# LIGO-Virgo Computing

Peter Couvares (LIGO Laboratory - Caltech)

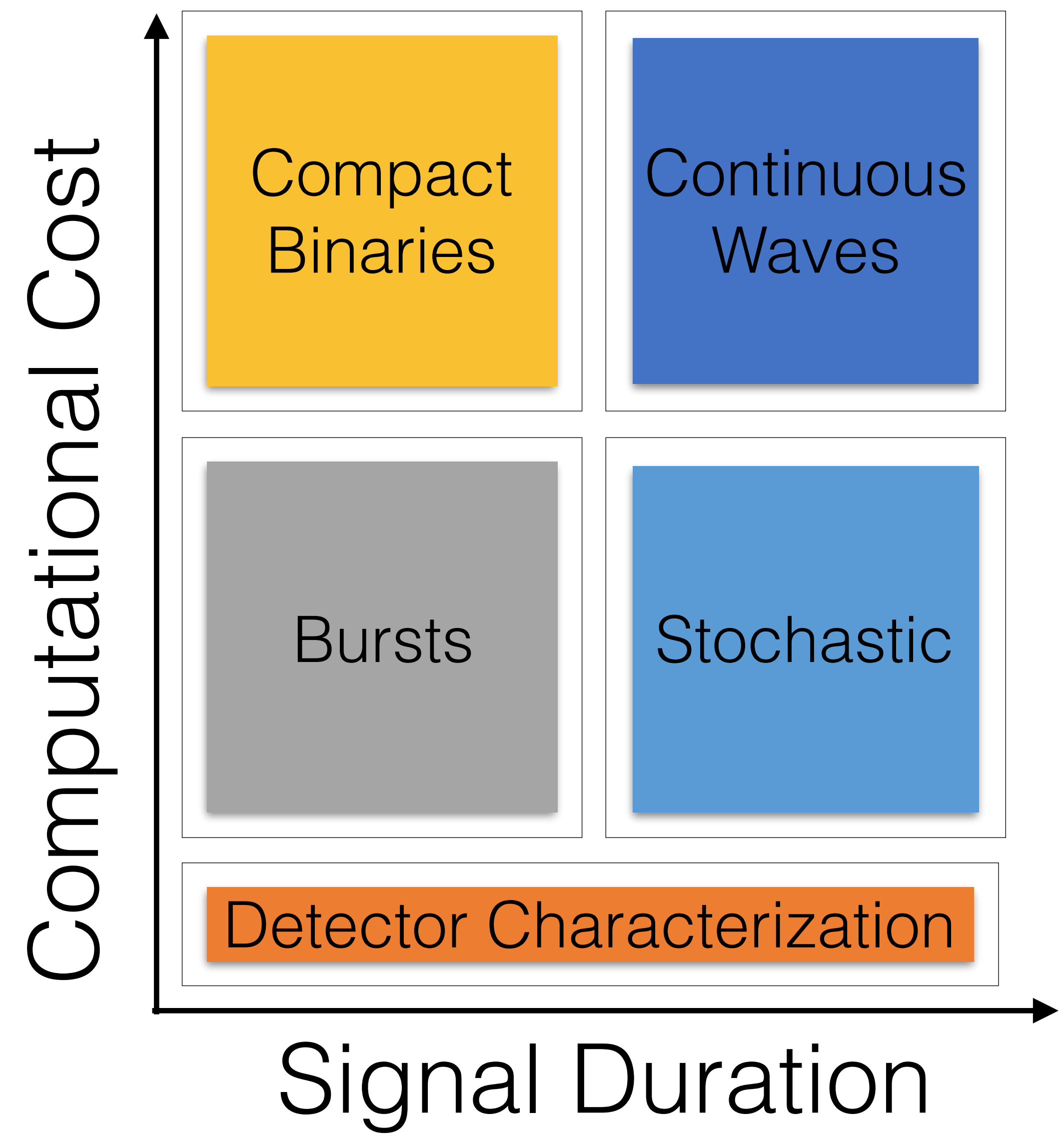
September 1, 2017
LIGO-G1701775-v1

# Outline

- Background on LIGO-Virgo Data Analysis Computing
- Challenges Facing GW Computing
- Concrete Opportunities?

# Data Analysis

- Four astrophysics groups: Bursts, Compact Binaries, Continuous Waves, Stochastic
  - Propose science goals, determine analysis algorithms, write the applications and run them
- Detector characterization group supports the commissioning teams and astrophysics groups
  - Determines analysis algorithms, writes applications, and runs them to identify instrumental artifacts
- Diverse algorithms and methods leading to heterogeneous demands on computing infrastructure

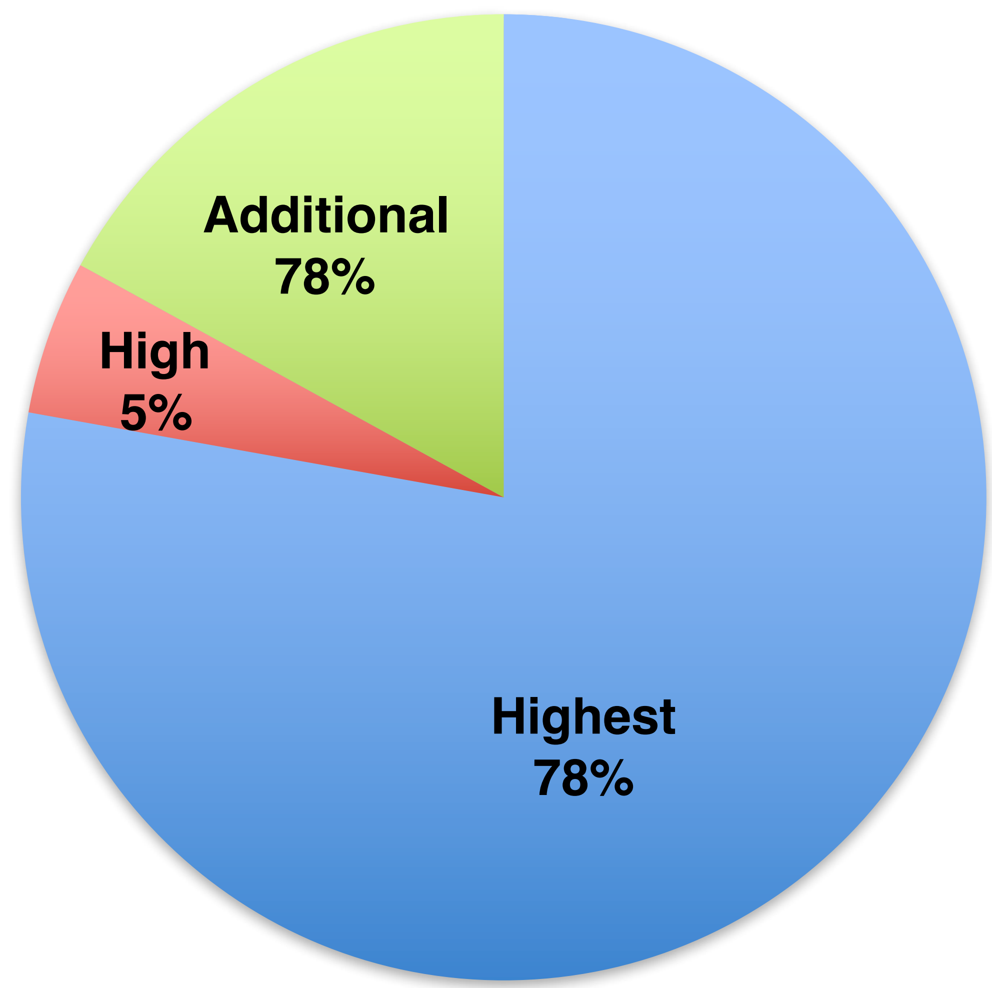


# Prioritized Science Goals Define LIGO's Computing Scope

- Three priorities of computing correspond to the priorities of LIGO science goals.
  - *Highest*: critical, core LIGO science. 78% of 2017-2018 (O3) computing.
  - *High*: valuable extensions to astrophysical sources and parameter spaces. 5% of 2017-2018 (O3) computing.
  - *Additional*: higher risk/reward. 17% of 2017-2018 (O3) computing.
- Each planned search is in one of these three categories.

**Computing Demand by Priority Category 2017-2018 (O3) Totals**

**1/2 Billion SUs!**



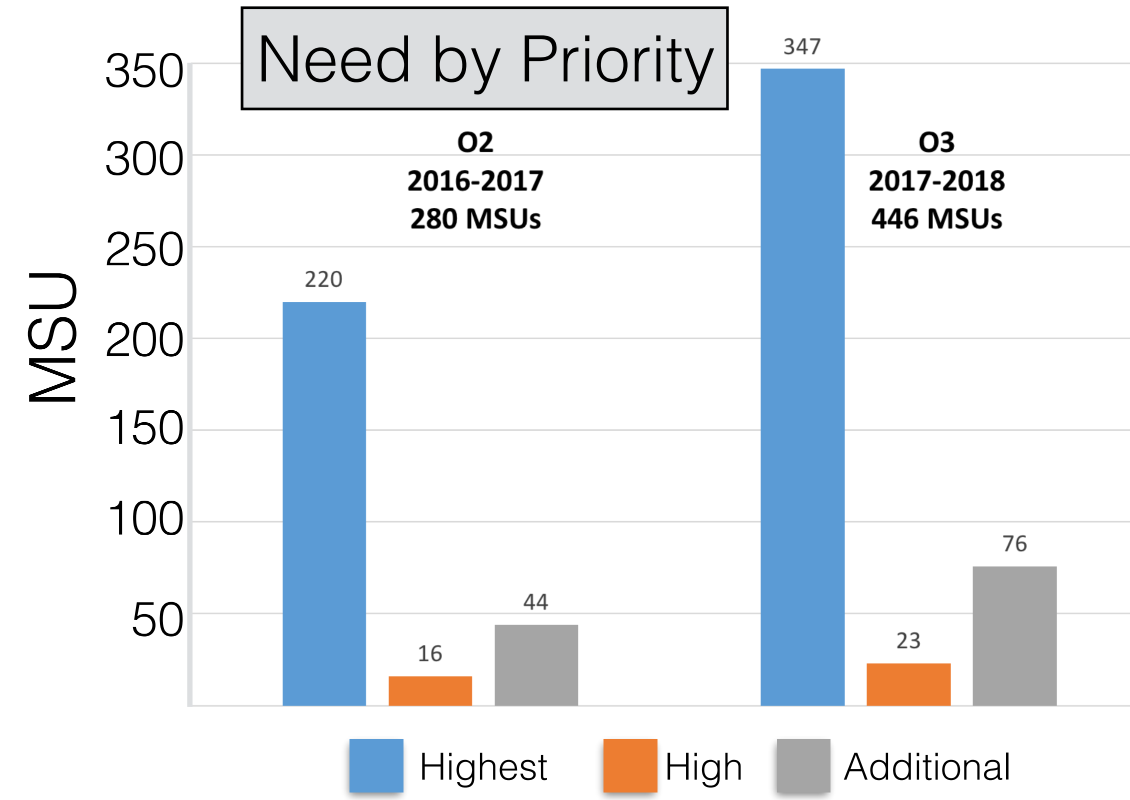

*1 SU = 1 aLIGO Service Unit = 1 Intel Xeon E5-2670 2.6Ghz CPU core-hour.*

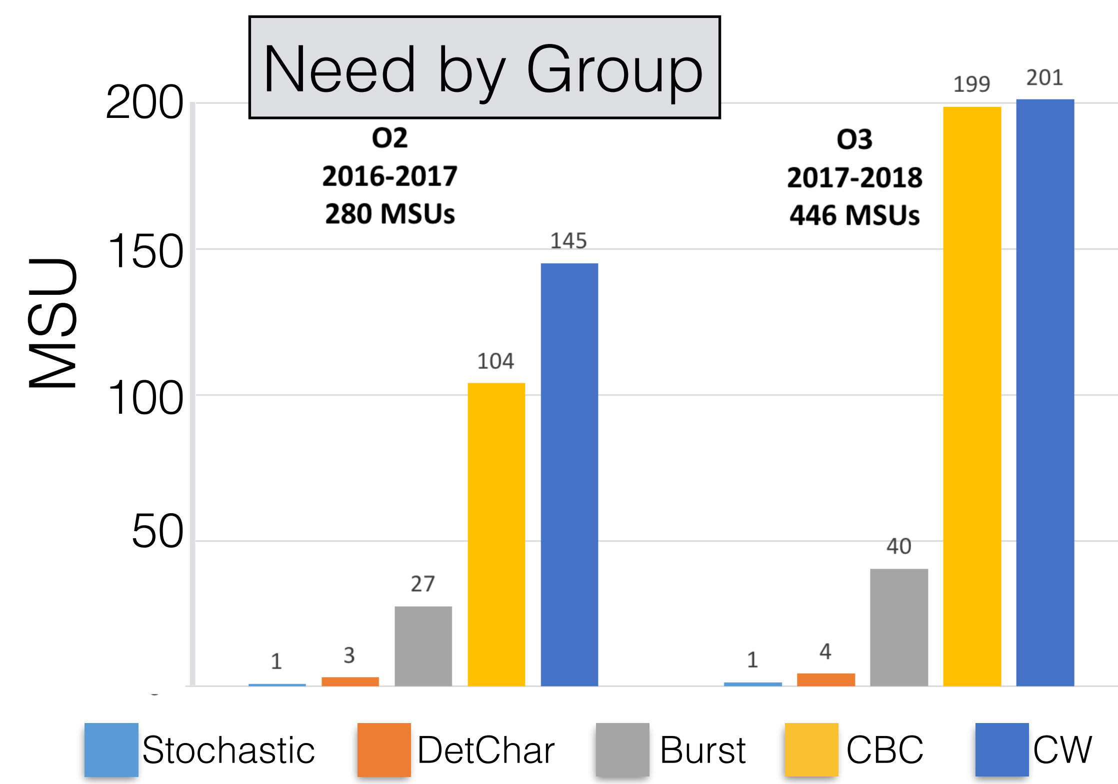
# Requirements

- Analyze data in unison from a worldwide network of detectors to reduce background and to enhance scientific output
- Low-latency analysis infrastructure - seconds to minutes
  - Transient alerts and data quality information within seconds of data acquisition
  - Rapid parameter estimation, verification, and follow-up automated (with some human vetting still in the loop at this time)
- Offline infrastructure - hours to months
  - Detector characterization feedback to commissioning and enhanced data quality generation
  - Deeper and broader searches for transients
  - Searches for continuous and stochastic signals
  - Parameter estimation, model selection and simulations

# Scale

- LIGO Data rates
  - Channels per site: ~200,000
  - Raw and Reduced data: 0.85 PB/yr
  - Strain per IFO: 0.12 MB/s
  - User data: 2.1 PB/yr
- LIGO-Virgo Computing requirements
  - SU=1 core hour on E5-2670
  - O1 actual: 170 MSU
- Users on LIGO-Virgo Computing Network
  - ~600 users, ~300 active past year
  - Top 20 users drive 70% of demand



# LIGO Data Analysis Computing: Demand

- >90 prioritized gravitational-wave and detector characterization analyses.
- >60 software pipelines implementing them.
- Distribution of data analysis computing demand by software pipeline: 10 pipelines drive 80% of demand, then a long tail.
- Contrast to distribution of data analysis computing engineering & operations support *effort* by software pipeline: flatter.

LIGO

# Executing analyses

- The bulk of our searches are *embarrassingly parallel.*
- All LIGO analyses and computing resources are managed using HTCondor, which schedules work and handles faults to ensure reliable execution of embarrassingly parallel jobs.
  - Broad use of single tools develops a knowledge base in scientific user community
- LSC computing staff and HTCondor team have a very close 15+ year-old collaboration.
  - regular meetings between senior staff
  - bug fixes / feature development and feedback

# LIGO Long-Term Challenge #1

- Increasing heterogeneity, complexity of computing platforms:
  - of **processing hardware** (CPU generations, GPUs, MICs) — due to the opportunities for cost savings, we MUST support multiple generations of CPUs, GPUs, MIC platforms and treat them each as distinct platforms — lowest common denominator code not good enough
  - of **providers** — internal to project, partners & collaborators, institutional, regional/national, commercial, volunteer
  - of target operating systems and **software environments** — containerization, etc. are tools to mitigate but aren't a silver bullet
  - of **batch/queueing** systems
  - of **storage and network** interfaces and capabilities
  - of **policies** for identity and access management, workflow prioritization
  - of **accounting** models and accounting systems
  - of **motivations and expectations** — mutual scientific/strategic interest, public or scientific recognition, financial or other compensation, etc. — and not everything is in a MOU, SLA, or contract

# LIGO Data Analysis Computing: Supply

- Many types of supply: **dedicated, allocated, opportunistic, volunteer**. Many providers in the US and abroad:
  - Dedicated LIGO Lab clusters (HTC)
  - Dedicated LSC clusters (HTC)

  LIGO Data Grid

  ~80% in O1

  - Virgo clusters (mostly allocated on shared resources, HTC)
  - PI clusters (shared, HTC and HPC)
  - Campus/regional shared clusters (allocated, HTC and HPC) e.g., OrangeGrid, PACE, SciNet
  - National shared supercomputers (allocated, HTC and HPC) e.g., XSEDE, Blue Waters
  - Opportunistic cycles (campus clusters, DOE labs, HEP clusters, etc.)
  - future: commercial cloud (EC2, Azure, Google, Rackspace, etc.)?

  Virgo + Open Science Grid

  ~20% in O1
- Two runtime software environments: LIGO Data Grid, Open Science Grid
- \+ Volunteer Einstein @ Home computing (~5 PetaFLOPS)

# Implications

- Increasing heterogeneity, complexity of computing platforms drives:
  - need for better software engineering and testing
  - need for additional organizational expertise+effort in optimization, distributed computing (architecture, engineering, support), and computing management
  - need for better tools, services, and processes for sustainable optimization
    - how do we help scientdsts write code that can be run efficiently on multiple platforms?
    - compilers aren't there yet, some higher level libraries can help
    - education and consulting for scientist/developers who are not first and foremost software engineers — provide value, avoid mandates
  - need to automate build and test for diverse h/w platforms — cloud testing is not there yet — not commoditized
  - need for more complex deployment, orchestration, instrumentation, and accounting of DA workflows

# Implications (cont)

- Increasing heterogeneity, complexity of computing platforms drives (cont):
  - need for funding — and budgeting — models for “metered” computing by the minute or watt, in addition to fixed capital investments every few years…
    - what information (and safeguards, and levers) do funders, project leaders, and computing users need to plan, execute, and abort metered computing, and on what timescales?
    - as a community, we have no idea how to do this — different world
    - the technology to enable metered computing will be easy compared to the policy, management, and scientific workflow changes needed to manage the spending of money on such a profoundly different timescale
  - need for aggregate accounting of work on disparate resources — very tricky, simplifications and approximations needed but each one makes someone unhappy — resources are tightly coupled to issues of money, scientific capacity, and recognition, things people care about enormously
  - need for management of the complex scientific, social, political, and financial aspects of shared computing

# LIGO Long-Term Challenge #2

- Uncertain and discontinuous funding streams for **computing labor embedded in the collaboration groups** outside the LIGO and EGO Laboratories:
  - The need to professionalize software development/engineering and to support increasingly complex computing environments demands more full-time professional computing expertise side-by-side with collaboration scientists (vs. part-time volunteer/service work by scientists).
  - Many of these are not strict software development or IT roles that can be outsourced beyond the project — they are hybrids of research computing, consulting, software engineering, and distributed systems development and administration roles.
  - These roles benefit enormously from institutional (project) “memory” — we pay dearly in time, money and quality when experience and relationships are lost.
  - Hard to recruit and retain career professionals on overlapping 1-3 year awards. Hard to find funding for this work; not always “transformative” science in and of itself, but needed to *enable* transformative science.
  - **This is an old problem** but is becoming more acute with the increasing need for these computing roles.

# LIGO Long-Term Challenge #3

- Computing Demand Uncertain — how much additional computing capacity may be needed to support post-O3 to 3G gravitational-wave data analyses?
- **We're trying to figure this out now, but it depends on what we see!**
- This period will include a potential network of LIGO, Virgo, Kagra, and LIGO-India reaching design sensitivity.
- How might searches be organized in a continuous observation era, and how will that affect the demand for computing over time? Will uneven demand driven by discrete detections become smoother as rates go up? How will search costs scale with sensitivity and network size?
- What new science might drive new demand?

# Concrete Opportunities

- Better coordination of Computing Security, and Identity and Access Management efforts between CERN and LIGO
  - Regular security call?
  - Unified voice on federated identity efforts.
    - Identity Provider (IdP) of last resort for research (one that supports Chinese researchers too)
    - Assurance profiles aimed at research institutions
    - EU privacy regulations and how to navigate them to allow Identity attribute release
  - Some of this interaction already happening via via FIM4R, but LIGO would like to expand those efforts and have a stronger voice in REFEDs and other forums where these matters are discussed
- More collaboration on optimization tools and technologies.
  - Informal working group on optimization, broadly defined (see next slide).
  - At first just to report mutual progress, let opportunities for collaboration emerge?
- More collaboration on software engineering tools and technologies.
  - GitLab, CERN Virtual Machine Filesystem (CVMFS), etc.

# LIGO-Virgo Optimization Approach: "The Whole Patient"

Iterate

- Scientific Prioritization and Scoping
- Estimation and Benchmarking of Computational Costs
- Optimization of Data Analysis Methods and Algorithms
- Optimization of Code Implementation and Libraries
- Compiler Optimizations
- Workflow Management Optimizations
- Development, Testing, and Simulation Process Optimizations
- LIGO-Virgo Computing Network Scheduling Optimizations
- Resource Supply Optimizations (make more cycles available)
- Workflow Portability Optimizations (expand usable resources)
- Hardware Procurement
- Pipeline Reviews including Computational Efficiency
- Documentation, Training, Collaboration and External Engagement

**What is CERN doing in each of these areas?**

Neglect nothing, focus on "bang for the buck" and where optimization effort can be most effective. Avoid adding burden where the payback is small.

# Questions?

- After the talk: email me <peter.couvares@ligo.org>.

# Extra Slides

# Why Bother?

- Is it worth it? Can/should we just manage everything on internal systems?
- Maybe, but...
  - not about cost
  - 100% internal clusters will have h/w heterogeneity that would drive some of these changes anyway
  - some of this work (understanding and encapsulating s/w dependencies, automated testing) improves the robustness and sustainability of DA pipelines and pays dividends in any environment
  - being prepared for things out of our control — opportunities, mandates, surprises
  - elasticity of supply (to handle peak demand, rapidly exploit scientific surprises and opportunities (something LIGO knows something about), cope with uncertain estimates
  - We don't want to be in the position where the science requires bursts of computing (and someone is willing to "pay" for it) and we can't deliver it, or where someone temporarily offers us substantial computing on a short timescale and our answer is, "we can't exploit it".