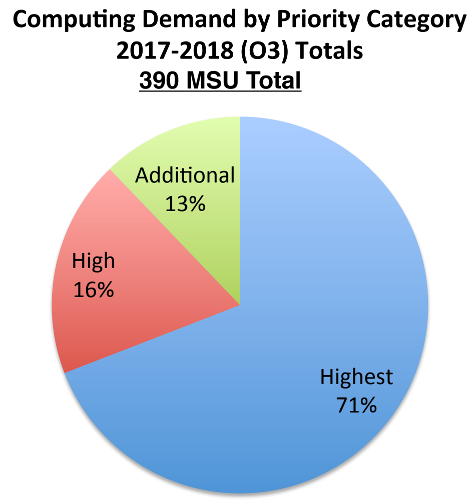

Prioritized Science Goals Define LIGO's Computing Scope

Three priority categories of computing have been defined to correspond to the three priority categories of LIGO science goals, and each planned search is in one of the three categories:

- *Highest*: critical, core LIGO science.
- *High*: valuable extensions to astrophysical sources and parameter spaces.
- *Additional*: higher risk/reward.

This enabled us to define a "bottom up" computing cost estimate that included each search, and as a result we are more confident that the results we are presenting this year accurately presenting LIGO's science needs.

Optimization Results

Dramatic gains in efficiency and estimated computational cost since May 2014:

Observing Run	CBC High Latency Deep Detection Search			CBC Low Latency Detection Search		
Estimated in:	2014	→	2015	2014	→	2015
O1 (2015–16)	115	→	5	5	→	0.5
O2 (2016–17)	369	→	39	27	→	4
O3 (2017–18)	896	→	133	104	→	17
Total	1380	→	177	136	→	22

- Factor of ~8 reduction in total estimated computational demand of high-latency CBC search.
- Factor of ~6 reduction in total estimated computational demand of low-latency CBC search.

Clear Plan for Computing Resource Development

Remaining aLIGO Project funds are sufficient to meet the estimated computational needs of LIGO through 2017-2018 (O3), given existing and committed LIGO Laboratory, LSC, and partner resources. **We do not expect to require or request substantial XSEDE resources**, although we will remain engaged with XSEDE.

LIGO's O1 computing hardware procurement is being driven by the measured performance of the search codes which dominate O1 computational demand. The LIGO Laboratory's ongoing hardware trade study will continue through 2017-18, and new trade study results will drive hardware acquisition prior to each observing run. Budget includes hardware and optimization team staff.

Uncertainty in demand estimates can be managed with continued optimization, scheduling flexibility, and management of science scope. Seeking additional LSC or shared computing resources are also options. **We can deliver the computing resources necessary to accomplish LIGO's science goals**.

LIGO Computing Optimization Project Staffed and Underway

Dedicated and expert LIGO Lab and LSC staff are devoted to optimization (4.5 FTEs total, 3.5 already hired). The project has LSC buy-in: the developers of every LIGO search pipeline are engaged; there is regular LSC Data Analysis Council attention; and the optimization team has good working relationships with key scientists and developers.

Given the high and low-latency CBC search optimizations achieved over the past year, we have a clear roadmap for further improvement of those and other computationally expensive pipelines. **We have the people and resources we need to succeed.**

The Data Analysis Computing Manager (Peter Couvares) is leading the LIGO optimization team and the ongoing LSC-wide effort to identify and realize further gains in computational efficiency. Optimization support is being focused on software pipelines based on their estimated computational demands. CBC low and high-latency search pipelines have already seen large efficiency gains, and have clear directions for further improvement. We will increase the optimization focus over the next year on improving the efficiency of all other software pipelines that individually drive ≥5% of estimated computational cost. The LIGO Optimization Team will also continue to facilitate "low-hanging fruit" optimization of all searches, regardless of cost.

Clear Optimization Roadmap

With effort and efficiency targets scaled to estimated computational cost, each LIGO software pipeline is being reviewed for potential for:

- Scientific optimizations: reconsideration of search parameters in terms of computational cost / scientific benefit.
- Algorithmic optimizations: FFT, BLAS, custom search code.
- Compiler optimizations: gcc, icc, native instructions.
- Vectorization, SIMD optimizations: both automatic and hand-tuned; looking ahead to AVX-512.
- Multi-threading optimizations: single-core vs all-core testing to estimate potential benefits of cache monopolization.
- GPU/MIC acceleration: high-level libraries and custom code.

The CBC high and low-latency searches have shown how fruitful each of these can be. **We have successfully demonstrated that we can optimize our two most critical pipelines, and we will continue to optimize other pipelines that drive our computing demand.**

GPU/MIC Acceleration

Preliminary PyCBC GPU results, soon to be formally reviewed: the best-performing GPU (GTX 980) delivers a factor of 3 higher search throughput than the best-performing CPU socket

(E5-1660 v3). **At equivalent search throughput, the most cost-efficient GPU card (GTX 750 Ti, $140) is a factor of 6 cheaper than the most cost-efficient CPU socket (E3-1220-v3, $820)**. Our early GPU memory reliability testing has been encouraging, but even with worst-case redundant execution, GPUs may hold a factor of 3 cost-efficiency advantage. Simple "offloading" of generic algorithms to co-processors via standard libraries may benefit other search codes, and custom implementations hold more promise at additional development cost. Co-processor optimization work has lead to a better understanding of optimization landscape more generally, leading to improvements in CPU code performance.

CIT has deployed a co-processor hardware test stand open to all LSC pipeline developers, with a broad market range of scientific/professional and consumer co-processors (Tesla K10, K80, 2090, GTX 580, 750 Ti, 970, 980 GPUs, Intel MIC, and others.) We perform systematic benchmarking of GPU codes under development to understand relative effects of GPU parallelism, speed, memory, bandwidth, and reliability on overall computational cost and efficiency. We have an **ongoing technical engagement with senior NVIDIA CUDA developers to overcome PyCBC GPU bottlenecks and improve CUDA performance beyond LIGO use-cases**. We are evaluating CPU/GPU density tradeoffs given our projected hybrid CPU/GPU workload.

<u>Shared Resources</u>

Given our latest demand estimates, **substantial shared resources are unlikely to be needed through 2017-18 (O3)**. However, given uncertainty, modest shared resources may be of benefit for short-term demand spikes, new searches, or hardware trade studies.
We will remain engaged with XSEDE and OSG:

- So we're prepared to scale quickly if we need to.
- So we remain engaged with XSEDE performance experts.
- So we're ready for (and can help define) evolving computing models, post-O3.

We hope to leverage the tools, services, and human expertise of OSG to facilitate a robust bridge between the production LIGO-Virgo Computing Network and campus grids that wish to contribute computing resources to LIGO. We also hope to be able to provide surplus LVCN resources to OSG if they become available.

<u>Summary</u>

- We have shown excellent results so far with CBC low and high-latency search optimization; we have established a paradigm to propagate the approach to optimize other searches.

- We are still ramping up our optimization team and wider LSC optimization effort, will continue to deliver improvements to both the computational efficiency of LIGO search software pipelines and the infrastructure in which they run.

- We have become well-engaged with the wider HPC and HTC computing communities, will continue to build collaborations with external facilities and teams.
- Preliminary hardware prototyping with GPUs is promising; we will bring this work rapidly to evaluation for procurement.