Scientific Computing at SLAC

Gregory Dubois-Felsmann
Chair, PPA scientific computing advisory committee

DOE OHEP facility review
8 July 2008
Scientific Computing at SLAC: a PPA Perspective

• Scientific computing at SLAC is supported by the “Scientific Computing and Computing Services” (SCCS) department, as well as by contributions from PPA (and other) programs themselves
  – The routine needs of scientific computing users are met...
    • Connectivity, access to relevant libraries and tools, archiving, etc.
  – ... As well as the specific needs of the programs:
    • Management of petascale datasets and their generation and analysis
    • Provision of large amounts of computation capacity and storage and its efficient and reliable application as a production facility
    • Parallel / MP computing for accelerator and astrophysical simulations

• There are both intellectual and physical components to this
An experienced and innovative team

• Scientific computing is inherent in two of the four core competencies of the laboratory, specifically:
  – Theory and innovative techniques for data analysis, modeling, and simulation in Photon Science, Particle Physics and Particle Astrophysics
  – Ultra-large database management for users and collaborations distributed worldwide

• This is thanks to the knowledge and experience of the teams of people who have worked on SLAC’s major programs and can bring their skills to meet new challenges

• Expertise covers the full range from the management of computational infrastructure to the design of scientific applications
## Capabilities and Projects

<table>
<thead>
<tr>
<th></th>
<th>BaBar</th>
<th>GLAST</th>
<th>ATLAS</th>
<th>LSST</th>
<th>SNAP</th>
<th>KIPAC</th>
<th>LCLS/PS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Capabilities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAQ/trigger</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Space systems</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Pipeline processing</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Petascale datasets</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Petascale databases</td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>MPI/SMP</td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔ (imaging)</td>
</tr>
<tr>
<td>Visualization</td>
<td>✔?</td>
<td>✔?</td>
<td>✔?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Advanced CPU/GPU use</td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>will need someday</td>
</tr>
</tbody>
</table>

**Projects**

- **established**
- **developing**
- **new R&D**
Program Details

• BaBar
• ATLAS
• GLAST
• SNAP
• LSST
• KIPAC
• Accelerator research
BaBar

• BaBar has been the dominant HEP computing task at SLAC for the past decade
  – Large dataset acquired 1999-2008, excellent physics productivity
    • 22 billion raw events acquired, 0.7 PB of raw data, ran 7-10 mon/year
    • 8.3 Gevents reconstructed and selected for physics analysis, 11 Gevents simulated in most recent pass
    • Most events “skimmed” to divide them into (overlapping) subsets for more efficient data access and distribution to remote sites
    • Current full dataset including skims: 0.7 PB; computing model requires “most” of this to be on spinning disk
    • Reprocessing and resimulation about every two years, reskim every year; total data produced so far: 2.9 PB (all kept in SLAC mass storage)
  – Better than 96% of all available luminosity recorded
  – Record of delivering results at every year’s major conferences within weeks of accumulating the last luminosity included
BaBar Resources

• BaBar computing divided among a set of “Tier-A” sites: SLAC, CC-IN2P3 (France), INFN Padova & CNAF (Italy), GridKa (Germany), RAL (UK; through 2007)
  – Responsibility for computing (CPU & disk) provision divided as ~2/3 SLAC, ~1/3 non-US
  – Tier-A countries delivered their shares largely on an in-kind basis at their Tier-A sites, recognized at 50% of nominal value
  – BaBar operating and online computing costs split ~50-50%
  – Planning to continue replacement, maintenance through 2010
  – Simulation also provided by 10-20 other sites, mostly universities

• Analysis computing assigned to Tier-A sites by topical group
  – Skimmed data & MC is distributed to sites as needed for this

• Specific production tasks assigned to some sites as well

• Roughly half of BaBar computing was off-SLAC 2004-2007
BaBar Post-Datataking Plans

• Data acquisition ended April 7, 2008

• Now in “intense analysis period” of ~two years
  – One final reprocessing and resimulation being carried out this summer
  – “Core” Y(4S) analyses being carried out on the full (1999-2007) data for completion this year or next summer; a large additional set of analyses will follow, some entirely new, some updates to the full data
  – Initial Y(3S) results being released this summer, with in-depth analysis of Y(3S) and Y(2S) data sets over the next 1-2 years

• “Ramp-down” period begins in 2010 (detailed profile TBD)

• In 2011 expect most Tier-A resources to be withdrawn and all remaining analysis to converge on SLAC
  – Total need in 2011 should be similar to or slightly less than current SLAC-only capacity, then decline fairly rapidly
BaBar Data Curation

• BaBar has a unique data sample
  – Very competitive Y(4S) dataset with an excellent detector and tools
  – By far the largest Y(2S) and Y(3S) datasets in the world
  – Very large and clean sample of tau lepton decays

• Many new-physics signals could be visible in this data
  – The mass scales being investigated at the energy frontier are accessible in many channels at BaBar, both at tree and penguin levels, producing enhancements in rare decays and/or distortions in the pattern of CP violation and the unitarity triangle parameters
  – We are already looking for those that seem most promising \textit{a priori}

However...

– Discoveries at the Tevatron or LHC (or beyond) could point in new directions
BaBar Data Curation - II

• Essential to be sure that the BaBar data can continue to be analyzed well into the future: data curation
  – Data access and simulation capability need to be preserved
  – Evaluating overlapping approaches:
    • Increasing portability and single-user usability:
      BaBar code and production systems need to be simplified and ported to the newest available OS versions and compilers, while experts are still available, in order to maximize their maintainable lifetime and accessibility
    • Freeze and virtualization:
      Beyond some point porting will become difficult to do and to fund. We are studying strategies for freezing the system, and its external dependencies, and running BaBar in a virtualized environment. [Security concerns!]
      • Simplified data model and parametric simulation [Limited scope]
  – Identified ~one FTE of staffing for this within current BaBar resources
    • Intend to seek grant support to help expand the project, seen as a pilot for what several other HEP experiments will soon face
ATLAS - current computing roles

- Host of “Western Tier-2” computing center
  - $400K M&S / year devoted to CPU and storage through FY2010
    - Current 244 TB of disk (xrootd cluster), ~715 cores (2007-equiv)
      - Expect four-year hardware replacement cycle, perhaps slightly shorter to avoid power/cooling constraints
    - Application (and CPU/disk ratio) is at US-ATLAS discretion
      - Current use is predominantly for (simulation) production
      - Expect to continue production dominance into early data-taking era
  - Accessed through OSG tools
  - Fully integrated with the general SLAC batch environment
    - ATLAS CPUs are actually a fair-share allocation of a fraction (~30%) of the general batch facility, shared with GLAST and BaBar and other users
    - Leverages investments in system management
    - Low marginal cost so far because of commonalities with BaBar, and because it’s still a small fraction of the total SLAC installed base
ATLAS - current computing roles II

- Significant scientific computing / software engineering roles
  - Data access and distribution
    - SLAC xrootd team is working on future developments driven, to a large extent, by ATLAS requirements
  - High-Level Trigger (HLT)
    - Trigger design
    - Scaling of configuration / initialization to 2000-node farm
  - Major effort in improving and speeding up simulation
    - Leverages Geant4 expertise at SLAC
    - But previous core G4 personnel are as a result now mostly working on ATLAS-specific simulation issues in the application of G4
  - Production management
    - Applying BaBar personnel / expertise to scaling of ATLAS production control system and database
ATLAS - exploring future roles

• “Western data analysis facility” - a major expansion
  – A major analysis center for West Coast ATLAS users, providing the resources for user-driven data-intensive analysis

Such a proposal would have a
  – collaborative component (viz. A. Schwartzman talk), a
  – physical component: provide extensive computing resources for calibration, analysis, and reco/sim R&D and a user-friendly computing facility around them, and an
  – intellectual CS/engineering component: co-locating strong new ATLAS expertise with broad and deep BaBar experience in efficient and innovative data analysis techniques, reliable high-efficiency production, agile support of user requirements
ATLAS - exploring future roles II

- Scope of physical component still under discussion
- Likely most interesting if a scale of a few thousand present-day cores can be reached (comparable to present BaBar installation)
  - Limited by BaBar runout and infrastructure constraints through ~2011 and by unresolved needs of LCLS for the same space/power/cooling
  - Opportunities available by replacing all pre-2007 BaBar hardware, rack unit for rack unit, or by adding additional Black Box datacenters (see below), at about 2000-4000 cores/BB
  - Already-established ability to share facilities with BaBar and GLAST minimizes costs beyond those needed for infrastructure
GLAST MISSION ELEMENTS

- GPS
- DELTA 7920H
- White Sands
- TDRSS SN S & Ku

- GLAST Spacecraft
- Large Area Telescope & GBM

- Mission Operations Center (MOC)
- GLAST Science Support Center
- LAT Instrument Science Operations Center
- HEASARC

- GRB Coordinates Network
- Alerts Data, Command Loads Schedules
- Schedules

- GBM Instrument Operations Center
GLAST LAT Computing at SLAC

- Instrument Flight Software design and implementation
- Pipeline processing of LAT data upon receipt from MOC
  - Housekeeping trending
  - Science data
    - Event reconstruction
    - High level science analysis
  - 300 cores, 0.5 TB output per day for 10 GB input
  - Heavy reliance on SCCS infrastructure - batch farm, Oracle, xrootd, fast networking, etc.
  - All GLAST bulk CPUs are shared with the general batch system, but with prioritization that favors GLAST’s latency requirements
- Data and Software distribution to the collaboration
  - Management of Science software
- Primary site for Simulations and Reprocessing
  - Allocating 900 cores by 2009 (including prompt data reserve)
  - Computing resources for Science Groups, using pipeline workflow system
  - Pipeline has been ported to Lyon CC. Possible room to expand in the future
- Plan is 250 TB disk per year + modest new core purchases, at least to replace old cores. TBD until we get experience on usage models, especially for reprocessing data.
SNAP

• Present roles in SNAP collaboration in
  – Electronics, DAQ, and flight software
  – Focal plane star guider device
    (See A. Roodman’s talk.)

• Beginning to explore whether a downstream computing role is possible and desirable
  – On the JDEM development and flight timescale, SLAC will have a large and experienced team who will have done pipeline processing and science data management in collaboration with NASA, on GLAST
  – One possibility: an overlap of a growing SNAP/JDEM role with the transition of GLAST to routine operations could produce a plausible profile for staff with relevant skills
  – Requires collaboration with institutions with existing roles in SNAP
LSST - existing efforts

• Camera design and construction: DOE
  – SLAC expertise and personnel in DAQ closely involved

• Data Management:
  database design and R&D
  – Multi-PB datasets envisioned (tens of PB of raw data); database queries from the entire astronomical community and the public; real-time alerts
  – DOE not initially assigned a central role, but...
  – Key SLAC personnel from the BaBar database and processing design, plus a strong new hire, have made a substantial impact and are developing leadership roles in LSST Data Management. This is still a small effort. Some funding ($200K/yr) from LSST project.
  – Looking for only modest short-term growth in this area, based on the availability and interest of other suitable personnel. Except...
LSST - SciDB project

• SLAC LSST database group’s R&D work, including a survey of existing technologies, has led in an interesting direction
  – Last fall: eXtremely Large DataBases (XLDB) workshop held here
    • Strong representation from the sciences, from the CS community, from database and hardware/software vendors, and from users in industry
    • Outcome: realization that the scale and some of the detailed requirements in petascale+ science are remarkably similar to current problems in industry; challenge from the CS community to clarify this
  – Smaller group met in March at Asilomar to try to do this, mainly by distilling out common requirements of science.
    • Attendance: HEP, astronomy, fusion research, genomics, ..., from DOE-SC and beyond; CS community; possible commercial users
    • Successful: 2.5 days produced a set of requirements, and...
    • A strong expression of interest by leaders in the CS database field (Stonebraker, Dewitt, et al.) to actually build a product: an open-source database optimized for the needs of petascale science

All organized by Jacek Becla et al., see JB talk in breakout
LSST - next steps with SciDB

• Requirements are challenging and interesting
  – Parallelism, fault-tolerance, multidimensional array-valued data, uncertainty & fuzzy joins, spatial and temporal features, provenance
  – LSST identified as a flagship application
    • Because of the scale of the science, the time scale of the project (soon enough to be serious, far enough off to allow for some risk-taking), and the leadership role of the SLAC LSST DM group

• Existing DB company, Vertica, to start an open-source division to coordinate development this fall
  – SLAC exploring technicalities of how to work together with a company; discussion of providing office space and some staff
  – Exploring funding from VCs, eBay, Microsoft
  – Design meetings already under way, aim is to start coding this fall, deliver a beta in 1.5 years, and see if it is well-aligned with LSST and other science needs
LSST - next steps with SciDB II

• A remarkable opportunity to influence the development of petascale database systems
  – The CS people involved have a track record of producing game-changing technologies

• Available SLAC personnel are just barely able both to pursue this and maintain existing commitments to LSST
  – LSST DM needs to proceed as it was before, for now, not counting on this project to produce a usable result.
  – Searching for additional funding to allow a more robust role in SciDB
    • Important that ongoing direction and involvement come from the scientific users, or the direction may become too responsive to commercial interests
    • Informal presentation to ASCR on June 26th, setting out the case for several FTE-years of effort over two years, starting as soon as possible
KIPAC

• KIPAC science program depends heavily on astrophysical and cosmological simulation and visualization

• Strong program in data analysis for GLAST, LSST, and other astronomical data sources also gearing up
  – Requires techniques well beyond “embarrassingly parallel”

• Simulation
  – Techniques: PIC, AMR-MHD, N-body, Radiation Transport all require massive parallel processing and high speed interconnects
  – Large scale simulations now routinely produce multi-TB data sets
  – Analysis requires further parallel computation and visualization
  – Storage and analysis require high bandwidths and new technologies for data access (parallel file systems); archival storage also needed
KIPAC - Visualization

- Data sets from simulation include many particles, length scales and time steps.
- Advanced visualization techniques and systems provide physical insight and communication of results.
  - Visualization is also key to outreach
- Visualization requires intensive advanced computation for data analysis
  - KIPAC research into effective application of modern GPUs
- Current facilities:
  - Recently installed a new 13’x7’ 3-D stereo projection system with full HD resolution.
  - Prototype tiled display (shown) which will be expanded to a 15 panel configuration.
KIPAC Facility Needs: a Vision

Present
- 20 node analysis cluster (coma)
- 10 node Apple Xserve
- 72p w/440 GB ram SGI Altix
- 90 node: 360 cores, 1.4 TB memory Infiniband cluster
- ~100 TB NFS storage
- 24 TB Lustre high speed storage
- 13’x7’ 3-D stereo projection
- Multi-LCD tiled display wall (soon)
- Prototype GPU parallel computing w/Nvidia graphics engines

Future
- Leadership computational astrophysics and large scale data analysis facilities
- Cycles: 200 nodes/yr over 4 year lifecycle (8 core today)
- Storage: 100 TB high speed, 200 TB long term storage, double every year
- Visualization: GPU computing cluster: 20 nodes, double/yr
- Network: 10 GBit between all major data consumers, double every year
- Steady-state scale ~1000 sq-ft with ~500 kW.
KIPAC challenge to SLAC computing

• A new kind of computationally intensive science for SLAC
  – In many cases this *cannot* share existing facilities
  – New technologies are required (new interconnects, parallel file systems, etc.)
  – Challenges the SLAC/SCCS tradition of doing our own systems integration
    • For HEP this often produced the best-suited designs and lowered costs
    • Parallel computing community relies much more on vendor integration
    • This may produce a significantly more heterogeneous environment for us to support
  
  – We’re still navigating the transition from an organization that had years to optimize itself to a single dominant computing model
    • Management and staff development are key challenges
    • Some existing things have to be given up in order to make room
Accelerator research

• Advanced computing has become critical to accelerator R&D
  – Massively parallel simulations are required in order to model accelerator structure, RF behavior, and thermal/structural stresses
  – Major inputs from CS/applied math in designing efficient algorithms
    • SLAC has developed world-leading expertise in applying and advancing adaptive-mesh finite-element techniques
    • Significant funding from SciDAC for computational research in this program at SLAC: 4.0 FTE of computational scientists, 0.5 FTE physicist
  – SLAC group maintains a number of leading-edge community codes
  – Resources have allowed breakthrough developments
    • Solving of “inverse problem”, working backward from measured RF behavior of a structure to identify construction imperfections (TJNAF)
  – Key involvement in design of future accelerators
    • LHC, LHC upgrade, ILC and high-gradient R&D, TJNAF upgrade
    • Unique capabilities lead to opportunities for work-for-others funding
Accelerator research II

- Strong relationships with other SciDAC centers and with the national supercomputing community
  - Significant input of code and tools from outside

- Needs of the SLAC group’s work require time at national SC facilities
  - Major allocations at NERSC and Oak Ridge: over 5.5M hours in FY08
  - Single ILC cryomodule simulation requires 4096 CPUs, many hours of run time

- SLAC’s existing MPI facilities somewhat facilitate development/debugging and are sufficient for some codes

- Visualization is becoming increasingly important
  - Requirements not too different from KIPAC
  - Collaboration might be very fruitful
Support for Community Tools

• SLAC has provided support for a variety of community tools over the years
  – Early years: graphics systems, analysis environments, EGS3/EGS4 electromagnetic simulations
  – Now: accelerator design, xrootd/Scalla, Geant4, EPICS, ...

• Long-term funding is a concern
  – Some already have community-oriented support
    • For example: SciDAC Compass funding for accelerator R&D
  – Some were funded through programs (mainly BaBar)
    • Geant4 is the key example
Geant4

• SLAC built up a very strong group from 2001 through 2007
  – Kernel architect, hadronic physics expert, efforts in electromagnetics, variance reduction, and visualization

• 2008 budget cuts forced layoffs and refocusing
  – G4 group is now primarily working on ATLAS-specific issues
  – Some efforts to seek external funding for other specific applications
  – This is very valuable to ATLAS, but...
  – In the long run this will result in a lack of investment in the fundamentals of simulation, a key tool with applications well beyond the LHC

• We are preparing a proposal for a coordinated US Geant4 support effort
  – Including collaboration with other laboratories
Research Projects

• SciDB
  – Above-mentioned outgrowth of LSST database R&D

• PetaCache
  – Grew out of recognition that trends in disk capacity and I/O scaling do not favor our scientific workload (head/GB, stagnant seek latency)
  – Solid-state disk (flash) offers a new path
    • True random access to large datasets could revolutionize analysis
  – ASCR support for an R&D program in this area
  – First: one-terabyte DRAM-based storage cluster (FY06)
    • Used for a variety of primarily engineering tests
  – Now: completing construction of a 5 TB modular flash-based system
    • Exploring R&D and application possibilities in ATLAS, LSST
    • Based on technology developed here for LCLS, LSST camera, ATLAS?

• Others...
  – Talk to us afterwards...
Personnel 2007-2010

• SCCS staffing was significantly cut in the 2008 layoffs
  – Unix-systems group and business computing were largely protected
  – Losses in many other areas: networking, infrastructure, telecoms, G4 support, desktop and user support
  – Capacity for “service R&D” was already low, and cut further
  – Many areas are now covered at one-person depth

• NB:
  – PPA program staff also provide much computing expertise - relationship is very collaborative!
Personnel breakdown within SCCS

• Note transition away from BaBar-dominance...
## Hardware Expenditures 2007-2010

### M&S for computing equipment, $K

<table>
<thead>
<tr>
<th></th>
<th>FY07</th>
<th>FY08</th>
<th>FY09</th>
<th>FY10</th>
</tr>
</thead>
<tbody>
<tr>
<td>BaBar &amp; Infrastructure</td>
<td>3250</td>
<td>1750</td>
<td>2300</td>
<td>2300</td>
</tr>
<tr>
<td>ATLAS</td>
<td>400</td>
<td>400</td>
<td>400</td>
<td>400</td>
</tr>
<tr>
<td>GLAST</td>
<td>530</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KIPAC</td>
<td>550</td>
<td>75</td>
<td>450</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4730</strong></td>
<td><strong>2225</strong></td>
<td><strong>3150</strong></td>
<td><strong>2700</strong></td>
</tr>
</tbody>
</table>

- Infrastructure spending previously allocated almost entirely to BaBar, as the dominant program.
  - Power, cooling, backbone switches, mass storage...
  - FY07: ~50% BaBar-specific end-user hardware; previous years ~75%
- BaBar has entered a replacement-only era, however...
Hardware Expenditures 2007-2010

M&S for computing equipment, $K

<table>
<thead>
<tr>
<th></th>
<th>FY07</th>
<th>FY08</th>
<th>FY09</th>
<th>FY10</th>
</tr>
</thead>
<tbody>
<tr>
<td>BaBar &amp; Infrastructure</td>
<td>3250</td>
<td>1750</td>
<td>2300</td>
<td>2300</td>
</tr>
<tr>
<td>ATLAS</td>
<td>400</td>
<td>400</td>
<td>400</td>
<td>400</td>
</tr>
<tr>
<td>GLAST</td>
<td>530</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KIPAC</td>
<td>550</td>
<td>75</td>
<td>450</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>4730</td>
<td>2225</td>
<td>3150</td>
<td>2700</td>
</tr>
</tbody>
</table>

- FY09 & FY10: major investment in replacing mass storage system and copying existing data
  - Oldest BaBar disk and CPU replaced at higher density
  - Additional Black Box installed to accommodate planned growth

- FY11 and beyond:
  - BaBar replacements drop; infrastructure costs fall to new programs
Future Challenges

An Infrastructure Challenge

• Equipment for non-BaBar programs has thus far been able to be offered at the marginal cost of the actual delivered computing hardware
  – A significant discount compared to a fully-costed budget; major benefits from this were received by ATLAS, GLAST, KIPAC
  – Possible in large part because until recently we had not reached hard limits of available capacity. Now we (nearly) have.

  – Demands:
    • BaBar capacity must be kept near present level through at least 2011
    • Needs of LCLS computing remain under study but could be great
    • KIPAC and ATLAS growth could be significant - perhaps ~1000 rack units
Future Challenges: power and cooling

• Significant infrastructure constraints:
  – Power delivery (aging infrastructure, lack of systematic dual-source)
  – Cooling (air-cooling capacity exhausted, max. +250 RU water-cooled)
• Substantially affected the ability of SCCS to deliver needed compute capacity in FY07, FY08; many installations delayed

• Interim solution: two Sun Black Box “portable datacenters”
  – 8 racks in a standard-sized shipping container with external hookups for power and chilled water; designed for very high density loads
  – Adequate power available from Substation 7 near Bldg. 50
  – Chilled water supplied by a new air-cooled electric chiller
  – Total of ~500 rack units available
Future Challenges: next steps

• Room for one or two more Black Boxes in the same area
  – Similar solutions may be available from other vendors
  – Power adequate; a second electric water chiller would be needed
    • Higher capacity required to support the higher-density hardware that is foreseeable for the next few years

• BaBar “electronics house” may also be available
  – Housed most BaBar data acquisition, triggering, and slow controls
  – Can be repositioned out of the way of BaBar D&D activities
  – Power and cooling limited to somewhat more than a single Black Box equivalent
Future Challenges: long-term infrastructure

• Needs:
  – More permanent structures and possibly much greater capacity
  – Breathing room for renovation of Building 50 power distribution, etc.
  – Desirable: “green” systems providing better than the present 1-to-1 ratio of power for computing equipment vs. power required to cool it

• Envisioned solution:
  – Collaboration with Stanford University to build a modular research computing facility, sited at SLAC and taking advantage of the large amounts of power and chilled water available in the research area
    • Extensive use of “green”, passive cooling, taking advantage of convection and natural airflows.
    • Several thousand rack units of capacity could be provided.
    • Funding model still being developed.
    • Possible to complete first module by 2011 if project initiated this fall.
Future Challenges: mass storage

- Mass storage: lower-cost storage for archival purposes and for infrequently-accessed data
  - Generally provided in the form of tape for the past 15-20 years
  - At SLAC: STK (now Sun) robotic silos, repeatedly upgraded over the past 15 years with new drives and tapes, but same basic mechanism
    - Over 20,000 tapes in use as of this week (3.8 PB of data stored), completely dominated by BaBar reconstructed and simulated data archive
    - Existing silos will no longer be supported by Sun in early 2010

- Imperative that a new mass storage system be on line by the end of maintenance of the existing system
  - In time to be able to copy the existing dataset (roughly 5 drive-years)
  - All current operations must be maintained during transition
  - Tape is still the answer (next generation may be MAID - disk-based)
Future Challenges: mass storage II

• A significant portion of the FY09 and FY10 budgets
  – Silo purchases, drive purchases, new tape costs
  – New server infrastructure
  – FY09: commission new system, copy data
  – FY10: fill out new system with drives, cut over all service

• Adapting the mass storage software and server design will be a significant human challenge
  – New drives are faster than existing systems can support
  – Architectural changes will be required to use them efficiently (network changes, additional disk buffering)
  – Running two systems in parallel will be challenging
Future Challenges: efficient use of CPUs

• After decades of exponential growth, Moore’s-Law scaling of CPU clock speeds has essentially stopped.
  – Virtually all recent scalar-computing progress in CPU/$ and CPU/watt has arisen from the move to multi-core CPUs.
    • 8-12 core chips, higher rack densities expected soon.

• Other technological changes are accumulating
  – Vector processing units in modern commodity-type CPUs
    • Typically poorly exploited by HEP’s very heterogeneous codes
  – Excellent floating-point processing performance available from GPUs
    • Now generally far in excess of that from general-purpose CPUs

• We must take advantage of these developments
  – “Just run more uncoordinated jobs per box” will not continue to scale
  – SLAC must prepare itself to meet this challenge and begin the relevant R&D - an issue both for scientific users and data centers.
    • Massive astrophysical and accelerator simulations will probably provide the first opportunities to exploit these technologies - and these are very much part of SLAC’s program in the coming years.
    • R&D on the use of GPUs is already under way in KIPAC
Outlook

• SLAC has a very strong team in scientific computing

• We are managing the complex transition from the BaBar era

• Our systems and people are being brought to bear successfully on new tasks:
  – GLAST, ATLAS, LSST, KIPAC, accelerator simulation, ...
  and we are preparing the ground for taking on further challenges