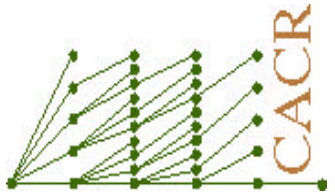
LIGO/CACR MOU

A Collaboration for
Development of Facilities for
Analysis, Archiving, and Networking
of LIGO Data

Prof. Thomas Prince
Member of the LIGO Laboratory
Associate Director CACR

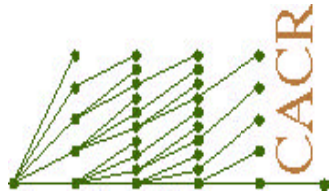
What is CACR?



- CACR: Center for Advanced Computing Research at Caltech
  - » A member group of the LIGO Scientific Collaboration (LSC)
    - – Relationship defined in LIGO/CACR MOU
  - » Significant expertise in parallel computing and networking
  - » A partner in the NSF NPACI (National Partnership for Advanced Computing Infrastructure), one of two national supercomputing consortia
  - » Major existing facilities for data intensive computing:
    - – Compute engines: 128 processor HP/Convex Exemplar.
    - – 300 Terabyte HPSS (High Performance Storage System)
  - » Caltech invests 1M$/yr of Institute funds in CACR

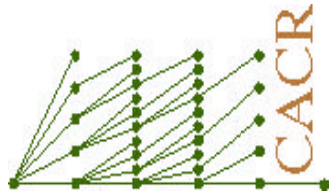
(http://www.cacr.caltech.edu)

LIGO/CACR MOU



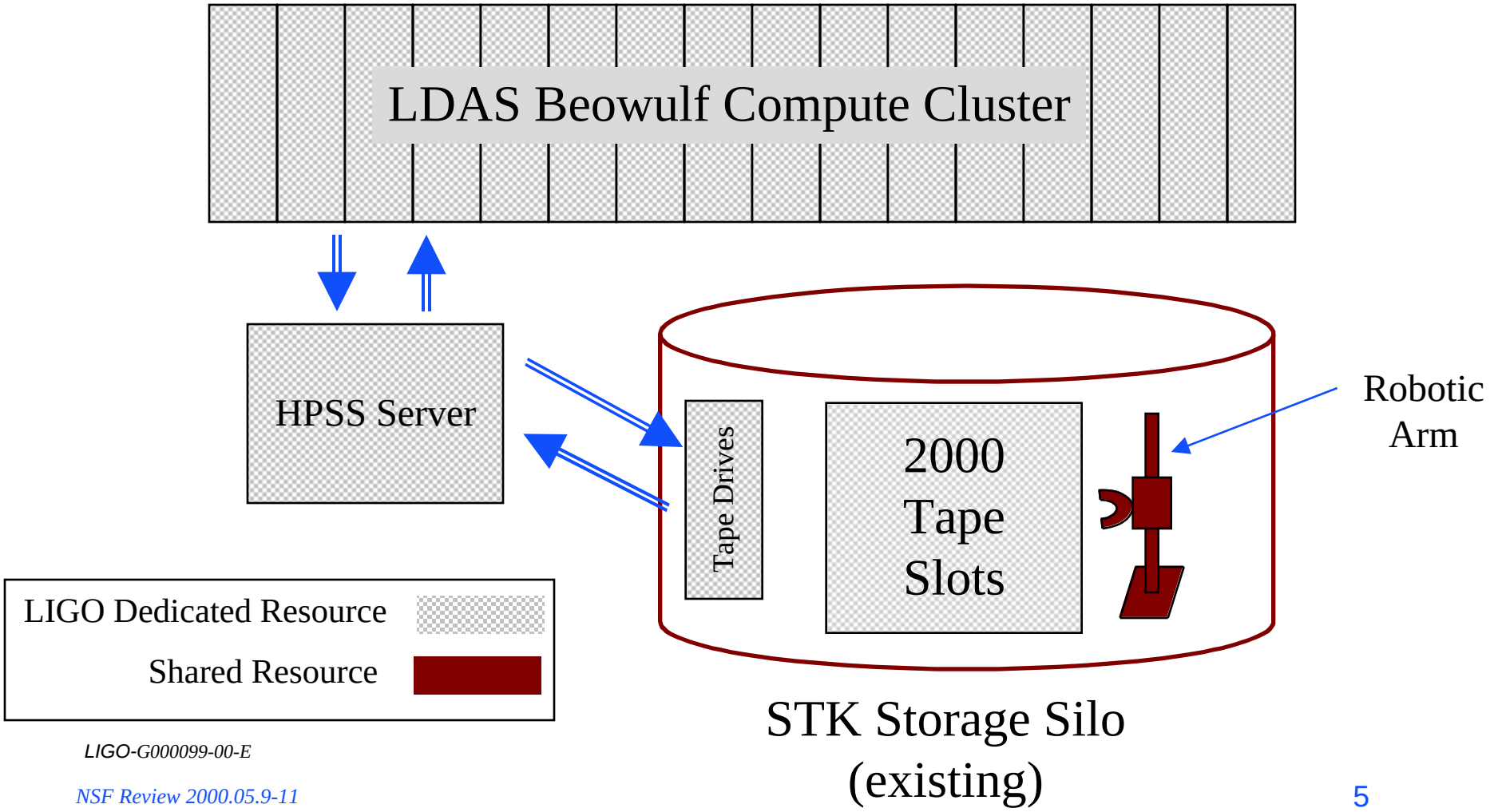
- Principal MOU Activities (keyed by MOU paragraph)
  - » Prototyping (Compute and storage systems): 8Ab,c,f
  - » Archival storage production system: 8Af
  - » Supercomputer Access: 8Ae
  - » Participation in GriPhyN: 8Ad
  - » Development of user interfaces and standards: 8Aa

LIGO/CACR MOU (8Af): Production Data Systems

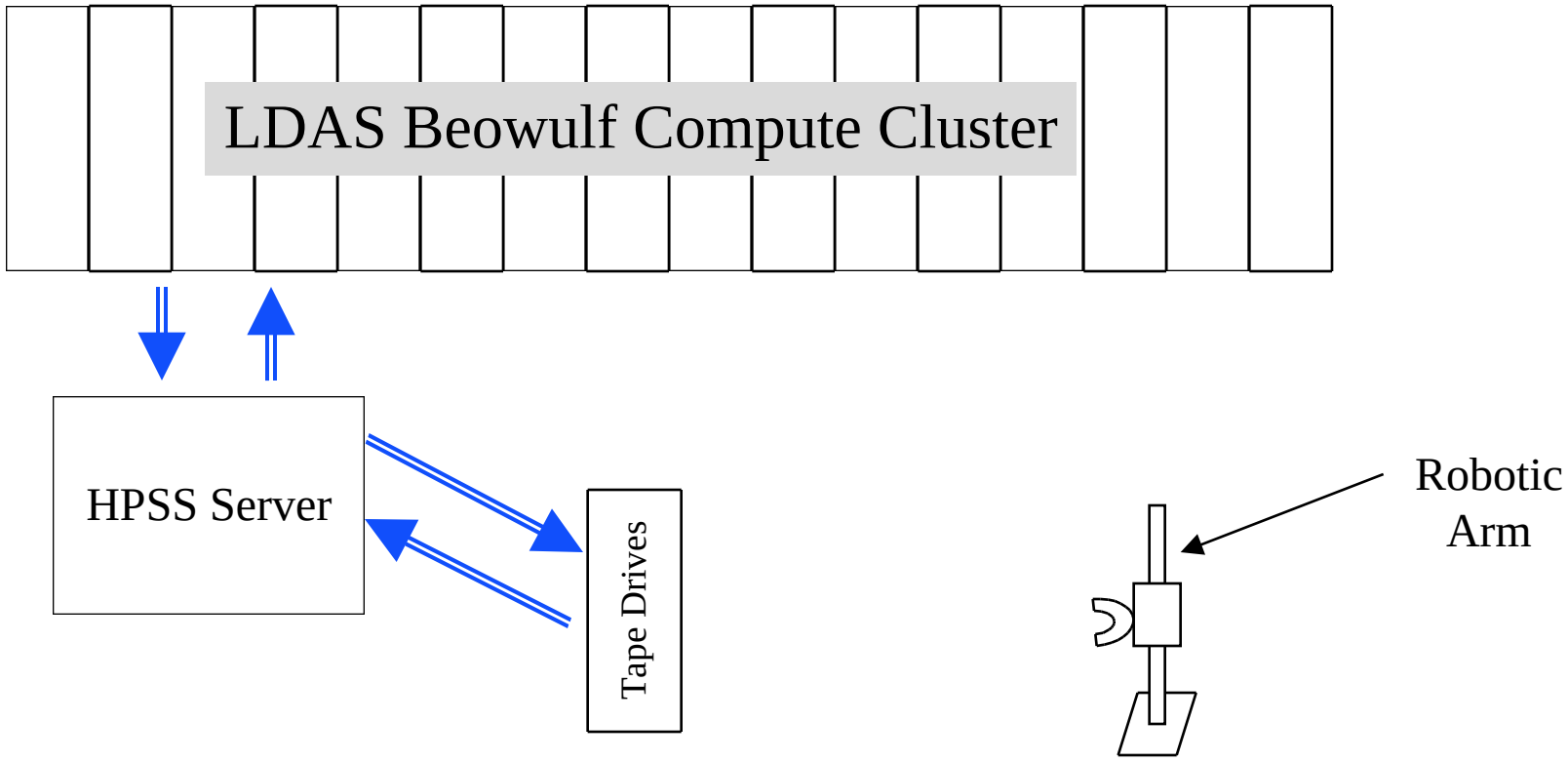


- LIGO will purchase and operate its own Beowulf cluster for production archival data analysis
- LIGO will purchase and operate its own server and tape drives for the HPSS archival storage system
- Area of collaboration: Tape robot
  - » CACR commits to providing 1/3 of existing STK tape storage silo and robot + HPSS licensing (100k$/yr)
  - » The LIGO LDAS system administrator at Caltech will support Beowulf compute and HPSS storage systems in a shared environment with CACR

LIGO/CACR MOU (8Af): LDAS Production System



LIGO/CACR MOU (8Af): LDAS Production System



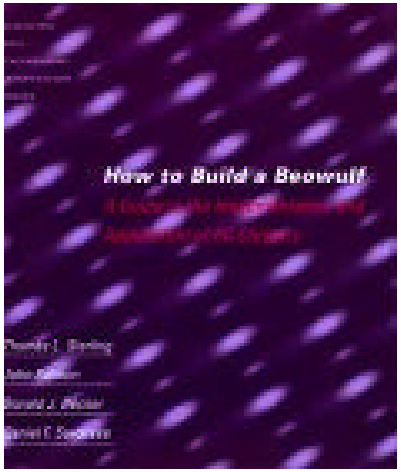
LIGO-Owned

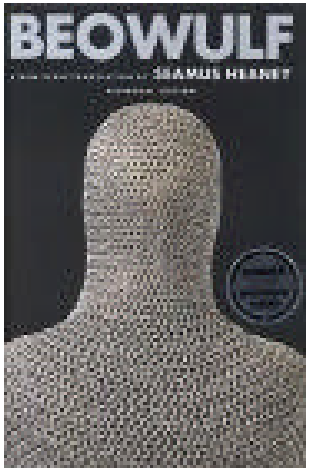
LIGO/CACR MOU (8Ab,e,f): Prototyping Activity

- Compute systems: PC cluster (Beowulf) chosen by LIGO as a cost-effective data analysis system
  - » CACR scientists “wrote the book” on Beowulf (which one?)
  - » Thomas Sterling of CACR is originator of Beowulf (the computer system)

    (http://www.cacr.caltech.edu/beowulf)
- Storage systems: HPSS identified by LIGO as viable archival storage approach

    (http://www.cacr.caltech.edu/resources/HPSS)
  - » LIGO Lab is archiving data daily from Hanford on CACR HPSS system
  - » Intensive data transfer test on CACR HPSS during 40m data run (Aug-Sept/99)
  - » 5-10 non-Lab LSC members currently using CACR HPSS

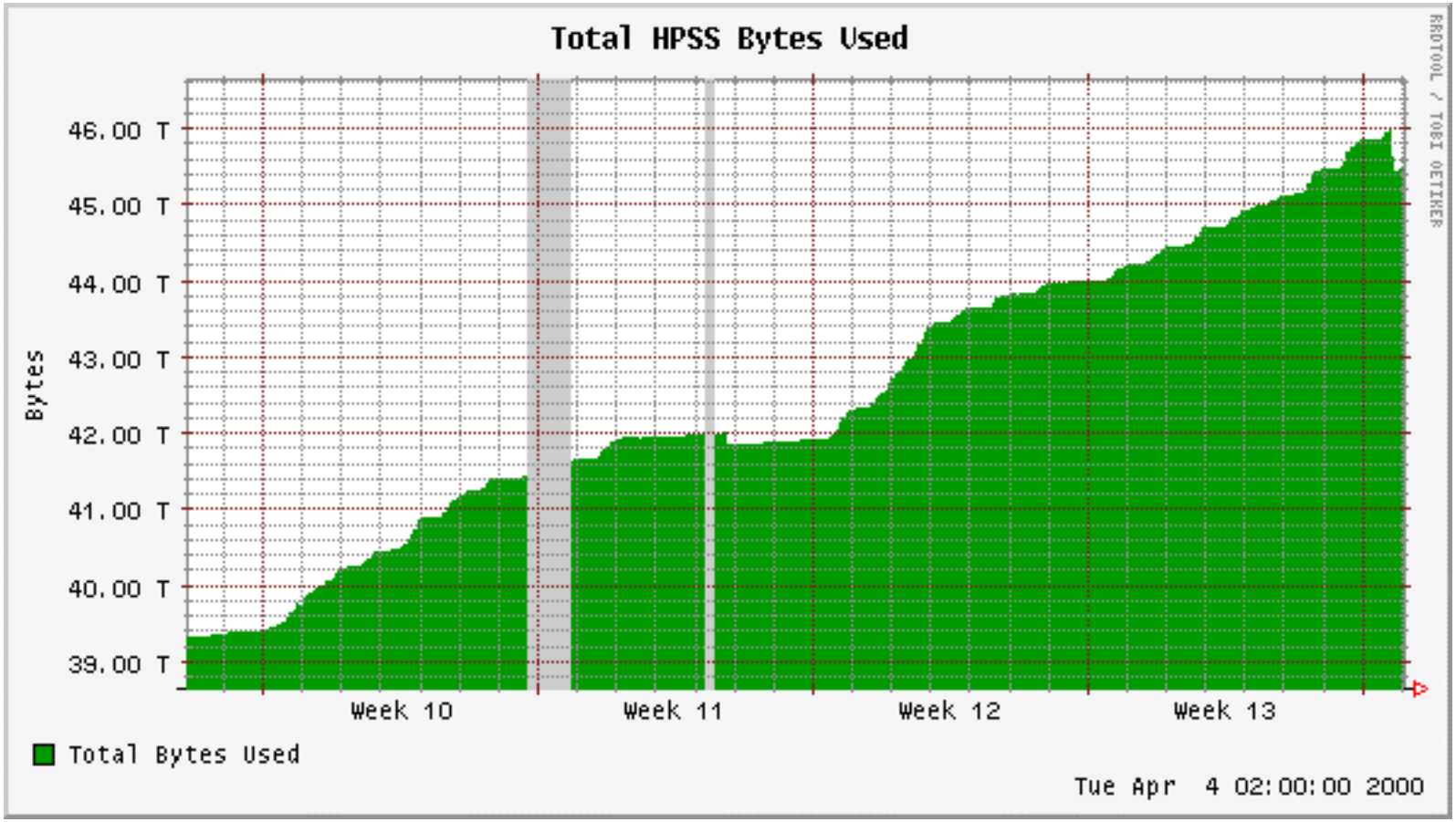




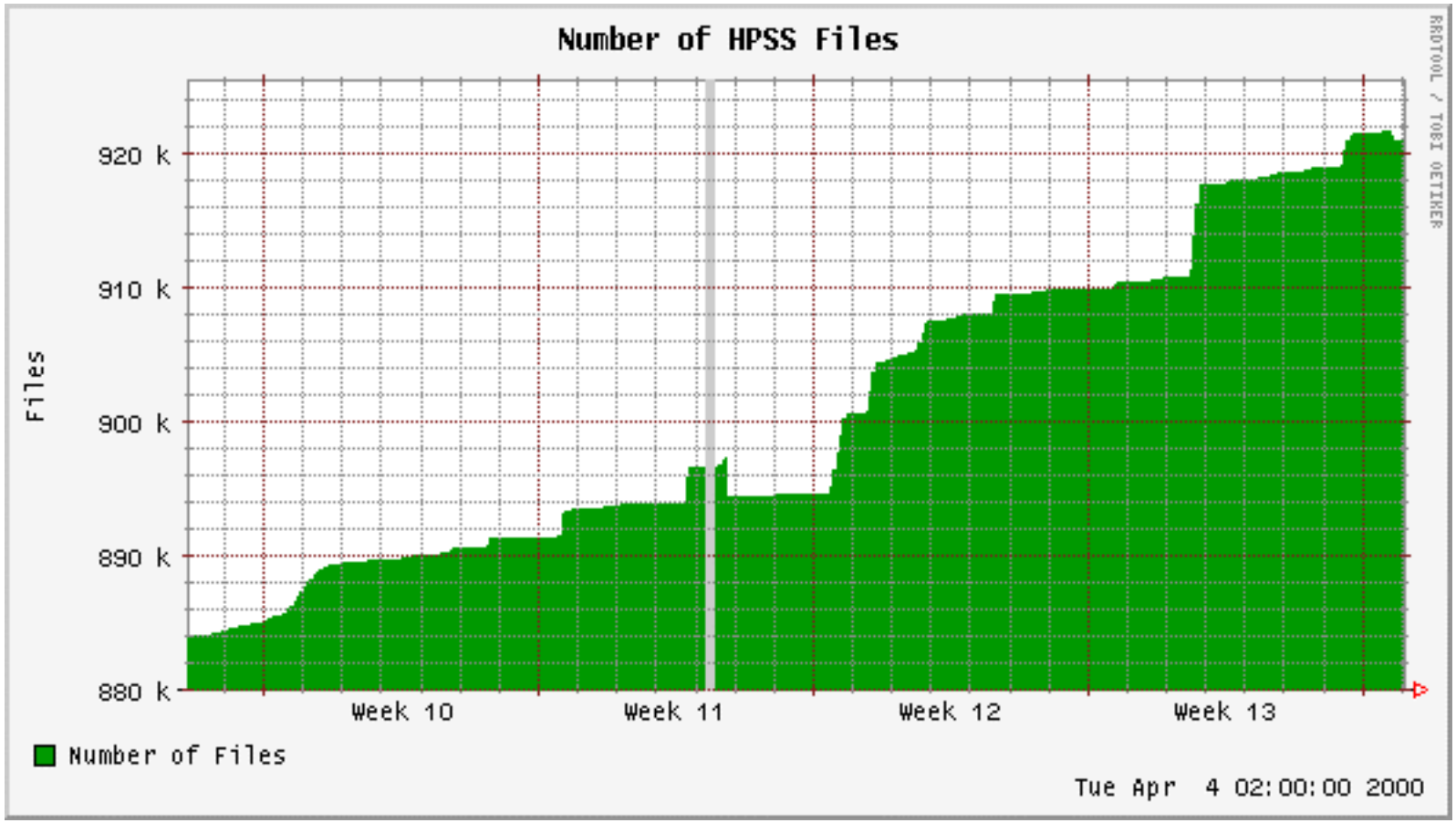
CACR Facilities: HPSS

- 300 Terabyte Tape Robot
  - » 4 Tape Drives
  - » 6000 Tape Cartridges
- 350 Gigabyte Disk Cache
- 6-node IBM SP-2 server
- HiPPI+ATM network interface

HPSS


Total HPSS Bytes Used
Bytes
46.00 T
45.00 T
44.00 T
43.00 T
42.00 T
41.00 T
40.00 T
39.00 T
Week 10
Week 11
Week 12
Week 13
Total Bytes Used
Tue Apr 4 02:00:00 2000
RRDTOOL / TOBI OETIKER

HPSS


Number of HPSS Files
RRDTOOL / TOBI OETIKER
920 k
910 k
900 k
890 k
880 k
Files
Week 10
Week 11
Week 12
Week 13
Number of Files
Tue Apr 4 02:00:00 2000

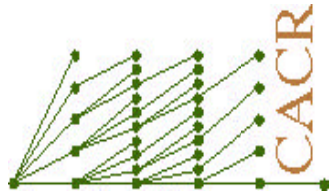
LIGO/CACR MOU: Additional Items

- LIGO/CACR proposal activities (MOU 8Ad)
  - » Recent example: GriPhyN (also UWM, UTB)
- Access to parallel supercomputers (MOU 8Ae)
  - » LSC members have access to CACR HP V-class and Beowulf parallel computers (formally through NSF NPACI)
  - » Managed as block grant to LIGO Lab
    - – LIGO Lab/LSC determines allocation policy
    - – Users obtain accounts at CACR
  - » It is CACR's intent that *all* use of CACR facilities for LIGO-related analysis will be subject to LSC review
  - » A resource allocation policy should be developed by the LIGO Lab and the LSC that covers significant compute facilities of *all* LSC member groups used for LIGO data analysis.

CACR Computing Facilities

- Primary Facilities
  - » 128 processor HP V-class
  - » 114 processor PC Beowulf
  - » Aggregate compute power of about 200 Gflops
- Other Facilities
  - » IBM SP-2
  - » SGI Origin systems
  - » Alpha Beowulf

LIGO/CACR Relationship



- CACR ***is not*** a computer center
- CACR ***is*** a federation of research projects
  - » In addition to LIGO, examples include:
    - – Astronomy (Digital Sky Project), Earth Sciences (Synthetic Aperture Radar Analysis),Computational Chemistry, Materials Science (ASCI), Physics (CMS), Biology, Fluid Mechanics
- Use of CACR facilities by outside users
  - » Track-record of providing advanced computing resources to nationwide consortia
    - – ASCI, SIO, CRPC, CSC, LSC
  - » CACR compute facilities currently available to NSF researchers nationwide through NPACI