The LIGO Science Program and iVDGL

Albert Lazzarini
LIGO Laboratory
California Institute of Technology
lazz@ligo.caltech.edu

NSF Review of iVDGL
February 10-11, 2004

LIGO:
Laser Interferometer Gravitational-wave Observatory

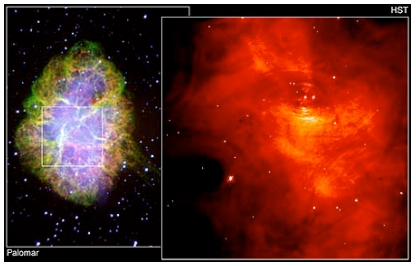
- LIGO is opening a new frontier in observational astrophysics
  - Detect & use gravitational waves (GW) to observe the Universe, provide a more complete picture of the Cosmos.
  - Complementary to radio/infrared/optical/X-ray/γ-ray astronomy (electromagnetic waves --- EM)
    - EM emitters not likely to be strong GW emitters & vice versa
  - Detect & observe cataclysmic events leading to death of stars, birth of neutron stars & black holes
  - Opportunity to observe gravitational radiation, study Einstein's Theory of General Relativity in the strong-field regime
    - Vicinity of massive compact objects, where GW are produced
- ***LIGO is now observing, acquiring science data and in full analysis production***

Realizing the full science potential of LIGO

The search for gravitational waves at hrms~ 10^{-21}

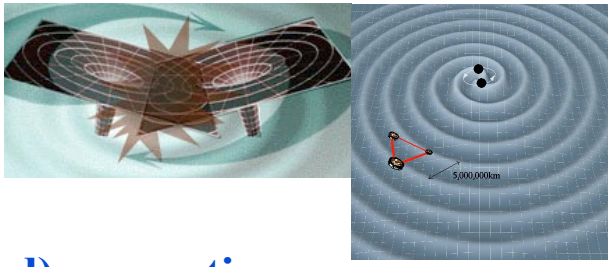
- Revealing the full science content of LIGO data is a computationally and data intensive challenge
  - Several classes of data analysis challenges require large-scale computational resources

❑ Search for gravitational wave (GW) analogs of electromagnetic (EM) pulsars



- ❑ GW sources not likely to have EM counterparts
  - ❑Fast (millisecond) EM pulsars are stable, old neutron stars (NS)
- ❑ GW emission likely to come shortly after birth of a rapidly rotating (deformed, hot) NS
- ❑ GW sky is **unknown**
  - ❑Searches will need to survey a large parameter space
- ❑ All-sky search for previously unidentified periodic sources requires > 10^{15} floating point operations per second (FLOPS)
- ❑ Follow up for Einstein@home candidates.

❑ Coalescence of compact binary systems ("inspiral chirps") which include spin-spin interactions will cover a huge parameter space (~ 10^6 greater than spinless systems)



- ❑Important for more massive systems
- ❑Massive systems have greater GW luminosities
- ❑Likely to be the first detected

❑ **These analyses are ideally suited for distributed (grid-based) computing**

LIGO -- iVDGL Participants

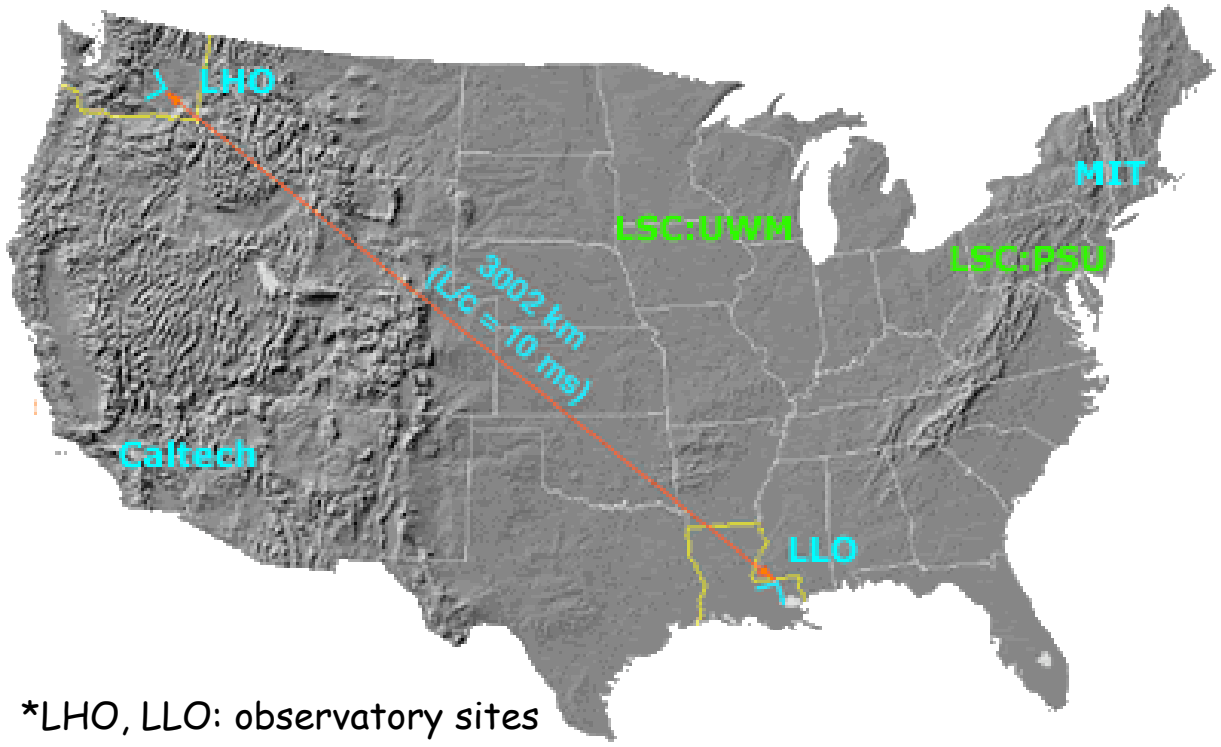
- LIGO Laboratory, Tier 1&2 operations
  - Stuart Anderson (CIT)
  - Keith Bayer (MIT)
  - Kent Blackburn (CIT)
  - Teviet Creightion (CIT)
  - Philip Ehrens (CIT)
  - Ben Johnson (CIT)
  - Erik Katsavounidis (MIT)
  - Dan Kozak (CIT)
  - Albert Lazzarini (CIT)
  - Edward Maros (CIT)
  - Greg Mendell (CIT)
  - Hari Pulapaka (CIT)
  - Peter Shawhan (CIT)
  - Igor Yakushin (CIT)
- UW Milwaukee -- LIGO Scientific Collaboration, Tier 2 operations
  - Bruce Allen
  - ***Paul Armor (20%)***
  - Patrick Brady
  - Scott Koranda
  - ***Brian Moe (100%)***
  - Alan Wiseman
  - 8 undergraduates (admin/system config.)
- Penn State -- LIGO Scientific Collaboration, Tier 2 operations
  - Vijay Agarwala
  - L. Samuel Finn
  - ***Michael Foster (100%)***
  - Jason Holmes
  - Jeff Nucciarone
  - ***John McNabb (50%)***
  - Murali Ramsunder
  - ***Keith Thorne (50%)***
  - Hannah Williams
  - 5 undergraduates (data analysis/applications)

Supported by iVDGL
Other support (off-project)

The LIGO Scientific Collaboration and the LIGO Grid

LIGO Grid: 6 US sites + 3 EU sites (Cardiff/UK, AEI/Germany)

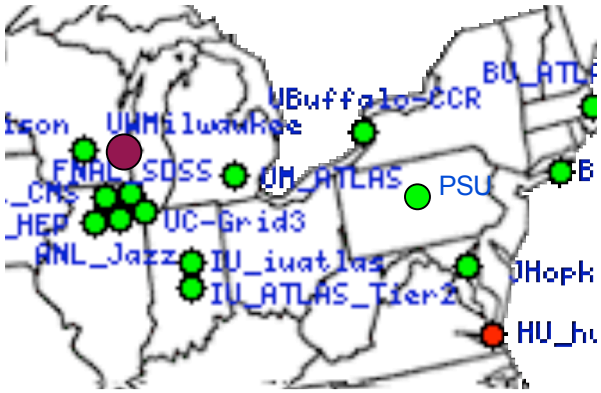
iVDGL has enabled the collaboration to establish a persistent production grid



*LHO, LLO: observatory sites

* LSC - LIGO Scientific Collaboration - iVDGL supported

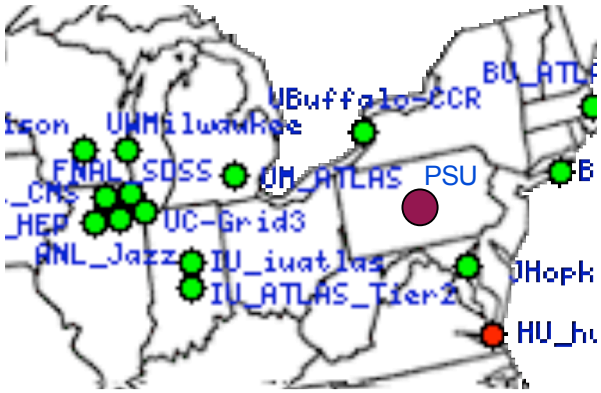
iVDGL Resources/Activities at UWM



- Original Tier 2 hardware from a prior NSF MRI grant
  - 296 single CPU (tower)nodes
    - New MRI upgrade proposal submitted
- iVDGL enabled ***upgrade*** of storage & ***1.2 FTEs*** for operation
  - Increased storage from 20 TB to 54 TB
    - Added 112 GB to each of 296 nodes in Linux cluster
  - Main iVDGL commitment to collaboration: provide complete LIGO S1, S2, and S3 reduced datasets to be available to collaboration with low latency
- UWM plays a key applications development role for iVDGL
  - Grid2003 Task Force
  - DOEGrids Registration Authority for iVDGL
  - VDT test harness development/testing/deployment support for iVDGL
- Grid3 participation:
  - Dedicated head node for iVDGL/Grid3
  - 3 levels of service in Condor pool set by user priorities
    - Highest (UWM) -> 1
    - Medium (LSC) -> 10^4
    - Lowest (iVDGL/Grid3) -> 10^6
- Recent usage:

User Name	Usage (node-hrs)	From	-	To
ivdgl@medusa.phys.uwm.edu	134357.09	12/03/2003 16:00		2/03/2004 01:12
btev@medusa.phys.uwm.edu	4514.92	12/03/2003 16:00		1/05/2004 16:44
usatlas1@medusa.phys.uwm.edu	4152.69	12/03/2003 15:59		1/29/2004 14:19
uscms01@medusa.phys.uwm.edu	4734.83	12/16/2003 22:40		2/03/2004 09:19

iVDGL Resources/Activities at PSU

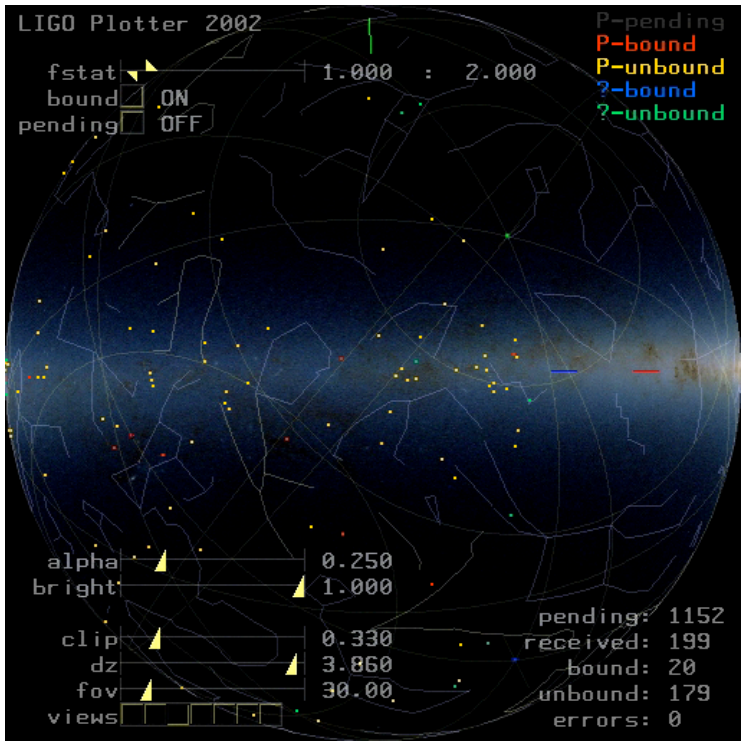


- Tier 2 ***facility*** & ***2 FTEs*** provided by iVDGL
  - 156 dual Xeon nodes
  - gigE switch & node-to-node connectivity
  - 34 TB RAID 5 storage
- SC2003/Grid3 -- System was still being commissioned during
  - Dedicated head node for LIGO data grid jobs
  - Single Grid3 test node for software and configuration tests within Grid3
- SOON -- Entire system will soon be registered under Grid3
  - Use policy: 3 levels of service in Condor pool set by user priorities
    - Highest (PSU) 50% of cycles on a weekly average
    - Medium (LSC) -> 40% of cycles on a weekly average
    - Lowest (iVDGL/Grid3) -> 10% on a weekly average
    - Per job limits:
      - 16 nodes/32 CPUs; 24 hrs; 1 GB user directories

SC03: Scale up of SC02 effort to production analysis

Blind search for periodic sources (GW pulsars) with S2 science data

- **Goal**: Implement a ***production-level*** blind galactic-plane search for GW pulsar signals
  - Run for 30 days on ~5X - 10X more resources than LIGO has -- using the grid (e.g., ***10,000*** CPUs for **1** month)
  - Would involve millions of individual jobs
    - Planning by Chimera/Pegasus
    - Execution by Condor DAGman
    - File cataloging by Globus RLS
    - Metadata by Globus MCS
- **Achieved**: Access to ~ 6000 CPUs for 1 week
  - ***~ 5% utilization due to bottlenecks***
- **Lessons learned:**
  - Performance bottleneck identified in job scheduling/management using head node
  - Globus jobmanager scaling issue on submit node
    - Could not handle 100s of jobs
    - Finer-grained control necessary in Chimera/Pegasus to handle local IO issues at cluster level
    - Job flow topology in LIGO application code (scripts) needed to be reworked
    - Later fixed using Condor GridMonitor
    - Recent performance measure: schedule ~1000 jobs: 296 jobs running,with 700 queued



Display of results in 3D projection of Milky way -- dots indicate pixels searched

Preparing for follow-on production run

SC2003/Grid3 participation

LSC DataGrid Service Details

Rows in blue are LSC resources. Click on the site name for information on getting access if you are a LSC scientist.

Rows in beige are sites for which special access has to be arranged by Albert Lazzarini.

Site	CPU count	login node	GridFTP server	batch jobmanager	RLS server	global scratch	GLOBUS_LOCATION
AEI	359	morgane.aei.mpg.de	morgane.aei.mpg.de	morgane.aei.mpg.de/jobmanager-condor	not yet	/home/NOBACKUP	/opt/ldg/globus
Birmingham	200	tsunami.sr.bham.ac.uk	tsunami.sr.bham.ac.uk	tsunami.sr.bham.ac.uk/jobmanager-condor	none	/raid/1/<login>	/opt/ldg/globus
Cardiff	160	mini.astro.cf.ac.uk	mini.astro.cf.ac.uk	mini.astro.cf.ac.uk/jobmanager-condor	not yet	none	/opt/ldg/globus
CIT	200	ldas-grid.ligo.caltech.edu	ldas-grid.ligo.caltech.edu	ldas-grid.ligo.caltech.edu/jobmanager-condor	not yet	?	/opt/ldg/globus
PSU	312	ligo-grid.aset.psu.edu	ligo-grid.aset.psu.edu	ligo-grid.aset.psu.edu/jobmanager-pbs	not yet	user home	/opt/vdt/globus
UTB	73	lobizon.utb.edu	lobizon.utb.edu	lobizon.utb.edu/jobmanager-condor	not yet	/home/<login>	/opt/ldg/globus
UWM	296	hydra.phys.uwm.edu	hydra.phys.uwm.edu	hydra.phys.uwm.edu/jobmanager-condor	rls://hydra.phys.uwm.edu:39281	/scratch	/opt/ldg/globus
AGT-Wisc	8	agt-login.cs.wisc.edu	agt-login.cs.wisc.edu	agt-login.cs.wisc.edu/jobmanager-pbs	none	/home/<login>	/vdt/globus
ISI	35	birdie.isi.edu	smarty.isi.edu sukhna.isi.edu sultan.isi.edu bindas.isi.edu pisa.isi.edu	columbus.isi.edu/jobmanager-condor birdie.isi.edu/jobmanager-condor skywalker.isi.edu/jobmanger-condor	rls://smarty.isi.edu rls://sukhna.isi.edu rls://skywalker.isi.edu	/nfs/cgt-scratch/griphyn	/nfs/v6/globus/GT2/linux/STABLE
Grid2003	~2400	details	details	details	details	details	details
LSU SuperMike	?	mike4.lsu.edu:2222	mike4.lsu.edu	mike4.lsu.edu/jobmanager-pbs	none	/exports/local6/sc2003_01	/usr/local/globus
Madison Condor Pool	800	none	beak.cs.wisc.edu	beak.cs.wisc.edu/jobmanager-condor	none	/shared/scratch/griphyn-ligo /shared/scratch2/griphyn-ligo/	/scratch/vdt

1600 (LIGO-owned) +
3243 (iVDGL-owned) +
~ 1024 (private - LSU/LA)
~ 5900 (total available)

- http://www.lsc-group.phys.uwm.edu/lscdatagrid/resources/index.html

Summary and Conclusions

To date LIGO has …

- ✓ Developed data replication, distribution capabilities over a data analysis grid
  - ✓ Near real time production of reduced data sets, transmission to Tier 2
    - ✓ Tens of TB over the internet
  - ✓ Provides access to reduced data, data replications, data mirrors
- ✓ Implemented the use of virtual data catalogs for efficient (re)utilization of data
  - ✓ Tracking data locations, availability with catalogs
  - ✓ Data discovery, data transformations
- ✓ Implemented a ***persistent data grid*** for the **international collaboration**
  - ✓ Access to distributed computing power - US & EU
  - ✓ Will eventually enable CPU-limited analyses as background jobs
    - ✓ Challenge: making full use of inherent CPU capacity
      - ❑ These efforts possible due to significant contribution made by the UW Milwaukee group to iVDGL deployment/testing

Plans

- Redeploy the GW pulsar search using more efficient script topologies
  - Saturate distributed grid resources
- Complete/extend the use of virtual data catalogs for efficient (re)utilization of data
  - Publish data as they become available automatically
    - Prototype exists in non grid-enabled internal code
    - Develop API to this module
- Enhance/extend ***persistent data grid*** for the collaboration
  - Add sites (Tier 3)
  - Add additional LIGO Laboratory (Tier 1) resources
- Actively participate in continued deployment and evolution of Grid3+
  - LIGO focus is necessarily on ***production*** issues
  - Open Science Grid (OSG) ?

FINIS