



# *BABAR*

## *STATUS AND DATA PRESERVATION*

Tina Cartaro  
*SLAC*

---

DPHEP Collaboration Meeting  
March 14<sup>th</sup>, 2017



# *BABAR COLLABORATION*

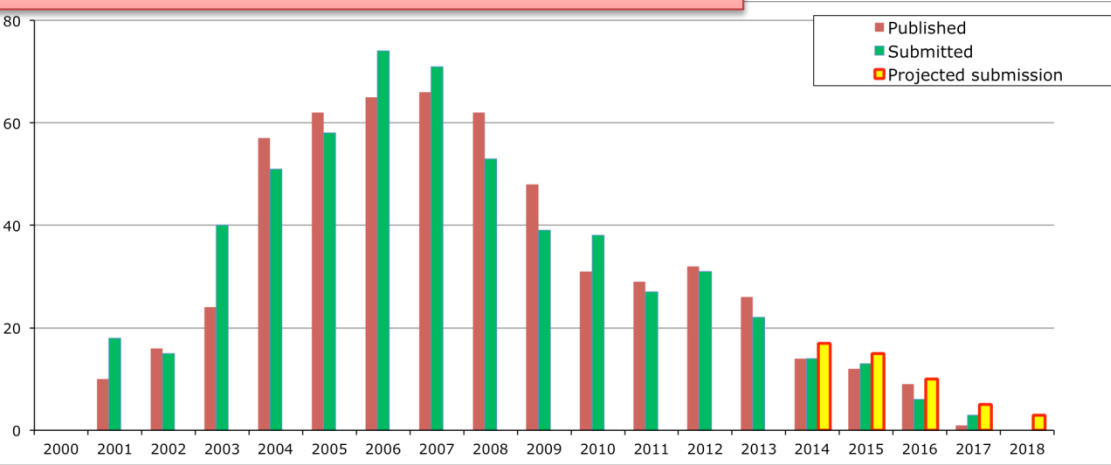


- ~270 members from 67 institutions in 13 countries
  - Plus associates
- Collaboration structure for the long term
  - Consolidate the governance structure
  - Main roles in the collaboration become permanent
  - Merge analysis working groups
  - Frozen author list (easier to maintain)
    - additions/removals/one-time authors always possible
- The Collaboration will not dissolve



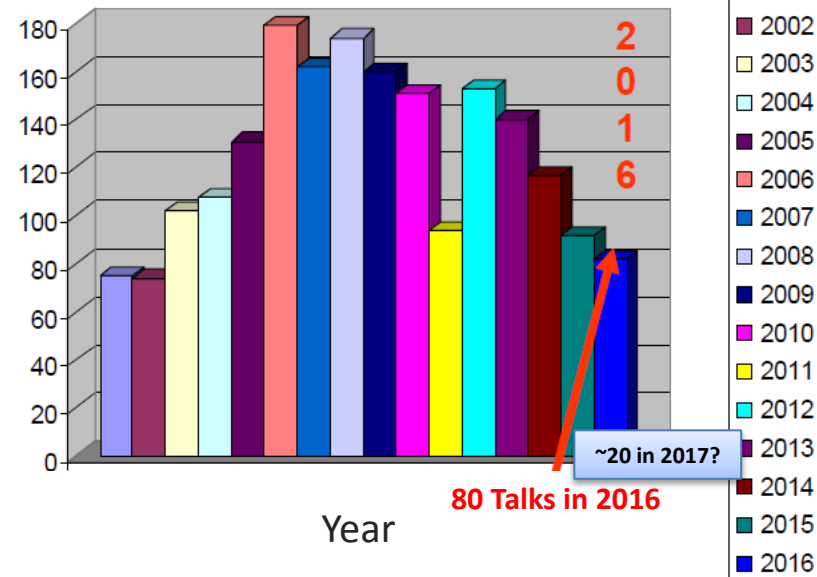
# BABAR SCIENCE PRODUCTIVITY

Papers published/submitted/projected by year



- 564 published
- 5 submitted/accepted
- 2 in final review before submission

Conference talks by year

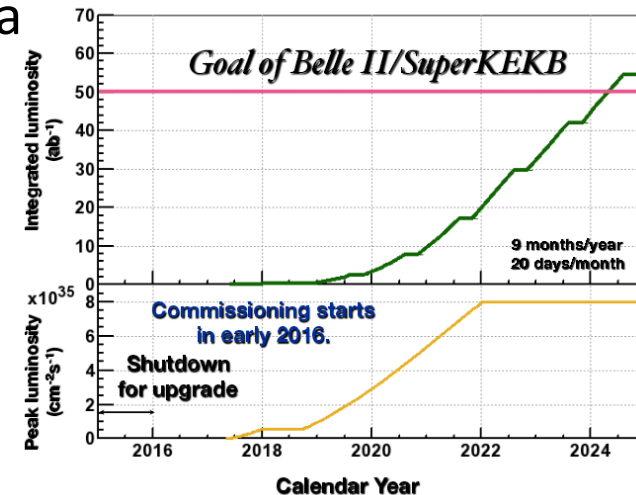


- 20+ analyses on track for publication, ~10 analyses have uncertain future
  - 10 new analyses started since January 2016 and all progressing
- Accepted conference talks are limited by funding



# BABAR DATA

- *BABAR* has collected data from Oct 22<sup>nd</sup> 1999 to Apr 7<sup>th</sup> 2008
  - Total data on tape (including raw and last two processings) : 2.7 PB
  - 2 PB will be being migrated to new tape media (T10000D)
  - The oldest processing will be left on T10000C
    - T10000C still readable by the new T10000D drives but they will eventually be ejected, stored at SLAC for 5 years then moved offsite
  - Note: Oracle dropped the T10kE drives production so SLAC will consider moving to LTO or IBM TS-series for the 12TB tapes. Most likely we will not have funds for the new media
- No support for *BABAR* foreseen beyond 2020
  - Already a challenge going past next year
    - *BABAR* data will not be superseded by LHC data
      - Good match for Belle II data taking schedule
      - Y(3S) dataset expected to remain unique for longer





# LONG TERM DATA ACCESS

- Ensure the ability to support analysis of the *BABAR* data until at least 2018 (or 2020 based on Belle II schedule?)
  - In order to maximize the physics output we aim for “Level 4 preservation”
    - Preserve data, conditions, calibrations, releases, tools, databases, capability of running production and user jobs including new Monte Carlo models
    - Provide a stable virtual environment (SL4,5,6 VMs ) running the *BABAR* Framework
    - Provide accurate documentation (the *BABAR*Wiki) to preserve know-how on physics analysis, OSs, Framework, ...
- Data storage
  - Data is stored on tape at SLAC and CC-IN2P3 (back-up)
  - Most used data sits on disk (XRootD)
- Dedicated hardware to support the *BABAR* environment
  - 1.3 PB of disk for data and users
  - SL4, SL5, SL6 platforms available (VMs) - no support for RH/SL7
  - 1668 batch job slots



The LTDA enclave



# *HARDWARE*

- Most of our hardware is reaching EOL (>5 years)
  - The LTDA batch machines will survive 2 or 3 more years without many problems
  - The Sun Fire X4500/4540 servers that host LTDA user directories, repositories, and much more, will be retired at the end of the year as SLAC will drop SunOS support
  - This does not involve central support (SLAC home directories)
- All the web services (wiki, Hypernews,...) running on physical hardware are being moved on VMs
- This will enable us to prolong the availability of the services beyond the lifetime of the hardware itself
  - OpenStack adopted at SLAC and this could open a number of possibilities for managing centrally all our virtual machines



# COMPUTING SUPPORT

- The official computing manpower supported at SLAC for *BABAR* is 0.15 FTE
  - The effort needed to keep a large and complex amount of data alive is somewhat independent from the number of users
  - Already now the capacity of delivering services and user support is extremely limited
    - No more centralized data production cycles, only Monte Carlo production
    - User tools for data skimming and Monte Carlo production are already available
- We rely on SLAC Computing Division for hardware support and infrastructure
  - Lab specific agreement: at SLAC “everything under B50 [computing building] roof” is under maintenance contract [for as long as parts are available]
- GridKa and CC-IN2P3 still support *BABAR* as TierA sites
  - CC-IN2P3 has a full copy of data
  - GridKa hosts the XrootD redirector for our Xrootd Federation reachable from any of our VMs (preconfigured with the latest software release and xrootd client)



# DATA ACCESS

- Public access problem not faced by the Collaboration
  - Formally and practically very complex
    - Even if each institution agreed to open the data, all or in part, we would still need manpower and resources to support public access (Framework and documentation, possibly in a simplified version, a portal, etc.)
    - Small data samples are already used by affiliated groups for education (undergrad students)
- Only collaborators can access *BABAR* data but anyone can join
- In the last years many papers have been published in collaboration with theorists
  - The external collaborators only sign their paper(s) as Associates and then they may decide to become a permanent Member (Collaboration service requested – typically internal reviews of papers and analyses) and sign all the papers





- Long (“*forever*”) time archival of our physics documents
  - All our BADs (BaBar Analysis Document) and Notes
  - Each document can have versions and auxiliary files, like talks, and other relevant material
    - Not “raw data” or root files
- Collaborations retain full ownership and control on the documents
  - Password protected
  - We decide what is public and what is not
- Full search capabilities of the open INSPIRE archive extended to the private collections
- Collection of relevant metadata completed, working on the XML format for ingestion



# SUMMARY

- International community expectation is that SLAC/DOE will continue to support *BABAR* but the future of *BABAR* is not clear beyond 2018
  - We are trying to stretch support until 2020 matching Belle II schedule
- Official computing manpower supported at SLAC for *BABAR* is 0.15 FTE
  - The capacity of delivering services and support is extremely low
    - No more centralized data production cycles, only Monte Carlo production (thanks to a volunteer member of the collaboration)
  - The extended level of virtualization will hopefully keep documentation, data, and infrastructure accessible on the long run
- Preparing to migrate all internal documentation to inSpire



***BACKUP***



# BABAR DATASET

As of 2008/04/11 00:00

- 433/fb at the peak of the  $\Upsilon(4S)$ 
  - about 470 millions BB pairs
  - mainly for B-physics analyses also used for charm, charmonium, tau, light quarks
- 54/fb 40 MeV below the peak
  - background control sample for B-physics studies
  - alternative (B-meson free) data set for other physics topics
- Below the  $\Upsilon(4S)$ 
  - 122×106  $\Upsilon(3S)$  decays (x10 Belle)
  - 99×106  $\Upsilon(2S)$  decays (x10 CLEO)
  - $\Upsilon(1S)$  decays accessed via  $\Upsilon(2S,3S) \rightarrow \Upsilon(1S) \pi^+\pi^-$
  - Studies of bottomonium states
  - Ideal data set for direct search for low-mass-beyond-SM particles
    - Light Higgs in Next-to-MSSM
    - Dark matter bosons
- Energy scan above the  $\Upsilon(4S)$ 
  - 132 points (25pb-1/point) plus 8 points (100pb-1/point) around the  $\Upsilon(11120)$
  - Measurement of b anti-b production cross section
  - Search for exotic bottomonium states

