



# *SLAC, BABAR AND DATA PRESERVATION*

Concetta Cartaro  
SLAC  
*BABAR* Computing Coordinator

DPHEP 7 – CERN, March 22<sup>nd</sup>, 2013



# *OUTLINE*

- Introducing the Collaboration
- Addressing the questions
  - Ingest – Archive – Environment
- How long is “Long Term”?
- SLAC overview (partial)
- Conclusions



# BABAR Membership Numbers

73 institutions in 12 countries

	Faculty & Staff	Postdocs	Grad Student	ALL	Student Assoc.
CANADA	9	4	7	20	5
FRANCE	22	1	3	26	3
GERMANY	9	4	7	20	3
INDIA	1		1	2	
ISRAEL	1		3	4	
ITALY	50	10	5	65	6
NETHERLA	1	1		2	
NORWAY	2			2	
RUSSIA	8		2	10	
SPAIN	3		1	4	
UK	15	4		19	
USA	86	24	11	121	25
<b>TOTAL</b>	<b>207</b>	<b>48</b>	<b>40</b>	<b>295</b>	<b>42</b>
<i>cf Oct 2012</i>	219	51	55	313	37
<i>cf May 2012</i>	219	51	56	326	30
<i>cf Nov 2011</i>	219	50	54	323	22

6 New BaBar members since 1 July 2012

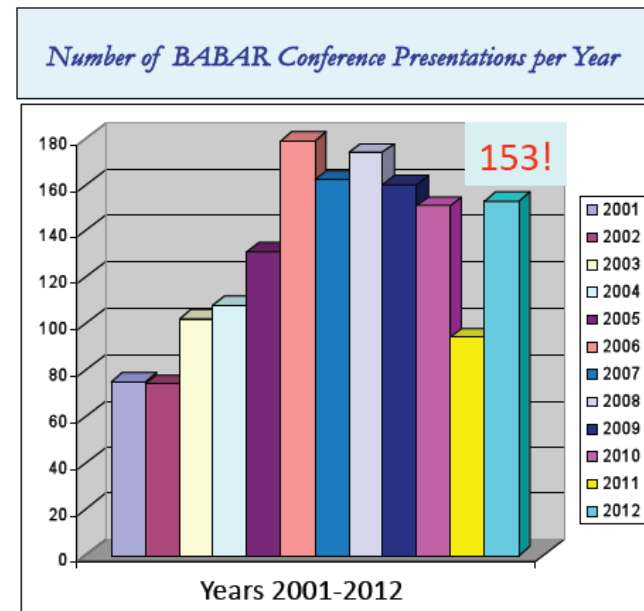
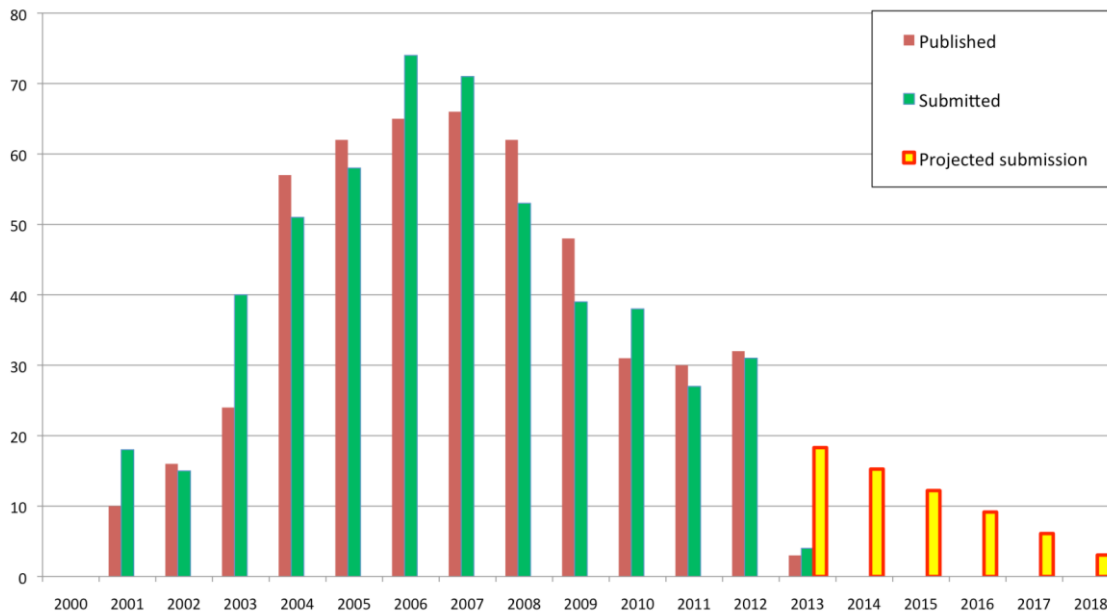
16 People joined BaBar as Associates Since Oct 2012 Collab Meeting

5 Grad students  
 0 Undergraduates  
 2 Postdocs  
 4 Faculty  
 4 PhD Staff  
 1 non-PhD Staff (SLAC)



# PRODUCTIVITY OF THE COLLABORATION

- 506 published papers
  - plus: 4 accepted, 6 submitted, 9 ready for submission
  - more in the pipeline to be submitted/published in 2013
  - Again 2013 will exceed projections
- 153 conference talks in 2012, >100 expected in 2013





# INGEST

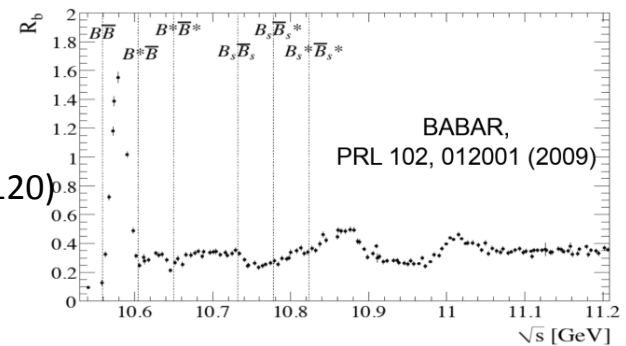
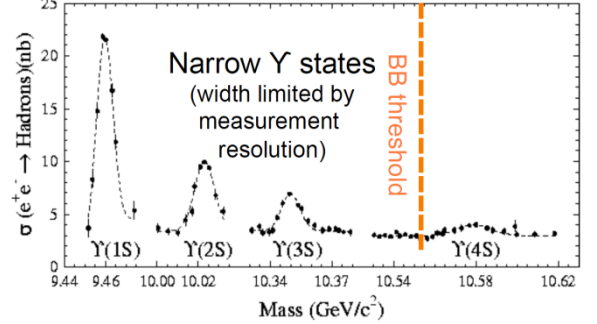
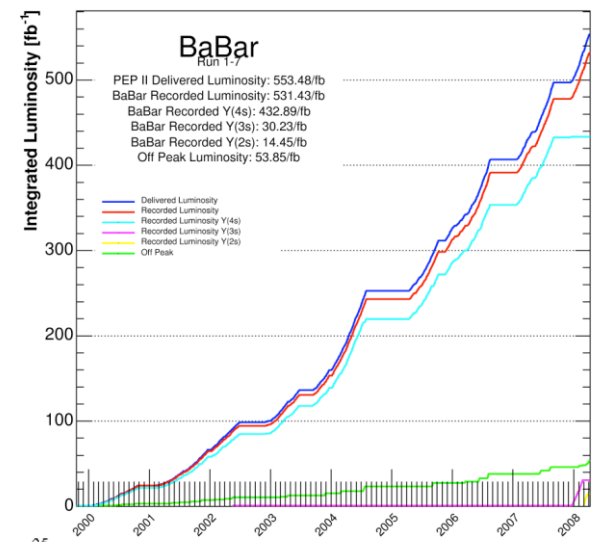
- “Ingest Issues”
  - How did you (the experiment) decide what data to save, how to make it discoverable / available, how is it documented, where is the data / meta-data etc. What are the access policies and target communities?
  - What tools do you use?
- Quick answer
  - Physics output must be maximized → level 4 preservation of two most recent data processings
  - Prepare accurate documentation that takes a green users from the start to the end of an analysis
  - Metadata/bookkeeping, internal notes, talks, are all stored/tracked in our databases accessible with tools (command line or web based) well described and easy to use while the code (core and user) is maintained in cvs.
  - Data: disk+tape
  - Target: the collaboration



# 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





# *BABAR DATA*

- BaBar has collected data from Oct 22<sup>nd</sup> 1999 to Apr 7<sup>th</sup> 2008
  - 700TB of raw data (~800TB including calibration runs), 1.2 PB (and counting...) from the last two data reprocessings
  - 65 on track analyses
    - Plus ~30 analyses progressing slower (generally lacking manpower)
    - Possibilities for new previously unforeseen analyses including discovery analyses
- $9 \times 10^9$  physics events,  $27 \times 10^9$  simulated events generated with the last code release (namely R24)
- Previous release, R22, still widely used when the data preservation effort started and still used by many ongoing analyses
  - Need to keep R22 processed data in order to maximize physics output



# LONG TERM DATA ACCESS

- Insure the ability to support analysis of the BaBar data until at least 2018
  - Data, conditions and calibrations, releases and tools, databases, capability of running production and user jobs
    - This means that in 5 years from now it will be possible to add a new decay mode, produce the MC events and the relevant skims, and perform a completely new analysis developing new selection code, fitting procedures, etc.
  - Accurate documentation
- Providing a stable environment
  - Validated OSs (SL4, SL5, SL6) enclosed in a virtualization layer able to support the BaBar Framework
    - Need to address: hardware support, security risks, keep know-how on OS, Framework, ...
- Open formats
  - Data format is ROOT based, databases work on Oracle and MySql, code is written in C/C++, Tcl, Perl, Python.
- Data Storage
  - 2PB (including raw data will be stored on tape in two Tier A sites (SLAC, CC-IN2P3)
  - Specific datasets are available at other TierA sites
  - Most used data sits on disk at SLAC





# LTDA



The LTDA cluster is in production mode since March 21<sup>st</sup> 2012

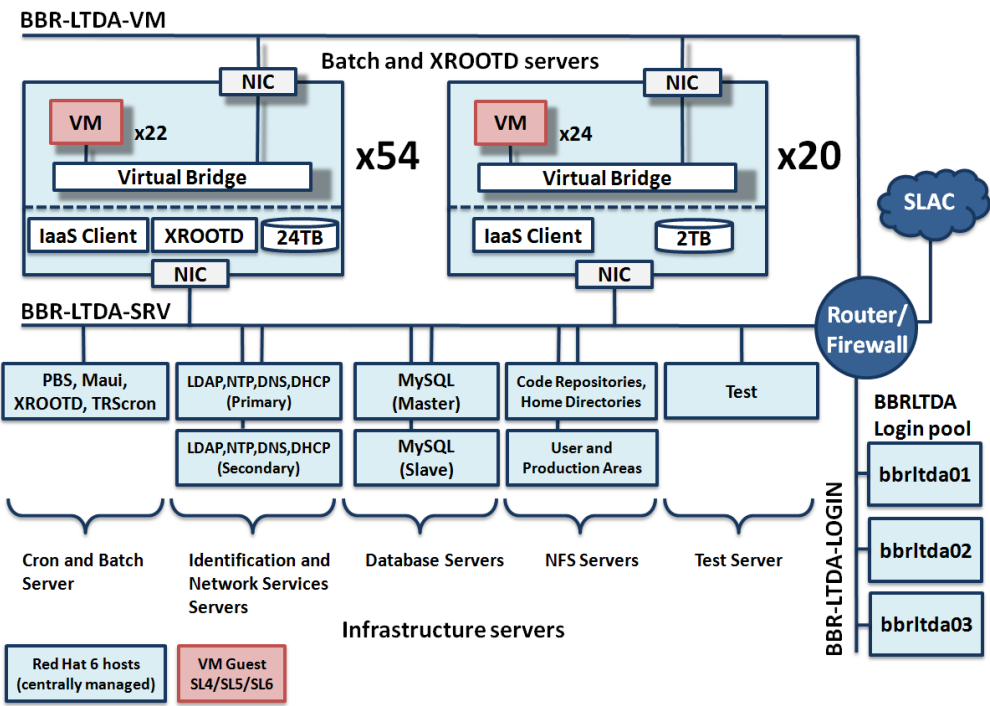
- 1.33 PB of disk space for data and users
- 1668 job slots
- SL4, SL5, SL6 platforms available

- Cisco 6506 network switch with 2x10Gb link card and 192 x 1Gb ports
- 9 infrastructure servers (Dell R410/R510)
- 54 batch and storage servers
  - Dell R510: dual 6-core Intel Xeon X5675, 3.07GHz, 48GB RAM, 12x2TB disks (no RAID)
    - 11x2TB disks used to stage data through XROOTD
    - 1x2TB used as local scratch
    - 24 cores with hyper threading
      - One VM per core up to 22 VMs/host
    - 1 physical core used for the host and XROOTD
- 20 batch servers (no XROOTD)
  - Dell R410: dual 6-core Intel Xeon X5675, 3.07GHz, 48GB RAM, 2x2TB disks mirrored (for OS + local scratch)
  - 24 cores used to run batch jobs (VMs)
- 2 NFS servers
  - Sun X4540 Thor server, 12 cores, 32 GB RAM and 32TB
  - One for local home directories and code repositories and one for user data



# VIRTUALIZATION & NETWORK

- Security threat associated to a VM connected to a network running old OS
- Risk based approach assuming that the VMs are compromised
  - Isolation of back versioned components with firewall rules
  - Physical hosts centrally managed by SLAC CD
  - Images are read-only, qcow2 produces a temporary file with changes to OS and it is deleted when the VM's shut down

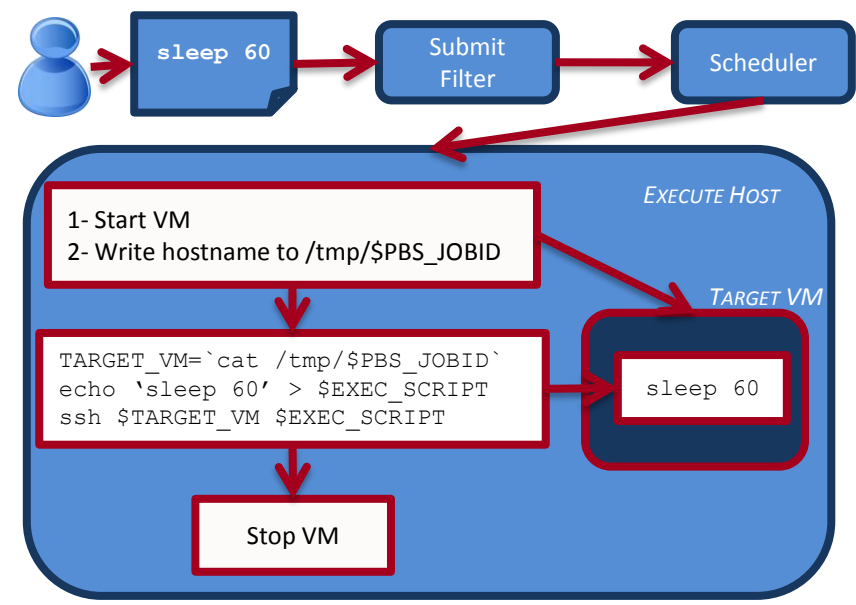


- VMs are not allowed to connect to SLAC network or the world
- The Login network is protected from the VM network
  - Allow one way ssh from Login to VM network
  - VMs are not allowed to write over the Login network
- Well defined services between VM network and SRV network
  - Infrastructure (DNS, LDAP, NTP), file service (Xrootd, nfs), batch scheduling
  - LDAP is a subset of the SLAC Kerberos list mapped on /nfs internal home directories
- Allow SRV and Login networks use SLAC infrastructure



# JOB SUBMISSION

- PBS/Torque is used to manage the batch resources and Maui is the batch scheduler
- The virtualization layer uses qemu with kvm support directly
- Need to create the network interface for the VMs
  - 24 MAC addresses per host and usage status stored in local db
- PBS Prologues and Epilogues scripts are used to create and destroy the VM's and the needed network environment





# *IF YOU WANT TO KNOW MORE*

- Homers' talk at DPHEP 7 on data, code, and storage details
  - <https://indico.cern.ch/getFile.py/access?contribId=28&sessionId=5&resId=1&materialId=slides&confId=233119>
- Tina's talk at DPHEP 6 including LTDA technical details, performance tests, and problems & solutions
  - <https://indico.cern.ch/getFile.py/access?contribId=28&sessionId=5&resId=1&materialId=slides&confId=233119>



# DOCUMENTATION WORKING GROUP

- All most used and fundamental info are being checked, updated and moved to a Media Wiki server, the *BABAR WIKI*
  - The effort needed is not trivial and the pages which are likely never to change again (Ex: detector web pages) are left in their original location
- The Documentation Working Group is coordinating the migration effort aided by an advisory committee
  - There are about 10 members (many are students) in the DWG but we promote the migration to the wiki as a Collaboration effort
  - Experts sign-off on the content of migrated pages
- Longer Term
  - Considering INSPIRE for long term storage of the internal notes and talks
    - A lot of help received by Zaven Akopov and Bernard Hecker but need to find time to convert our db tables with the metadata into suitable XML format
  - No choice made by the Collaboration yet



# *HOW IS THE WIKI USED*

- Workbook
  - A step by step tutorial from the login to the analysis of the final rootuples
    - Includes also BaBar Analysis Schools talks and videos with many tutorials on analysis techniques and statistical approaches
- Technical pages
  - Many technical pages in place and signed off by experts and senior members
    - Tracking, trigger, PID, Neutrals, ..., data and datasets, ...
- Personal pages and Notebooks
  - Used as logbooks during an analysis to show plots, make to-do lists, write/receive comments
- The DWG members act as moderators to avoid duplication of information and proliferation of cross references

https://bbr-wiki.slac.stanford.edu/bbr\_w

### Workbook Contents

Here is a list of the contents of the BaBar Workbook, following sections should be completed in order.

- [1. Account Setup](#)  
How to set up your SLAC Account.
- [2. Quick Tour](#)  
A quick tutorial in which you perform all the b the core of the Workbook; most other pages r
- [3. Packages and Releases](#)  
Packages and releases: what they are, how BaBar's file system. How to use SRT/CVS co
- [4. Modules and the Framework](#)  
Introduction to modules. How modules are pu
- [5. Event Information](#)  
How to access the event information, and w main particle candidate lists and event variab
- [6. Tcl: Run-time job control](#)  
How to interact with the Framework and talk
- [7. Editing Module Code](#)  
How to write code in modules. Using the main BtaCandidate, and HepAList.
- [8. Compile and Link](#)  
All about gmake, BaBar's compile-and-link util
- [9. Run the Job](#)  
How to run BaBar executables. Program flow
- [10. Debugging](#)  
Using the debuggers to track run-time errors.
- [11. Run-time Parameters](#)  
How to make module parameters into run-time
- [12. Tcl Files](#)  
Basic Tcl commands. How Tcl files work. Tcl
- [13. Find Data](#)  
Using BbkDatasetTcl to find data, and to prod analysis job.
- [14. Batch Processing](#)  
How to use the batch system. Batch comman Computing CPU time for batch system log files
- [15. Analysis Tools](#)  
An overview of important analysis tools, inclu composition, and more.
- [16. Sim/Reco](#)  
Monte Carlo (MC) simulation, and reconstruct
- [17. Make CM2 Ntuples](#)

Firefox

Physics/Physics notebooks/TDBC 1001 -...

https://bbr-wiki.slac.stanford.edu/bbr\_



™ and © Nelvana, All Rights Reserved.

Page Discussion Read Edit View h

## Physics/Physics note

< Physics | Physics notebooks  
(Redirected from Physics analysis/TDBC 1001)

Search for T, CP and CPT Violation In B update

Contents [hide]

- 1 Analysts
- 2 Other documentation
  - 2.1 Current experimental status on Ac
- 3 Outstanding issues
- 4 Solved issues
- 5 Datasets
- 6 Blinding scheme
- 7 Analysis flow
- 8 Detector asymmetry
- 9 Meetings
- 10 Canonical plots
- 11 Notebook
- 12 Code cheat sheet
- 13 Notes

### Analysts

- [redacted]
- [redacted]
- [redacted]
- [redacted]

### Other documentation

- Supporting document [BAD #2384](#) [#P](#) in B
- Previous paper from BaBar <sup>[1]</sup>
- Collaboration Meeting, Jan 2011 Present
- Update from Daniel at June 2011 Collabo
  - Summary of topics covered.
  - "The asymmetric Kerr metric as a source

https://bbr-wiki.slac.stanford.edu/bbr\_wiki/index.php/Physics

Firefox

Physics/Tagging - Bbr\_wiki

https://bbr-wiki.slac.stanford.edu/bbr\_

Cartaro Talk Preferences Watchlist Contributions L

Page Discussion Read Edit View history ☆ Search

## Physics/Tagging

< Physics

### B Flavor Tagging

Static HTML web page [#P](#)

- This page is under construction. In the meanwhile, please see the [Static HTML web page](#) [#P](#).

Contents [hide]

- 1 What is B flavor tagging?
  - 1.1 Definitions
- 2 Tagging algorithm
  - 2.1 Lepton
  - 2.2 Kaons
    - 2.2.1 Kaon1
    - 2.2.2 Kaon2
  - 2.3 Pions
    - 2.3.1 KaonPion
    - 2.3.2 Pion
  - 2.4 Other
  - 2.5 Notag
- 3 Determining Tagging Performance and Parameters
- 4 How to use Tag08?
  - 4.1 Release required
  - 4.2 Additional tags required?
  - 4.3 Recipes
  - 4.4 Documentation
- 5 Performances and fit parameters
  - 5.1 Results of the () fit
    - 5.1.1 Performance summary tables
    - 5.1.2 "Tag08": MONTECARLO BFlav sample (SP-2222)
    - 5.1.3 Performance summary tables

### What is B flavor tagging?

!!!



# INTERNAL DOCS

Firefox

BAIS: TDBC-10/02 Details

https://www.slac.stanford.edu/babar-internal/BAIS/info/analyses.html?bat\_id=1123

AWG Code	Analysis Name	Description	Data Info															
TDBC-10/02		T-reversal asymmetry, e.g., B <sub>even</sub> → B <sub>0</sub> and B <sub>0</sub> → B <sub>even</sub> , which cannot be achieved by CP transformation.	Sources: CM2 Samples: Run1 Run2 Additional comment:															
Publication Status <b>PUB</b>	Conference Status <b>NA</b>	T-violation in B → (ccbar)K <sub>0</sub>																
Analysts (*primary)	[redacted]																	
Review Committee	comm525 [HN: rev-TDBC-10-02] (07 Jan 2012...21 Nov 2012) Members: [redacted] article RClor]																	
Target Publication Period																		
Target Journal	Physical Review Letters (PRL)																	
Target Conference																		
Keywords	beta, CP, T, CPT																	
PAC Keywords																		
CVC Keywords																		
CWR #1	<p><b>Result type:</b> Journal article</p> <p><b>Review Committee:</b> comm525 [HN: rev-TDBC-10-02] (...)</p> <p><b>Contact author:</b> Martinez-Vidal, Fernando</p> <p><b>CWR period:</b> 26 April 2012 to 10 May 2012</p> <p><b>Final notice period:</b> 14 July 2012 to 21 July 2012</p> <p><b>Final reader(s):</b> [redacted]</p> <p><b>Primary BAD:</b> BAD 2465, version 5</p> <p><b>Authors list:</b> [redacted]</p> <p><b>Supporting BAD(s):</b> BAD 2398</p> <p><b>PubDb document:</b> PUB-12-011</p> <p><b>Institutional Reading Group(s)</b> 5a. Anney, Cincinnati, Johns Hopkins, LAL, Nikhef, Notre Dame, UT Dallas 6b. Edinburgh, Liverpool, Milan, Stanford, UT Austin, Wisconsin</p> <p><b>Target:</b> Physical Review Letters (PRL)</p> <p><b>Collaboration-wide talk:</b> [redacted] 16 May 2012 <a href="#">Direct Observation of Time Reversal Violation</a></p>																	
BADs	<table border="1"> <thead> <tr> <th>BAD #</th> <th>Doc Type(s)</th> <th>BAD Title</th> </tr> </thead> <tbody> <tr> <td>BAD 2398</td> <td>Supporting Document</td> <td>Time Reversal Violation using B → (ccbar)K<sub>0</sub> decays</td> </tr> <tr> <td>BAD 2465</td> <td>Journal Draft</td> <td>Observation of Time Reversal Violation in the B<sub>0</sub> meson system</td> </tr> </tbody> </table>			BAD #	Doc Type(s)	BAD Title	BAD 2398	Supporting Document	Time Reversal Violation using B → (ccbar)K <sub>0</sub> decays	BAD 2465	Journal Draft	Observation of Time Reversal Violation in the B <sub>0</sub> meson system						
BAD #	Doc Type(s)	BAD Title																
BAD 2398	Supporting Document	Time Reversal Violation using B → (ccbar)K <sub>0</sub> decays																
BAD 2465	Journal Draft	Observation of Time Reversal Violation in the B <sub>0</sub> meson system																
PubDb Doc	<table border="1"> <thead> <tr> <th>BaBar #</th> <th>Preprint #</th> <th>SLAC PUB #</th> <th>Accession Code</th> <th>PubDb Title</th> </tr> </thead> <tbody> <tr> <td>PUB-12-011</td> <td>1207.5832 [hep-ex]</td> <td>SLAC-PUB-15192</td> <td>[redacted]</td> <td>Observation of Time Reversal Violation in the B<sub>0</sub> meson system</td> </tr> </tbody> </table>			BaBar #	Preprint #	SLAC PUB #	Accession Code	PubDb Title	PUB-12-011	1207.5832 [hep-ex]	SLAC-PUB-15192	[redacted]	Observation of Time Reversal Violation in the B <sub>0</sub> meson system					
BaBar #	Preprint #	SLAC PUB #	Accession Code	PubDb Title														
PUB-12-011	1207.5832 [hep-ex]	SLAC-PUB-15192	[redacted]	Observation of Time Reversal Violation in the B <sub>0</sub> meson system														
Other Docs																		
Internal Meetings	<table border="1"> <thead> <tr> <th>Date</th> <th>Name</th> <th>Title</th> <th>Speaker</th> <th>URL</th> </tr> </thead> <tbody> <tr> <td>03-NOV-2009</td> <td>Collaboration Meeting</td> <td>Time-reversal violation measurement in B → J/ψ K<sub>0</sub></td> <td>[redacted]</td> <td><a href="http://www.slac.stanford.edu/BFROOT/www/Organization/CollabMtg/2009/detNov09/Collaboration/Tues3c/villanueva.pdf">http://www.slac.stanford.edu/BFROOT/www/Organization/CollabMtg/2009/detNov09/Collaboration/Tues3c/villanueva.pdf</a> [Edit]</td> </tr> <tr> <td>22-APR-2010</td> <td>-</td> <td>Topic: Time-reversal violation measurement in B → J/ψ K<sub>0</sub></td> <td>[redacted]</td> <td><a href="https://www.slac.stanford.edu/BFROOT/www/doc/BbrMeetingOrganizer/vol04/pro/mtr001679/itm0005046/">https://www.slac.stanford.edu/BFROOT/www/doc/BbrMeetingOrganizer/vol04/pro/mtr001679/itm0005046/</a> [Edit]</td> </tr> </tbody> </table>			Date	Name	Title	Speaker	URL	03-NOV-2009	Collaboration Meeting	Time-reversal violation measurement in B → J/ψ K <sub>0</sub>	[redacted]	<a href="http://www.slac.stanford.edu/BFROOT/www/Organization/CollabMtg/2009/detNov09/Collaboration/Tues3c/villanueva.pdf">http://www.slac.stanford.edu/BFROOT/www/Organization/CollabMtg/2009/detNov09/Collaboration/Tues3c/villanueva.pdf</a> [Edit]	22-APR-2010	-	Topic: Time-reversal violation measurement in B → J/ψ K <sub>0</sub>	[redacted]	<a href="https://www.slac.stanford.edu/BFROOT/www/doc/BbrMeetingOrganizer/vol04/pro/mtr001679/itm0005046/">https://www.slac.stanford.edu/BFROOT/www/doc/BbrMeetingOrganizer/vol04/pro/mtr001679/itm0005046/</a> [Edit]
Date	Name	Title	Speaker	URL														
03-NOV-2009	Collaboration Meeting	Time-reversal violation measurement in B → J/ψ K <sub>0</sub>	[redacted]	<a href="http://www.slac.stanford.edu/BFROOT/www/Organization/CollabMtg/2009/detNov09/Collaboration/Tues3c/villanueva.pdf">http://www.slac.stanford.edu/BFROOT/www/Organization/CollabMtg/2009/detNov09/Collaboration/Tues3c/villanueva.pdf</a> [Edit]														
22-APR-2010	-	Topic: Time-reversal violation measurement in B → J/ψ K <sub>0</sub>	[redacted]	<a href="https://www.slac.stanford.edu/BFROOT/www/doc/BbrMeetingOrganizer/vol04/pro/mtr001679/itm0005046/">https://www.slac.stanford.edu/BFROOT/www/doc/BbrMeetingOrganizer/vol04/pro/mtr001679/itm0005046/</a> [Edit]														
Related Analy	<table border="1"> <thead> <tr> <th>AWG Code Relation</th> <th>Related Analysis Name</th> </tr> </thead> <tbody> <tr> <td>TDBC-07-03</td> <td>Sin(2beta) from (c cbar) K<sub>0</sub>, Run1-6 update</td> </tr> </tbody> </table>			AWG Code Relation	Related Analysis Name	TDBC-07-03	Sin(2beta) from (c cbar) K <sub>0</sub> , Run1-6 update											
AWG Code Relation	Related Analysis Name																	
TDBC-07-03	Sin(2beta) from (c cbar) K <sub>0</sub> , Run1-6 update																	

BAIS: Analysis Index

https://www.slac.stanford.edu/babar-internal/BAIS/info/index.html?VIA\_FORM=1&SEARCH

[BaBar Home](#)
[Computing](#)
[Detector](#)
[Documentation](#)
[Hypernews](#)
[Organization](#)
[Physics](#)
[BaBar Web Search](#)
[Meeting Organizer](#)

[Help](#)
[Analysis Index](#)
[Analysis Detailed Index](#)
[AWG Index](#)
[IRG Index](#)
[Admin & Maintenance](#)

[Create new analysis entry](#)
[Create new AWG](#)
[Review Committees](#)
[Analyst List](#)
[Reviewer List](#)

[Potential Reviewer List](#)
[RClor Time Alerts](#)

HN userid: "cartaro"

### Analysis index for all AWGs

BAIS contains [redacted] analyses.

Please either enter a search query here :

We are testing a new way of selecting BAIS entries using a pseudo-SQL command to be entered by the user. If you decide to try it, please send comments or problem reports to [redacted]@slac.stanford.edu. The usual BAIS sort and selection box remains unchanged. Thank you.

**Analysis Search Form**

Enter search query : find code=TDBC-10/02

For example : find (awg = charm or cwrstart > 01-jan-1900) and name = mixing  
or find awg = charm order by pubstat (note: order by can be used with awg, pubstat, name, confstat, cwrstart, cwend and sub)

Search by	AWG	AWG code	Analysis Name	Analyst(s)	Publication Status	Target Pub Period	Target Journal	Conference Status
Code	awg	code	name	analyst	pubstat	targetpub	journal	confstat
Options	breco	awg-yy/xx	anything	anything	Free	4/13	PRL	None

Search by	Target Conference	CWR start	CWR end	Keyword	RC first meeting	PAC sign-off	RClor EET	BAD #	IRG	Submission	Final Notice Start	Final Notice End
Code	targetconf	cwrstart	cwend	key	firstmtg	signoff	eet	bad	irg	sub	fnstart	fnend
Options	RICH 2013	dd-mmm-yyyy or 0 (empty date)	dd-mmm-yyyy or 0 (empty date)	anything	dd-mmm-yyyy or 0 (empty date)	yes/no	yes/no	number	ex:1a	dd-mmm-yyyy or 0 (empty date)	dd-mmm-yyyy or 0 (empty date)	dd-mmm-yyyy or 0 (empty date)

Submit Query Reset form

Or enter selection & sort criteria here :

**Analysis Selection & Sort Form**

Sort by:  AWG Code  Publication Status  Analysis Name  Conference Status





# ACCESS POLICY

- You need to be a BaBarian to access *BABAR* data
- Point is... anyone can be a BaBarian
- If you have an idea to test, a theory to verify, or a new measurement that can be done with our data come and join the Collaboration
  - You sign only your paper(s) together with the Collaboration as an Associate and then you may decide to become a Member (Collaboration service requested) and sign all the papers
  - Very fruitful interaction with theorists
    - Example: T-Violation measurement (see later) done by a theorist (now BaBar Associate) and a BaBar PhD student
- Public access problem still not faced by the Collaboration
  - Even if we made the data public without the Framework is quite useless
    - Need manpower to support
    - Small data samples are used by affiliated groups for education (undergrad students)



# ARCHIVE

- “Archive issues”:
  - How is the archive managed? How are errors detected and handled? What is the experience?
  - What storage system / services are used?
  
- Quick answer
  - The archive is managed by the BaBar Computing Group with infrastructure support from SLAC Computing division
  - There are SLAC centralized monitoring tools but we developed our own tool that notifies via email any failure. We also implemented batch queue monitoring and system usage stat reports.
  - Storage is a mix of disk (XROOTD) + tape



# *KEEP LTDA RUNNING*

- Administrative issues and user support
  - 1 FTE needed so far
  - Development still ongoing
    - Validation system
      - Shield against kernel updates bugs, only tested updates are released to the cluster
    - Monitoring
      - Web based tools
    - Backup system
      - ZFS snapshots + tapes
- Everything else falls under general BaBar support
  - Software and hardware



# *BABAR COMPUTING SUPPORT*

- Need professional support (0.5 FTE) and BaBar support (~2FTE for BaBar core computing) plus Collaboration service work
- We also rely on SLAC Computing Division for hardware support and infrastructure (including Storage Tek tape system and LSF batch queues)
  - Lab specific agreement: at SLAC “everything under B50 roof” is under maintenance contract
- Most of the infrastructure is operational but downsizing will come soon enough
  - In 2013 at least two TierA sites will run out of BaBar funds
  - Hardware past end of life (not under B50 roof) and without any maintenance contract



# BABAR COMPUTING HARDWARE

## Tape technology

- BaBar has 2700 x 1TB tapes (T10kB, 1TB each)
  - Includes user data backup
  - Doesn't include users' personal backup (yet) ...
    - Traditionally provided by SLAC
- T10kC (5 TB) technology has been out for quite some time now and T10k-D (6 – 10 TB) will be probably announced soon
  - Oracle's roadmap also includes T10k-E in 2015 at 12-20TB capacity
- Drives are guaranteed to read/write only the current generation and to read only the previous generation
  - Good news: T10k-C are reformattable as T10k-D
- Hardware (non)refreshment
  - User data is stored on old NFS servers (often >7years old) and backed up to tape every 3 months (assume slowly changing data)
  - Data distribution and HPSS buffer space
  - Web servers, db servers, production servers, ...



# OFFLINE ENVIRONMENT

- “Offline environment issues”:
  - What have been the key challenges in keeping the offline environment alive? What are the key lessons learned / pitfalls to be avoided? What would you have done differently if long-term preservation had been a goal from the early days of the experiment?
- Quick answer
  - Stumbling stone: migrations (code, platforms, media)
  - Simplicity and portability
    - If common formats are defined stick to them
  - Use of open formats
    - Sooner or later you’ll run out of funds and licenses are a very nasty chapter especially when you would like to move away from proprietary tools and you’re lacking manpower
  - Have faith that storage will become inexpensive in the right timeframe



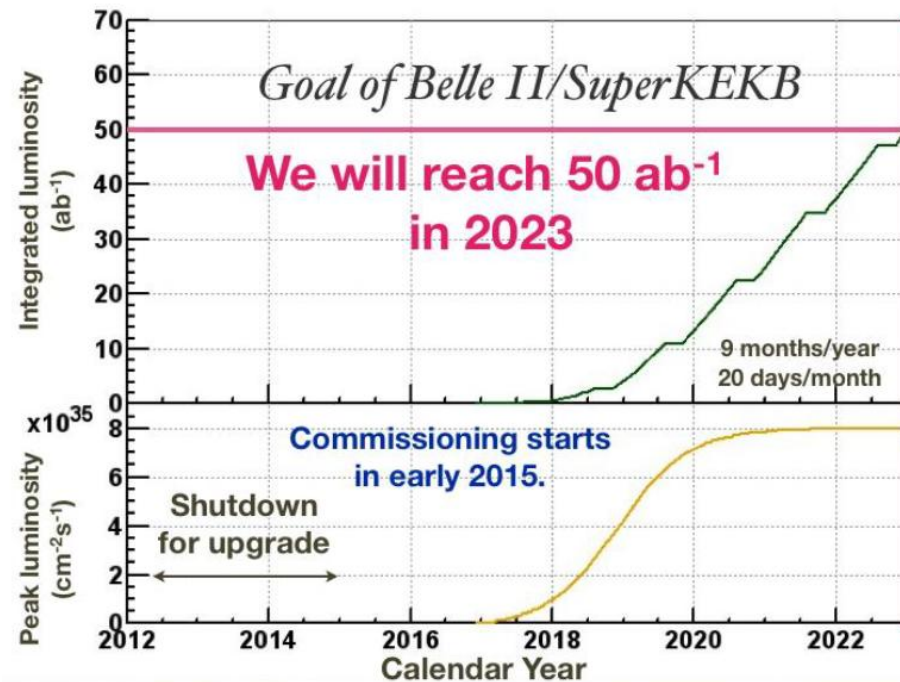
# *PITFALLS AND WHATIFS*

- Define key strategies for the “long term” before the long term arrives
  - Well, yes... easy to say but it is hard to imagine 10-20 years in the future and find resources when it is either too late or too early and forces are either lacking or needed elsewhere
- Adopt a simpler (is it possible?) or common self describing data format and look for portability
  - LTDA is a portable design, VMs are very portable too and will save us from very painful platform migrations
- When it is too late then try to understand what “long term” means and make a strong case



# LONG TERM ? HOW LONG?

- BaBar (and Belle) data will not be superseded by LHC data
  - Some datasets expected to remain unique for longer (Y(3S) dataset)
- Long Term Data Access well matched with Belle II timescale
- Now that the SuperB case is clear we need to understand the impact on our program
  - Need to embark on a long term planning process that will enable the BaBar data set to be exploited for as long as is practical and useful
  - Consider modifying BaBar Governance Rules to facilitate the implementation of those plans and to ensure that the integrity and reputation of the collaboration are maintained



K. Akai, SuperKEKB accelerator status, Nov 12, 2012





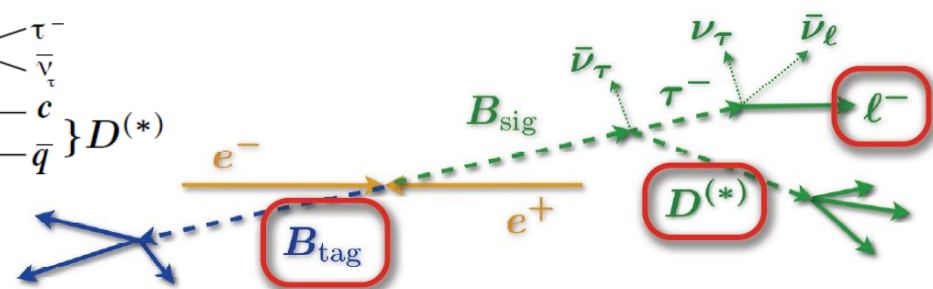
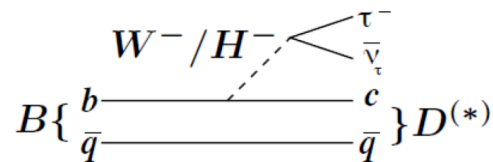
# *MAKE A STRONG PHYSICS CASE*

- BaBar made a *Beyond 2012* plan back in 2010 but we are running (and spending) way above expectations
  - Need DOE support
- BaBar being repeatedly in the news in 2012 and publishing new results (not updates of old analyses) make our case stronger in a very desolated landscape
  - Sequestration has not yet hit SLAC in a clear way but the new DOE baseline (-5%, that is -10% at this time of the fiscal year) will give us undoubtedly another squeeze

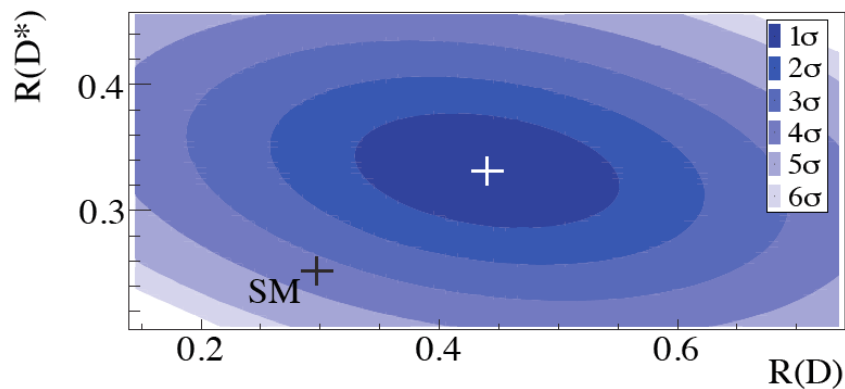


# TESTING SM AND CHARGED HIGGS

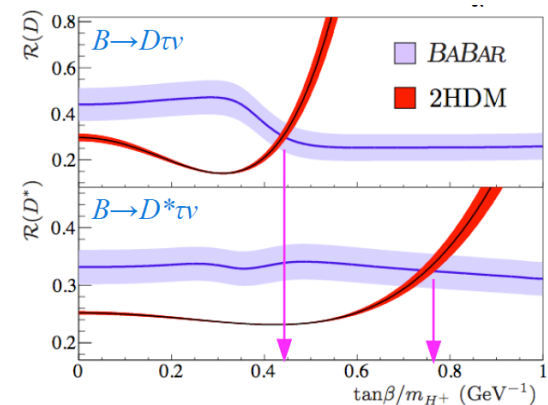
- $B \rightarrow D^{(*)} \tau \nu$



$$R(D^{(*)}) = \frac{BF(B \rightarrow D^{(*)} \tau \nu)}{BF(B \rightarrow D^{(*)} l \nu)}$$



Combined deviation from SM at  $3.4 \sigma$   
 2HDM-II excluded with probability 99.8%  
 all over the  $(\tan\beta, m_H)$  parameter space



$$R(D) \Rightarrow \tan\beta / m_H = 0.44 \pm 0.02$$

$$R(D^*) \Rightarrow \tan\beta / m_H = 0.75 \pm 0.04$$

### Synopsis: More tau leptons than expected



Evidence for an Excess of  $B \rightarrow D^{(*)} \tau \bar{\nu}_\tau$  Decays  
 J. P. Lees et al. (BABAR Collaboration)  
 Phys. Rev. Lett. 109, 101802 (2012)  
 Published September 6, 2012

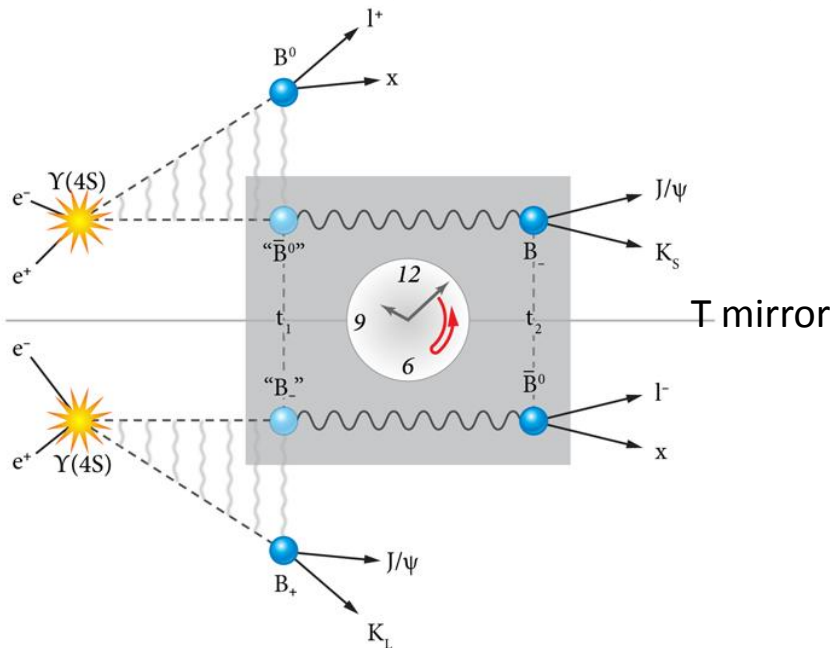
As reported in *Physical Review Letters*, the BaBar collaboration at SLAC has analyzed a large data set and found an excess of events containing tau leptons in the decay of bottom mesons that doesn't agree with the predictions of the standard model of particle physics.

**PRL Highlight as well as Editor's Suggestion  
 in September 2012 Vol 109, Issue 10**



# T-VIOLATION

- Identify a T-conjugate pair of processes that can be experimentally distinguished and measured
- Measure time-dependent asymmetries of T-conjugate pairs of B decays, making use of the EPR entanglement from  $\Upsilon(4S)$
- T-violation observed at  $>14\sigma$  !
  - Consistent with expectation from CPT theorem and measured  $\sin 2\beta$ . Measurement independent from any assumption on CP or CPT

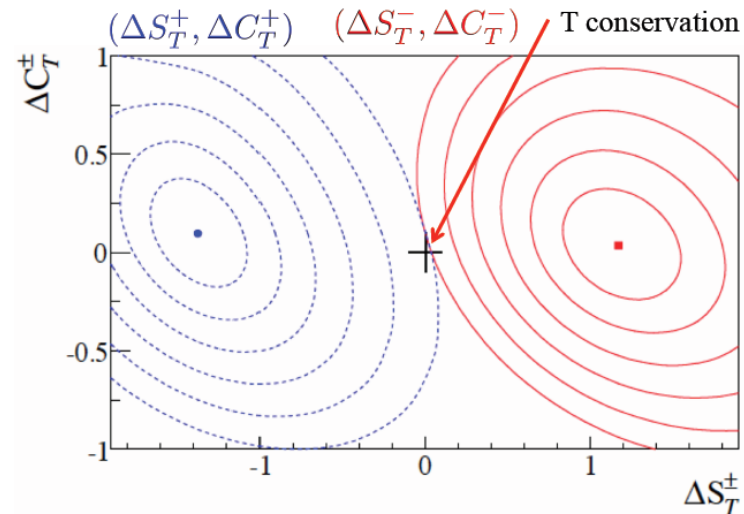


$$A_T = \frac{P(|i\rangle \rightarrow |f\rangle) - P(|f\rangle \rightarrow |i\rangle)}{P(|i\rangle \rightarrow |f\rangle) + P(|f\rangle \rightarrow |i\rangle)}$$

$$|i\rangle = \frac{1}{\sqrt{2}} [B^0(t_1)\bar{B}^0(t_2) - \bar{B}^0(t_1)B^0(t_2)]$$

$$= \frac{1}{\sqrt{2}} [B_+(t_1)B_-(t_2) - B_-(t_1)B_+(t_2)]$$

Phys. Rev. Lett. **109**, 211801 (2012)



Semileptonic decays project a B-flavor state:  $l^+ \rightarrow B^0$  ( $\bar{B}^0$  flavor tag) and  $l^- \rightarrow \bar{B}^0$  ( $B^0$  flavor tag)  
 Decays to  $J/\psi K_{L,S}$  project a CP eigenstate:  $J/\psi K_L \rightarrow B_+$  ( $B_-$  CP tag) and  $J/\psi K_S \rightarrow B_-$  ( $B_+$  CP tag)



# T-VIOLATION IN THE PRESS

Firefox  
Particle physics: Time's arrow in B mesons...  
www.nature.com/nature/journal/v491/n7426/full/491640a.html  
nature International weekly journal of science  
Full text access provided to CERN by Periodicals Section  
Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors  
Advanced search

Firefox  
The arrow of time: Backward ran sentences...  
www.economist.com/node/2156111/print  
The Economist  
World politics | Business & finance | Economics | Science & technology | Culture | Blogs | Debate | The World in 2013 | Multimedia  
Print edition  
**The arrow of time**  
**Backward ran sentences...**  
To the relief of physicists, time really does have a preferred direction  
Sep 1st 2012 | from the print edition  
TIME seems to flow inexorably in one direction. Superficially, that is because things deteriorate with

Firefox  
physicsworld.com  
IOP A website from the Institute of Physics  
Sign in | Forgotten your password? | Sign up | Contact us  
Search  
Filter by topic: Please select... Filter  
Home | News | Blog | Multimedia | In depth | Jobs | Events  
Buyer's guide

Firefox  
Time-reversal asymmetry in particle physics...  
www.physicstoday.org/resource/1.phtoad/v5/i11/pi16\_sl  
physicstoday  
Home | Print Edition | Daily Edition | Advertising | Buyer's Guide | About us | Jobs | Sign In | View Cart  
Keyword Advanced  
enter search here... All Site Content GO  
FOLLOW US: FACEBOOK TWITTER RSS E-MAIL ALERT  
Physics Today / Volume 65 / Issue 11 / Search and Discovery  
Previous Article | Next Article  
Alerts Tools Share  
**Time-reversal asymmetry in particle physics has finally been clearly seen**  
Bertram M. Schwarzschild  
November 2012, page 16  
DIGITAL OBJECT IDENTIFIER  
http://dx.doi.org/10.1063/PT.3.1774  
Bedrock theory has insisted since 1964 that the weak interactions should look slightly different when the movie is run backwards.  
PUBLICATION DATA  
ISSN  
0031-9228 (print)  
1945-0699 (online)  
PUBLISHER  
American Institute of Physics  
PACS  
11.30.-j  
Symmetry and conservation laws  
11.30.Qc  
Spontaneous and radiative symmetry breaking  
14.40.Nd  
Bottom mesons (IB>0)  
Full Text: Download PDF | View Cart  
When it was discovered in 1957 that the weak interactions of elementary particles are not symmetric under the parity operation  $P$ , theorists retreated to the seemingly safe presumption that the combined operation  $CP$  remained an inviolate symmetry. (The charge-conjugation operation  $C$  replaces all particles by their antiparticles.) But seven years later came a second rude awakening: The decay of neutral  $K$  mesons revealed a minuscule but undeniable violation of  $CP$  symmetry. So particles viewed in a mirror don't behave exactly like their antiparticles.

NATURE | RESEARCH HIGHLIGHTS  
PARTICLE PHYSICS...  
Time's arrow in B mesons  
BOTANY  
Plant fertilization protein found  
CLIMATE CHANGE  
Carbon drop in snail shell shock  
ZOOLOGY  
Blue whales...

Firefox  
Physics - Particle Decays Point to an Arrow of Time...  
physics.aps.org/articles/v5/i129  
Physics  
American Physical Society  
spotlighting exceptional research  
Log in | Create Account (what's this?)  
Home | About | Browse | APS Journals | SEARCH  
**Viewpoint: Particle Decays Point to an Arrow of Time**  
Michael Zeller, Department of Physics, Yale University, New Haven, CT 06520, USA  
Published November 19, 2012 | Physics 5, 129 (2012) | DOI: 10.1103/Physics.5.129  
An experiment studying  $B$  meson decays makes a direct observation of time-reversal violation without relying on assumed relationships with other fundamental symmetries.  
Time moves irrevocably in one direction. Things get old, decay, and fall apart, but they rarely ever reassemble and grow young. But at the particle level, time's arrow is not so clearly defined. Most collisions and other particle interactions look the same whether run forwards or backwards. Physicists have, however, identified a few reactions that appear to change when time is reversed, but the reasoning has assumed certain relations between fundamental symmetries of particle physics. The BaBar collaboration has now observed time-reversal violation directly and unambiguously in decays of  $B$  mesons. The measured asymmetry, reported in *Physical Review Letters* [1], is statistically significant and consistent with indirect observations.  
Observation of Time-Reversal Violation in the  $B^0$  Meson System  
J. P. Lees et al. (The BABAR Collaboration)  
*Phys. Rev. Lett.* 109, 211801 (2012)  
Published November 19, 2012 | PDF (free)  
+Enlarge image  
In trying to understand the nature of particle interactions, observing the behavior of those interactions under different symmetry transformations has proven invaluable in formulating and verifying the fundamental theory. It is well known, and has been experimentally shown, that the strong and electromagnetic interactions are unchanged when viewed in a mirror world, in which particle positions are reflected ( $T$  to  $-T$ ). In contrast, experiments in 1956 [2] demonstrated that the weak interaction is not invariant under such spatial inversion.  $P$  &  $C$  decade  
Article Options  
Printable Version  
Share/Email This  
Download PDF  
Export Citation (BibTeX)  
Export Citation (RIS)  
Subject Areas  
Particles and Fields  
Related Articles  
More Particles and Fields  
Observing Matter-Antimatter Oscillations  
Viewpoint: Mar 5, 2013  
Higgs-like Particle in a Mirror  
Synopsis | Feb 21, 2013  
All Particles and Fields  
New in Physics



# $a_\mu^{HAD}$ CONTRIBUTION TO $g-2$

$$a_\mu^{SM} = \left( \frac{g-2}{2} \right)_\mu = a_\mu^{QED} + a_\mu^{had} + a^{weak}$$

$a_\mu$  precisely measured at BNL E821:  
 $a_\mu^{exp} - a_\mu^{SM} = (28.7 \pm 8.0) \times 10^{-10} \quad (\sim 3.4\sigma)$

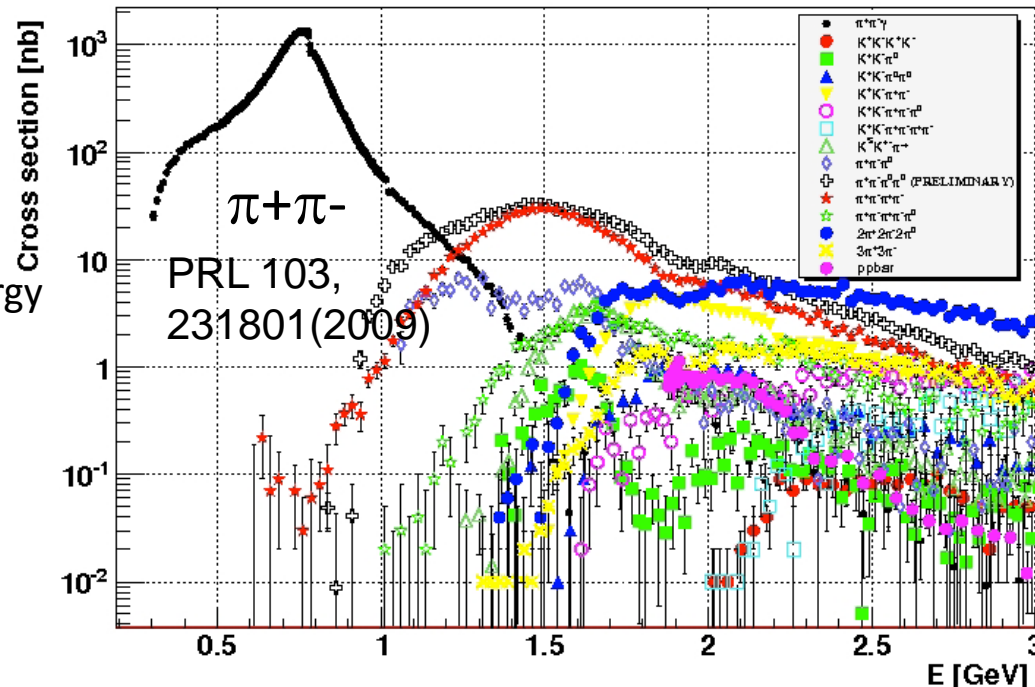
Dominant uncertainty from hadronic vacuum polarization. Cannot be calculated by QCD “first principles” so determine it via dispersion relations, by measuring the total hadronic cross section

$$a_\mu^{had} = \frac{\alpha^2}{3\pi^2} \int_{4m_\pi^2}^{\infty} \frac{K(s)}{s} R(s) ds$$

BaBar is the only experiment to measure low energy cross section from threshold up to ~4-5 GeV.

- In all cases most precise measurements

More results to be added from  $\sigma(e+e- \rightarrow \pi+\pi-\pi+\pi-)$ ,  $\sigma(e+e- \rightarrow K+K-)$  and other channels





# SLAC AND DATA PRESERVATION

- Excluding BaBar no other current project at SLAC seems to be concerned with preservation
- Fermi/GLAST relies on NASA support
  - High level data is preserved in FITS and simulations and performance measures are all parameterized. There is no plan (yet) for the low level ROOT data.
  - High level simulation and analysis, and the lower level reconstruction/simulation. The NASA science support center is committed to supporting the high level code and data products for a long time (20 yrs)
  - No real plan for low level code life past the mission end
- LCLS has a 10 years policy
  - User home (up to 20GB) stored on disk + tape for indefinite amount of time
  - Tape archive (unlimited size) has two copies for 10 years
  - Committed to have the XTC file parser compatible with all data ever taken at LCLS
  - After end of LCLS program they will maintain the code for more than 2 years but less than 10
- New experiments and projects are too young to be worried about preservation but DOE will require soon a Data Management plan
- Older experiments like SLD and LASS have data saved on tape and code partly migrated to newer platform while documentation is on paper only (when still existent) but their survival is endangered (see DPHEP2)



# CONCLUSION

- Long Term may not be as long as you think...
- Ramp down planning, freezing, validation,..., yes to all you want, but external factors could make the choice for you
  - Economic crisis, sequestration,...
- We need a robust foundation for data preservation that is really able to help on many levels, from technology to \$ if needed
  - Can DPHEP / DASPOS become such an entity ?



# ACKNOWLEDGMENT PAGE

- Thanks to the LTDA developers
  - Coordinator: Tina Cartaro
  - BaBar software expert: Homer Neal
  - Development and system administration of LTDA: Marcus Ebert
  - Network design: Steffen Luitz
  - Virtualization expert: Marcus Ebert
  - System performance and CDB: Igor Gaponenko
  - Databases, tools and production: Douglas Smith and Tim Adye
  - Computing Division experts
    - System setup and administration: Booker Bense, Lance Nakata, Randall Radmer and all the Unix-Admin team
    - Xrootd expert: Wilko Kroeger
    - Network setup: Antonio Ceseracciu
    - BaBar-SLAC Computing Division liaison: Len Moss → Andrew May