Big Science with Big Instruments

Science at Petascale: Advancing Global Scientific Projects Through Distributed Computing

Frank Würthwein
OSG Executive Director
UCSD/SDSC

Abstract

We will provide an overview of three global science drivers, the Large Hadron Collider, gravitational wave detection, and the IceCube neutrino observatory, elucidating common patterns in the architecture of their global computing infrastructure.

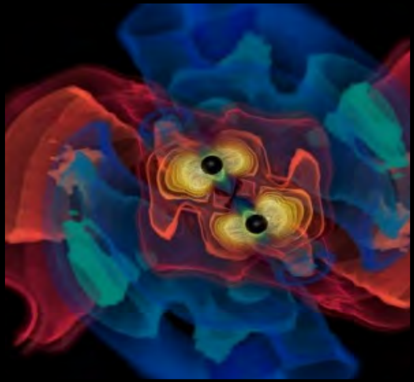
We outline the role AWS can play in these large international projects using advancements in computing technologies as their data volumes explodes over the next decade.

Disclaimer

- While I am an expert in experimental particle physics, doing my own research with the CMS experiment at the Large Hadron Collider.
- I am not much better than an educated lay person when it comes to the physics of LIGO, Virgo, … and IceCube.
- My group provides global compute and data integration for LIGO, Virgo, …, and IceCube.

The Science

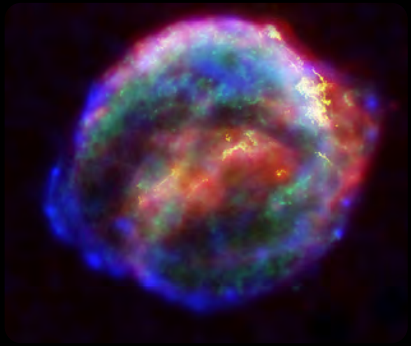
Gravitational Wave Astrophysics



Credit: AEI, CCT, LSU

Coalescing Binary Systems

Neutron Stars,
Black Holes



Credit: Chandra X-ray Observatory

'Bursts'

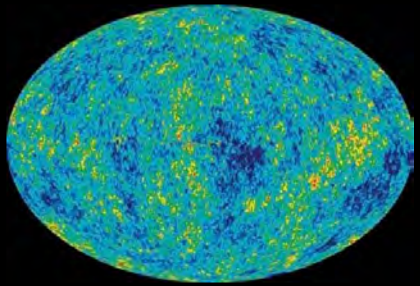
Core collapse supernovae
Cosmic strings
Unknown

Casey Reed, Penn State

Continuous Sources

Spinning neutron stars

crustal deformations, accretion



NASA/WMAP Science Team

Stochastic GW background

cosmological or astrophysical

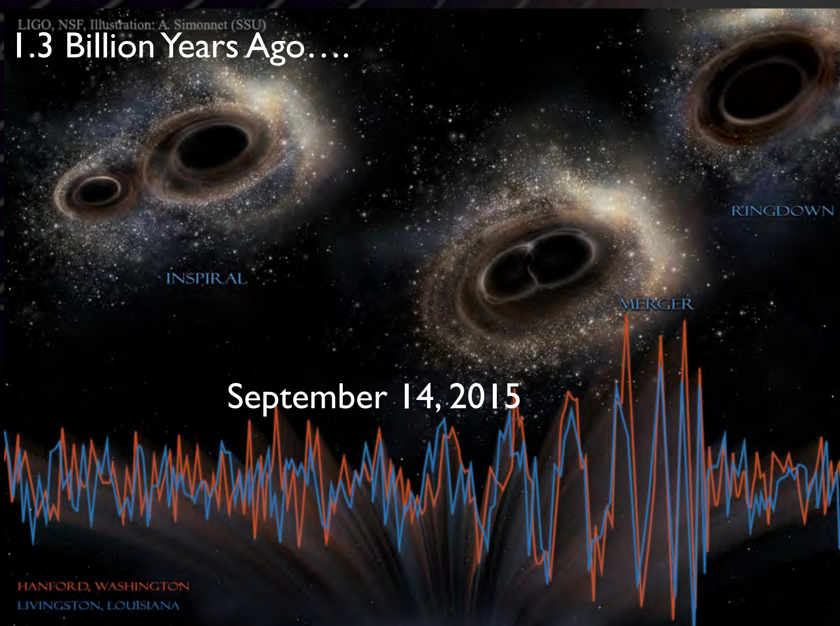
Broadly speaking, 4 different types of wave phenomena are searched for.

Coalescing Binary Detections

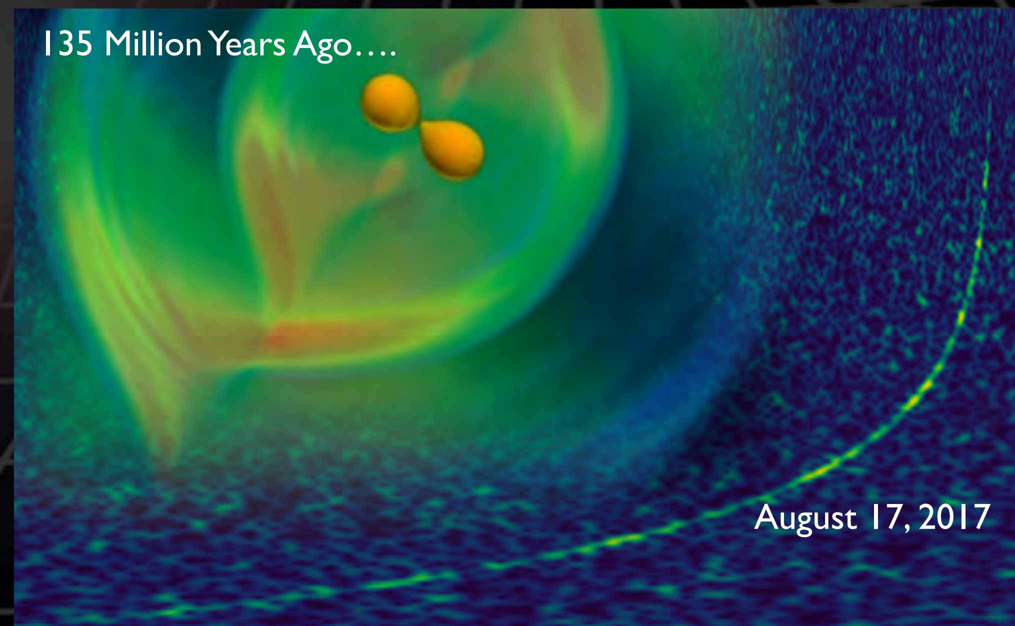
Open Science Grid

UCSD

GW150914 and GW170817:
Two ground-breaking discoveries that opened a new era in Gravitational Wave Astronomy



Binary Black Hole Coalescence

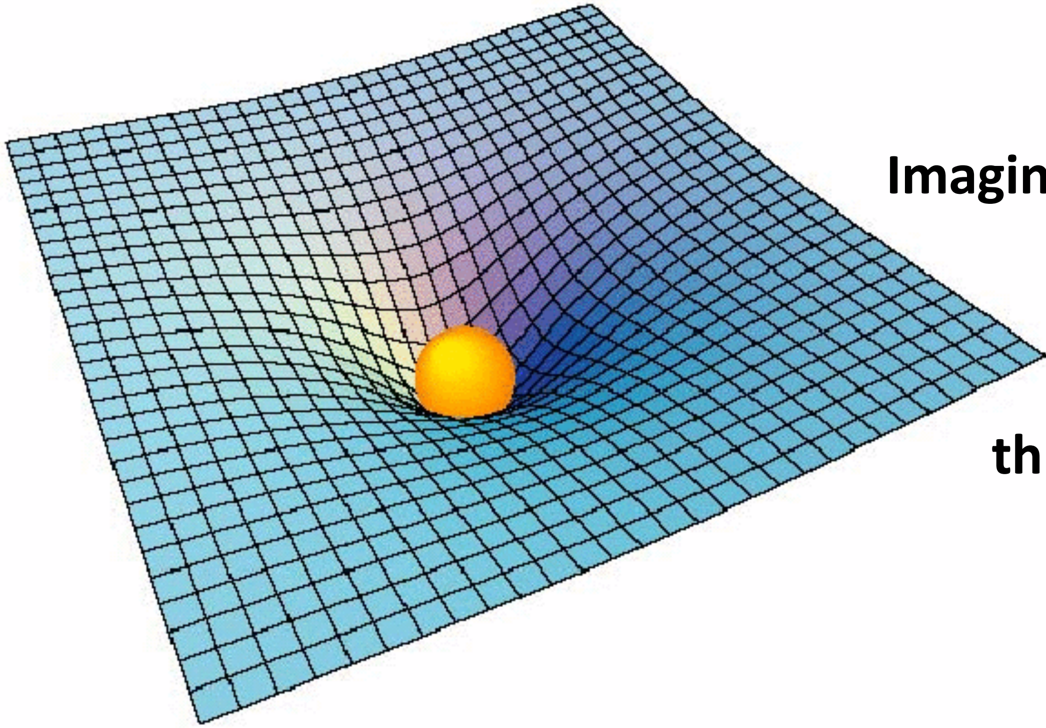


Binary Neutron Star Coalescence

Both come from just one of the 4 signatures that are being searched for.

Intro Physics for gravitational waves

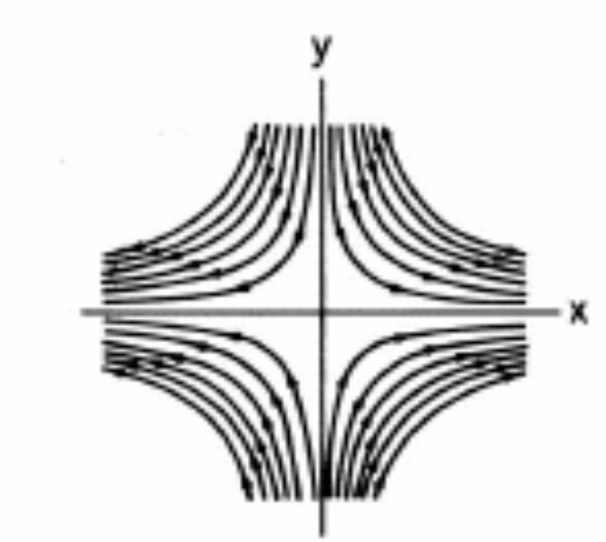
Making gravity waves



Large masses (yellow) can distort the shape of space-time (blue fabric)

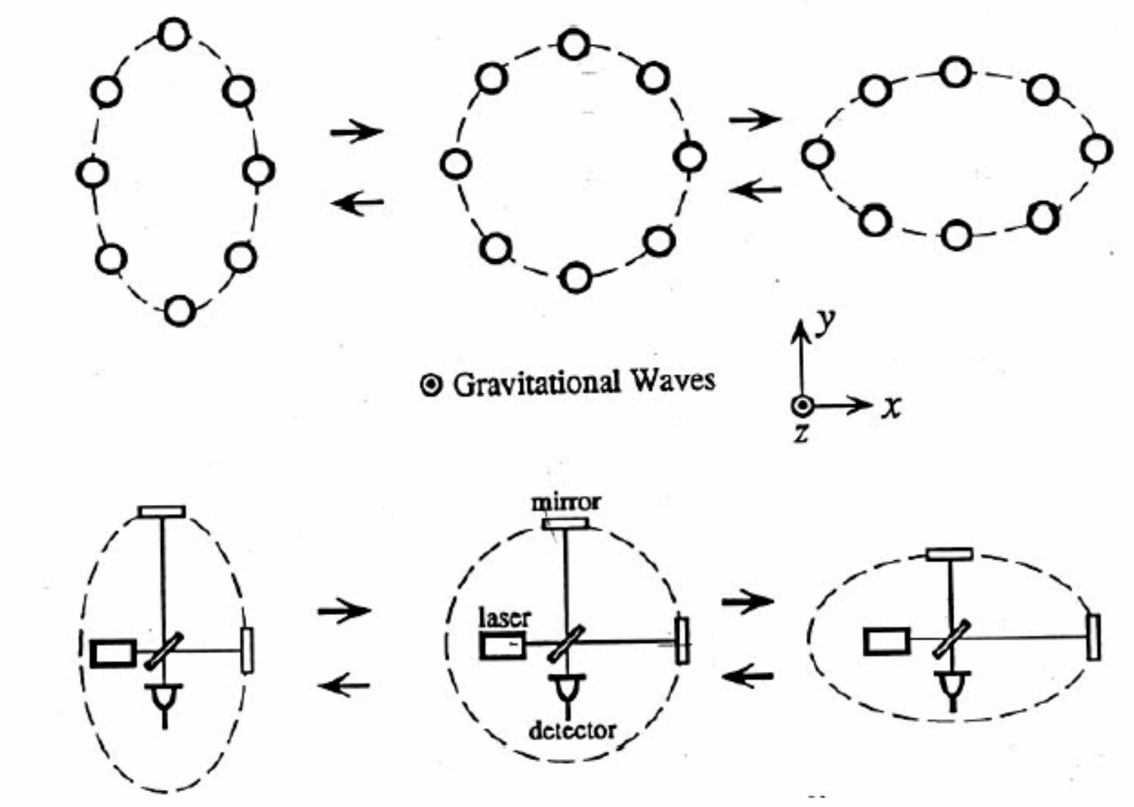
Imagine the yellow mass to explode. As it explodes, the distortion of space time disappears, and a wave ripple propagates through space like a wave on water.

Gravity waves



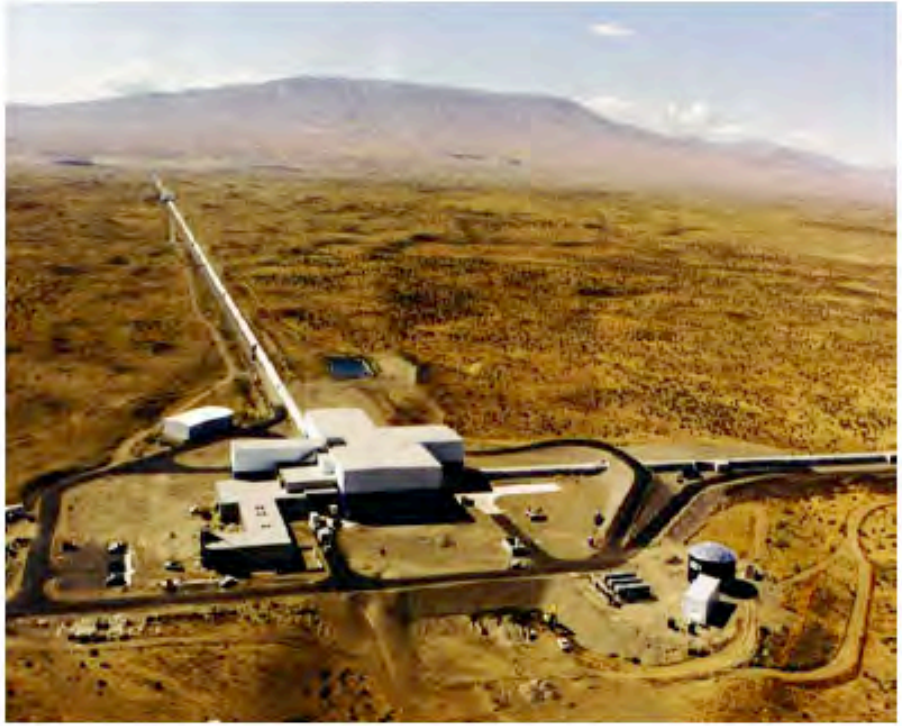
A gravity wave passing you stretches space.
Think of it as an oscillatory pulling of a square cloth along its two diagonals.

Gravity wave detection



A gravity wave passing you stretches space.
This is measured as a distance difference between the two arms of the Michelson interferometer.

LIGO Observatory Facilities



LIGO Hanford Observatory [LHO]

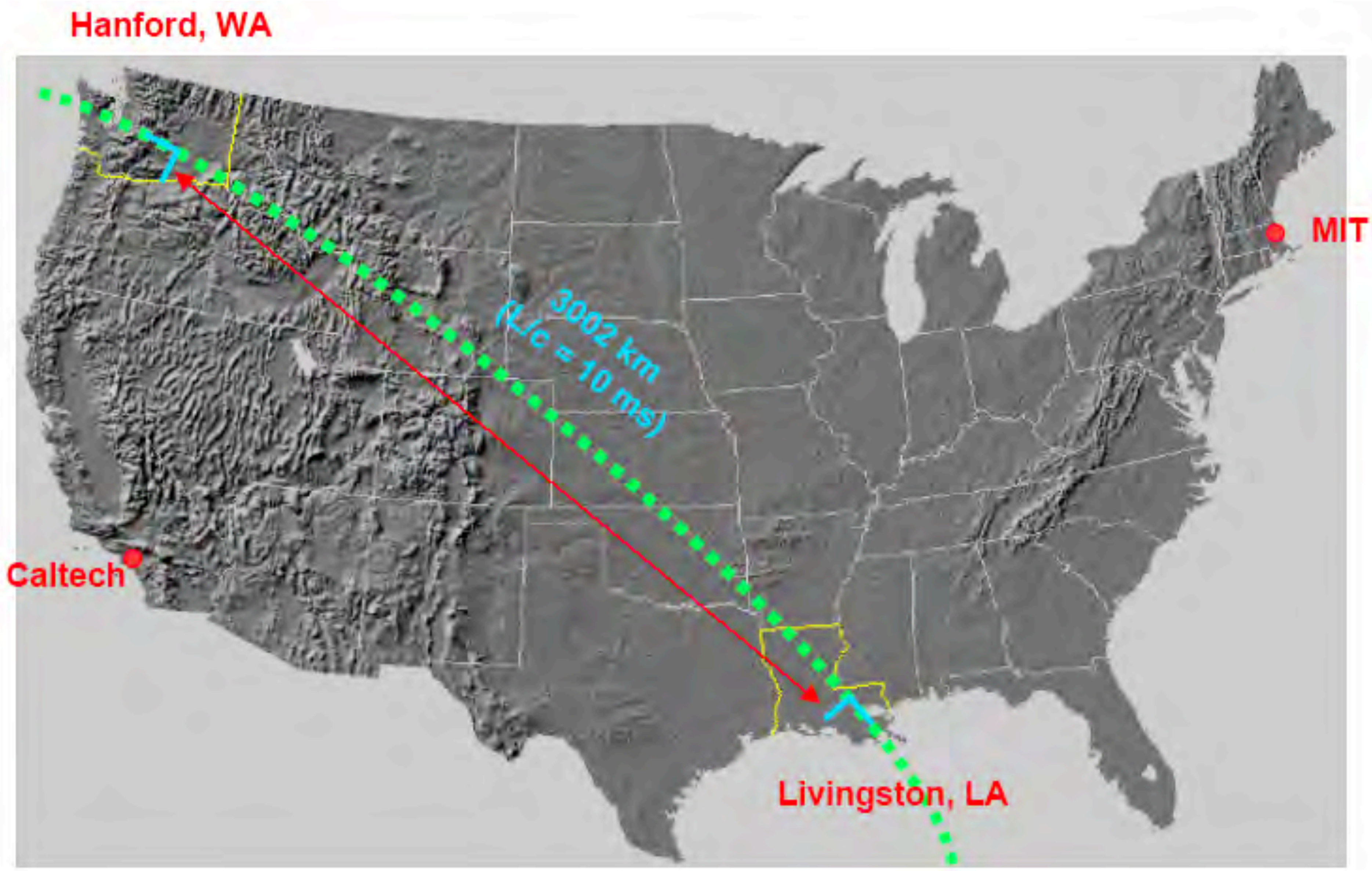
26 km north of Richland, WA

2 km + 4 km interferometers in same vacuum envelope

LIGO Livingston Observatory [LLO]

42 km east of Baton Rouge, LA

Single 4 km interferometer



In addition, there’s another interferometer in Italy, called VIRGO, and more are being built in Japan, India, ...

The Nobel Prize in Physics 2017

Photo: Bryce Vickmark
Rainer Weiss
Prize share: 1/2

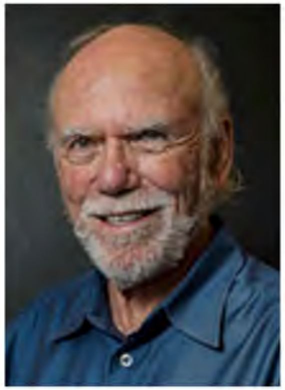
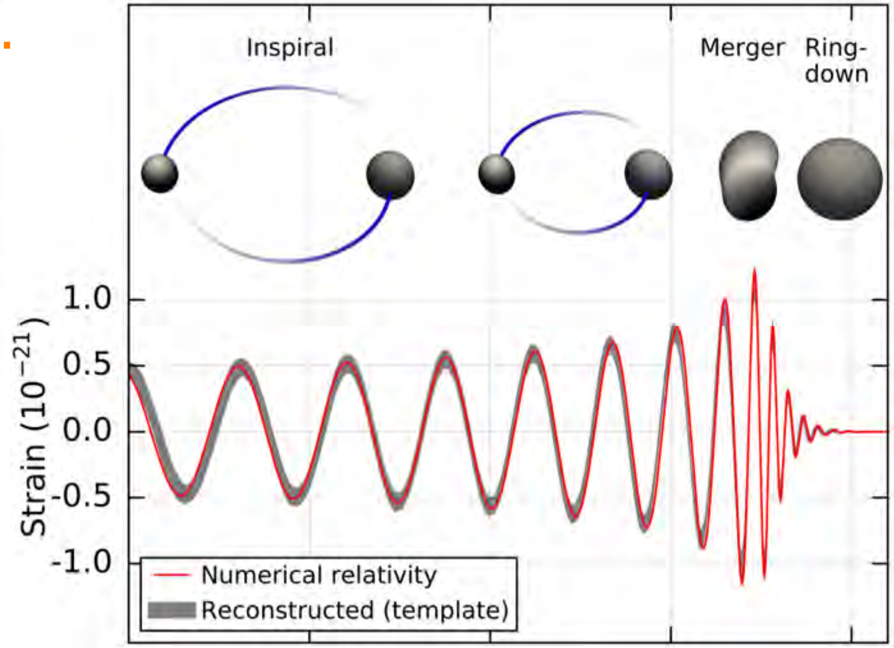


Photo: Caltech
Barry C. Barish
Prize share: 1/4

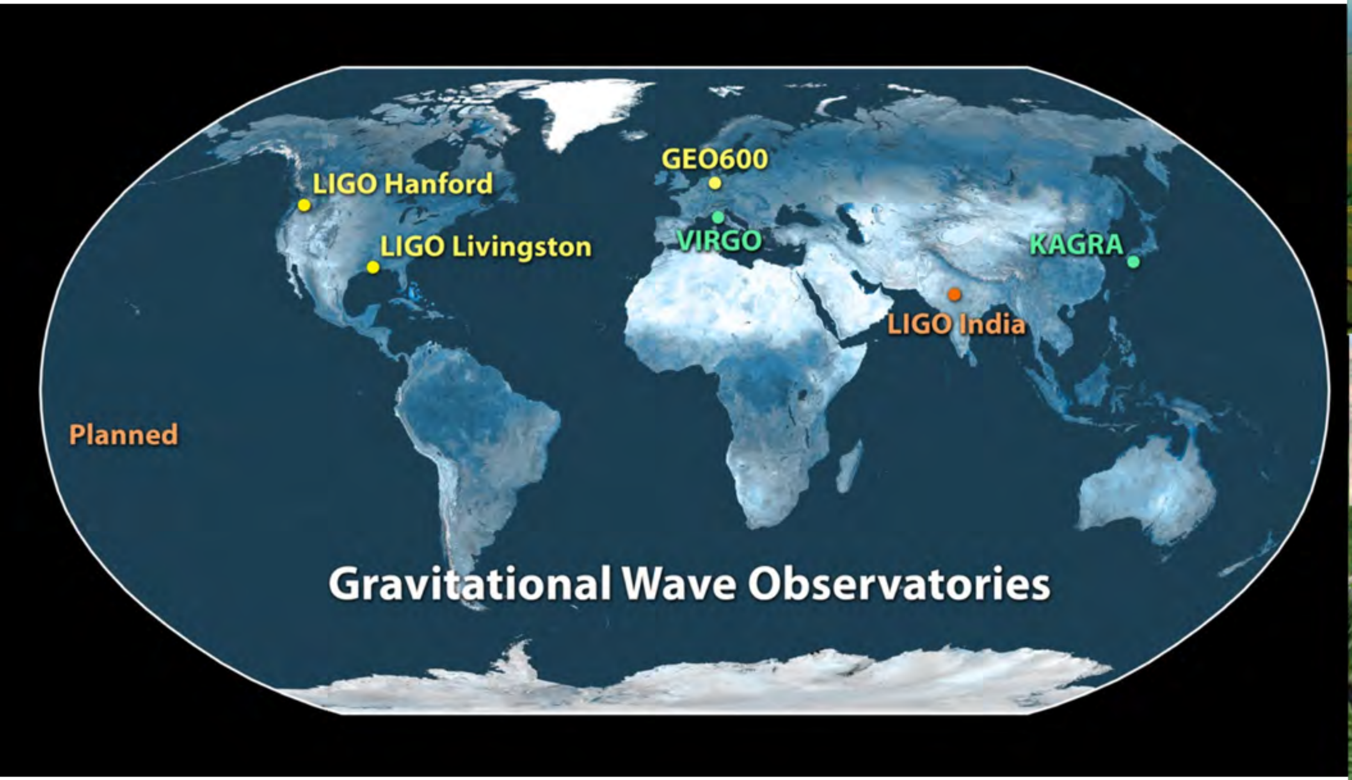
Photo: Caltech Alumni Association
Kip S. Thorne
Prize share: 1/4

The Nobel Prize in Physics 2017 was divided, one half awarded to Rainer Weiss, the other half jointly to Barry C. Barish and Kip S. Thorne *"for decisive contributions to the LIGO detector and the observation of gravitational waves"*.



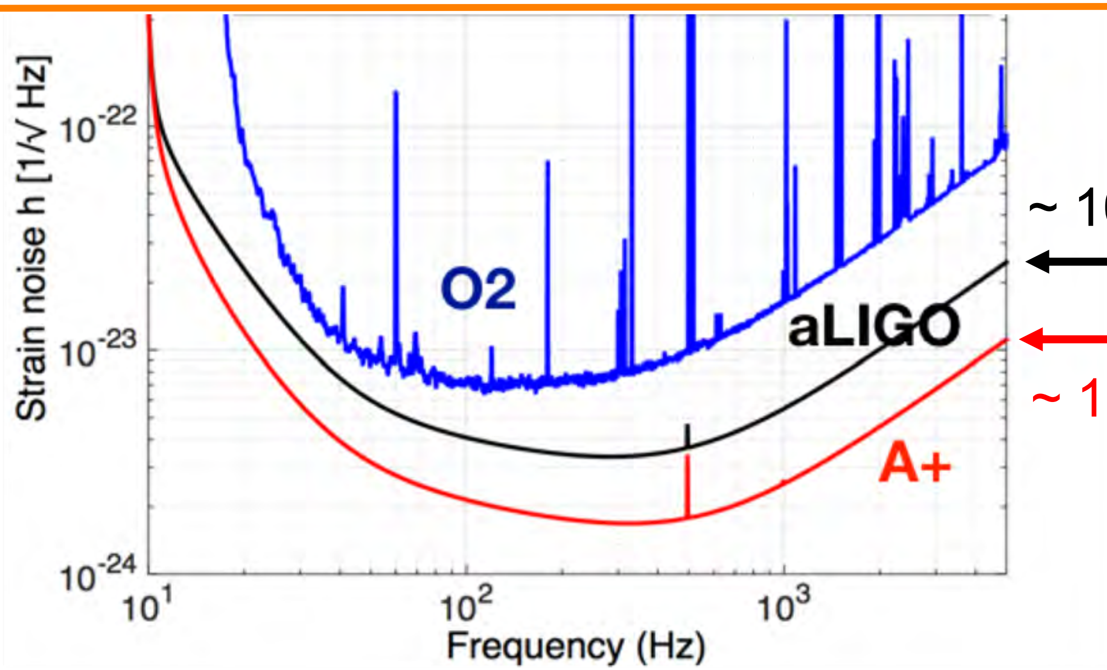
Since the first detection in 2015, LIGO has seen a dozen waves

Multiple Instruments Working Together



A Global Quest

Evolution of Strain Sensitivity

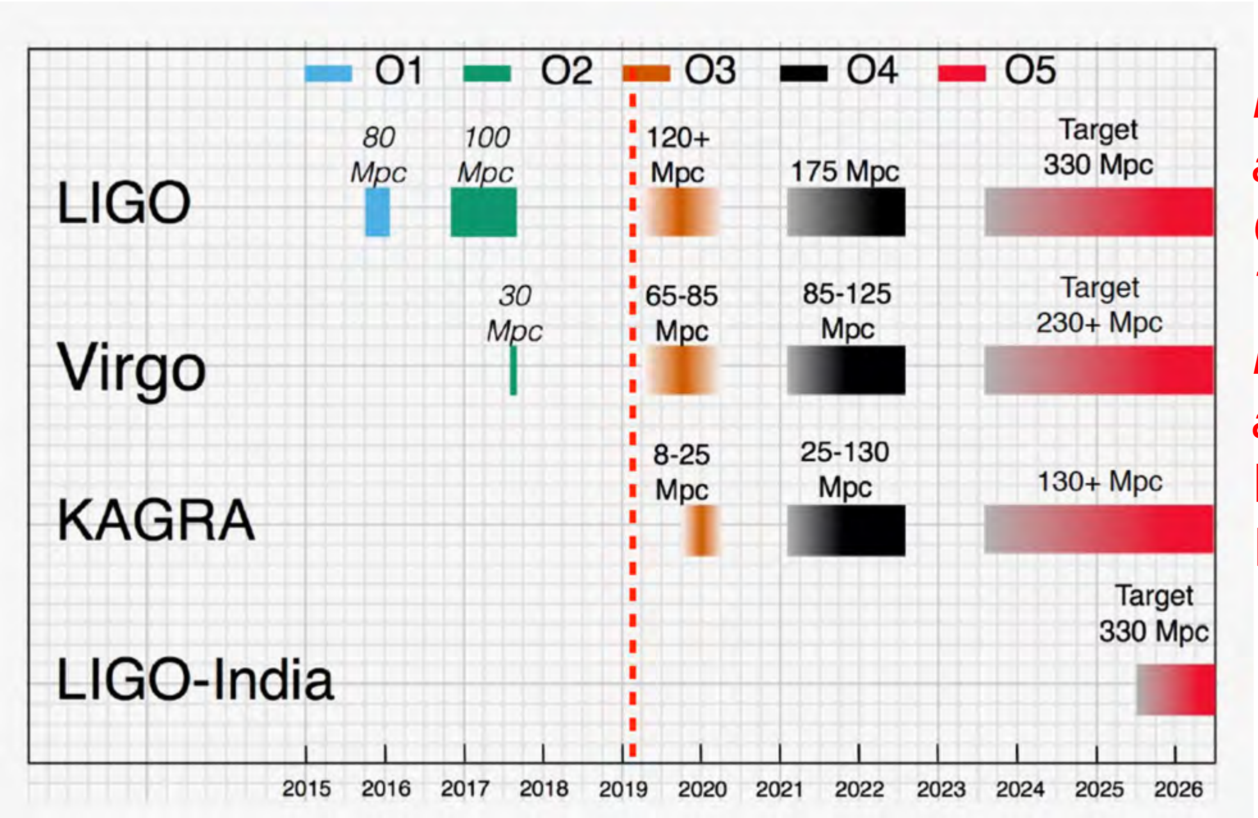


~ 10^2 binary coalescence per year (2021)

~ 10^3 binary coalescence per year (~2024)

Rapid Reduction of Strain Noise within the next few years leads to massive increase in detections because of R^3.

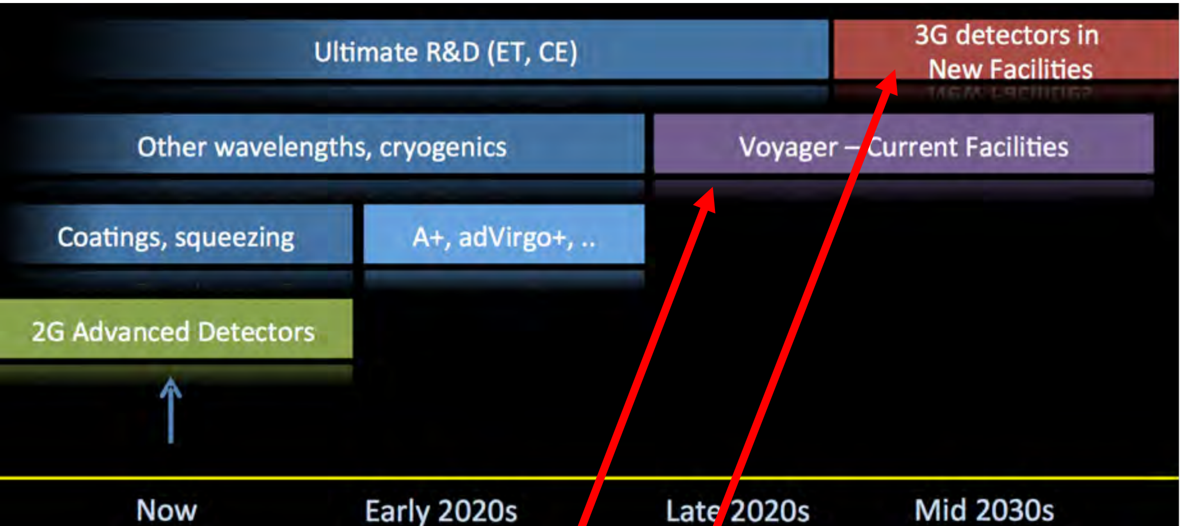
Near Term Observing Plan

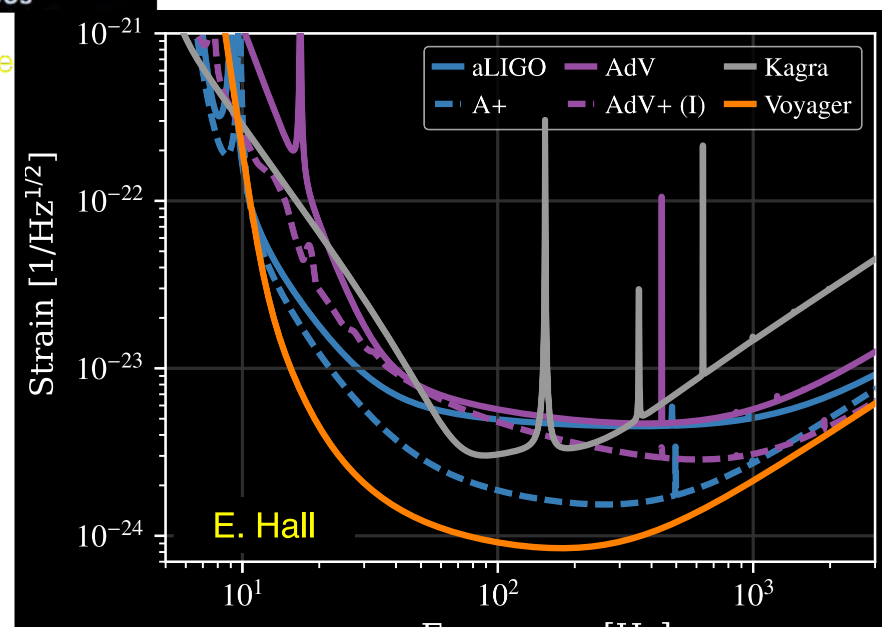


Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO and Advanced Virgo and KAGRA — https://dcc.ligo.org/LIGO-P1200087/public

Number of detected compact binary mergers ~ R^3

Longer Term Roadmap

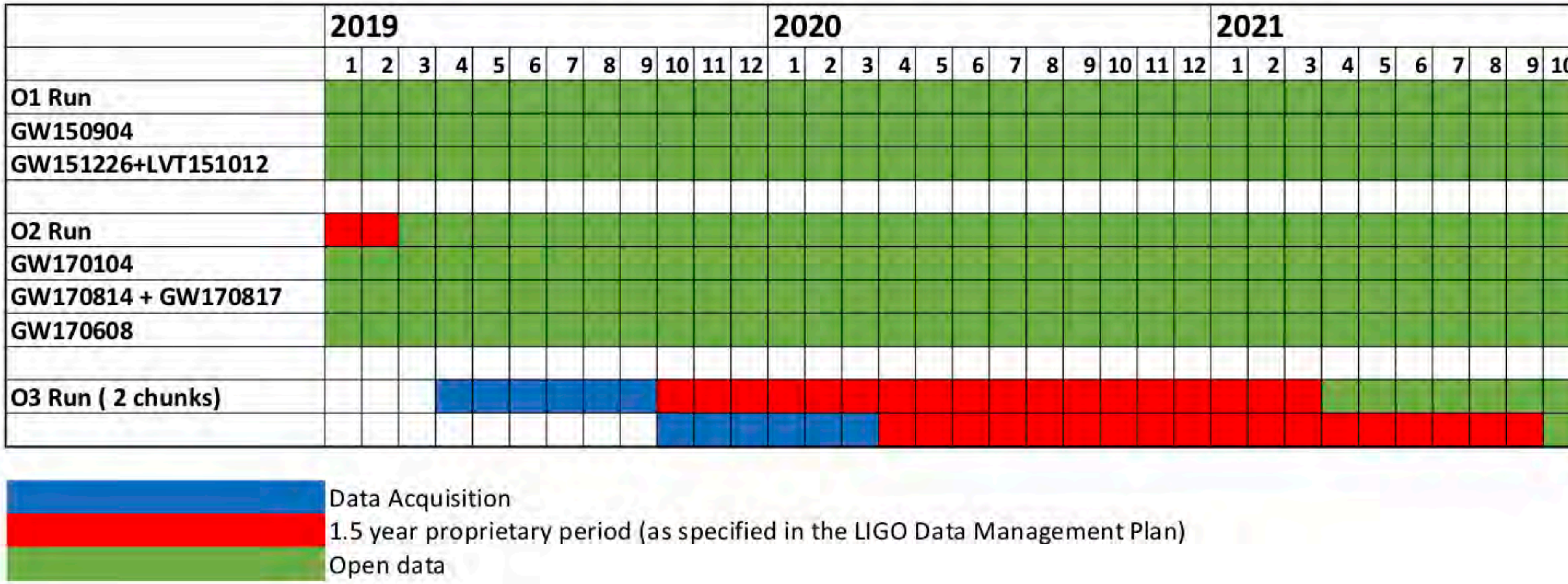




Voyager ~ 10^4 binary coalescence/year

3[rd] gen ~ 10^5 binary coalescence/year

Commitment to Open Data

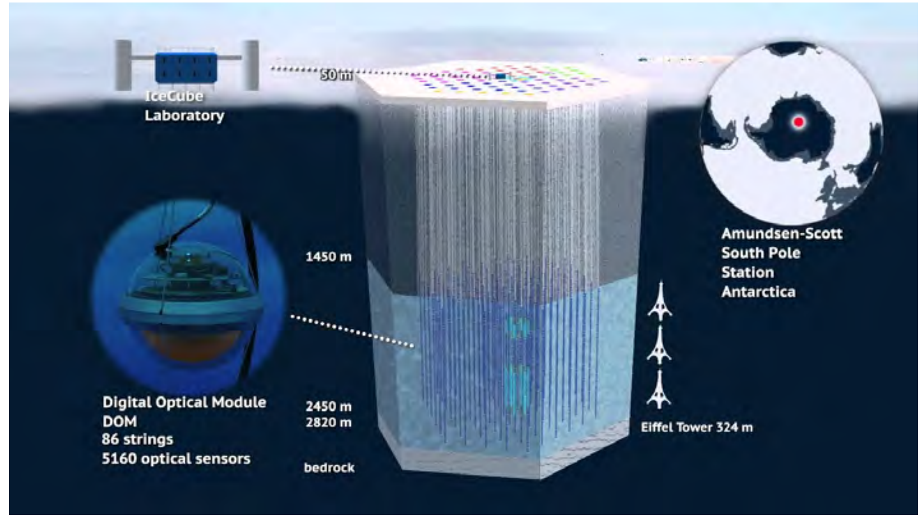


A community of scientists that exploit the open data is slowly emerging.

Also, starting with O3: Open Public Alerts.
https://www.ligo.org/scientists/GWEMalerts.php

IceCube

A cubic kilometer of ice at the south pole is instrumented with 5160 optical sensors.



A facility with very diverse science goals

Astrophysics:

- Discovery of astrophysical neutrinos
- First evidence of neutrino point source (TXS)
- Cosmic rays with surface detector

Particle Physics:

- Atmospheric neutrino oscillation
- Neutrino cross sections at TeV scale
- New physics searches at highest energies

Earth Science:

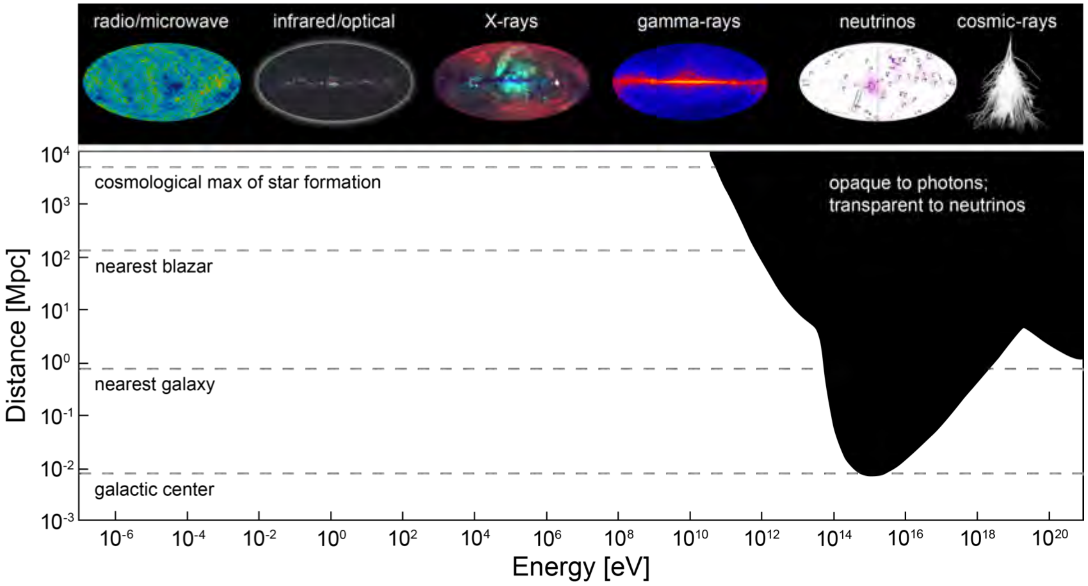
- Glaciology
- Earth tomography

Restrict this talk to high energy Astrophysics

High Energy Astrophysics Science case for IceCube

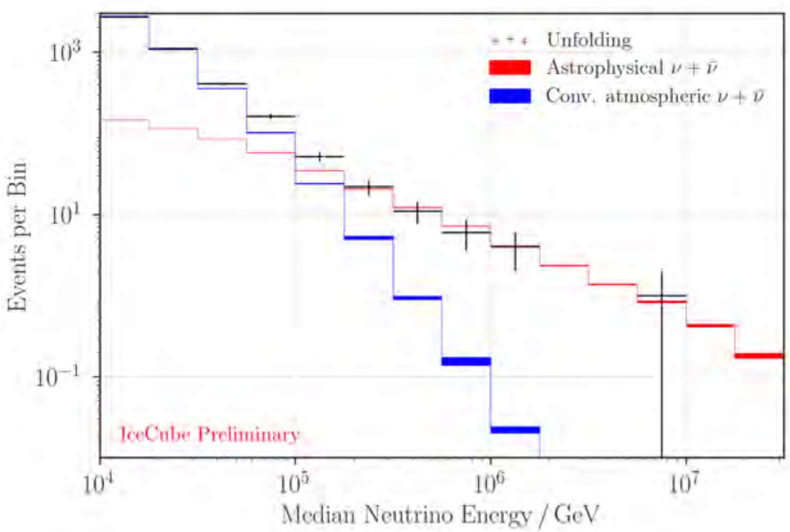
Open Science Grid

UCSD



Universe is opaque to light at highest energies and distances.

Only gravitational waves and neutrinos can pinpoint most violent events in universe.

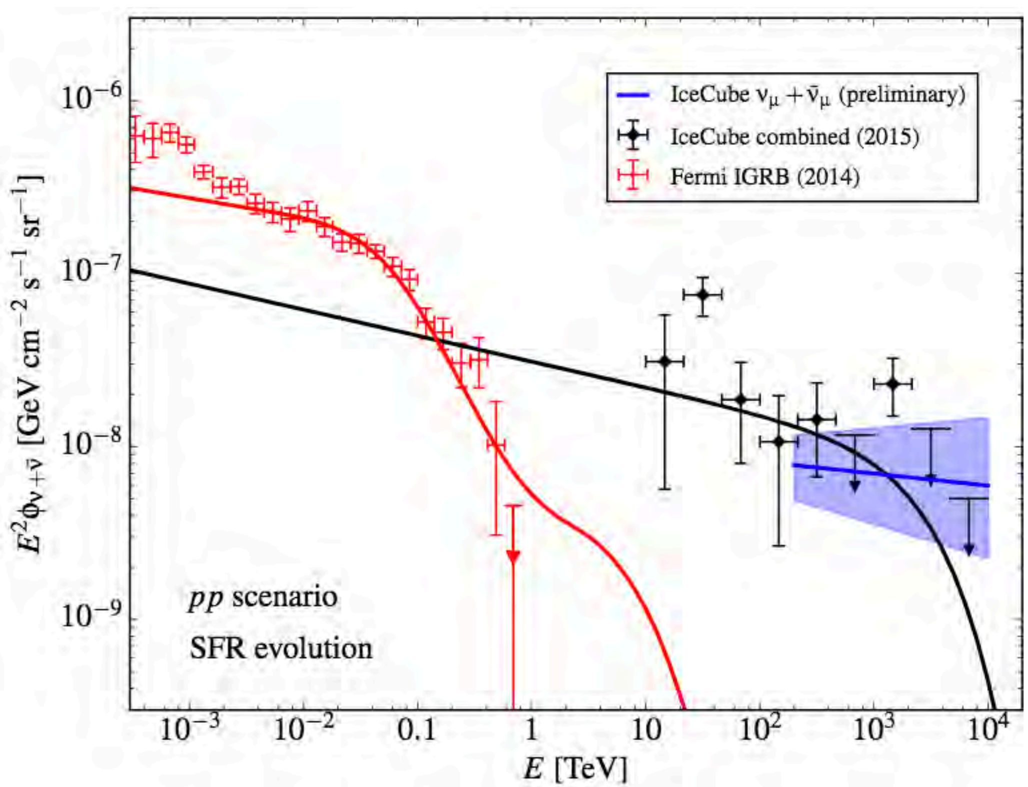


Fortunately, highest energy neutrinos are of cosmic origin.

Effectively "background free" as long as energy is measured correctly.

High energy neutrinos from outside the solar system

First 28 very high energy neutrinos from outside the solar system



Science 342 (2013). [DOI: 10.1126/science.1242856]

Red curve is the photon flux spectrum measured with the Fermi satellite.

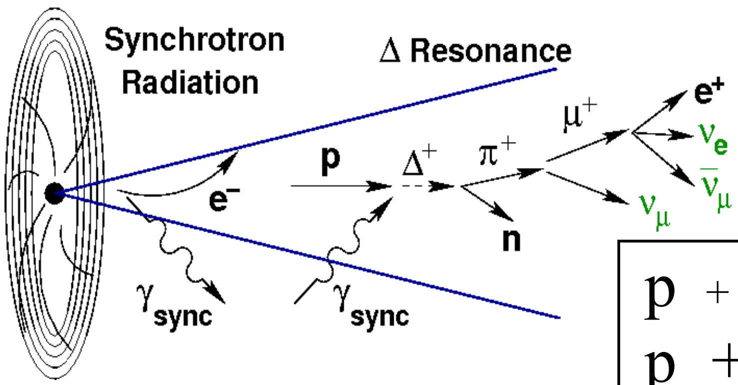
Black points show the corresponding high energy neutrino flux spectrum measured by IceCube.

This demonstrates both the opaqueness of the universe to high energy photons, and the ability of IceCube to detect neutrinos above the maximum energy we can see light due to this opaqueness.

Understanding the Origin

We now know high energy events happen in the universe. What are they?

The hypothesis:



The same cosmic events produce neutrinos and photons

$$p + \gamma \rightarrow \Delta^{+} \rightarrow p + \pi^{0} \rightarrow p + \gamma\gamma$$
$$p + \gamma \rightarrow \Delta^{+} \rightarrow n + \pi^{+} \rightarrow n + \mu + \nu$$

We detect the electrons or muons from neutrino that interact in the ice.
Neutrino interact very weakly => **need a very large array of ice instrumented** to maximize chances that a cosmic neutrino interacts inside the detector.
Need pointing accuracy to point back to origin of neutrino.
Telescopes the world over then try to identify the source in the are IceCube is pointing to for the neutrino.

Multi-messenger Astrophysics

The pointing resolution Challenge

Observed pointing resolution at high energies systematics limited.

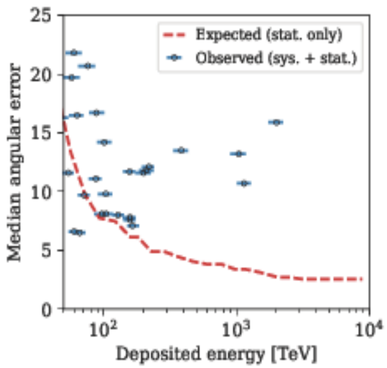


Figure 2 Systematic effect on angular resolution on electron and tau neutrinos and neutral-current muon neutrino events.

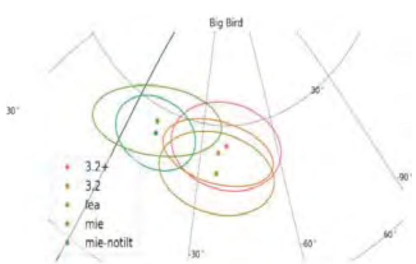


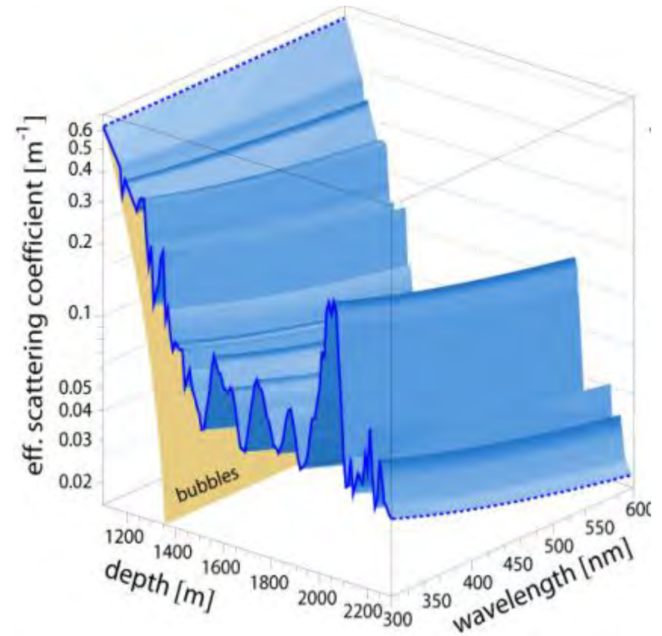
Figure 3. Reconstructed position on the sky for 2 PeV event with various generations of ice model

Central value moves for different ice models

Photon propagation through ice runs efficiently on single precision GPU.

Detailed simulation campaigns to improve pointing resolution by improving ice model.

Ice properties change with depth and wavelength



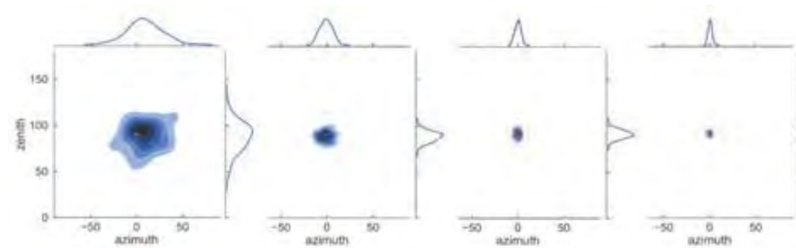
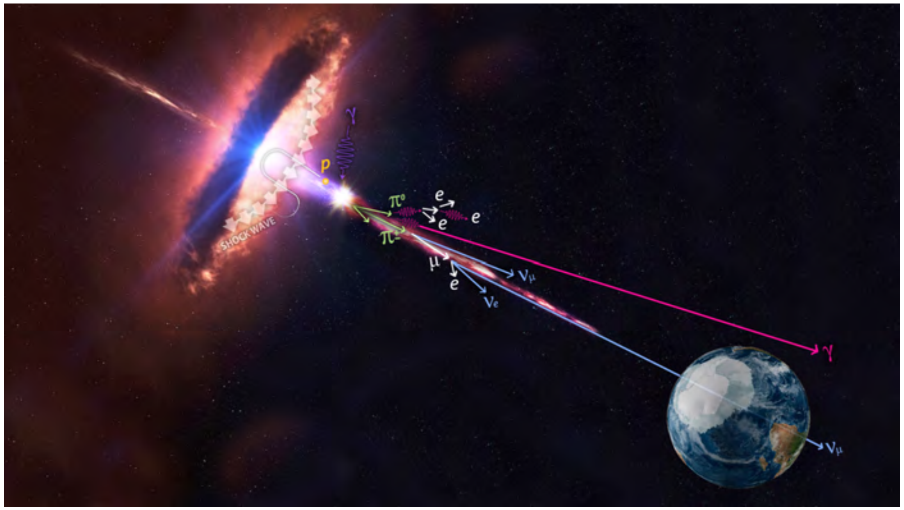


Figure 4. Possible improvements to angular resolution with better understanding of ice model

First evidence of an origin

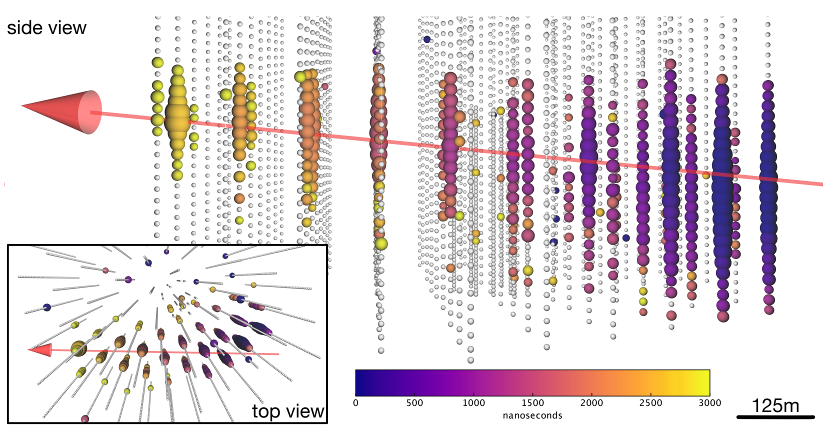


IceCube alerted the astronomy community of the observation of a single high energy neutrino on September 22 2017.

A blazar designated by astronomers as TXS 0506+056 was subsequently identified as most likely source in the direction IceCube was pointing. Multiple telescopes saw light from TXS at the same time IceCube saw the neutrino.

Science 361, 147-151 (2018). DOI:10.1126/science.aat2890

First location of a source of very high energy neutrinos.



Neutrino produced high energy muon near IceCube. Muon produced light as it traverses IceCube volume. Light is detected by array of phototubes of IceCube.

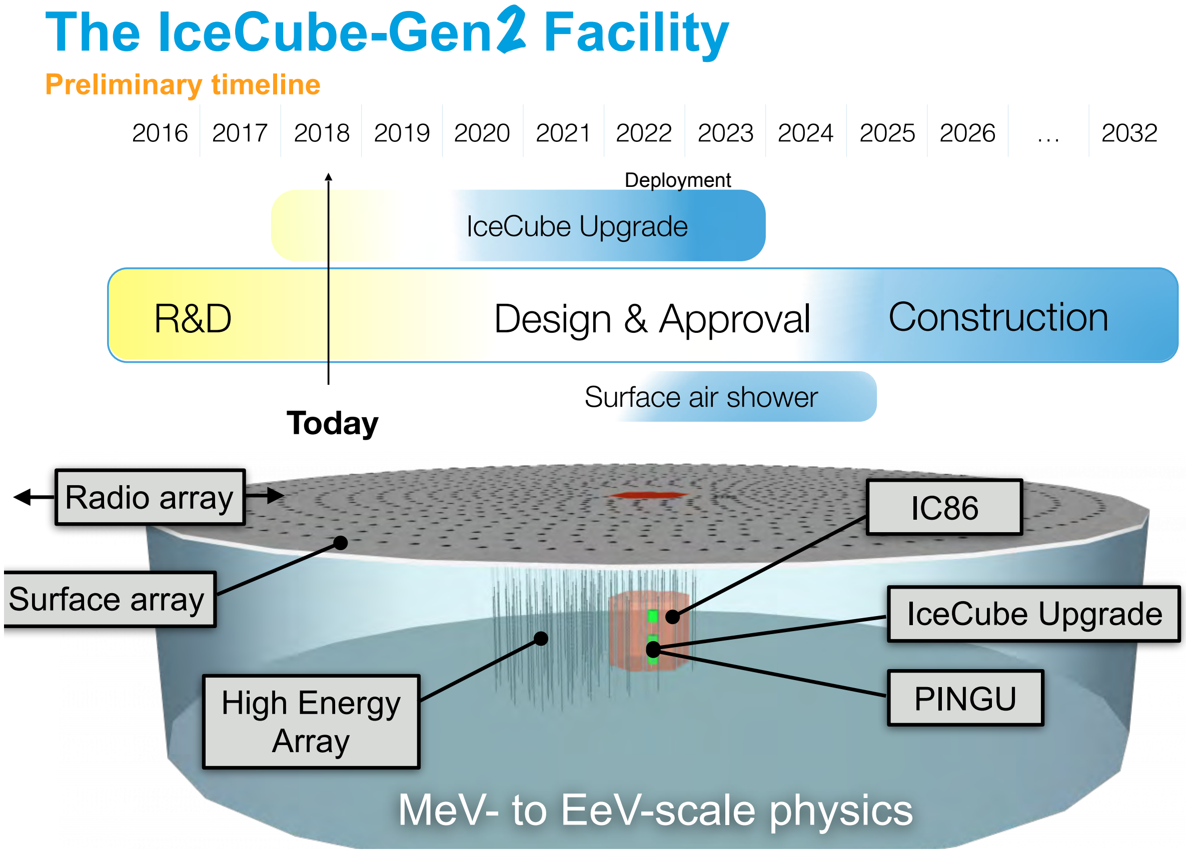
IceCube's Future Plans

Near term:

add more phototubes to deep core to increase granularity of measurements.

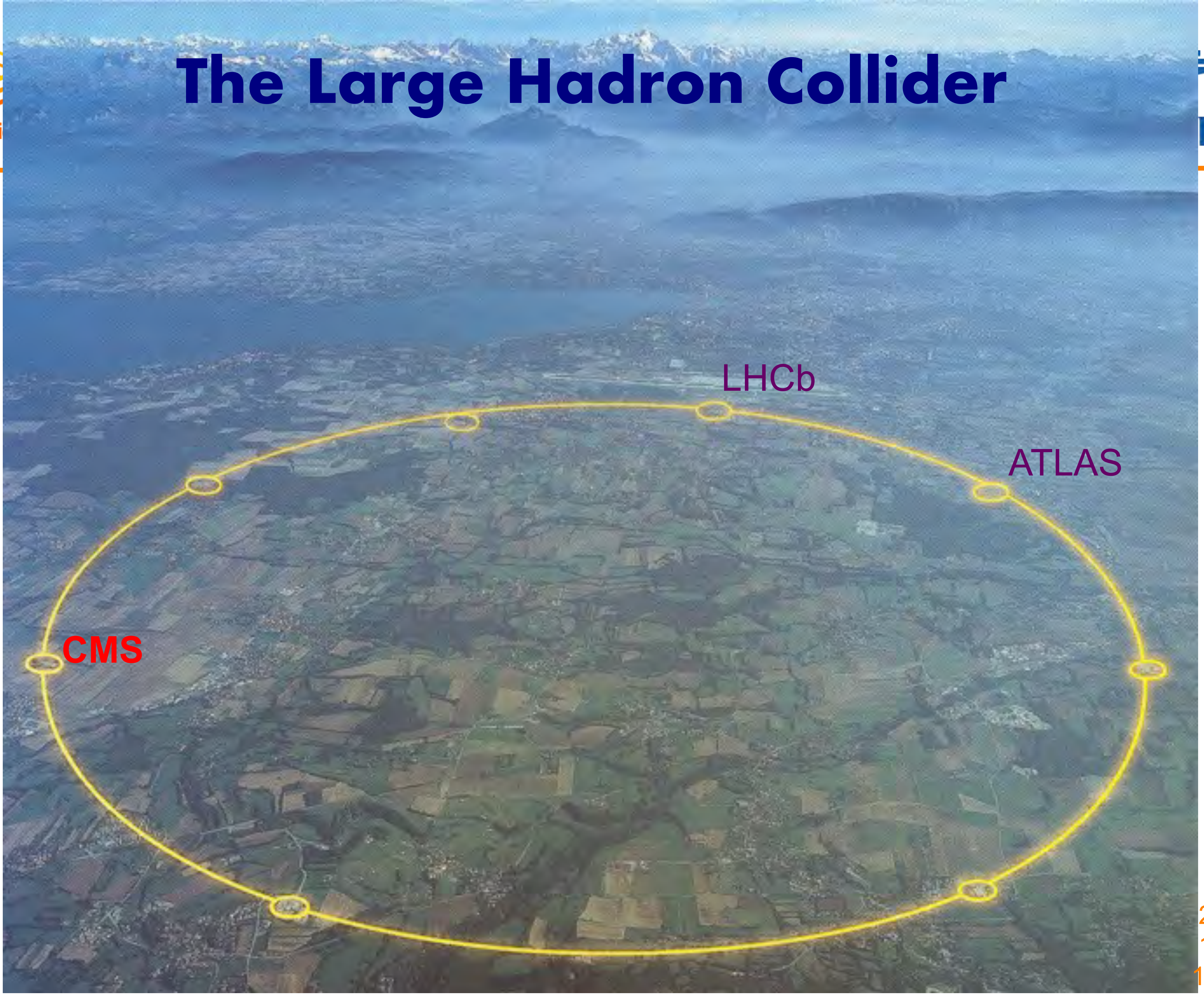
Longer term:

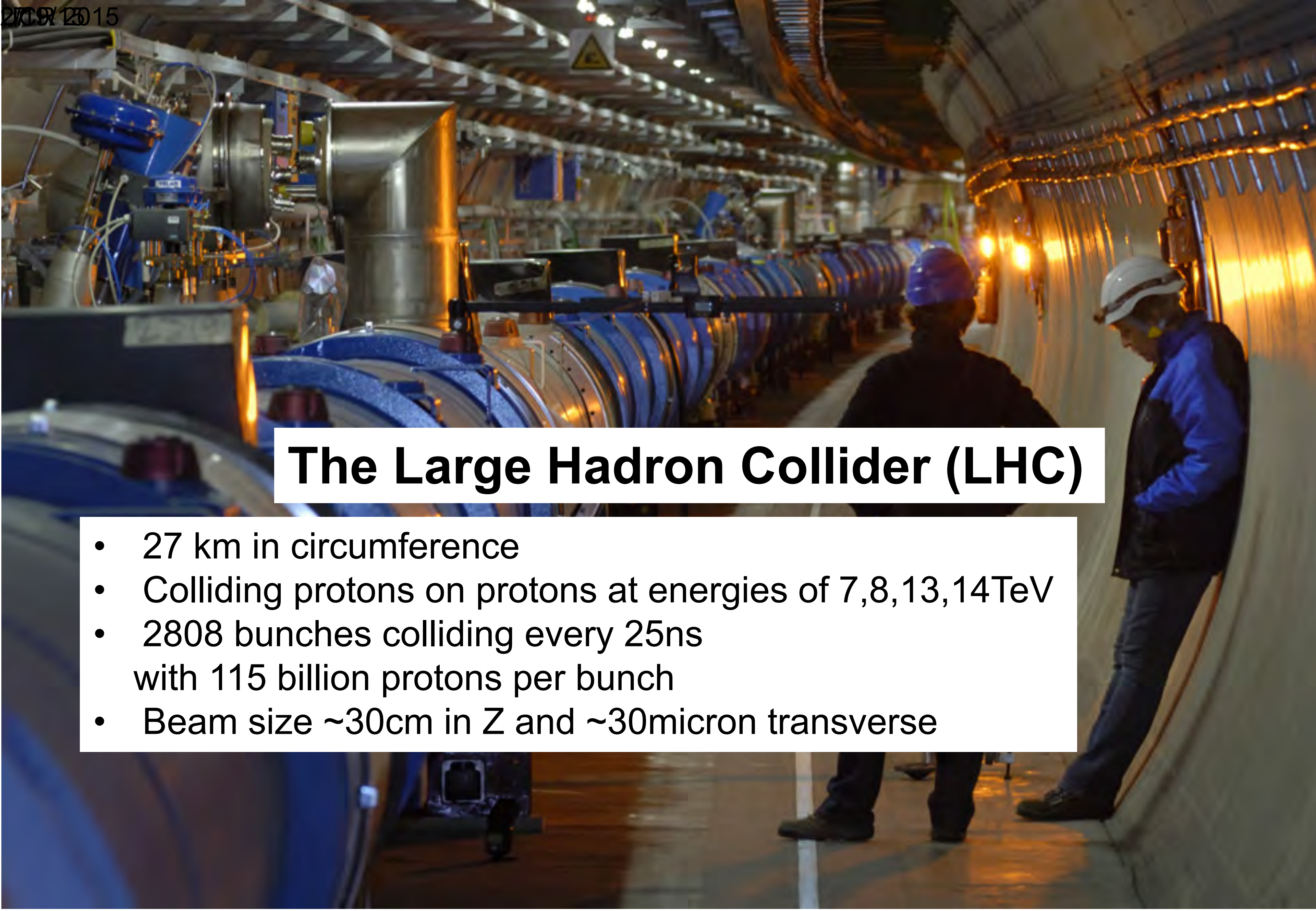
- Extend instrumented volume at smaller granularity.
- Extend even smaller granularity deep core volume.
- Add surface array.



Improve detector for low & high energy neutrinos

The Large Hadron Collider





The Large Hadron Collider (LHC)

- 27 km in circumference
- Colliding protons on protons at energies of 7,8,13,14TeV
- 2808 bunches colliding every 25ns
  with 115 billion protons per bunch
- Beam size ~30cm in Z and ~30micron transverse

Open Science Grid

"Big bang" in the laboratory

UCSD

- We gain insight by *colliding protons at the highest energies* possible to measure:
  - Production rates
  - Masses & lifetimes
  - Decay rates
- From this we *derive the "spectroscopy" as well as the "dynamics" of elementary particles*.
- Progress is made by going to higher energies and more proton proton collisions per beam crossing.
  - More collisions => increased sensitivity to rare events
  - More energy => probing higher masses, smaller distances & earlier times

Spectroscopy and Dynamics

- Spectroscopy:
  - What are the particles that exist ?
  - What are their properties ?
- Dynamics:
  - What are the forces ?
  - How do the particles couple to the forces ?
  - How do these depend on energy and angular momentum ?

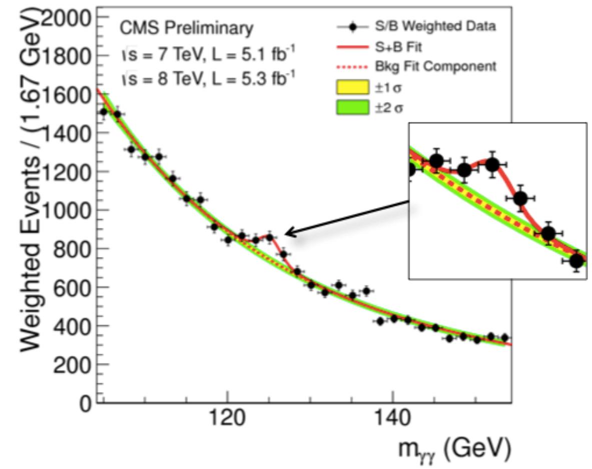
The Higgs Boson Discovery

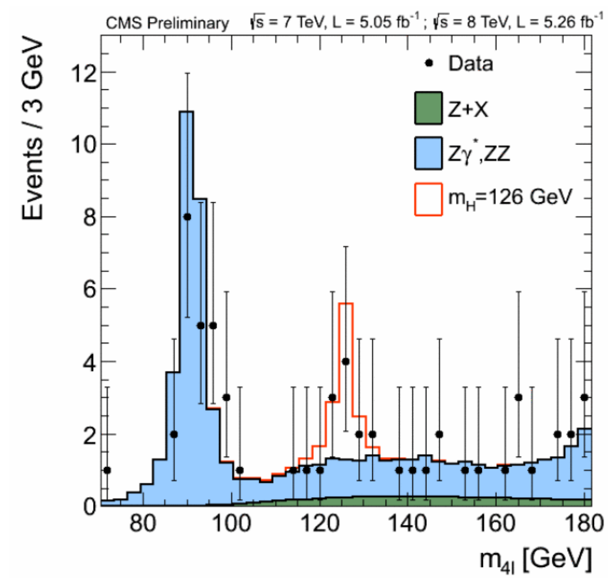
The Nobel Prize in Physics 2013

Photo: A. Mahmoud
François Englert
Prize share: 1/2

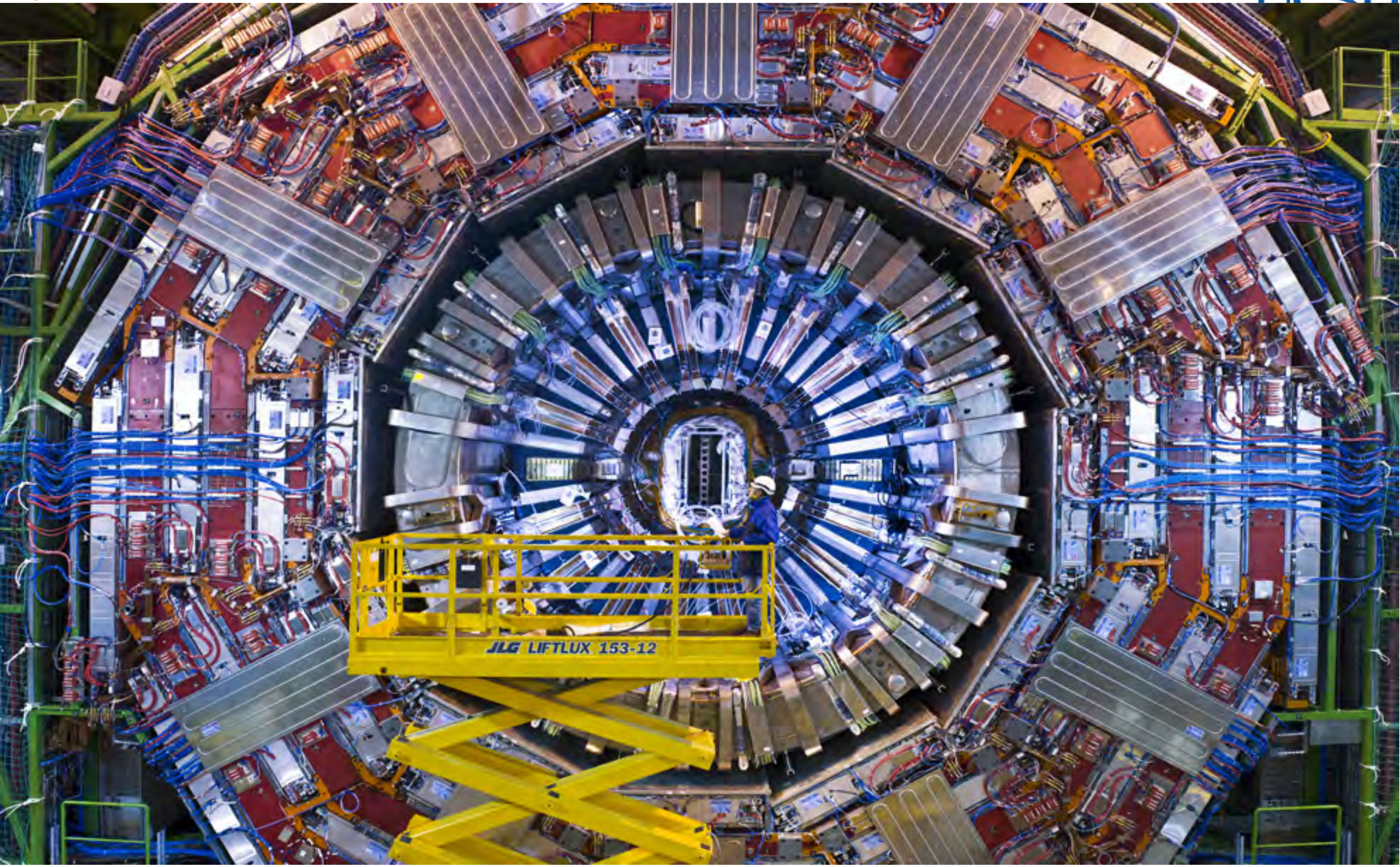
Photo: A. Mahmoud
Peter W. Higgs
Prize share: 1/2

The Nobel Prize in Physics 2013 was awarded jointly to François Englert and Peter W. Higgs *"for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider"*

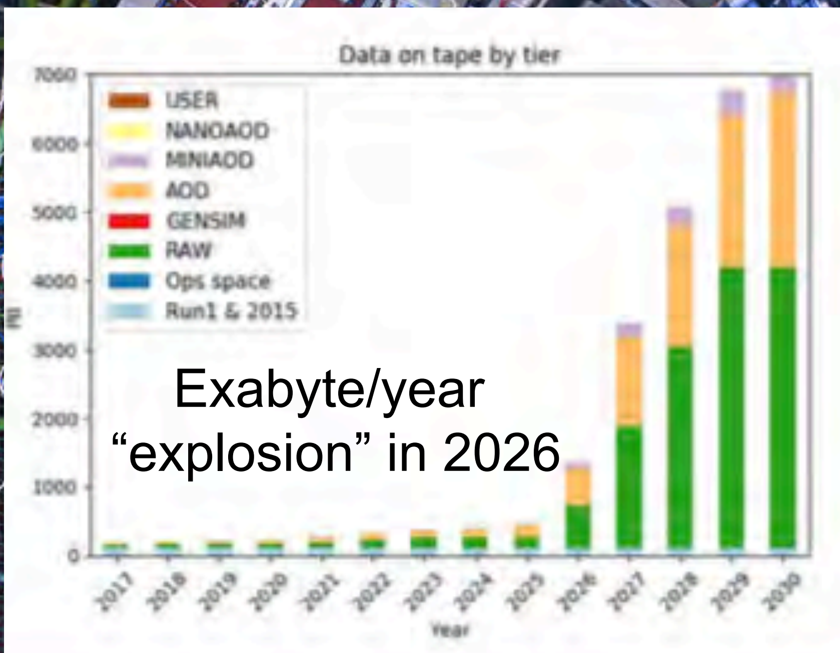




The CMS Experiment (R-Φ view)



The CMS Experiment



- **80 Million electronic channels**
  x 4 bytes
  x 40MHz

  ~ **10 Petabytes/sec** of information
  x 1/1000 zero-suppression
  x 1/100,000 online event filtering

  ~ 1000 Megabytes/sec raw data to tape
  ~10 Petabytes of raw data per year written to tape, not counting simulations.

- 2000 Scientists (1200 Ph.D. in physics)
  - ~ 180 Institutions
  - ~ 40 countries

- 12,500 tons, 21m long, 16m diameter

Commonality Across Experiments

Thousands of collaborators across 100's of institutions in 10's of countries

Open Science Grid

UCSD

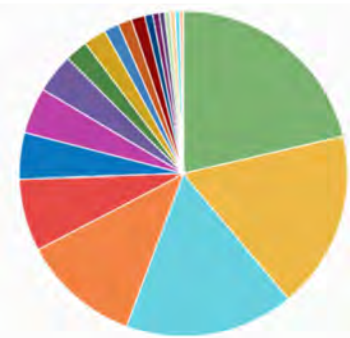
LIGO Scientific Collaboration

~ 1200 members ~ 100 institutions, 18 countries

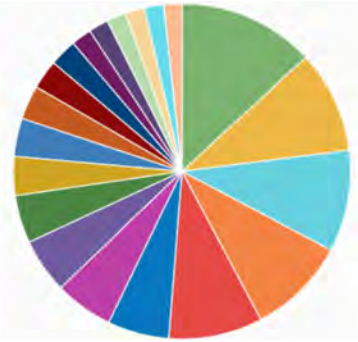
Global Computing & Data Infrastructure

Open Science Grid

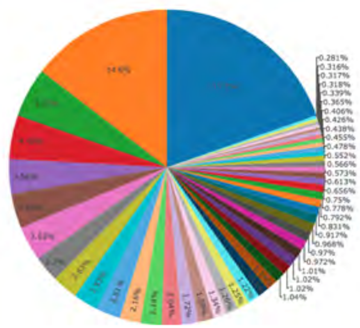
UCSD



LIGO computing



IceCube computing

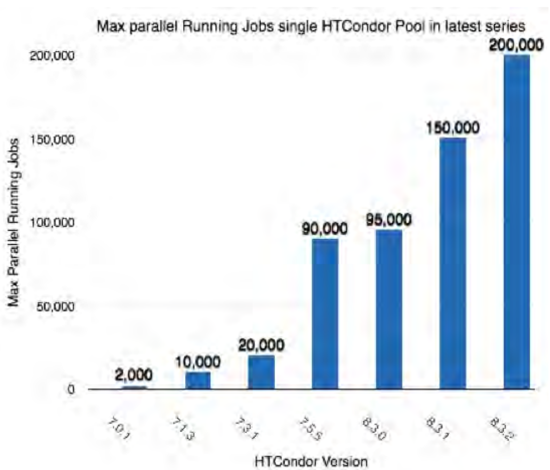


CMS data distribution

Many funding agencies contribute hardware in many places.

AWS integration is a natural extension of the globally integrated infrastructure

Applications are ingeniously parallel



Gedanken Experiment: (off-the-shelf hardware)
200k VMs of Dual 20 core CPUs = 8M CPU cores
200k VMs of 8 GPUs = 1.6M GPUs

Clearly, limits of scalability are very large, and depend on integrated capacity more than scalability of the underlying applications, or the middleware.

OSG Compute Federation

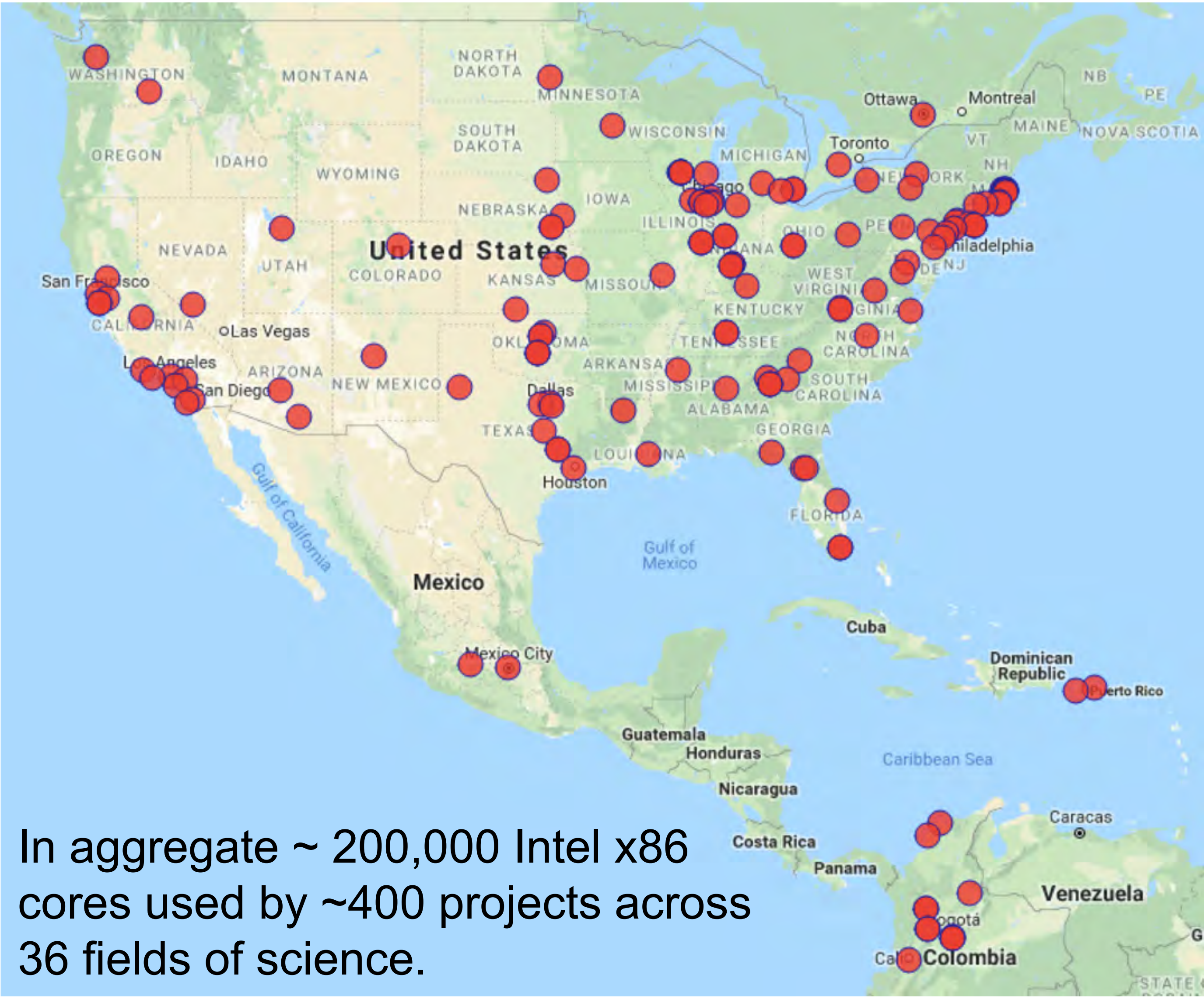
Open Science Grid

UCSD

OSG federates ~100 clusters worldwide

Owners determine policy of use.

Many allow opportunistic use of spare capacity.



In aggregate ~ 200,000 Intel x86 cores used by ~400 projects across 36 fields of science.

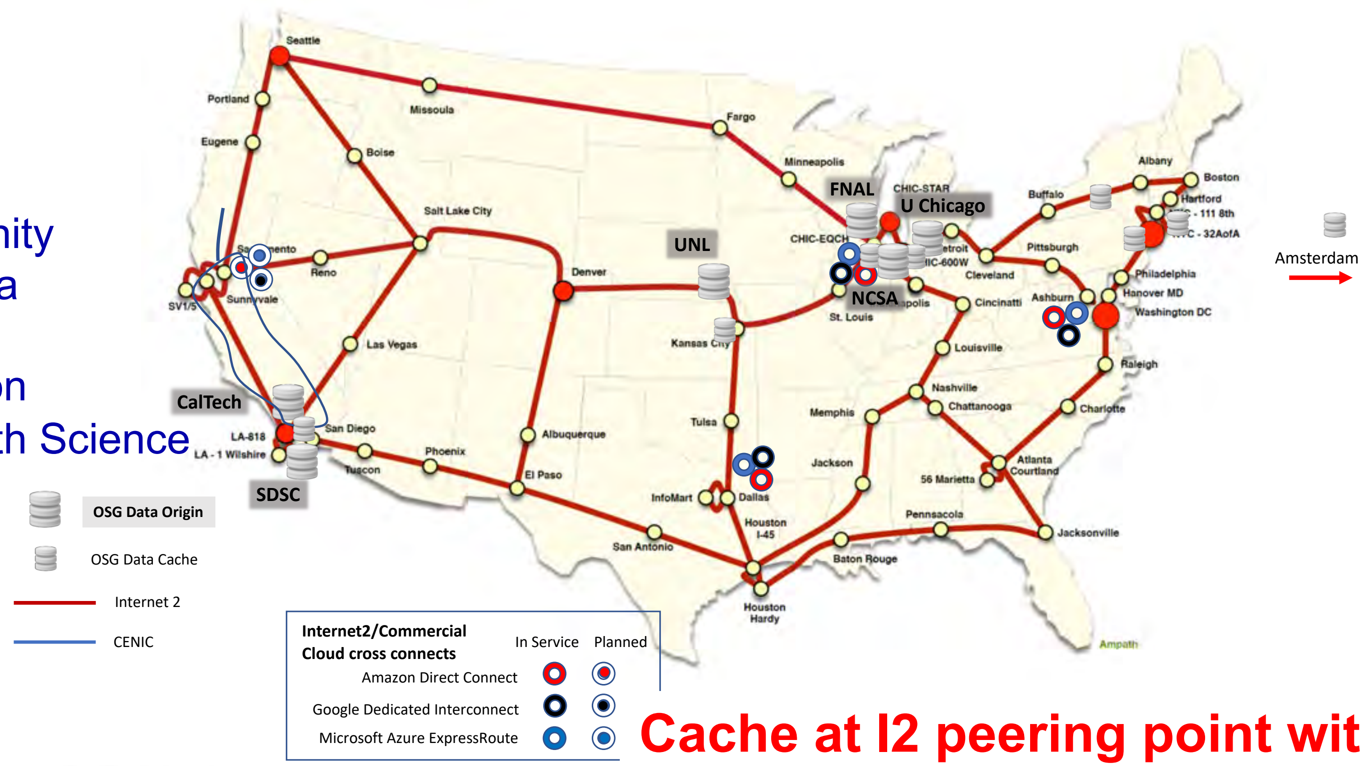
Open Science Grid

OSG Data Federation

UCSD

6 Data Origins
9 Data Caches

FNAL: HEP experiments
U.Chicago: OSG community
Caltech: Public LIGO Data
UNL: Private LIGO Data
SDSC: Simons Foundation
NCSA: DES & NASA Earth Science



Cache at I2 peering point with Cloud providers in Chicago

Directory	Working Set	Total Read
/pnfs/fnal.gov/usr/dune	13.107GB	395.537TB
/pnfs/fnal.gov/usr/minerva	255.266GB	270.994TB
/gwdata/O1	169.585GB	258.341TB
/pnfs/fnal.gov/usr/des	193.57GB	120.993TB
/user/ligo	5.612TB	83.564TB
/pnfs/fnal.gov/usr/nova	162.632GB	18.841TB

30k

1k

15

100

Depending on community, files were read 10-30,000 times during last 60 days.

Elasticity is valued

- IceCube and LIGO have “cosmic events” that are of high value, and elastic scale out to get best possible parameters quickly can be of value.
- IceCube and CMS are driven by annual simulation and/or reconstruction campaigns.
  - E.g. reprocessing data with better calibrations & reconstruction software to improve quality of data, and thus science reach.
  - Researchers analyze data when campaign is (mostly) complete. Tails are annoying, and delay science.
    - Scientists would much prefer shorter turn around time.
- Fundamentally, most of the processing is pre-emptable, thus perfect for spot market.

Open Science Grid

Summary & Conclusion

UCSD

- Humanity has built extraordinary instruments by pooling human and financial resources globally.
- The computing for these large collaborations fits perfectly to the cloud due to its "ingeniously parallel" nature.
- The scientists would much prefer elastically scaling out their operations in order to reduce turn around time by scaling out.
  - Run less time in the year but at a larger scale.

Contact us at: help@opensciencegrid.org
Or me personally at: fkw@ucsd.edu