data analytic computing, modeling and simulation. Since this initiative is co-led by the NSF, it is very appropriate that the NSF Leadership Class supercomputer, Blue Waters, has been at the forefront of this effort by creating environments that are highly efficient for both large parallel modeling, and for large data pipelines for observation and experiment.

The NCSA Gravity Group, the Blue Waters Application and Systems Team, the LIGO Lab at Caltech, the San Diego Supercomputing Center (SDSC) and Open Science Grid Project worked for a year to connect the LIGO Data Grid to the Blue Waters supercomputer.

Supporting high throughput LIGO data analysis workflows concurrently with highly parallel numerical relativity simulations and many other complex workloads is the most recent success and most complex example of successfully achieving convergence on Leadership Class computers like Blue Waters, which is much earlier than was expected to be possible.
Scientific Discovery

Present: black hole and neutron star collisions

Future: supernovae, exotic objects...

Models and simulations

Big data analytics

Observations

Fusion of HPC & HTC, containers, OSG, LDG, CVMFS to distribute datasets

Open source software stacks for HPC numerical relativity simulations and gravitational wave discovery

Theory

\[ G_{\mu\nu} = 8\pi T_{\mu\nu} \]

Present: black hole and neutron star collisions

Future: supernovae, exotic objects…
The Royal Swedish Academy of Sciences has decided to award the
2017 NOBEL PRIZE IN PHYSICS

Rainer Weiss
Barry C. Barish
Kip S. Thorne

“for decisive contributions to the LIGO detector and the observation of gravitational waves”
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: $1 billion 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
Trends in simulation and **data driven science**

**The Big Data Revolution**

![Map showing the growth of the Big Data Revolution in 2018](image)

Source: IMF, Yahoo! Finance