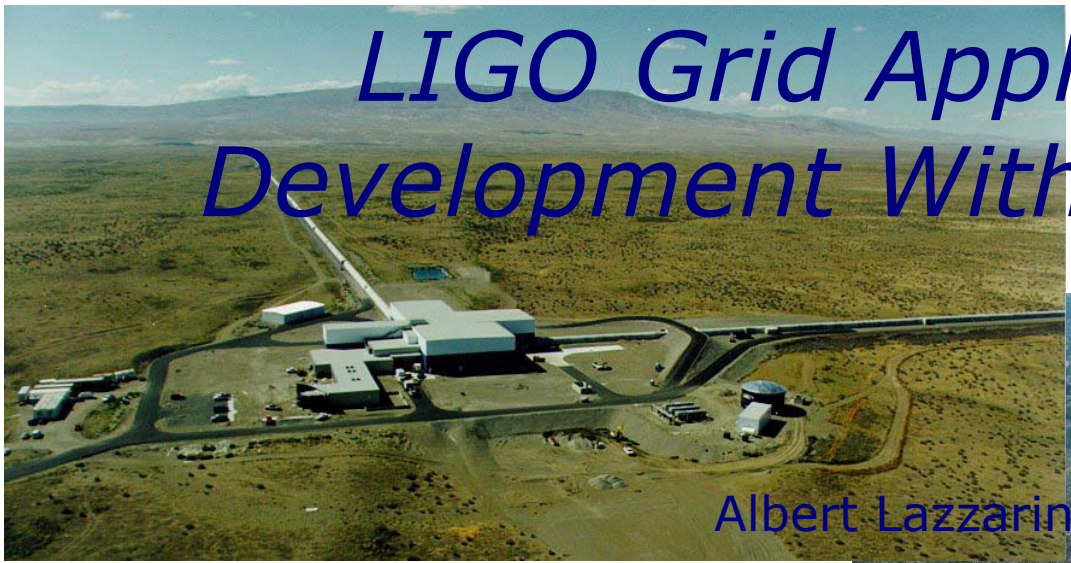
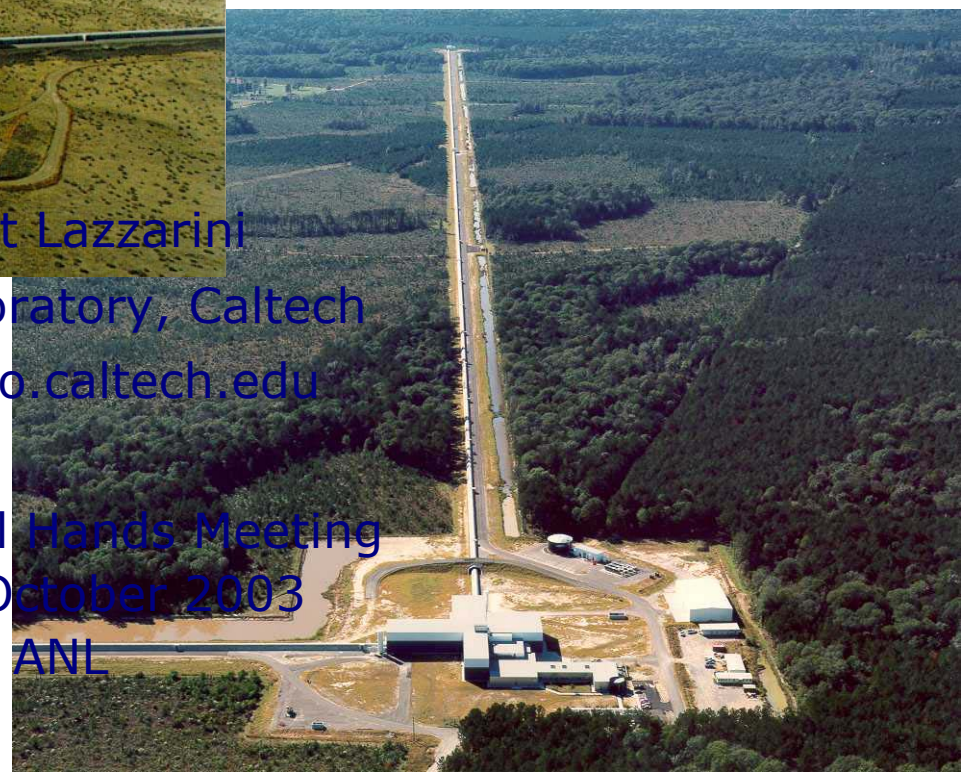
LIGO



LIGO Grid Applications Development Within GriPhyN

Albert Lazzarini
LIGO Laboratory, Caltech
lazz@ligo.caltech.edu

GriPhyN All Hands Meeting
15-17 October 2003
ANL



LIGO- G030537-01-E

LIGO Applications Development Team (GriPhyN)

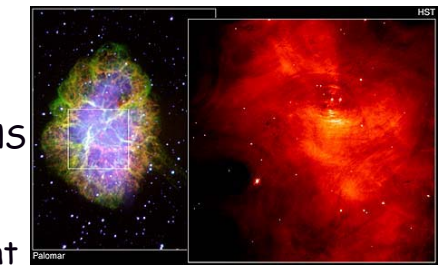
- LIGO Scientific Collaboration
  - LIGO Laboratory
  - UWM
  - AEI (Germany)
- ISI

LIGO and Grid Computing

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

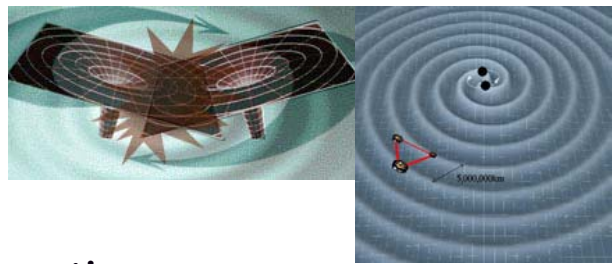
◆Search for gravitational wave (GW) analogs of electromagnetic (EM) pulsars

- ◆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)



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

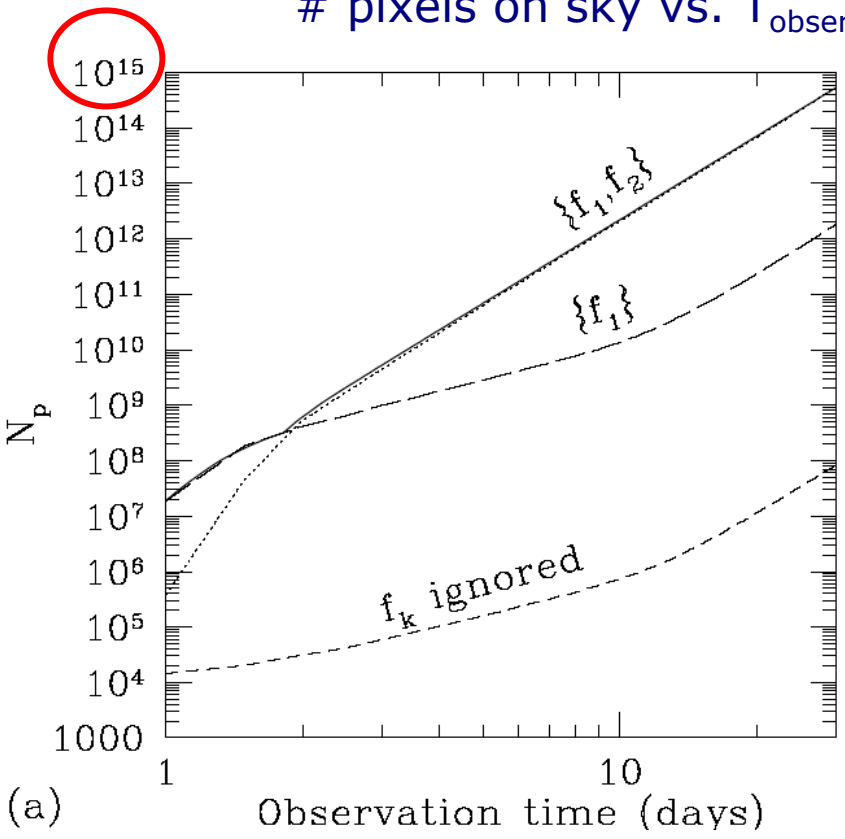
- ◆Important for more massive systems
- ◆Massive systems have greater GW luminosities
- ◆Likely to be the first detected



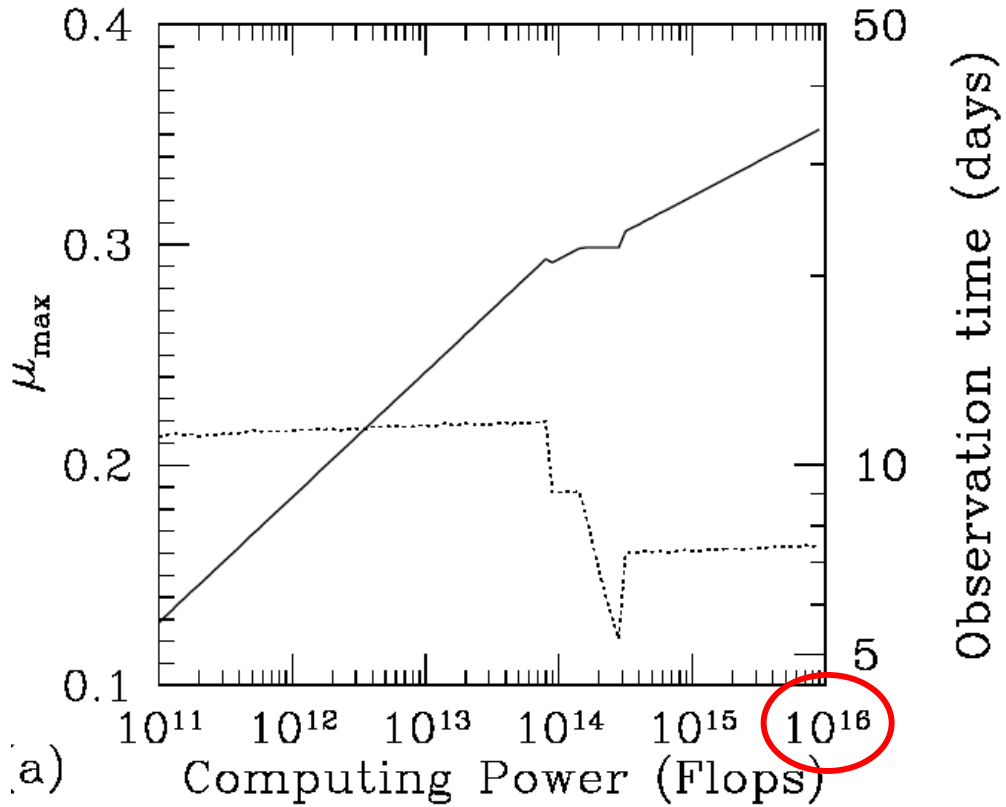
◆**These analyses are ideally suited for distributed (grid-based) computing**

Efficient search of entire sky for unknown GW pulsars (with no EM counterpart) requires > 10^{15} FLOPS

pixels on sky vs. $T_{observation}$



#FLOPS vs.



Figures taken from Brady et al., Phys. Rev. D 57, 2101 (1998)

LIGO GriPhyN Application:

Grid-deployed production-level GW pulsar search

The pulsar search conducted at SC 2002 -- TOY DEMO

- Used LIGO data subset collected during the first scientific run of the instrument
- Targeted a set of 1000 locations including both known EM pulsar as well as random locations in the sky
- Results of the analysis are available via LDAS (LIGO Data Analysis System) DB queries
- Performed using LDAS and compute and storage resources at Caltech, University of Southern California, University of Wisconsin Milwaukee.

QuickTime™ and a YUV420 codec decompressor are needed to see this picture.

SC 2002

- Over 58 directions in the sky searched
- Total of
  - 330 tasks
  - 469 data transfers
  - 330 output files produced.
- The total runtime:11.4 hr

Goal: Use SC2003 to initiate a ***production*** analysis run for GW pulsars

- Stand-alone search codes to be used for the SC2003
  - Perform frequency-time transformations of data
  - Parameters of transform depend on {θ,ϕ,(f, f ‘, f “,…)}
- Code originally developed by GEO project as part of LIGO collaboration
- The code has been modified to enable a user to specify input and output file names and parameters on the command line
  - Stage data to/from compute resources as needed
- Auto generator will approximately search the galactic core for pulsars
  - Specifies the range of parameters to be search
- Results will be shown on a 3D Visualization

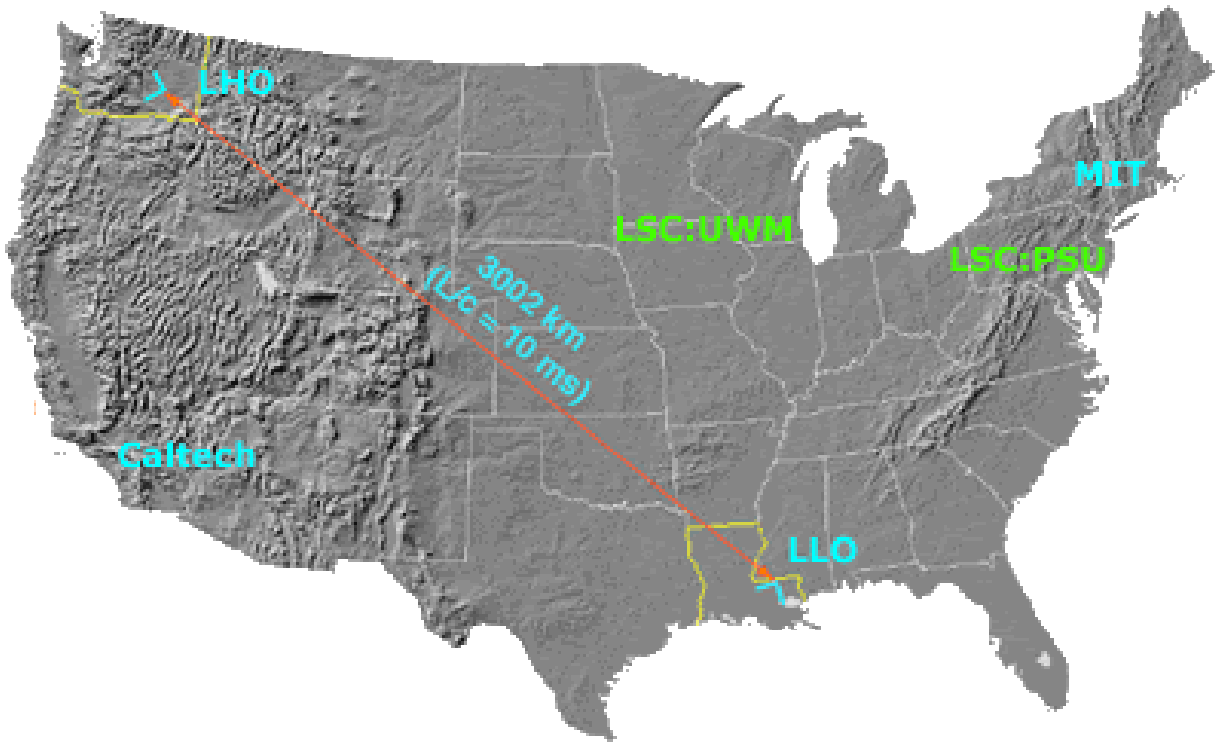
Grid-deployed ***production-level*** GW pulsar search

- Use SC2003 to implement ***production-level*** scalable version of SC2002 demonstration
- Original goal: Implement a ***production*** GW pulsar search over a patch (less than 4π of sky) search running for ~30 days on O[10x] more resources than LIGO has => use the grid (***10000*** CPUs for 1 month)
  - Subscale implementation using ONLY LIGO-owned resources in development
  - Joint development with ISI CS team.
- As additional resources become available, will scale production run to expand beyond LIGO resources.

LIGO Grid Sites

LIGO Grid: 6 US sites + 2 EU sites (Cardiff/UK, AEI/Germany)

Collaboration: 35+ institutions world wide; 400 members



SC2003 production run

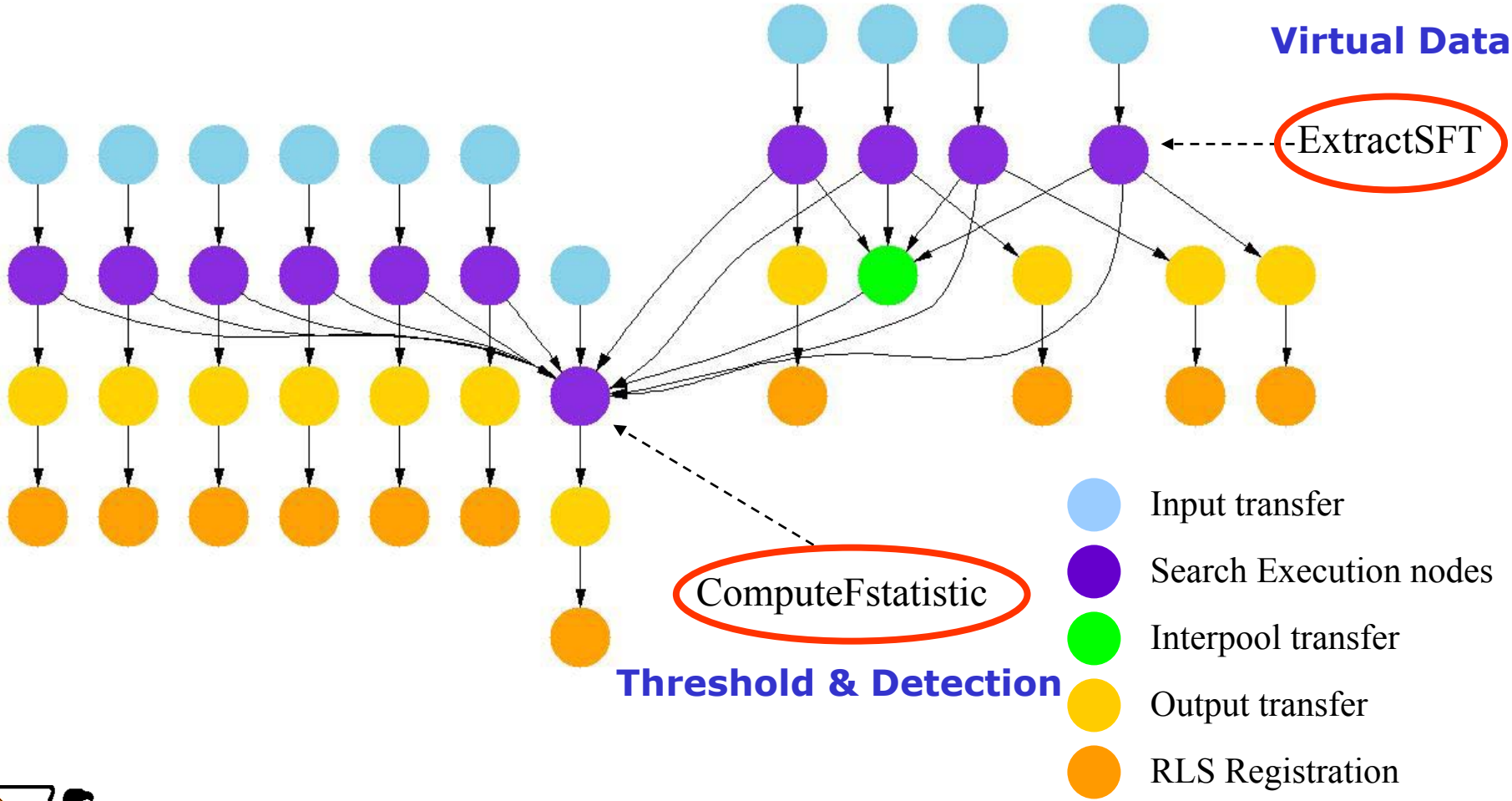
- Sites to be used
  - ✓ ISI Condor and LSF Clusters (035 CPUs)
  - ✓ LIGO CALTECH condor pool (220 CPUs)
  - ✓ UWM condor pool (300 CPUs)
  - ✓ AEI (Germany) condor pool (300 CPUs)
  - Cardiff condor pool (046 CPUs)
  - UW Madison condor pool (800 CPUs)
  - *?Teragrid resources (~1024 CPUs)*
  - *?Grid3 resources*

LIGO

Implementation details

- Web portal based authentication using MyProxy and request submission.
- Resource discovery and information using Globus MDS
- S2 science run data from 3 interferometers to be used
- Each input file contains data for 1800s approximately 25Mb in size
- Executable staging for sites which don't have geo-code installed
- Site monitoring being done using GANGLIA
- Globus Replica Location Service used for registering **raw** and **materialized** data products

GW Pulsar Search - Dag Structure

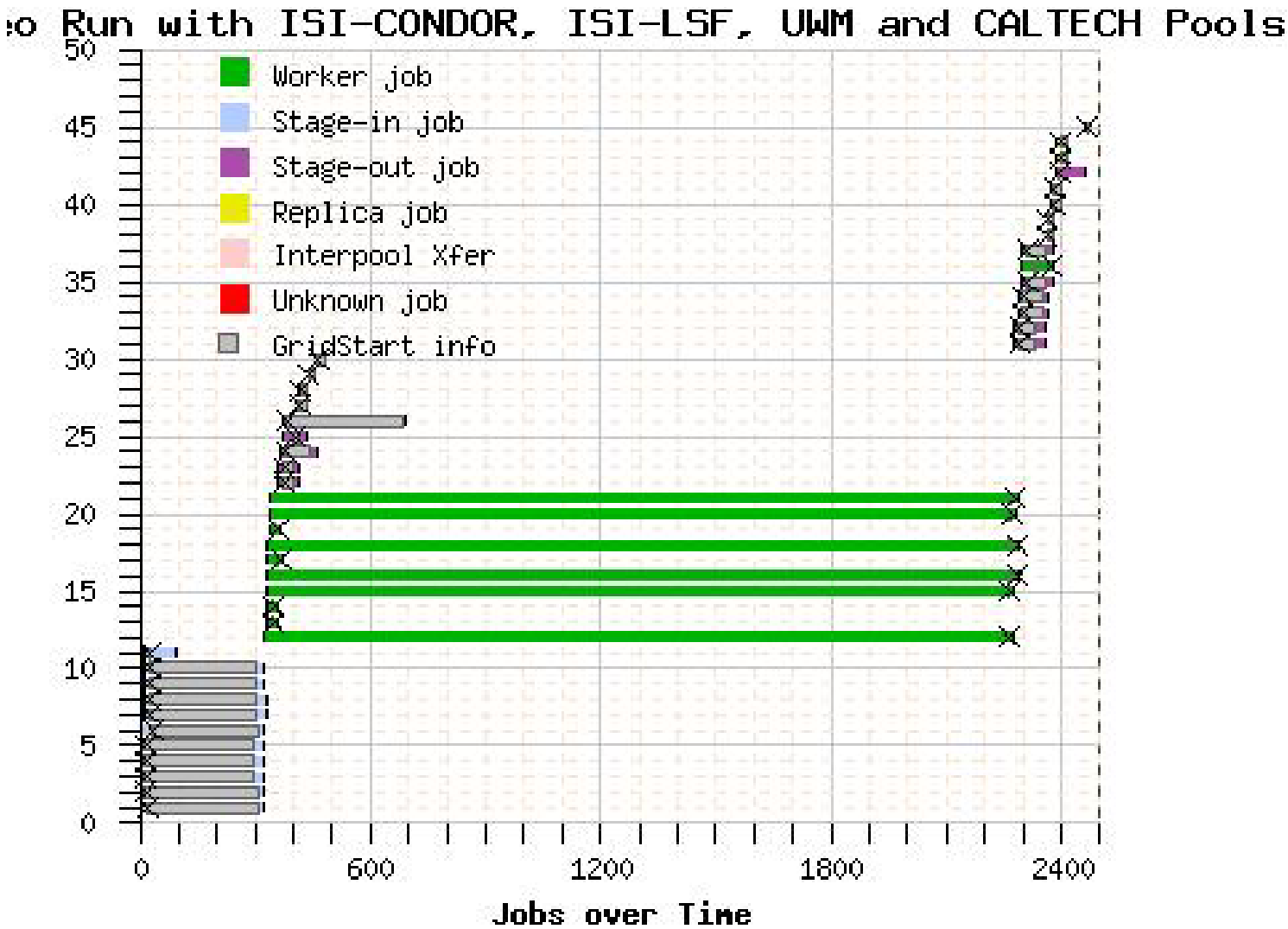


Status

- First dry run week of 2003.10.07 for testing out the search code and the SC demo/production infrastructure.
- The jobs were run at the ISI condor pool, ISI lsf cluster, CALTECH condor pool and UWM condor pool
- The outputs transferred to the ISI condor pool
- The jobs breakdown is as follows
  - 11 input transfer jobs (bringing in 1002 input files) (1000 raw files and 2 calibration files earth.dat and sun.dat)
  - 10 Extract SFT jobs each crunching on 100 input files
  - 1 ComputeFstat job which takes the 1000 extracted SFT files and computes fstatistic.
  - 1 interpool transfer job which transfers data from runs at CALTECH, UWM and ISI LSF cluster to the ISI condor pool (The computeFstat was running at ISI LSF cluster)
  - 11 output transfers (All the output data was stored in the storage location associated with the ISI condor-pool)
  - 11 RLS registration jobs (All data was registered at rls://smarty.isi.edu associated with the ISI condor-pool)

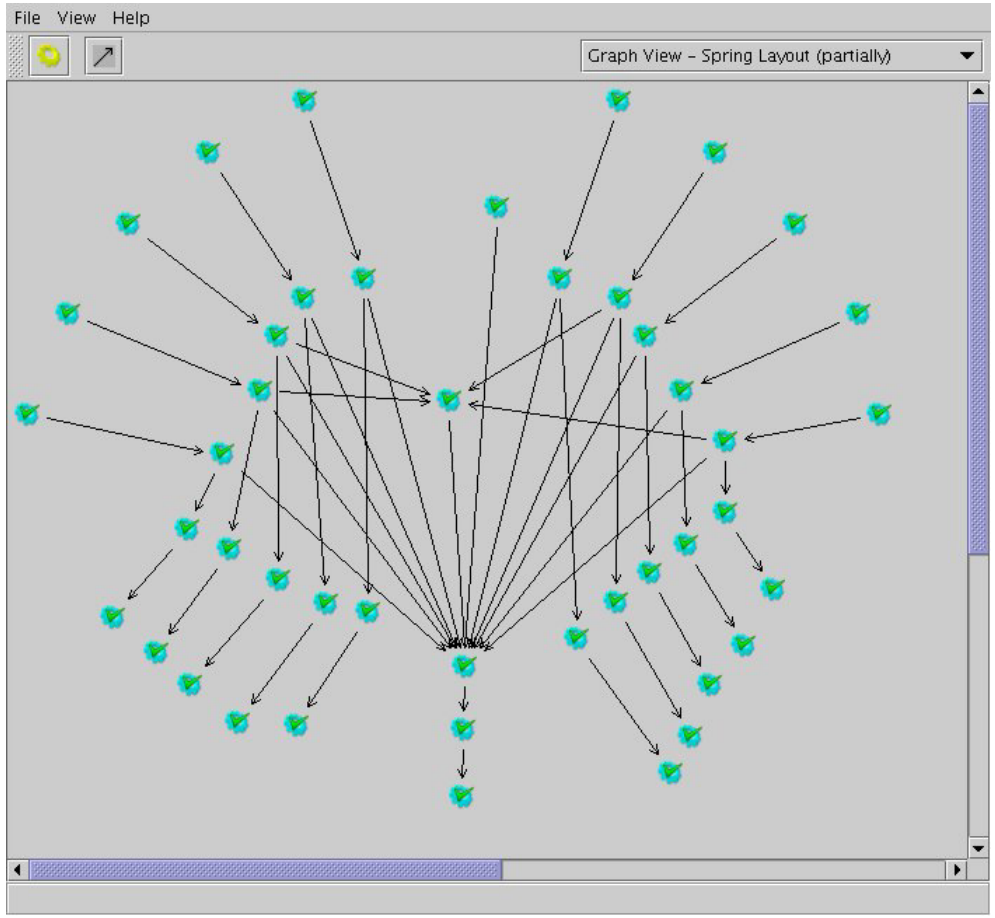
LIGO

GW Pulsar Search - Dag Over Time



GriPhyN

Run time Dag Visualization



Visualization code by Gregor von Laszewski, and Michael Hategan (ANL)

LIGO Science Has Started

- **First Science run** ("S1"): 2002.08.26 - 09.09
  - First LIGO scientific results submitted for publication
- **Second science run** ("S2"): 2003.02.14 - 04.14
  - Sensitivity was ~10x better than first run
  - Duration was ~ 4x longer
- **Third science run** ("S3"): 2003.10.31 - 2004.01.05
- LIGO is analyzing ***real*** data ***today*** with emerging grid technologies
  - Need to balance priorities:
    - > data analysis challenges vs. grid R&D challenges
  - Opportunity to provide "case study" feedback to middleware development activities
    - > robustness, QA
    - > timeliness -- tracking of opensource software releases (e.g., linux kernels, …)
    - > what works, what does not work, what needs improvement, …

GW Pulsar Search - Dag Over Time by Hosts