

The *BaBar* Computing and Analysis Model

Gregory Dubois-Felsmann
Caltech

BaBar Technical Review
11-12 October 2000

Outline

- Status
- Need to determine future directions for computing
- Computing and analysis model
 - Essential features of the model
 - Strategic and management issues
 - Data model and access
 - Analysis strategy
 - *Covered in other talks:*
 - Online computing (Luitz)
 - Simulation production (Ryd)
 - Computing at data centers and remote sites (Gowdy)
 - Data distribution (Hasan)
 - User perspective (Olsen)
 - Technology and costs (Mount)
 - Cost and schedule (Young)

BaBar computing status

- Covered in C. Young's talk in the plenary session
- Basically,
 - We were able to collect and analyze data sufficiently well to present results this summer including data after just a few weeks' latency
 - And to continue running at the PEP-II design daily integrated luminosity
- The computing system:
 - Supported data acquisition at the required rates and deadtime
 - Allowed us to validate the quality of the data as it was acquired, and find and fix problems promptly when needed
 - Reconstructed the data with a typical 24 hour latency after acquisition
 - Provided for skimming and distribution of event samples to SLAC users and remote institutions
 - Supported the development of the production and analysis software required to accomplish all these goals

Status, cont.

- **There were some serious problems, but they were solved**
 - A variety of sources of data acquisition rate limits, deadtime were eliminated
 - Reconstruction (OPR) initially was not able to keep up with DAQ
 - Cured after major efforts in finding and relieving bottlenecks – some in software, and some in hardware requiring unexpected expansions to the SCS infrastructure
 - User data access was unacceptably slow and inconvenient in the beginning
 - Addressed for all sites with development of *Kanga* data format
 - Data access from Objectivity database at SLAC was improved with hardware and software changes
- **Some problems remained at a non-critical level**
 - Mostly associated with managing change rates, complexity
 - Difficulties in introducing updates to the online system – need a test stand
 - High rate of changes to reconstruction code, calibrations led to fragmentation of the data set, multiple reprocessings
 - Many pieces of the system were just “good enough,” need attention as we go forward to higher luminosities and increasingly factory-like operation

Future needs – the basic parameters

- We have presented results based on $\sim 9 \text{ fb}^{-1}$ of data
- We have collected a total of $\sim 20 \text{ fb}^{-1}$ at up to $2.5 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$
- But we are expecting $\sim 280 \text{ fb}^{-1}$ by the end of 2003 at luminosities up to $\sim 1.48 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$

- We have performed about 20 analyses on the existing data
- We have ~ 600 collaborators and will surely have many more analyses going in parallel

- We are finding ourselves short-staffed in many areas of computing
 - We need to make it increasingly easy for collaborators to contribute

Computing model study process

- Working group formed August 2000
 - Representatives from many institutions and constituencies
 - Charged to consider the next three years of *BaBar* computing
 - Defined a set of *prioritized requirements* (q.v.)
 - Discussed and settled on a *strategy* (q.v.) to address the requirements
 - Subjected to internal review on 13 September; comments under consideration

Essential features of the model

- High-level considerations
 - We currently have a successful computing implementation (as demonstrated by results presented at Osaka and elsewhere)
 - Assume *evolution* over the next few years, not starting over from scratch
 - Computing to be understood as an *integral* part of *BaBar*, not as a isolated department
 - Some “computing” problems can be solved with work in other areas: setting physics goals, tightening triggers and OPR filters, refining skims
 - We must be *adaptable*. There are large error bars on the projected luminosity and budget
 - The experiment as a whole must *supply manpower* to support and improve computing
 - Those who are given *responsibility* for a task must also be given the *authority* to carry it out

Setting physics priorities

- With luminosity of 1.5×10^{34} , physics rate (hadrons, taus, and mu-pairs) will total ~ 90 Hz
 - We will need to *prioritize* treatment of the data on two levels
 - *Trigger/filter*: need to define requirements for efficiency and rate in broad physics categories (hadronic, tau, two-photon, etc.). Algorithms will have to be designed to meet the resource constraints subject to these requirements
 - *Analysis*: need to guide allocation of offline computing resources – especially disk space – at a more fine-grained level. (e.g. CP events, B mixing events, Bhabha, 3-1 tau, etc.)
- **Setting these priorities is a task for the collaboration as a whole**
 - **Must be done soon to provide guidance for trigger and filter development**

Organization and management

- The experiment must clarify its *organizational structure*
- Must establish and adhere to a *computing schedule* for
 - Releases
 - Deployment of production executables
 - Reprocessing and skimming
 - When serious problems are discovered in a production executable after deployment, a fix may be needed; but then we must then revise our schedule and **account for the impact of failures to identify problems** before deployment – and improve procedures to prevent recurrences
- Must establish a mechanism for acquiring the manpower needed for computing maintenance and development
 - **Suggest assigning various tasks to institutions for defined time periods. The institutions can then allocate individuals accordingly**

Software administration

- Two key challenges
 - Make necessary changes easy in development and deployment
 - Minimize unnecessary changes
- Change rate must be limited to avoid reprocessings
 - Needs ongoing attention to QA/QC – notably *pre-deployment testing* – ... and *planning* – notably of release schedules
- Must reduce the cost of entry for contributing code
 - ... *while ensuring its quality*
 - We are significantly short-staffed for expected future computing needs in all areas (detector subsystems, reconstruction, online, etc.)
 - Tools:
 - *Documentation* (including human-written and auto-generated documentation)
 - *Training* (including professionally taught OOAD classes)
 - *Consistency* (across reconstruction, physics tools, online) of interfaces, standards
 - Large-scale *dependency management* and *physical design* improvements

High level strategy – processing

- Implement a filter equivalent to <10 nb at $L=3 \times 10^{33}$; tighten it further as luminosity improves to 1.5×10^{34}
 - We are at 13 nb now
 - 10 nb corresponds to about one-third the current Level 3 trigger rate
 - Events passing this filter are subjected to full reconstruction in OPR
 - **Baseline**: still include the entire s -channel hadronic cross section; note hadronic fraction increases with time as calibration streams are held at constant rates
- Refine skimming strategy
 - Define levels:
 - Tight: $< 1\%$ of hadronic cross section. These skims will be used for most analyses
 - Loose: $< 10\%$ of hadronic cross section
 - Very loose: larger than 10% of hadronic cross section
 - To minimize duplication, correlated *skims* to be grouped into *streams*
 - These (probably ~ 20) will be the fundamental units of data transfer and deletion
 - Manage resource constraints in terms of streams
 - Especially disk space for keeping data readily accessible
 - Implement stream-based reprocessing for 2002+

High level strategy, cont.

- A useful “*Mini*” data format should be defined and implemented
 - Event hierarchy would become: Raw, Reco, **Mini**, Micro, Tag
 - Some quantities now in Micro may go here
 - **But only if the information is really not needed at Micro level**
 - Will be useful for more detailed analyses, event display
 - Expected to allow track refitting but not full re-reconstruction of events
 - **But detailed design – including review of the existing Micro – remains to be done**
 - Must make sure that it can be used efficiently in conjunction with streams

High level strategy – data storage model

- Thus far:
 - All levels of event store and support databases implemented in Obj'y
 - Micro, Tag, and minimal Conditions database also in “Kanga” format
 - Developed following August 1999 computing review
 - ROOT-based persistent storage
 - Fully exploited transient-persistent interface abstractions developed for Objectivity
 - Obj'y used for all sites' MC production and for physics analysis at three
 - SLAC, IN2P3, LBNL
 - Kanga data used at these and all other sites for analysis
 - Duplication of data and support of two systems raises questions of scalability and maintainability
 - We have to review this now!
 - Should we and can we eliminate the duplication, and if so how and when?
 - Remember, we are OK right now – *first do no harm!*

Data model choices

- We considered two notional models:
 - **Model 1:**
 - Maintain Objectivity store for all levels; maintain Kanga Micro, Tag
 - Improve Objectivity usability for analysis and for remote sites
 - Implement Mini in Objectivity
 - Implement disk staging for all Objectivity levels but Tag
 - Use physics priorities to help manage staging space
 - Provide hooks for non-HPSS staging systems, to be implemented at remote sites
 - Improve Objectivity data distribution tools
 - When ready, deploy Objectivity for analysis at more sites
 - At first in parallel with Kanga
 - Replace Kanga when Objectivity gives same or better usability and maintainability
 - Decision made site-by-site
 - **Model 2:**
 - Retain Objectivity for production tasks (OPR, reprocessing, simulation)
 - Move analysis exclusively to Kanga
 - Mini implemented in Kanga only
 - Data distribution via Kanga only

Data model choice issues

- Model 1
 1. Represents ideal case (Objy does what we want and everybody uses it)
 2. Potential for transparent access to all levels of an event
 - ✘ Requires significant Objy development
- Model 2
 1. Uses largely proven technology in Micro-based analyses
 2. Places fewer demands on Objy by allowing better control of access
 - ✘ Requires significant Kanga development, especially for scaling
 - ✘ A Kanga-only mini may be less functional than people desire
 - Problems with database access, links to the raw hits
 - ✘ Institutionalizes separation between reconstruction producers and physics analyzers

Data storage model – strategy chosen

- Models considered in detail by the working group, decisions made
 - Shown to internal review committee (report available, q.v.)
 - Presentations to recent collaboration meeting
- Proposed:
 - Retain Objectivity for production
 - Continue Objectivity development to address deficiencies for analysis, use at remote sites
 - Project for CY 2001
 - Roll out Objectivity to a small set of new test sites
 - Three by January 2002; expect initial sites to need extra TLC during shakedown
 - Move most sites to Objectivity analysis by end of 2002
 - Kanga is maintained, tracking Tag, Micro changes, as long as needed
 - Until Objectivity is widely deployed and productive for analysis, meeting specified milestones and requirements

Objectivity development issues

- Must remedy the deficiencies which make it difficult to use for analysis, at remote sites
 - Believed not inherent to Objectivity and addressable by end of 2001
- What needs to be done (main points):
 - Test and deploy important new features from vendor:
 - Employ large DB ID's to reduce DB size to ~250 MB from ~10GB, improve staging behavior
 - Employ read-only databases to limit lock management load, improve reliability
 - Account for and reduce storage overhead throughout event store
 - Set up automated network-based Objy data distribution system
 - This will be a major effort, approximately two FTEs for one year
 - Continue to improve the scaling behavior of OPR with luminosity
 - OPR/reco groups also need to improve code efficiency
 - Undertake a detailed survey of uses of the various data levels
 - May result in redefinition of Rec and Micro when deploying Mini

Objectivity roll-out to remote sites

- Now only SLAC, IN2P3, LBNL use Objy event store for analysis
- The computing group will assist in the installation of Objy at additional remote sites, which must have:
 - Specified minimum hardware (CPU and disk)
 - As given in the Requirements document
 - Sufficient staffing to participate in the installation
(1 FTE to start, decreasing as experience is gained)
 - Requirement is for steady-state operation to absorb only 0.1 FTE
- Goal is to install Objy at three new sites by January 2002
... and have it actually used for physics analysis
 - By end of 2002 run, most sites should have Objy event stores
 - Not all sites in a given Tier* need to make transition at the same time
 - * (see subsequent slide, Data Distribution talk)

Continuation of Kanga support

- Kanga has made possible convenient global access to Micro-level data ... and the physics analyses we have undertaken to date
- Need for Kanga will continue at least into 2002
 - Maintenance must continue until Objy is widely deployed
 - Means that changes in Tag and Micro must be tracked in Kanga
- The collaboration must provide the personnel necessary for Kanga maintenance (and improvement)
 - This is greatly assisted by extensive common underpinnings w/ Objy
 - Effort on either of Objy or Kanga should not dilute the other

Kanga support, cont.

- How can we tolerate the Objy/Kanga duplication? *We must.*
 - We believe long-term scalability and full exploitation of BaBar data require the use of Objectivity
 - We cannot abandon Kanga until we can accomplish our physics goals without it

Data model – distribution & availability

- Covered in some detail in a separate presentation
- Main ideas:
 - Define three-tiered hierarchy
 - Tier A: SLAC now, IN2P3 ramping up into 2002, perhaps one more later
 - Master sites for data distribution
 - Union of all Tier A sites has all *BaBar* data, each has at least 30%
 - Tier B: Regional centers, e.g., RAL, INFN
 - Staging areas for streams and their redistribution or narrowing for small sites
 - Focal points for national, regional efforts and accumulation of resources
 - Tier C: Analysis sites
 - Minimum hardware requirements established
 - Minimal support burden for the event store *is a requirement*

NB: All classes of sites may contribute to simulation production

Data availability, cont.

- Anticipate that most likely *no site will have all the data!*
 - We won't be able to afford the resources to duplicate all bulk data indefinitely
 - Distributed processing tools (c.f. GRID technology) must be provided (start development now) for large-scale analyses for which narrow skims are inadequate
 - More generally, of course, the storage and processing resources needed for the whole of our physics analysis effort will have to be assembled from all the collaboration's sites
 - Each site must manage the staging of that data which it does have
 - Policies set by physics priorities, enabled by common staging hooks
 - *We are looking at keeping a duplicate tape-only copy of the original raw data to allow disaster recovery*

Data availability, cont.

- Notional model for availability on disk of beam data at Tier A sites

	2001	2002	2003	2004	2005
Tag	100%	100%	100%	100%	100%
Micro	100%	70%	50%	50%	50%
Mini	20%	20%	10%	10%	10%
Reco	5%	3%	2%	1%	0.7%
Raw	5%	3%	2%	1%	0.7%

Analysis strategy

Applicable whether using Kanga or Objectivity...

- Skims will be collections of events within streams
 - Implemented either as pointers or by filtering on tag bits
 - Notion is ~100 skims and ~20 streams
- Events may appear in more than one stream
 - Average duplication factor not to exceed 2
- **The stream will be the unit of data transfer**
 - Objy streams written by OPR, Kanga by a dedicated production task
- To facilitate analysis at smaller sites, tools will be provided to produce self-contained skims (narrowed streams) from production streams
 - Done at a site after a production stream has been imported
 - Bookkeeping is up to the requesting user/site
 - But centrally maintained tools will be provided to do the work

Resource constraint management

- We will not be able to do everything,
 - at least not right away, yet *BaBar* fundamentally must retain its character as a **general-purpose instrument for studying *B* physics**
- We envision that the critical constraints we can tune against will be:
 - **disk space** for retaining fast access to data
 - **CPU capacity** required for *re*-processing, *re*-simulation

Resource constraint management, cont.

- We envision that we will manage these constraints by
 - Setting physics priorities and reflecting them in a disk staging strategy oriented around streams
 - With appropriate levels of resident Micro, Mini data (Tag is always 100% on disk)
 - ... and by developing the ability to reprocess data by stream
 - But we must as a prerequisite
 - Limit the pre-OPR filter to as close to the s -channel cross section as possible
 - Less Bhabhas, of course
 - Agressively lower the event size in all levels of the (Objectivity) event store
 - We will pursue further reductions in rate, but...
 - The techniques will not become clear until we have had more experience with analysis – and evaluation of systematics. Some possibilities for additional filtering:
 - Continuum **uds** background (would require sensitive shape cuts, problematic for D physics)
 - Two-prongs (detailed evaluation of consequences for systematics needed)

Conclusions

- We are doing well already
 - Next round of results expected this spring
- We are planning for having to do much better...
 - Because PEP-II upgrades look to be on a very fast track
- ... and for how to do it
 - A lot of development work will be required to realize our model
 - And then we'll still need the tangible resources to exploit its capabilities
- So we look forward to your advice!