LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY
- LIGO -
CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Technical Note	LIGO-T050053-A-Z	2006/06/08

The LSC Computing Plan

LSC Computing Committee, on behalf of the LSC

This is an internal working document
of the LIGO Scientific Collaboration

California Institute of Technology
LIGO Project, MS 18-34
Pasadena, CA 91125
Phone (626) 395-2129
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Project, Room NW17-161
Cambridge, MA 02139
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

LIGO Hanford Observatory
Route 10, Mile Marker 2
Richland, WA 99352
Phone (509) 372-8106
Fax (509) 372-8137
E-mail: info@ligo.caltech.edu

LIGO Livingston Observatory
19100 LIGO Lane
Livingston, LA 70754
Phone (225) 686-3100
Fax (225) 686-7189
E-mail: info@ligo.caltech.edu

WWW: http://www.ligo.caltech.edu/
http://www.ligo.org/

Contents

1 Executive Summary **3**

2 Introduction **7**

3 LSC Computing Model **9**

3.1 Data Model Input Parameters 9

3.2 LSC Tiered Structure and the Roles of the Various Tiers 10

3.3 Data Flow 12

3.4 Data Processing 13

3.5 Data Analysis 14

3.6 Simulation Process 16

3.7 Calibration 17

4 Commissioning of Datagrid Upgrades and Improvements **17**

5 Resource Requirements **18**

6 Networking Requirements **23**

7 Summary **24**

A Details of the Resource Usage **25**

List of Figures

1 Map of the LSC Datagrid 3

2 LSC Datagrid resources and needs by year 4

3 Map of the LSC Datagrid 11

4 Apportionment of resources by year and by search class. 15

5 Graph of projected resource requirements to support science running through 2012. 22

6 Table of current and expected LSC Datagrid resources 2005/2006 - 2011/2012 26

7 Table of current and expected LSC Datagrid resources 2005/2006 - 2011/2012, continued from Figure 6 27

List of Tables

1 LIGO LCU capacity, use & unmet need in future years. 5

2 LIGO data rates and volumes for S5. 10

3 LIGO LCU capacity, use & unmet need in future years. 21

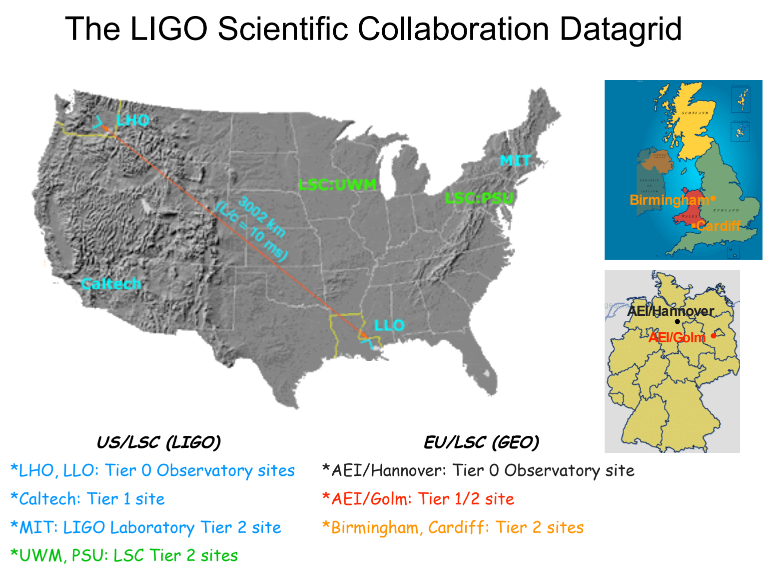

Figure 1: The current LSC Datagrid consists of 6 sites in the U.S. and 4 sites in the EU. Data are generated at the three observatories (two LIGO plus GEO). The main deep data archive for LIGO is at Caltech. Backup (100%) copies of raw data are maintained at the observatories. See http://www.lsc-group.phys.uwm.edu/lscdatagrid/

1 Executive Summary

This is intended to be a living document that will be revised as experience continues to accrue within the collaboration.

Sufficient experience across the entire collaboration now exists from the analysis of the first four science runs to enable a near and longer term Computing Plan for the LSC that spans both LIGO Laboratory and the associated centers located at a number of LSC institutions, both in the U.S. and Europe. The LSC has established a collaboration-controlled distributed computing environment termed the **LSC Data-grid**, (hereafter referred to simply as the *Data-grid* or *LDG*). Figure **??** presents the current LSC Datagrid. The purpose of this document is to present current capacity and to determine the scope and requirements for future data analysis activities within the LSC over the course of the next 4 - 5 years.

The primary focus of the current version of the plan is the U.S. based resources, with a goal to inform both the collaboration and our funding agency on current and future needs. This plan provides the current best estimate of the resources required. The plan addresses data volume generation, data levels, their distribution, and the data analysis that takes place within the analysis working groups of the LSC that were established to address different classes of astrophysical sources [1].

The model described herein has evolved from previous concepts and now reflects the "as-built" implementation and use of the ten existing Datagrid centers. The advent of the World Wide Grid spawned new ideas on how to organize the Datagrid as a worldwide distributed computing resource for the collaboration. This document is intended to identify projected needs and to thereby help the LSC and LIGO Laboratory management in its planning processes.

The LSC has completed four science runs, S1 - S4. The first two runs have been completely analyzed, and most LSC search analysis working groups have nearly completed the S3 and S4 run analyses. S5 run began

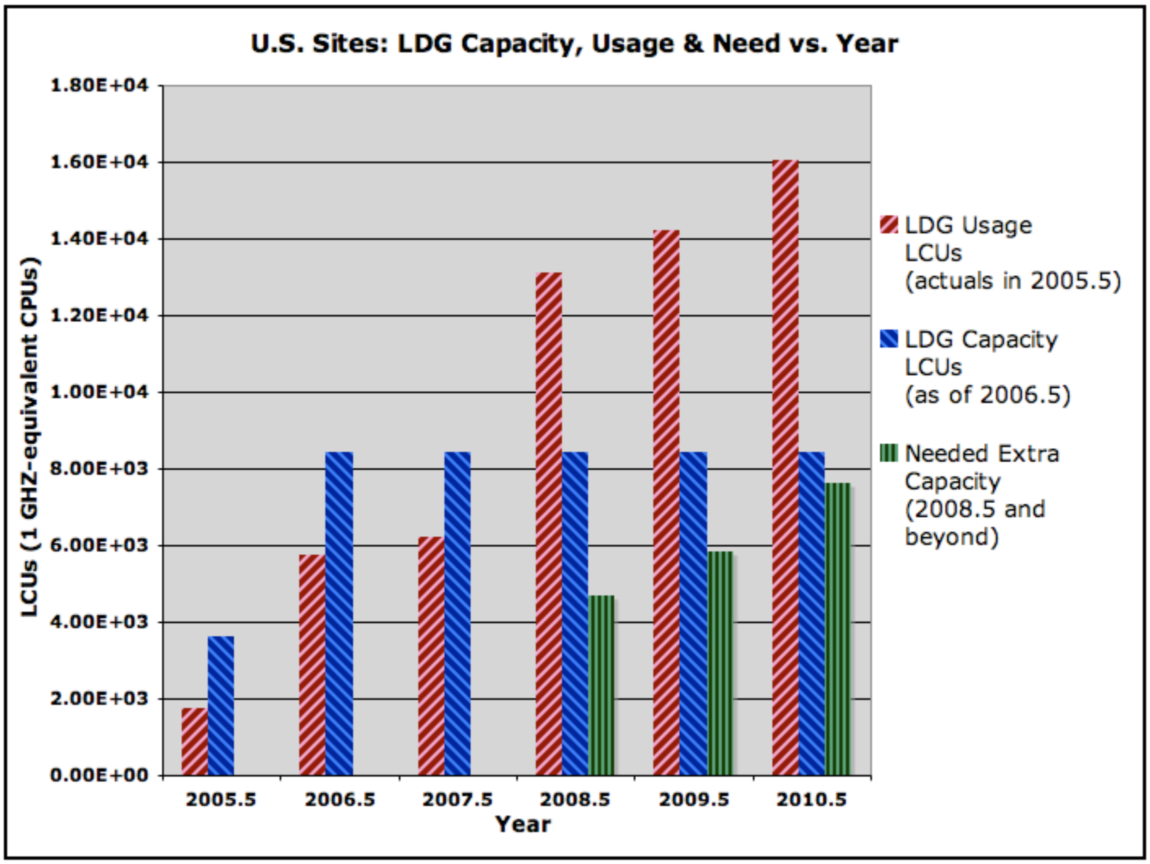

Figure 2: LSC Datagrid resources available during 2005/2006 and 2006/2007 (blue). The extant 2006/2007 level is projected through 2010/2011, also in blue. The 2005/2006 actual use is indicated in red, corresponding to $\sim 50\%$ of the resources. Projections for future years, also in red, were provided by each analysis working group. 1 LCU-yr $\equiv$ 1 GHz-yr of CPU clock speed $\sim$ 1 GFLOPS-yr.

Year	LDG Capacity LCUs	LDG Usage LCUs	LDG Unmet Need, LCUs
2006	3621	1727	-
2007	8419	5746	-
2008		6201	-
2009		13087	4668
2010		14233	5813
2011		16051	7631

Table 1: LDG capacity, use, and future unmet needs in units of LCUs.

in November 2005 and is planned to be the longest run to date, at least 18 months in duration. As of this writing, S5 has been under way for over 6 months. It provides a first opportunity for the LSC to confront its computing model in a continuous production-level mode of operation.

The principal requirement for LSC computing facilities is to enable the science potential of LIGO data to be exploited to the greatest extent possible. This includes, for example, (i) providing speedy access for all members of the LSC to reduced data sets for analysis during and after data-taking periods, (ii) appropriate access to raw data for follow up analysis, and (iii) providing the computational to ensure that the so-called *flagship searches* can be completed in a timely manner. This document outlines a model that makes substantial use of Grid Computing concepts, thereby allowing the same level of data access, and making available an equitable amount of computing resources to all members of the LIGO Scientific Collaboration.

Referring once again to Figure 1, data are generated by the interferometers at the two LIGO and the GEO observatory sites. While these are functionally the Tier 0 sites for the LSC; however, unlike the high energy physics model, the computing resources at the remote observatories are modest by comparison. Caltech serves the role of the collaboration Tier 1 site and the university centers at UWM, PSU, and MIT serve the collaboration as the U.S. Tier 2 sites. In the EU, AEI/Golm may be considered GEO's Tier 1 site and the UK sites are Tier 2 centers. This document addresses the collaboration's needs in the U.S. At a future date, GEO's plan may be included when it becomes available.

Recent experience further informs these needs, leading to the resource requirements trend shown in Figure 2. This shows the estimated resources for the period spanning the current S5 and future (planned) S6 run, leading up to the period prior to the availability of Advanced LIGO interferometers. The trend corresponds to the U. S. portion of the LDG only.

The primary findings of this plan may be summarized as follows:

- Based on the projections presented here, the known and planned LSC Datagrid resources available during CY2006/2007 will be adequate to meet the expected need through 2008. Margin for code performance and scheduling inefficiency is reflected by using actual experience during the recent past. Interferometer up-time less than 100% also represents margin, since the volume of data to be analyzed scales with up-time. Historically, this has hovered around 60% - 70% for single interferometer up-time.
- As quantified in Table 1, beyond this time, the LSC must provide additional computing that represents a ∼56% increase in 2009 (an additional ∼4700 LCUs), with a ramp up beyond that time.
  - To set the scale, 4800 LCUs today would represent ∼ 600 2X dual-core 2.2 GHz CPU nodes, costing a projected ∼$1M in FY2008.
- As LDG cluster facilities are upgraded, storage capacity can be accommodated without incurring specific mass-storage expenses.

- The LIGO WAN is expected to provide sufficient bandwidth needed for the foreseeable future without requiring dedicated, specific upgrades.

2 Introduction

The LSC computing model has evolved over the past four science runs to embrace the Grid paradigm and the associated high degree of decentralization and sharing of computing resources. However, as the Datagrid computer facilities differ among themselves, they are also suited to play different roles. Thus a degree of hierarchy, with specific responsibilities at each level, is natural. All roles described are critical to the success of the distributed model. The required level of computing resources means that non LIGO Laboratory facilities are critical to the science mission of the LSC.

The full raw data frames are transferred to Caltech from the two LIGO observatories and to AEI/Golm for the GEO interferometer. The raw data for LIGO are archived at Caltech. Copies of the raw data are also written to tape and maintained at the observatories: LHO and LLO each maintain copies of their respective raw data.

Several levels of Reduced Data Sets (*RDS*) are produced for broader distribution across the datagrid. For LIGO, these are produced at the Tier 0 observatories immediately after the data are acquired. They are streamed to Caltech, whence they are replicated to the rest of the datagrid sites. For GEO, they are generated at AEI/Golm and replicated to the Datagrid from there.

Beginning with S4, on-line analysis is performed using the Tier 0 computing facilities. These analyses provide first-look astrophysical analysis and provide the collaboration with near-real time end-to-end assessment of detector performance and sensitivity. As LIGO ramps up to nearly continuous running beginning with S5 (late 2005) and beyond, It is expected that the observatory facilities will become dedicated to on-line analysis. However, offline analyses will be permitted if there is sufficiently unused computational capacity.

The Tier 1 and 2 computing facilities provide resources to perform explorational, developmental, and production analysis of science run data. Each Datagrid facility is responsible for funding the required resources for its declared role. A Computing Committee, which includes representatives from the Tier 0/1/2 Datagrid centers, facilitates the planning and management of computing resources for the LSC. Datagrid center roles are negotiated in this forum to optimize the smooth running of the system and to maximize availability for scientific investigations. In addition, the LSC Computing Committee works to coordinate future proposals for computing resources in order to ensure that the needs of the LSC as a whole are considered in such proposals. This plan provides an important input into this process.

The LSC data analysis activities are breaking ground in this new field of gravitational-wave astronomy. Low-latency analysis is needed if LIGO wants the opportunity to provide alerts to the astronomical community in the future. Maximum scientific exploitation requires data analysis to proceed at the same rate as data acquisition. The natural unit for the rate of computation for LIGO is the FLoating point OPerations per Second, or **FLOPS**. This document quantifies computational capacity or need in terms of the *LIGO Computing Unit* or **LCU**. The LCU is defined as the computing power of a 1 GHz clock speed CPU. We have found that 1 LCU $\equiv$ 1 GHz of CPU clock speed $\sim$ 1 GFLOPS is a good approximation for the kind of FFT-intensive analyses undertaken in LSC data analysis. As of this writing, a current-generation 2X dual-core 1U node with 2.2 GHz clock speed represents 8.8 LCUs.

By their nature, template-based matched filtering techniques require the most computation per datum. Searches for inspirals – binary neutron stars (NS/NS), neutron star - black hole systems (NS/BH), and black hole - black hole systems (BH/BH) are exemplars of this type of search. Blind all-sky searches for periodic waves from rotating deformed massive objects that have no known EM counterpart also fall into this class of search and, indeed, will exceed the computational requirements for inspirals.

It is estimated that matched filtering searches for signals from compact coalescing binaries with individual masses $\geq 1\mathrm{M}_\odot$ may require as much as $\sim$ 2700 LCUs to analyze data in the time it takes to acquire it.

Deep searches for continuous sources of gravitational waves can utilize more than a few ExaFLOPS (10^{18} FLOPS) or more than 3×10^9 LCUs to keep up with data taking. The former challenge is accessible today; the latter remains unachievable for the foreseeable future. This means that searches for periodic sources will be computationally limited to sub-optimal algorithms (e.g., partial-sky, few-parameter, or short duration FFTs) constrained by existing resources.

In addition to the Datagrid resources available at the Tier 0/1/2 centers, the LSC has launched a distributed computing effort called *Einstein@Home*. This project was part of the APS 2005 World Year of Physics. *Einstein@Home* distributes searches for continuous sources of gravitational waves over donated computer cycles all around the world. This computing model is well suited to the continuous wave searches: searches over long time durations can be divided into narrow frequency bands, thus limiting the amount of data that must be transmitted to the remote machines. Analysis of the data can proceed asynchronously and without impacting follow-up studies. Recent performance metrics indicate that the LSC has been able to obtain $\sim$ 40 TFLOPS ($\sim$40,000 LCU equivalents) through this screensaver-based computing technology.

On the other hand, the *Einstein@Home* computing model it is not so well-matched to other classes of computationally intensive searches, such as that for compact coalescing binaries. This latter type of search requires low-latency, guaranteed analysis and the data are not readily broken out into small chunks for analysis by different computers. Based on involvement and growth to date, we plan on a significant part of the computational load for CW searches to continue to be made up by this different but effective computing environment.

3 LSC Computing Model

3.1 Data Model Input Parameters

Data Analysis

Data analysis inputs for the LSC Computing Plan were derived as follows.

1. Each U.S. Datagrid site provided the following facilities information:
   - Top-level hardware configurations during 2005/2006: # of nodes, # of CPUs/node, CPU clock speeds, etc.;
   - Estimated site up-time during the period of measurement;
   - Major upgrades made during the measurement period were factored into the available resources calculation.
2. Actual Datagrid-wide usage data were derived from the Condor logging information:
   - CPU-hrs used by user and by Datagrid facility;
   - User affiliation within the four LSC data analysis working groups –
     - Bursts Sources (*Bursts*)
     - Binary Coalescencing Sources (*Inspirals*)
     - Periodic Sources (*CW*)
     - Stochastic Background Sources (*Stochastic*) .
3. Each of the working groups were asked to certify the usage statistics from 2005/2006 and were then asked to provide estimated increases in use and/or needs for future years in terms of factors over 2005/2006. It is important to note that the usage captured from the Condor logs is *complete* in the sense that all work required to complete a data analysis, including CPU use for Monte Carlo simulations and multiple passes through the datasets are included in the logged usage.
4. This information was used to develop a spreadsheet database model for the LSC data analysis activities.

LIGO Data Volumes

LIGO data are stored as frame files, as specified in the LIGO-Virgo Frame Specification[3]. A number of different LIGO data sets are written soon after data are acquired. Various levels of data reduction are performed and written as Reduced Data Sets, or *RDS*s. These are described the LSC Data Analysis White Paper[1]. Reduction involves both decimation by removing less interesting channels and resampling channels of interest at reduced rates.

- **Level 0** (L0) Full raw data set, used for permanent archival and for generation of subsequent levels of RDS. Data are always available from Caltech and may be used at any time to create "custom" RDSs for specific analyses by individuals or groups.
- **Level 1** (L1) Approximately 15% - 20% of the L0 data volume, containing the primary physics channel and a large number of ancillary channels and vetoes useful for environmental investigations.

Data rates & volumes for the first 4 months of the S5 run		
Run: S5 Year: 2005	MB/s	Raw Volume/ Year (TB)
Level 0 (2 copies)	24.0	757
Level 1 per copy	2.3	73
Level 2 per copy	-	-
Level 3 per copy	0.18	5.7
Level 4 per copy	0.05	1.4
Totals	27	1399

Assumes copies of L1, L3 at each of 8 LDG sites.

Table 2: LIGO data rates generated during the first four months of the S5 science run. L2 RDSs are not generated during S5.

- **Level 2** (L2) Further reduced data set, still containing the strain channel and fewer ancillary channels. This was not generated during S4.
- **Level 3** (L3) A lightweight data set, containing the strain channel.
- **Level 4** (L4) The lightest weight data set, consists for the L3 data downsampled by 4X.

Table 2 shows the S5 run data volumes contained in these RDS levels.

3.2 LSC Tiered Structure and the Roles of the Various Tiers

The LSC Datagrid is both Grid-oriented and hierarchically structured. Different facilities have different responsibilities that may be characterized in the following ways. All sites and their roles are critical to the successful exploitation of LIGO data and a sensible balance among the resources in the various tiers of the hierarchy is required for the operation of the model. This discussion will focus on the U.S. facilities. Refer to Figure 3.

Tier 0 Facilities

LIGO Laboratory maintains and runs two Tier 0 facilities, one at each of the two LIGO Observatory sites. These remote sites are responsible for acquiring the L0 data and writing them locally to tape for permanent archival of the data produced at that site. Each site has a 14 TB robot plus off-line tape storage. Additionally, L0 data are transfered to the Tier 1 facility at Caltech for off-site archival of the complete set of all LIGO data from both observatories.

The Tier 0 sites also process the L0 data to generate L1 - L4 RDSs. To the extent that the WAN bandwidth between the Tier 0 and Tier 1 centers permit, all levels of data are streamed to Caltech. When bandwidth

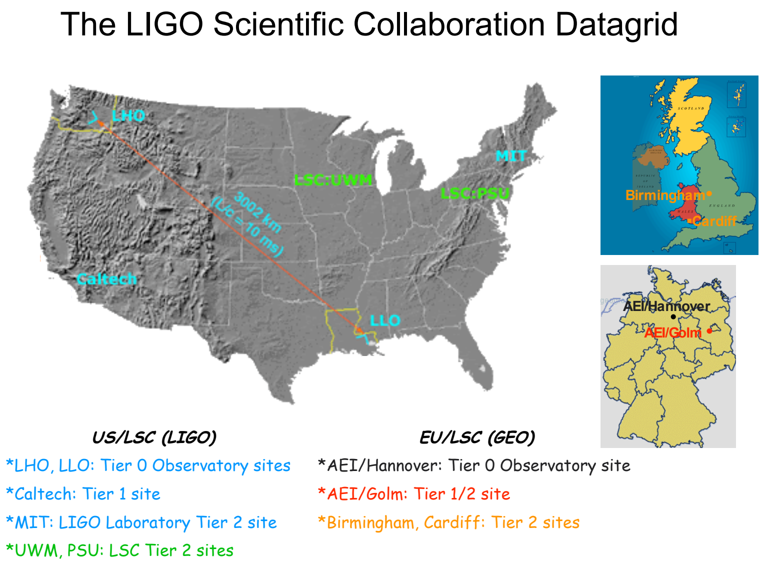

Figure 3: The current LSC Datagrid consists of 6 sites in the U.S. and 4 sites in the EU. Data are generated at the three observatories (two LIGO plus GEO). The main deep data archive for LIGO is at Caltech. Backup (100%) copies of raw data are maintained at the observatories. See http://www.lsc-group.phys.uwm.edu/lscdatagrid/

does not permit, the largest volumes are written to a second tape copy at the Tier 0 sites and sent via commercial carrier to Caltech.

The Tier 0 computing facilities are dedicated during the science runs to providing near real-time, low-latency end-to-end analysis results. Excess capacity is allocated to off-line processing at lower priority. The facilities are used by (i) the interferometer science team to assess machine performance and stability and (ii) by the analysis groups to monitor the data stream. The sites provide high availability and response time in the case of problems. In the event of prolonged down time, L0 processing to produce RDSs is taken over by the Tier 1 facility described next. To account for local data access demand, each site has an HSM combined disk-plus-tape buffer corresponding to about 21 days (1 year) of disk (tape) access. Data older than $\sim$ 1 year are archived off-line in a tape cassette storage library. Human intervention is required to retrieve specific data cassettes and to reinsert them into the robotic archive. To date, the continued improvement of interferometer sensitivities over time have led to essentially no requirement to provide lookback capability on-line at the Tier 0 facilities (such automated access to the full complement of LIGO data is available, however, at the Tier 1 facility).

Tier 1 Facility

LIGO Laboratory at Caltech maintains the Tier 1 facility for the collaboration. It has the responsibility to host and provide long-term access to and archival of L0 data for all major science and engineering runs. In addition trend data and so-called AstroWatch data acquired in the background between runs are also archived to provide long-term historical look back to on-site environmental data. The Tier 1 center also provides the capacity to perform the reprocessing of the L0 data, and to provide LSC-wide access to the derived RDSs with short latency (on disk) and longer latency (on tape). All the datasets are considered to be for the collaboration as a whole, and the storage and CPU are also for that purpose.

The Tier 1 site allows access to and provides capacity to analyze all the hosted samples. The Tier 1 facility has a sufficient level of service in terms of availability and response time to enable the collaboration to have access to its facilities. Given the vital role in receiving the L0 data and reprocessing, down times in excess of 12 hours become problematic in terms of catching up with processing and storage of L0 data. Should this ever happen, the fallback option is to receive L0 from the Tier 0 sites via commercial carrier on tape media.

Tier 2 Facilities

The LSC in the U.S. currently has two collaboration Tier 2 facilities at PSU and UWM, and one LIGO Laboratory Tier 2 facility at MIT. These represent a computational resource comparable to the Tier 0 + Tier 1 resources. The Tier 2 facilities play a number of significant roles for the LSC. They provide locally accessible RDSs for simulation and analysis. The Tier 2 sites represent production analysis capabilities. The Tier 2 sites are of different sizes and hardware generations. They provide simulation and data analysis capacity for data analysis working groups and subgroups. This analysis activity is generally organized by working group or source search class, but often consists of multiple and different analyses being undertaken at any one time. They typically will host a full set of RDSs from the most recent science run and can also provide complete look-back capability over all past science runs for L3 and L4 RDSs.

The Tier 2 centers also host sample data for code development. Some Tier 2 sites may take a significant role in calibration, especially the generation of so-called time-domain calibrated $h[t]$ data following the local analysis interests and involvements. Due to the size of the LSC Datagrid, each tier in the hierarchy is critical and the level of service expected from the few Tier 2 sites in terms of availability and response time is comparable to that for the Tier 1.

3.3 Data Flow

The source of the input data for the computing model comes from actual experience during the 14 month period January 2005 - February 2006. During this period S3, S4 and S5 data analyses were taking place concurrently with different admixtures at the various Datagrid sites. The data volumes and data set types reflect the most recent specifications available. The data volumes generated and recorded during science runs and RDS definitions have stabilized over time. It is expected that the volume of data generated per unit time is accurate; what is less certain is the number of copies the collaboration chooses to maintain across the datagrid and the volume of processed results individuals and groups will want to maintain. The current values quoted by the four groups that provided information were modest compared to the acquired data volume.

The L0 data can be streamed from the Tier 0 sites to the Tier 1 archive over OC3 with margin capacity. For several reasons, it is also advantageous to write archive tapes at the Tier 0 centers and to then bulk-ship them to Caltech. Thus, by LHC standards, LIGO's bandwidth requirements are very modest. The Tier 2 centers are all on Internet 2 and Abiline and thus have a minimum of OC48 or higher access to the WAN. Initial problems with tuning of transmission parameters for data streaming across the datagrid have been resolved and at the present time, near real-time data access by Tier 2 centers is not WAN-bandwidth limited; access by smaller university groups who are connected to the WAN with less than OC3 bandwidth, while being an issue, is not the subject of this Plan.

L1, ... , L4 data are also presently streamed from the Tier 0 sites to the Tier 1 site, which then redistributes it to the Tier 2 sites. End-to-end data transmission latency, from their creation to when they become available at the Tier 2 sites, has been measured during the current run to be typically $\lesssim$2 hours from Tier 0 to all Tier 2 sites.

Upon the availability of new frame data from the CDS Data Acquisition System, the L0 frames are processed as follows:

a) L0 data are copied to permanent SAM-QFS tape at the Tier 0 sites;

b) L0 data are either copied to a second tape or streamed to permanent SAM-QFS mass storage at the Tier 1 facility;

c) L1, ... ,L4 data are created from the Tier 0 disk copies of the L0 frames and are stored locally at the Tier 0 site;

d) L1, ... ,L4 data are then streamed to the Tier 1 facility for local storage and redistribution to the Tier 2 sites.

Step b), the transfer of the L0 data to the Tier 1 facility, is an important requirement. This can be done either by commercial carrier shipment of tapes from the Tier 0 site (as it was done in the past) or by streaming over the WAN when suitable bandwidth is available, as was first achieved during S4.

3.4 Data Processing

The S5 LIGO data flow relies on calibration data generated periodically by the calibration team during science operations. Calibration xml files are propagated across the Datagrid separately and are available at each site for local use. At the same time, the calibrated time domain $h[t]$ data produced on the fly at the Tier 0 centers are propagated to all LDG sites. Historically, the calibration of the interferometers has been an iterative process providing a number of releases over a long period of time; this implies a need to reprocess all the strain channel data repeatedly to generate $h[t]$ as new calibrations are published. At the present time, it is envisioned that these subsequent steps will occur in a distributed fashion across multiple Datagrid sites. To set the scale for this task, one year of strain data from the three LIGO interferometers can be processed on a 300 CPU cluster in $\sim$ 2.5 days of continuous processing (*i.e.*, $\sim$0.7 CPU-yr per interferometer per year of data).

Latency

During science running, the latency of data availability is determined by two factors:

- L0 data are written in 32s frames, L1 in 64s frames, L3 in 256s, and L4 data are written as 1024s frames – for L3 this represents an intrinsic latency of $\sim$10 minutes while for L4 this latency is 17 minutes.
- Data propagation across the Datagrid, including multiple steps for Tier 0 $\rightarrow$ Tier 2. This produces an end-to-end latency of $\lesssim$ 2 hr.

Reprocessing

Frequency-domain calibrations take place as part of the data analysis pipelines. Thus new calibration releases require full reprocessing of science run data. As mentioned above, when time-domain $h[t]$ calibrations become generally available, the need to generate communally available calibrated strain data for multiple searches may require that end-to-end reprocessing of the strain channel be performed at multiple sites.

Typically, three iterations of calibration information are published over the course of many months, and would thus require the full regeneration of the $h[t]$ data three times. This will be an off-line process. Iterations such as these were included in the actual performance metrics upon which future projections were provided by the analysis working groups.

3.5 Data Analysis

Data analysis activities within the LSC are organized into four working groups as defined in the LSC Data Analysis White Paper[1]. Briefly, the working groups are dedicated to the development and execution of data analysis strategies looking for different classes of astrophysical sources of gravitational waves:

- Deterministic signals with well parametrized phase and amplitude evolution
  - Continuous wave (periodic) sources – rotating neutron stars;
  - Inspiraling (chirping) sources – binary compact object coalescences;
- Burst or transient sources – searches for unmodeled or poorly understood waveforms of short duration, $\Delta t \lesssim 1$ s.
- Non-deterministic signals – stochastic signals

Each working group has developed one or more analysis pipelines implementing different search algorithms. The different pipelines run in the analysis environments that are discussed in reference [4]. Each analysis can be run on a number of LSC Datagrid sites that have resident local caches of the RDSs needed for the analysis.

The work of the groups proceeds in a combination of short development and test runs and longer production runs. Development runs typically use much smaller data subsets that are set aside specifically for algorithm development.

The results of the production analyses consist of data products that are later post-processed, typically on personal workstations or laptops using commercial tools such as Mathematica or Matlab and noncommercial tools, such as CERN's ROOT package. These end-product analysis machines are not considered in the current model, which is limited to Tier 0 - 2 production centers. Searches for transient and events of finite duration (*i.e.*, bursts and inspirals) generate events and associated metadata. Continuous signal searches (*i.e.*, periodic sources and stochastic background) generate intermediate products consisting of spectra that are then post-processed outside the production runs. While much of the post-processing takes place on individual workstations or laptops, the advent of the much longer duration S5 run has shown that an adjustment to this model is needed. Some of the post-processing has been migrated into the cluster-based analysis in order to become more efficient.

Analysis procedures and data flow

The resources required by various classes of production analysis jobs vary widely. Figure 4 shows the relative proportion of resources used during the period Jan. 2005 - Feb. 2006 by the different data analysis working groups. Also shown are projections for future years based on extrapolations provided by the individual groups based their experience during 2005-2006. The CW demand is counterbalanced by effort in the EU and using the Einstein@home screen saver environment.

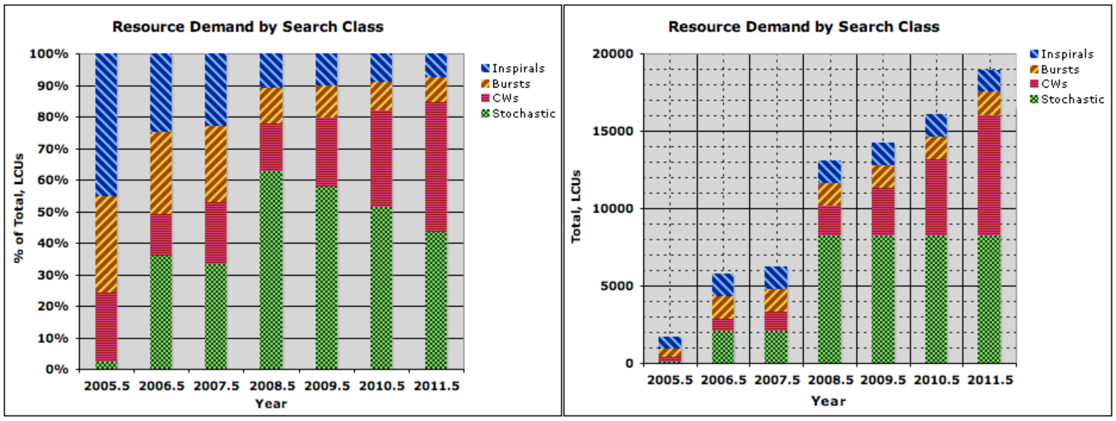

Figure 4: Bar graphs showing apportionment of resources by year and by search class. The fraction used by common infrastructure computation shared by all searches (*e.g.*, RDS production, data distribution, etc.) is negligible by comparison and is therefore not shown. The total resources needed each year increase monotonically over time.

LSC users are required to register their x.509 certificates for access to the LSC Datagrid via its website [5]. Once they are registered, they have unrestricted access to all Datagrid resources. Resource allocation will be performed by the LSC Computing Committee based on recommendations from the Data Analysis Committee. Up to the present, the LSC has not had to impose quotas on local storage and resource usage; however this is likely to change as competition for resources increases. Highest priority will be assigned to production runs from the analysis working groups in accordance to their projected work loads. Users developing algorithms and testing or validating their code will be assigned lower priority. Alternatively, a number of Datagrid sites maintain smaller development and test resources that may be used for non-production work.

The LSC grid computing system enables processing to take place at remote sites; it is the intent to eventually permit the use of multiple sites for a single job using VDT components. The grid model makes the system extensible: non-LSC grid resources available on the Open Science Grid (OSG)[6] may also be utilized by LSC applications that have been fully grid-enabled.

Resource Model for Analysis

The required resources for analysis are dominated by the production analysis runs. While at the present time, the post-processing analysis by individuals is frequently performed locally on resources directly accessible or controlled by individual analysts, it is the case that some post-processing analyses have reached a level of complexity that has required inclusion of the post-processing into the cluster-based production analysis. The 2005-2006 actual resource demand and future projections factor this into the estimates.

As stated earlier, the present Computing Model is based on accrued experience with production-level data analyses that were performed for all classes of searches for the S1 through (the ongoing) S5 science runs. Thus the inputs to the model are based on empirical data and thus do not rely on theoretical estimates of MFLOPS per analysis. The end-to-end parameter values include all code inefficiencies, data I/O inefficiencies, and GFLOP/GHz conversion factors that are specific to each of the analyses. It can be reasonably expected that with increased competition for finite resources, that all the analysis pipelines can be further refined to improved computational efficiencies. This potential margin derived by performance increases is not explicitly taken into account. If the various pipelines are successful in achieving performance enhancements, this represents an additional margin that can be applied to the estimated resource requirements.

The LSC Computing Committee conducted a survey of historical use based on mining Condor log files. Thus the historical perspective is accurate. In addition, each of the data analysis working groups was asked to provide as accurate as possible information on the resource usage by each group. The information gathered for the current usages included the following:

- A measure of end-to-end CPU-days used for the S3, S4, and early S5 analyses based on Condor log files. Since the LSC is now fully in production analysis, the quantity *CPU-days per science run day* could be determined.
- The fraction of the analysis that was run at each of the LSC Datagrid sites. This information, along with the hardware characteristics at each of the sites (*e.g.*, # of CPUs, CPU clock speed, CPU RAM, CPU local disk storage, etc.) provides the resource requirements in fundamental units, denoted in terms of **LCU**s (LIGO Computational Unit). The LCU is a measure computational efficiency of the hardware in terms of *e.g.* GFLOPS/GHz.
- The approximate volume of data products produced by the analysis (in the current version of this Plan, this was determined to be sufficiently small to not be included in the resource requirements projections)
- Common infrastructure production needs. This accounts for hardware resources needed to support control room (DMT) analysis, RDS production, and near real time data distribution across the Datagrid and simulated data sets.

Distributed analysis system

The LSC Datagrid distributed analysis system enables users to submit jobs from any location with processing to take place at remote sites. Data product results are available after the job finishes and are then staged by the users to local resources for post-processing.

The distributed analysis system is continuing to evolve at the present time. Several efforts are under way within the collaboration to fully grid-enable analysis pipelines to use the VDT more fully and to thereby allow the abstraction of the job submission process so users do not need to log on to specific sites to run their jobs. Currently, the inspiral and stochastic groups are working on this.

3.6 Simulation Process

Software simulation for production analysis serves two purposes, namely, (i) pipeline validation to demonstrate that known signals can be detected and extracted with parameters consistent with the injected ones, and (ii) astrophysical source modeling to determine search algorithm detection efficiency. Simulation analyses are integrated into the production pipelines for each of the searches and are included in the end-to-end resource utilization numbers provided by the groups. Note that the simulations considered here serve to quantify detection efficiencies of our searches and are not intended to simulate instrumental performance of the LIGO interferometers: this latter task is addressed by the commissioning activities of the Laboratory and are not in the scope of this document.

3.7 Calibration

Calibration information was originally generated in the frequency domain in the form of data files containing the (complex) response function relating $AS\_Q(f)/h(f)$. Frequency domain calibration computing overhead is borne separately by each pipeline. Data storage required associated with the calibration files is of order few $\times$ 10 MB.

In addition, near-real time generation of calibrated $h[t]$ time series frame data are generated. Several analysis groups are migrating from frequency to time domain calibration. The computation of $h[t]$ takes place as part of the common infrastructure production analysis and is included in the estimated resource need.

Calibrated data

As discussed above, $h[t]$ data are written to frames. The associated data volume was presented in Table 2. The latency for the availability of calibrated $h[t]$ frames is determined by there 1024s duration. There is an additional latency to publish and distribute these data over the WAN.

Offline calibration

Experience with all science runs to date has been that the final calibrations for all interferometers require two to three iterations. Intermediate versions of calibrations are published periodically after the completion of the science runs, typically over the course of the 12 months following a run. It is expected that convergence to final calibrations will require fewer and faster iterations.

Each iteration that is published requires, in principle, an end-to-end re-analysis of the production data. These iterations using frequency domain calibrations were included in the resource usage information provided by the analysis groups. As the collaboration migrates to using time domain calibrations more extensively, there will be required a concerted production run to produce new calibrated RDSs off-line that will then be distributed for re-analysis. The trade-off in cost between distributed frequency domain calibration inside multiple analysis pipelines and a one-time cost of producing calibrated time domain frames commonly used by multiple analyses is a favorable one. It has not been explicitly included in the projections; however, as discussed earlier, the net computational overhead for $h[t]$ production is a small component of the end-to-end production resource utilization.

4 Commissioning of Datagrid Upgrades and Improvements

The LSC Datagrid has grown and evolved concurrently with its utilization as a production system over the course of the first four science runs. It will continue to evolve as more grid-enabled functionality is integrated into the Datagrid.

Validation of new features is typically performed between major science runs either as part of the production analysis use of the Datagrid. Features required to support science running are validated during periods before the science run, typically during preceding engineering runs. Interleaved engineering and science run operations have permitted, in the past, the gradual upgrade of resources without serious disruption of services.

As LIGO moves into longer science runs, this will become progressively more difficult to execute without disruption. Past experience has shown that rolling deployment of infrastructure software during actual data-taking periods is a painful experience that should be avoided if at all possible. In the future, the LSC Computing Committee, in coordination with DASWG, will need to plan carefully how upgrades and patches will be deployed during run time.

The goal will be to retain as much flexibility as possible during the early stages of a long science run in the software and processing chain and in the use of the Datagrid resources.

5 Resource Requirements

Computing requirements for the remainder of the Initial LIGO observational era

The purpose of this plan is to assess where the LSC Datagrid stands after four science runs and while it is in production with its longest ever run, S5. This allows us to estimate as closely as possible the magnitude of resources that will be needed to support data analyses for the period between now and the commencement of Advanced LIGO operations in 2012+. Experience during the most recent science run suggests that the interferometers should be able to operate in science mode $\lesssim$ 66% of the time: 1 year of data will require $\gtrsim$ 1.5 years of data taking. Thus, in practice, the science data acquisition rate is reduced from the full data rate proportionately, and the collaboration will have $\gtrsim$18 months to analyze 12 months of science data; this represents an overall safety margin that is not included in the projected resource requirements.

The computing needs that were considered in this analysis were those for the production analyses and common infrastructure data production requirements. As discussed earlier, post-processing is assumed to be accommodated by resources at the disposition of individual users at home institutions.

Compact Binary Inspiral Search Requirements

The searches for gravitational waves from the inspiral phase of compact binary coalescence use matched filtering methods to search the data stream. These matched filters depend on the masses and spins of the individual components as well as coalescence time and a constant phase factor. When multiple instruments are combined optimally, the parameter list increases to include polarization angle, inclination and sky location.

The working group on searches for coalescing binary inspirals reviewed their usage history using S4 and the presently ongoing S5 science run. They considered needs for four classes of inspiral searches, resulting in a total estimate of 759 LCUs:

1. Binary Neutron Stars: 237 LCUs to keep up with data for all three instruments;
2. Binary Black Holes: 244 LCUs to keep up with data for all three instruments;
3. Primordial Black Holes: 950 LCUs to keep up with data for all three instruments;
4. Spinning Black Holes: a new search not yet performed, estimated at $\sim$237 LCUs.

Burst Search Requirements

The searches for gravitational waves from transient or unmodeled sources cannot rely on optimal matched filtering techniques. Rather, simpler methods based on wavelet analyses, excess power detection and other forms of frequency-time searches of periodogrammed data are employed. During S5 and beyond, coherent techniques will be applied that analyze data from multiple instruments that are combined optimally. In this case, the character of the search becomes more complex because a coherent network based analysis requires pixelation of the sky, bringing with it greater computational complexity that lends itself well to cluster-based computing.

The burst working group reviewed its historical computer use for S4 and the ongoing S5 analyses and determined that, based on their growing sophisticated analysis pipelines, usage will grow in the near term and then stabilize during 2007 and beyond. The largest users within the burst group reviewed their usage totals for past runs, recognizing that there were learning curves and inefficiencies not likely to recur when looking forward. In September 2005, the burst group pulled together a data analysis and computing plan for the S5 run. A static copy of that plan from the following web site:
http://www.lsc-group.phys.uwm.edu/bursts/investigations/s5/misc/incoming/BurstS5Computing_20050930.html.

For the original plan in September, the group estimated their computing needs in units of "1 GHz CPUs", and the aggregate total in the plan was about 380 LCUs. However, since that time, the development of coherent analysis methods has progressed since that time, and it seems that we will have at least three such methods in use, requiring more CPU time. Including a 30% margin, they estimated their need as ~1250 LCUs, or about 15% of the 2006/2007 extant computational capacity. They were not able to estimate eventual need driven by joint analyses with Virgo in the coming years.

In summary, they estimate:

1. "Standard" S4-type analyses would require 380 LCUs to keep up with the data;
2. Cross-correlation techniques are estimated to required 300 LCUs;
3. Maximum likelihood coherent searches are estimated to require 300 LCUs;
4. Including a 30% margin, the total comes to 1250 LCUs.

Periodic Search Requirements

The searches for gravitational waves from rapidly rotating neutron stars represents the most computationally intense class of data analysis challenges facing the LSC. For a very long time to come, the ability to perform large-area, deep searches will be computationally bound and thus will always be conducted in a suboptimal manner (e.g., looking a very restricted subset of the full parameter space). Moreover, as mentioned elsewhere in this document, the periodic searches group has successfully deployed a completely different computing model based on opportunistic computing derived from a screensaver program that is freely available for download by the worldwide community of home PC owners.

The bulk of periodic search analysis is conducted by the GEO portion of the LSC and uses EU resources available to the collaboration. These are discussed briefly below. GEO operates three clusters in Golm, in Birmingham and in Cardiff. Generally, these clusters have been utilized as follows.

- Golm (Merlin cluster): is a very highly utilized cluster dedicated to CW searches. 50% of the Fstat isolated search was run on this cluster, and the entire Hough search. All the extensive Monte Carlo simulation and algorithm development for these two searches were also performed on Merlin. In general the Hough search and all the coherent Fstat search developments are done on Merlin. Note that this includes the current multi-IFO Fstat and Hough searches. Merlin was funded by MPG particularly for CW work. So CW work has priority on this cluster, then comes other LSC work and finally usage from people in the astrophysical relativity division at AEI (this adds up to less than a few percent of merlin's usage). E@h runs on the Merlin using its idle cycles.

- Birmingham: the Fstat Sco X-1 search and algorithm development was performed on the cluster at Birmingham. The facility is not fully dedicated to LSC data analysis.

- Cardiff: Pipeline tuning and algorithm development for inspiral searches have been performed on this cluster. This is also a shared facility with the Cardiff physics department.

The group has evolved to a hierarchical analysis approach that leverages Einstein@home for the heavy lifting work of blind wide parameter space searches. The group relies on LDG facilities for:

- extensive algorithm characterization and tuning;
- smaller parameter space searches: time-domain searches of known objects or small parameter space searches of interesting targets (bright X-ray sources, SN remnants...)
- high sensitivity follow-up studies;
- "Fast" scan of the data with incoherent methods
  * for data characterization needs (for monitoring of spectral artifacts, for checking the quality of the data close frequencies where interesting targets lie).
  * to identify candidates to look more closely at with coherent methods.

Based on a continued heavy reliance on GEO EU resources, the group predicts computing requirements for the U. S. portion of the LDG to be at a level of 2200 LCUs for 12 weeks every 6 months during the next year. The projection then assumes that in the out-years sustained Moore's Law doubling of CPU capacity every 18 months will enable the group to continue to expand its search of the vast parameter space.

The outlook over the next $\sim 5+$ years is predicated on the following assumptions:

- That Einstein@hone continues to flourish and provide sustained capacity that is unmatched by the LDG;
- That GEO continues to support computing infrastructure at the level that has been possible to date.

Stochastic Search Requirements

The search for stochastic gravitational waves of either cosmological or astrophysical origin began as a small few-workstation analysis and has grown over the course of the S4 and S5 analyses to represent a significant computational need looking to the future. The reasons for this are several: (i) the signal preprocessing and signal conditioning required to enable the LSC to use the collocated H1-H2 detector pair has been recognized to be a potentially large computational challenge; (ii) the transition from single-result 4π sky-averaged measurements to spatially and frequency resolved measurements of the stochastic GW sky (*à la* CMB analyses). Moreover, the stochastic search requires the direct cross-correlation of pairs of instruments. As the global network flourishes, the baseline pairs will continue to grow as $N(N-1)/2$. Thus, going from LIGO to LIGO+Virgo implies a doubling of interferometer pairs for this analysis.

In summary, the major factors estimated by the stochastic working group included,

Year	LDG Capacity LCUs	LDG Usage LCUs	LDG Unmet Need, LCUs
2006	3621	1727	-
2007	8419	5746	-
2008		6201	-
2009		13087	4668
2010		14233	5813
2011		16051	7631

Table 3: LDG capacity, use, and future unmet needs in units of LCUs.

1. ∼50x increase in computational power for the next 2 years due to performing more all-sky searches, using the current implementation of the gravitational wave radiometer search that does not require deconvolution, and allowing for calculation of strain-environmental coherences for an H1-H2 analysis. This also includes LIGO-Virgo and Virgo-GEO all-sky analyses;
2. ∼200x increase for the next period starting in mid 2008 and beyond once Virgo comes on line and a full spatially resolved analysis over multiple baselines is performed. Computational load also is expected to come from a fuller utilization of the cross-correlations with environmental and non strain interferometer channels in order to remove the H1-H2 environmental cross-correlations.

Summary of computing needs from all working groups

It is important to emphasize that the resource requirements discussed above represent scalings from demonstrated performances based on prior science data production experience. The model uses actual hardware inventories through mid CY2005 and includes upgrades beyond CY 2005 that are known to be likely. The deficit between available resources and projected needs in the out years reflects a need that must be addressed by the collaboration if it is to fully exploit the data from a long science run on the time scale that is required.

Figure 5 summarizes the current (2005/2006) and projected resource requirements for future years. The requirements for each of the search groups are shown, along with the totals. Shown for 2005/2006 and 2006/2007 are the existing LDG U. S. resources. Based on the projections presented here, the known and planned LSC Datagrid resources available during CY2006/2007 will be adequate to meet the expected need through 2008. As quantified in Table 3, beyond this time, the LSC must provide additional computing that represents a ∼56% increase in 2009 (an additional ∼4700 LCUs), with a ramp up beyond that time. To set the scale, 4800 LCUs today would represent ∼ 600 2X dual-core 2.2 GHz CPU nodes, costing a projected ∼$1M in FY2008.

Storage needs

The LIGO Laboratory Tier 1 Center at CIT houses the permanent archive of LIGO data. The Tier 2 sites rely on spinning storage for local persistent copies of RDSs. The model developed by the LSC is one in which storage is collocated with the computing resources, inside the cluster architecture. For example, the most recent CIT cluster purchase (end of 2005) included 220 TB of RAID 5 storage within the cluster. In the coming years, the trend in disk storage performance is such that it is not expected to be a resource issue for the LSC. It is possible today to purchase a node for a cluster that also serves as a RAID 5 3TB storage subsystem.

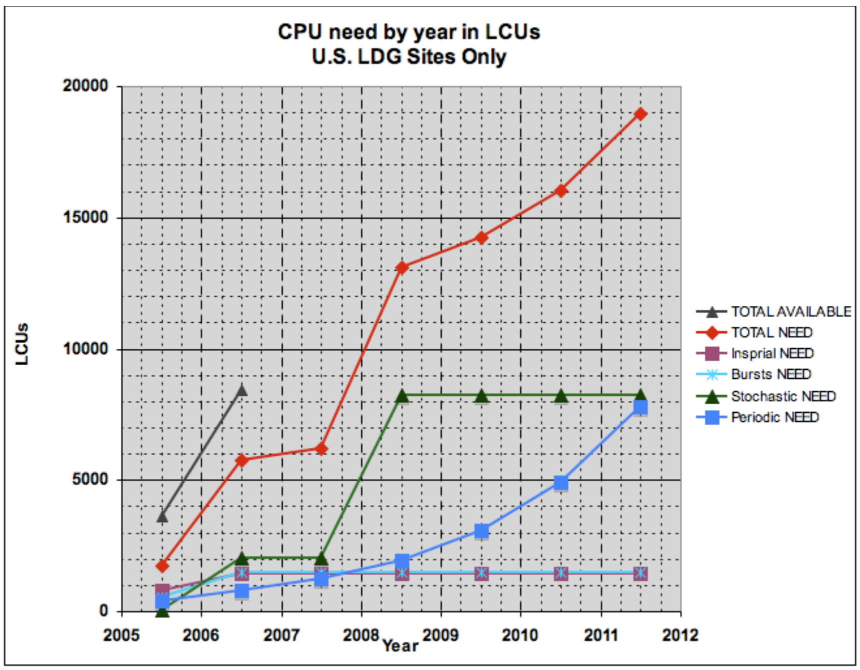

Figure 5: Graph of projected resource requirements to support science running through 2012. Bursts and inspiral searches expect to saturate their need in the next year or two. Periodic searches expect to follow a Moore's Law curve as they continue to perform larger sub-optimal searches of the parameter space. Stochastic searches are expected to increase in complexity with the availability of Virgo data (here assumed to be 2008+) due to the increase number of baselines and the implementation of full spatially resolved and frequency resolved searches.

6 Networking Requirements

Tier 0 $\leftrightarrow$ Tier 1 WAN

The LIGO science data bandwidth requirements are modest by comparison with large HEP (*e.g.*, LHC) experiments. The L0 - L4 data volume can be accommodated over 1000 Mbps Ethernet-over-fiber with $\sim 400\%$ headroom.

Tier 1 $\leftrightarrow$ Tier 2 WAN

The university Tier 1, 2 sites are all connected to the US NSF infrastructure, Abilene/Internet2. Eventually the National Lambda Rail will further increase the inter-university bandwidth.

Connections to European Tier 2 Datagrid sites from the U.S. is provided via the EU GEANT infrastructure.

7 Summary

The LSC Datagrid computing and data handling capabilities for 2005-2006 have been measured and quantified. These inform realistic projections based on actual performance metrics determined by each of the LSC data analysis working groups who require access to major Datagrid resources.

The recent computing facility upgrades at CIT, LHO, LLO, MIT and UWM provide the LSC with margin in computing capacity relative to need for the next year or two. The growth in requirements beyond 2008 correspond to an increase in Datagrid capacity $\sim$ 1.56x the current configuration. This factor is estimated by comparing existing and known, planned upgrades of Datagrid resources versus the projections provided by the analysis working groups.

In coming up with these estimates, the periodic sources group made reasonable assumptions regarding the availability of EU GEO resources and the continued flourishing of Einstein@home. These off-load the computational burden for the U. S. portion of the LDG.

The LSC is also actively pursuing the development and deployment of the national Open Science Grid (OSG). The OSG is a follow-on to the current NSF and DOE funded grid R&D projects (PPDG, GriPhyN, iVDGL). If successful, the OSG promises to provide a natural extension of the current LSC Datagrid that will provide additional computational capacity, especially for peak or excess capacity due to compute-intensive searches, when this capacity is not within the scope of supporting the collaboration's dedicated production capability.

Last, this Plan serves as the basis to develop more detailed procurement needs model for LSC resources at all hierarchies of the LSC Datagrid.

References

[1] Stuart Anderson, Keith Riles, Albert Lazzarini, Tom Nash, Alan Weinstein, and Alan Wiseman, *LSC Data Analysis White Paper Draft V*, LIGO Laboratory document T990104-05 available at http://www.ligo.caltech.edu/docs/T/T990104-05.pdf

[2] B. C. Barish and R. W. Weiss, Phys. Today, 52, 44 (1999).

[3] http://www.ligo.caltech.edu/docs/L/LIGO-T970130-F.pdf.

[4] http://www.ligo.caltech.edu/docs/L/L030155-00.pdf.

[5] http://www.lsc-grid.uwm.edu

[6] http://www.opensciencegrid.org

[7] SETI@home is is a scientific experiment that uses Internet-connected computers in the Search for Extraterrestrial Intelligence (SETI). You can participate by running a free program that downloads and analyzes radio telescope data. URL: `http://setiathome.ssl.berkeley.edu`.

[8] http://www.physics2005.org/events/einsteinathome/

[9] http://boinc.berkeley.edu/

A Details of the Resource Usage

This section contains printouts of a detailed spreadsheet model that was used to make the tables and graphics in this Plan.

The first page contains a summary of current and envisioned resources of the LSC Datagrid. It includes known and planned upgrades. Each of the U.S. tiers is listed separately, while the EU member tiers are grouped together.

The last three pages list the information provided by the analysis working groups (light blue areas of the spreadsheet), along with an extensive set of calculations that are based on these inputs. Each working groups is listed separately. The common infrastructure is shown last. Below that are compilations by analysis group and by site.

BURST SEARCHES										
Year	Factor w.r.t 2005 - 2006	GROUP NEED - LCUs in that year	% OF LCU USAGE OR NEED BY U.S. SITE							
			MIT	LLO	LHO	UWM	PSU	CIT	Comments	% of Total LDG use or need
2005 - 2006	1	530	3%	29%	12%	3%	14%	39%	Actual usage	31%
2006 - 2007	2.8X	1484	5%	13%	13%	25%	10%	34%	Provided by P. Shawhan	26%
2007 - 2008	2.8X	1484	5%	13%	13%	25%	10%	34%	Provided by P. Shawhan	24%
2008 - 2009	2.8X	1484	5%	13%	13%	25%	10%	34%	Provided by P. Shawhan	11%
2009 - 2010	2.8X	1484	5%	13%	13%	25%	10%	34%	Provided by P. Shawhan	10%
2010 - 2011	2.8X	1484	5%	13%	13%	25%	10%	34%	Provided by P. Shawhan	9%
2011 - 2012	2.8X	1484	5%	13%	13%	25%	10%	34%	Provided by P. Shawhan	8%

PERIODIC SEARCHES										
Year	Factor w.r.t 2005 - 2006	GROUP NEED - LCUs in that year	% OF LCU USAGE OR NEED BY U.S. SITE							
			MIT	LLO	LHO	UWM	PSU	CIT	Comments	% of Total LDG use or need
2005 - 2006	1	375	0%	17%	29%	2%	5%	46%	Actual usage	22%
2006 - 2007	2.1X	774	0%	0%	7%	23%	35%	35%	Provided by M.A. Papa	13%
2007 - 2008	3.3X	1228	0%	0%	8%	26%	26%	39%	Provided by M.A. Papa	20%
2008 - 2009	5.2X	1950	0%	0%	8%	26%	26%	39%	Provided by M.A. Papa	15%
2009 - 2010	8.2X	3095	0%	0%	8%	26%	26%	39%	Provided by M.A. Papa	22%
2010 - 2011	13.1X	4914	0%	0%	8%	26%	26%	39%	Provided by M.A. Papa	31%
2011 - 2012	20.8X	7800	0%	0%	8%	26%	26%	39%	Provided by M.A. Papa	41%

STOCHASTIC SEARCH										
Year	Factor w.r.t 2005 - 2006	GROUP NEED - LCUs in that year	% OF LCU USAGE OR NEED BY U.S. SITE							
			MIT	LLO	LHO	UWM	PSU	CIT	Comments	% of Total LDG use or need
2005 - 2006	1	41	####	5.3%	27.2%	0.0%	0.0%	42.5%	Actual usage	2.4%
2006 - 2007	50.0X	2055	5%	13%	13%	25%	10%	34%	Provided by J. Romano & A. Lazzarini	35.8%
2007 - 2008	50.0X	2055	5%	13%	13%	25%	10%	34%	Provided by J. Romano & A. Lazzarini	33.1%
2008 - 2009	200.0X	8220	5%	13%	13%	25%	10%	34%	Provided by J. Romano & A. Lazzarini	62.8%
2009 - 2010	200.0X	8220	5%	13%	13%	25%	10%	34%	Provided by J. Romano & A. Lazzarini	57.8%
2010 - 2011	200.0X	8220	5%	13%	13%	25%	10%	34%	Provided by J. Romano & A. Lazzarini	51.2%
2011 - 2012	200.0X	8220	5%	13%	13%	25%	10%	34%	Provided by J. Romano & A. Lazzarini	43.4%

Figure 6: Table of current and expected LSC Datagrid resources 2005/2006 - 2011/2012

INSPIRAL SEARCH										
Year	Factor w.r.t 2005 - 2006	GROUP NEED - LCUs in that year	% OF LCU USAGE OR NEED BY U.S. SITE							
			MIT	LLO	LHO	UWM	PSU	CIT	Comments	% of Total LDG use or need
2005 - 2006	1	781	0.9%	10.6%	28.8%	13.8%	1.4%	44.5%	Actual usage	45.2%
2006 - 2007	1.8X	1434	5%	13%	13%	25%	10%	34%	Provided by P. Brady	25.0%
2007 - 2008	1.8X	1434	5%	13%	13%	25%	10%	34%	Provided by P. Brady	23.1%
2008 - 2009	1.8X	1434	5%	13%	13%	25%	10%	34%	Provided by P. Brady	11.0%
2009 - 2010	1.8X	1434	5%	13%	13%	25%	10%	34%	Provided by P. Brady	10.1%
2010 - 2011	1.8X	1434	5%	13%	13%	25%	10%	34%	Provided by P. Brady	8.9%
2011 - 2012	1.8X	1434	5%	13%	13%	25%	10%	34%	Provided by P. Brady	7.6%

ALL SEARCHES									
Year	Factor w.r.t 2005 - 2006	GROUP NEED - LCUs in that year	% OF LCU USAGE OR NEED BY U.S. SITE						
			MIT	LLO	LHO	UWM	PSU	CIT	Comments
2005.5	1	1727	2%	18%	24%	8%	6%	43%	Actual usage
2006.5	3.33X	5746	4%	11%	12%	25%	13%	34%	Derived from inputs
2007.5	3.59X	6201	4%	10%	12%	25%	13%	35%	Derived from inputs
2008.5	7.58X	13087	4%	11%	12%	25%	13%	35%	Derived from inputs
2009.5	8.24X	14233	4%	10%	12%	26%	14%	35%	Derived from inputs
2010.5	9.29X	16051	3%	9%	11%	26%	15%	36%	Derived from inputs
2011.5	10.96X	18937	3%	8%	11%	26%	17%	36%	Derived from inputs

Figure 7: Table of current and expected LSC Datagrid resources 2005/2006 - 2011/2012, continued from Figure 6