Overview, future & computing model change ...

Covering for some sub-systems who could not make it & DAQ100 ...

BROOKHAVEN NATIONAL LABORATORY

Quick overview of STAR-trek ...

- sub-systems's report
- DAQ100 summary
- Other software issues
- Mid-term software evolution and motivations
- Computing model change

Jérôme Lauret, Collaboration Meeting, February 2004

Sub-system status & progress

Still 12 sub-systems, only some will be covered here

- **PMD**
- **TOFr / TOFp**
- **E-EMC**
- **B-EMC**

- Trigger (*Akio Ogawa*)
- L3 (*Thorsten Kolleger*)
- FTPC (*Janet Seyboth*)
- SVT (*Helen Caines*)
- TPC (in regards of DAQ100 and ITTF) (*Javier Castillo*)

Trigger (*Akio Ogawa*)

`From StXXXTriggerDetector() to StTriggerData()`

- Code spanned over several makers, several people, some not maintained + and BAD Year2 & 3 experiences, asked Akio then for oversight and coordination, **one class does it all** (with help from Mirko Planicic)
- At least 1 year old class in `StEvent` i.e. **StTriggerData()**
- It allows accessing **ALL** trigger data

Easy to deal with year by year differences, while keeping backward compatibility, easy-to-use access functions

After ITTF integration (most tables gone), this will be THE way to access to trigger data. Migrate to **`StTriggerData`** NOW !!!

Caveat: Need some fixes in common micro DST

Trigger

http://www.star.bnl.gov/webdatanfs/dox/html/StTriggerData_8h-source

```
StEvent * event=(StEvent *)GetInputDS("StEvent");
StTriggerData* trgdata= event->triggerData();
If(trgdata){
   for(int pmt=0; pmt < 16; pmt++) cout << trgdata->bbcADC(east,0) <<" ";
   cout << ", sum=" << trgdata->bbcADCSum(east) <<endl;
}
```

More access function

- `unsigned short` year`()` **returns 2003 or 2004...etc**
- `void` dump() **print out a dump**
- `char*` getTriggerStructure`()` **get pointer to the trigger data structure (for experts)**

Do NOT worry about year to year difference (taken care of it for you)
Name should be obvious... (documentations lacking for now but under work)
Detector experts need to contact akio(akio@bnl.gov) if your detector need mapped access function. Once done it will be available without re-productions.

L3 (*Thorsten Kolleger*)

Initial myrinet routing problem from DAQ to L3, running stable since end of January. Currently running 3 express streams

- Upsilon (Thorsten , Soeren, ...) : EMC high tower > 3.0 GeV **&&** L2 invariant mass 7.2 < m < 50 **&&** L3 track-tower match (reject high towers from neutrals)
- strangelet (Zhangbu, Aihong, ...) central **&&** ZDC strip + ZDC neutron **&&** L3 select on CTB-ZDC
- high-pT (Tom, ...) (1 EMC high tower > 9 GeV) || (1 TPC track > 5 GeV and 1 tower > 6 GeV)

Level-3 express streams are running since Fri Feb 13, processed through FastOffline, looks reasonable.

Tom, Thorsten and Soeren will return to Germany **April 1st** .
Only remote support from Frankfurt until end of 2004.
An institution responsibility (council meeting report) ??

FTPC status (*Janet/Peter Seyboth*)

Physics Contributions (+ discussion with PWGC)

- Spectra and flow analysis (20 M events should be enough)
- Charged/Neutral particle fluctuations using PMD and FTPC (no data yet)
- Ultra peripheral collisions (no data yet)
- Centrality and reaction plane determination independent from TPC
- Charged particle correlations (back to back, forward rapidity) with FPD triggered high p_T p^0 (FPD readout not ready yet)
- Strangeness at large eta (forward Lambda)

FTPC status – then ...

FTPC in the run

- 18 M events mainly in `productionMinBias`

Calibration/ Rotation

- Presently using 2003 calibration, basically OK
- Small adjustments will be made when TPC calibration is available

- Compare vertex positions determined separately in TPC and FTPC.

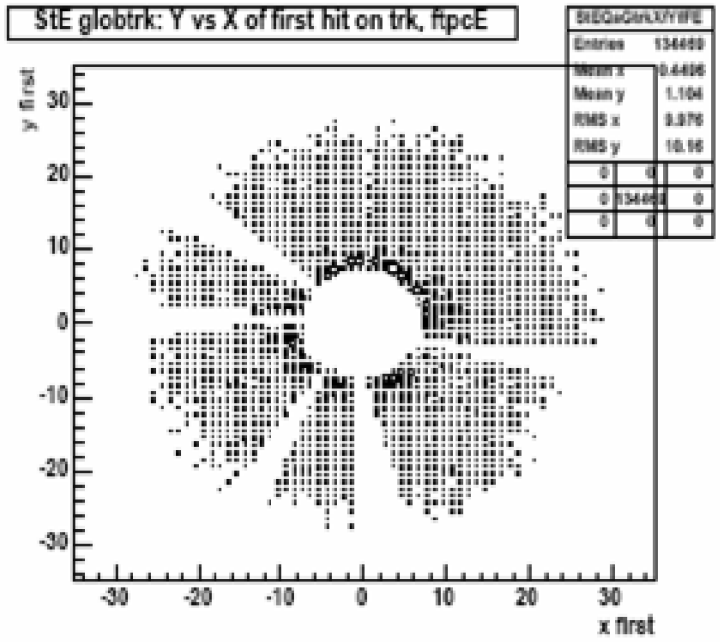
After TPC calibration :

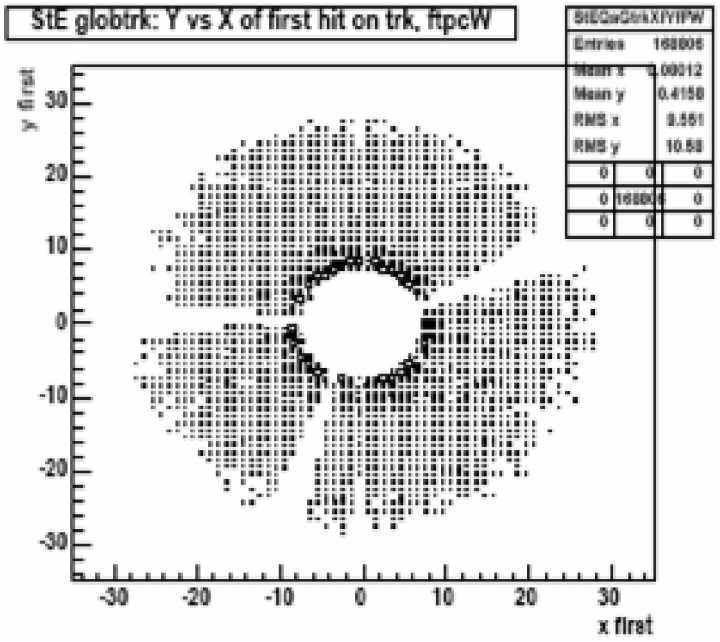
~ 10 K reconstructed events with no FTPC rotations to estimate 2004 rotation angles

~ 10 K events to verify the result

FTPC status – now ...

- unfortunately some of the readout electronics were lost
- ~ 20 % in FTPC East progressive tracking inefficiencies





~~FTPC currently turned OFF~~ **:-(, back on now with lower voltage**

Jérôme Lauret, Collaboration Meeting, February 2004

FTPC status – future ...

Medium term

- Will run with lower voltage and monitor electronic stability
- Alexei looking for along term plan

Integration of FTPC into ITTF chain:

- Work in progress, to be completed by mid-March (thanks to Markus Oldenburg)

FTPC software transfer to Purdue group:

- Technology transfer in progress
- 2004 calibration being done by Terry Tarnowsky (and Brijesh Srivastava)

TPC status & DAQ100 (*Javier Castillo*)

Offline code mainly done and working.

A few items since last analysis meeting (November 2003)

- After burner and comparison online clusters vs. offline raw + FCF done (adjustment needed)
- Required a switch to apply (not apply) TPC gain corrections (*H.Ward*) [done]

Calibration

- Synchronization of online and offline (DataBase) tpc gain files "By hand" from file produced by Tonko, once every now and then (can change only if FEEs died or are replaced) Need to be automated !!

TPC status & ITTF

A lot ~~needed~~ to be done

- Corrections needed to be moved outside of TPT (ExB, PadRow, ShortedRing, etc ... corrections and Sector alignment
  `StTpcHitMoverMaker` (*T. Kollegger*)
- Tracker tuning for Real Data (*B. Hippolyte*)
- *Worry?* Hit Errors parameters used for tracking differ from hit residuals
  - Residual studies ongoing
  - If changes, we need to redo efficiencies study for cross-check

- Other issues with TPC/ITTF (*Manuel Calderon*)

DAQ100 summary

Sent a long summary in starsoft after closing of the the chapter with embedding studies.

http://www.star.bnl.gov/HyperNews-star/protected/get/starsoft/5288.html
http://www.star.bnl.gov/protected/hbt/mercedes/daq100/nov03.html
http://www.star.bnl.gov/protected/highpt/jklay/DAQ100/25Nov2003/embedding/DAQ1(

The team

HBT	***Mercedes Lopez Noriega***
Strangeness	Betty Bezverkhny
Spectra	Levente Molnar / Raimond Snellings
Estruct	Duncan Prindle & Jeff Porter
High Pt	***Jennifer Klay***
UPC	Janet Seger
Spin	Jan Balweski
EbyE	Aihong Tang
Heavy Flavor	Jingguo Ma

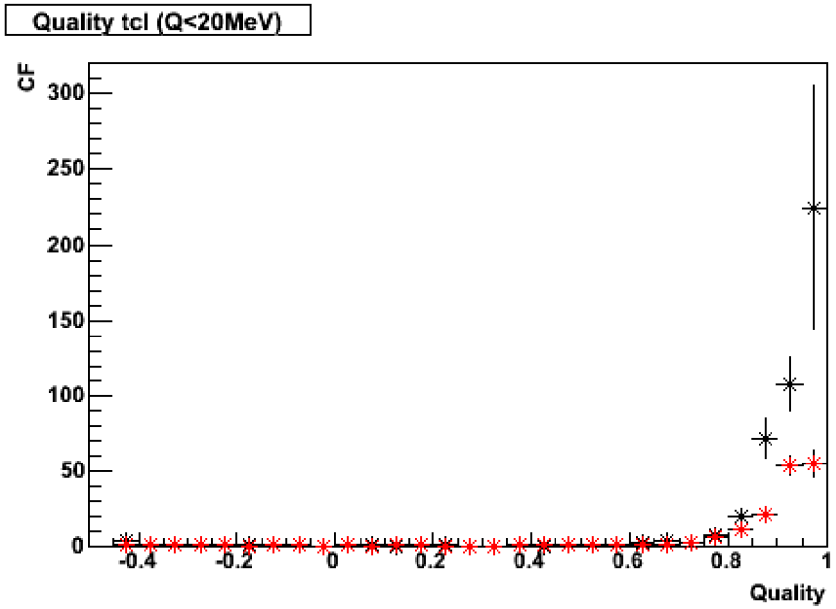
DAQ100 summary

All looked, most signed-off. The remaining points

- Divergence at large eta (> 1.0-1.2) comparing to TPT (low Pt)
- NFIT issues for nfit > 42 (stronger effect at ½ field)
- Low Pt (0.3-0.5) did not show Lambda pic (low stat and dE/dx ?)
- Better in terms of splitting, merging comparable
- UPC have less events (factor of 2) with FCF, 1/3 with current standard cuts

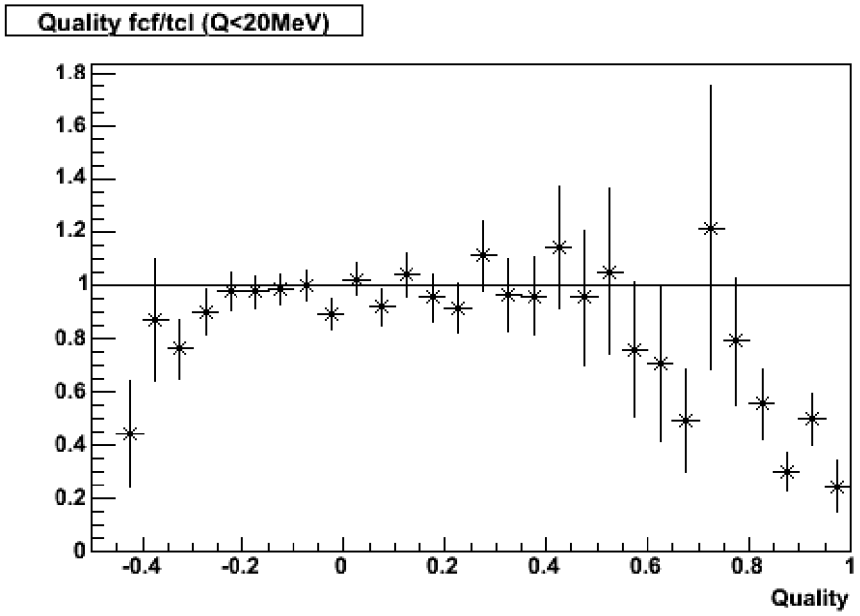
Would not be too fair to the good work from all to select results, nonetheless ...

DAQ100 - HBT

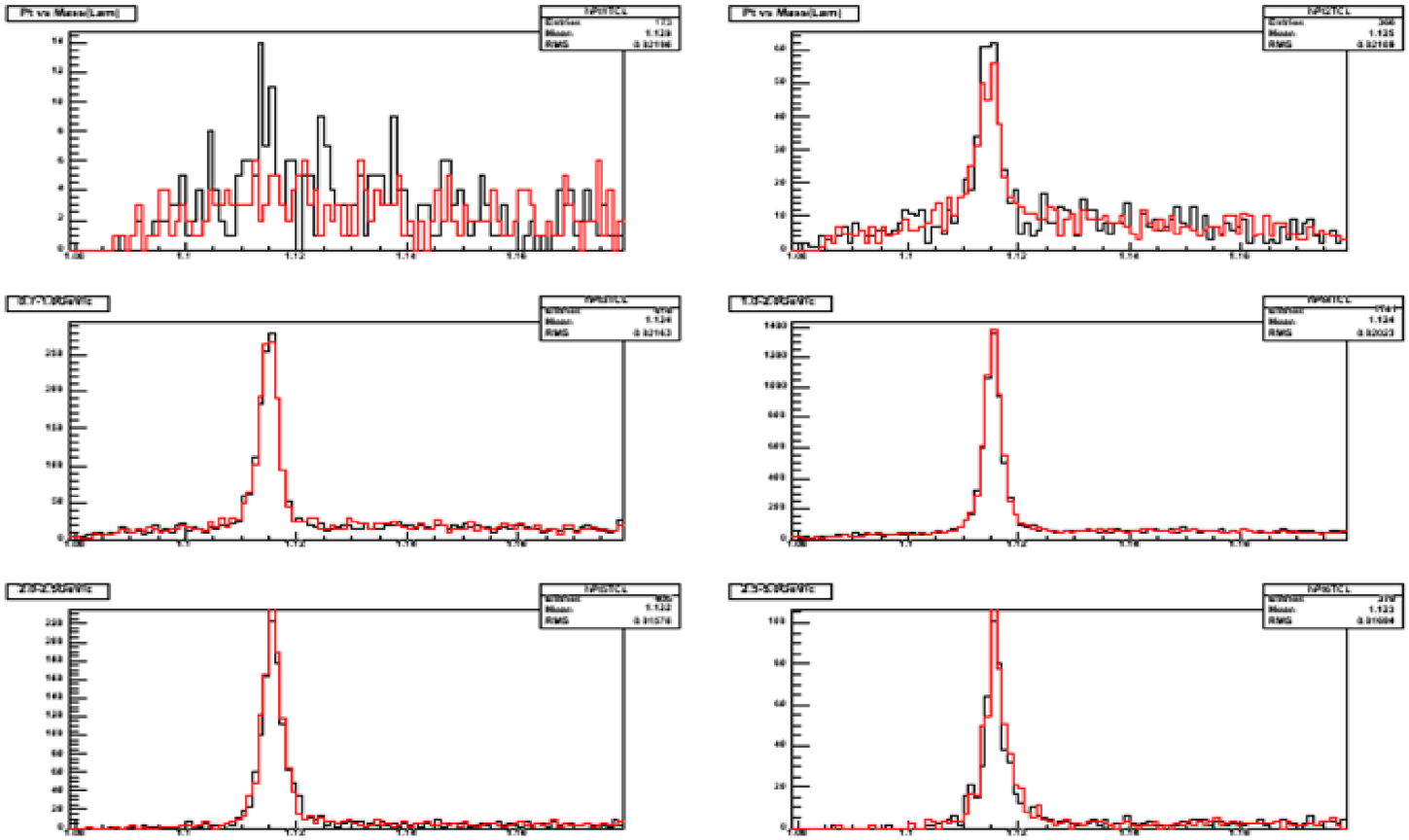


ratio of real pairs over mixed pairs; ideal tracking ratio = 1

Red, FCF
Black TCL



DAQ100 - Strangeness

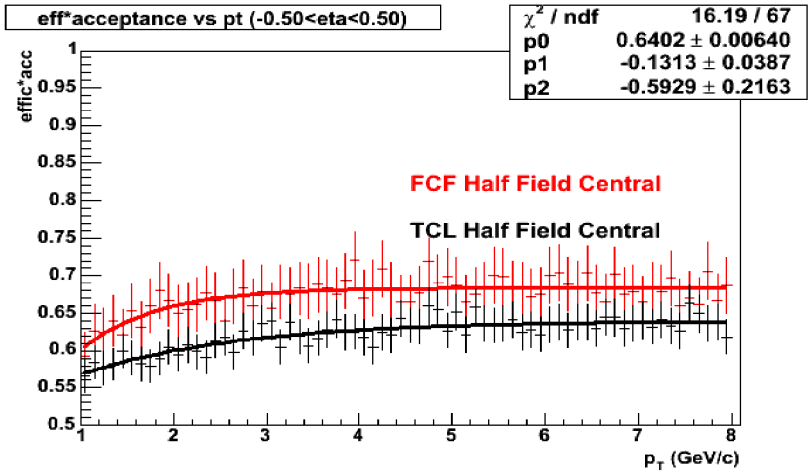


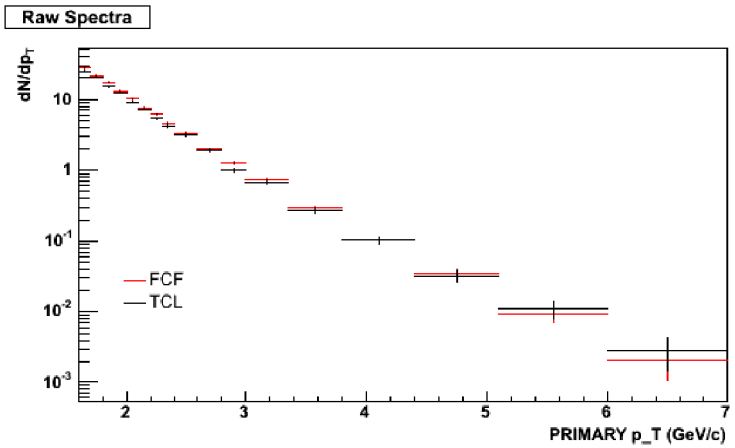
Lambda invariant mass, only lowest bin is doubtful

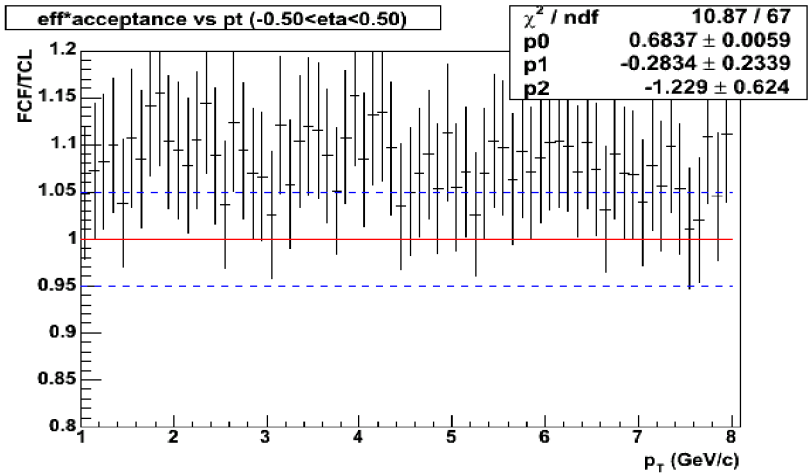
Jérôme Lauret, Collaboration Meeting, February 2004

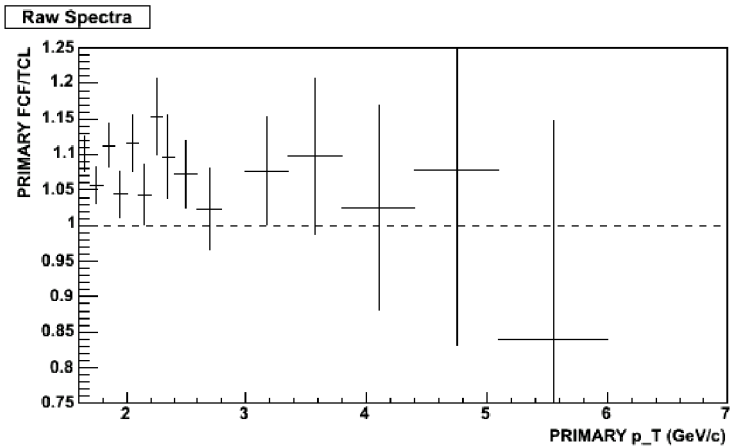
DAQ100 embedding – eff vs Pt

Spectra Comparison



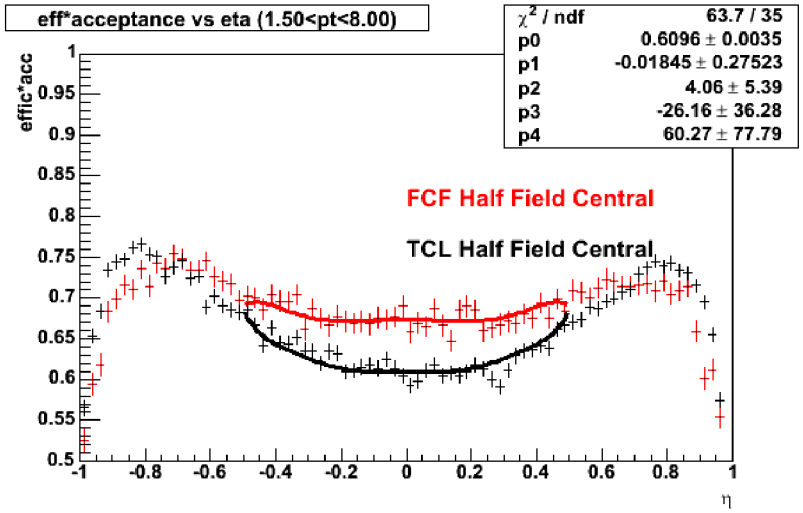


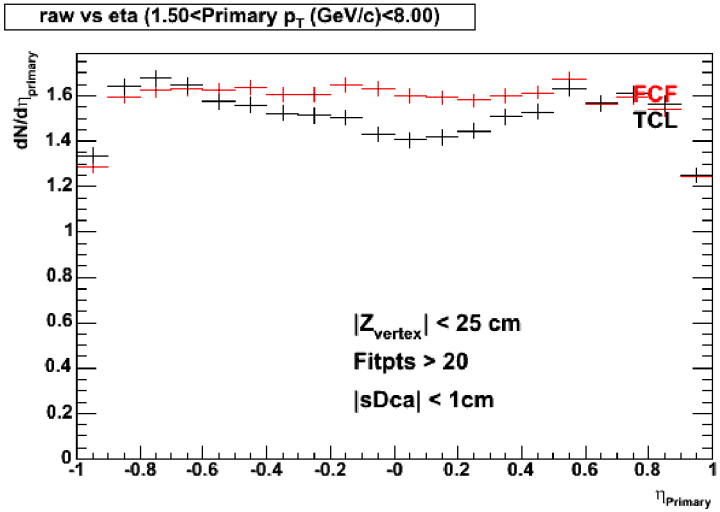


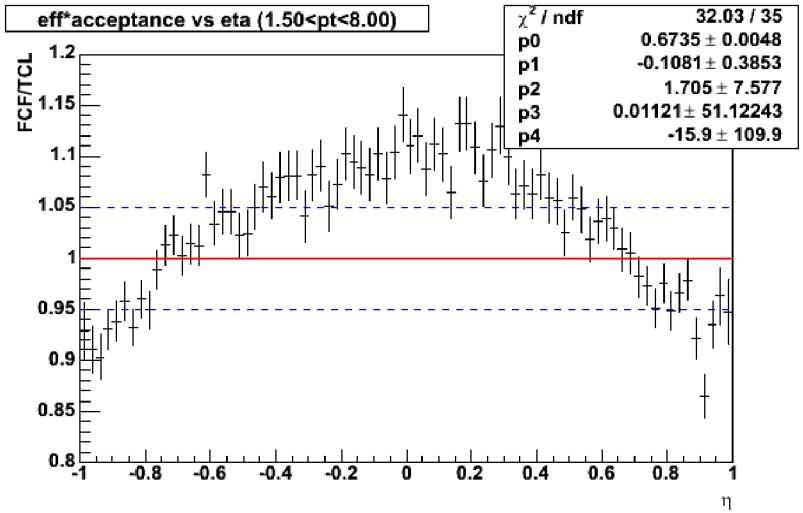


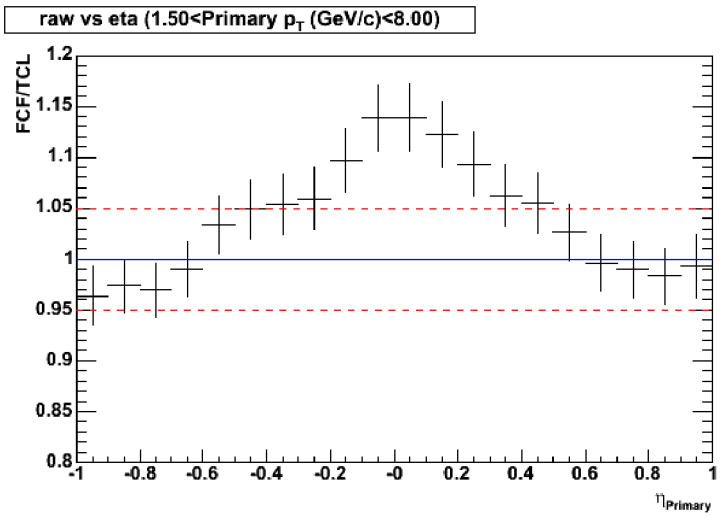
Jérôme Lauret, Collaboration Meeting, February 2004

DAQ100 embedding – eff vs eta

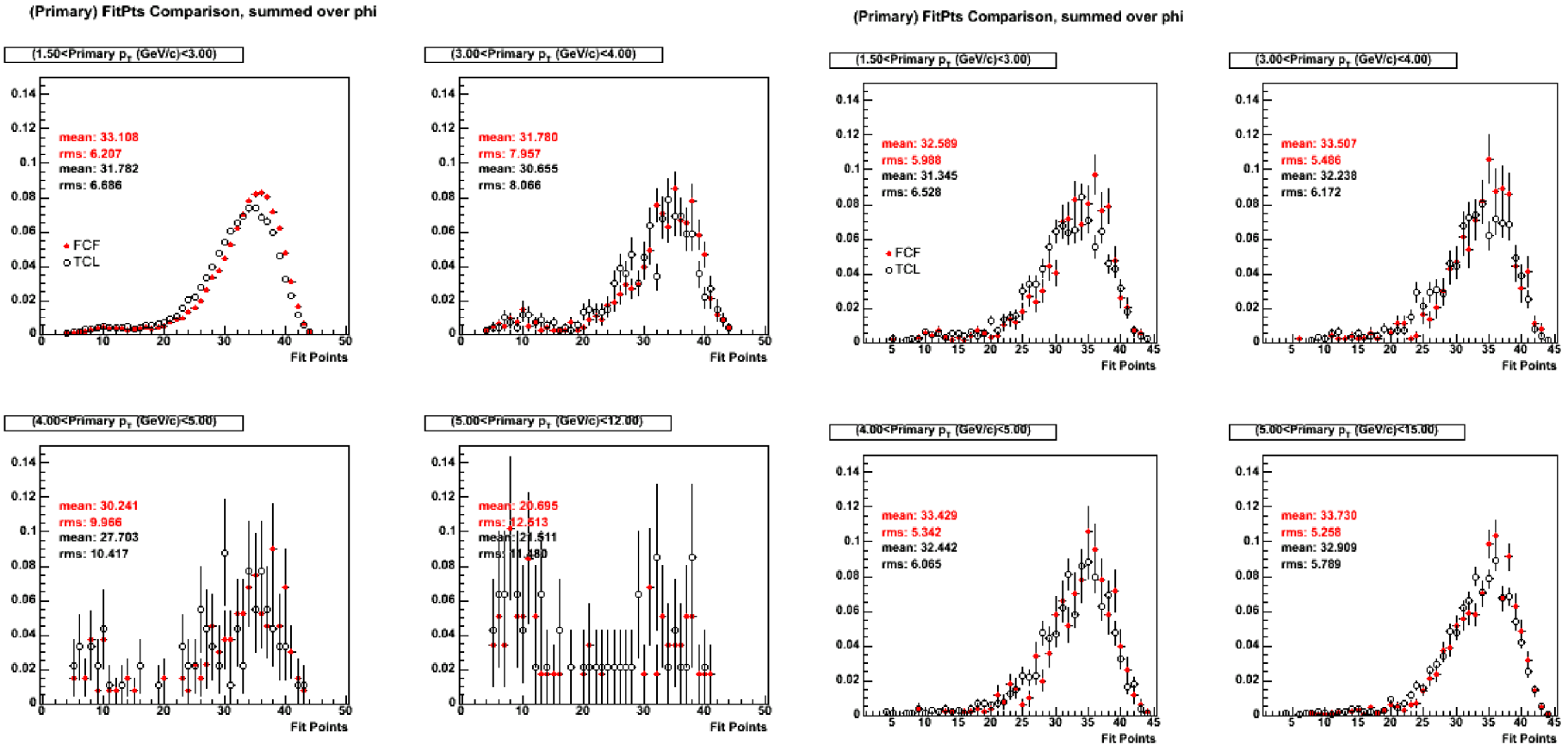








DAQ100 embedding – nfit points, central half field



Other – online and database

Database and online (Micheal DePhillips)

- RunLog browser was extended : new prescales in trigger Summary, added detector present list (adjustment needed)
- As presented at the last collaboration meeting, database scales for run conditions

`http://www.star.bnl.gov/STAR/comp/meet/CM200308/`

HOWEVER

- MySQL is NOT a professional database

  Shigeki's dogma : "cheap, fast/scales, reliable"

- Several cases recently where > 10 MB / events requested (broke)

In 2/3 instances, could simplify and reduce by factors > 100

Other – details later ...

- Run support: FastOffline is running relatively well
- A few infratructure problems : but gain is immediate: several sub-systems catching problems as they appear.
- Allows for ongoing calibration / alignment (SVT, EMC, FTPC, TPC to some extend)

- QA in Run4 (and beyond) (*Lanny Ray*)

- Reconstruction developments (*Yuri Fisyak*)
- (New) Calibration issues (Shorted TPC) (*Gene V. Buren*)
- Simulation developments (*Maxim Potekhin*)

Other - ITTF

Reminder (last analysis meeting)

- New coming data would at maximum be processed in Fast Offline old chain mode

Production WILL NOT occur before (and until) Sti is ready !!!

After the final integration week, in April, **we break away from old code.**

Un-movable dates

March 15th is the target date for the latest version of the library with the old code

In parallel, final integration WILL occur (TBA)

- Progress and status (*Manuel Calderon*)

lots of changes coming our way ...

Other – Future (½ - 1 year)

What we want to avoid

A cruise on that boat ...

But however

- Database is at the limit of usability for calibration
- STAR production is crumbling under heavy Physics program
- STAR has a longer term plan relying on detector developments and R&D
- ...

We need to get serious with several new Computing projects

Other – Future (will hear about it again)

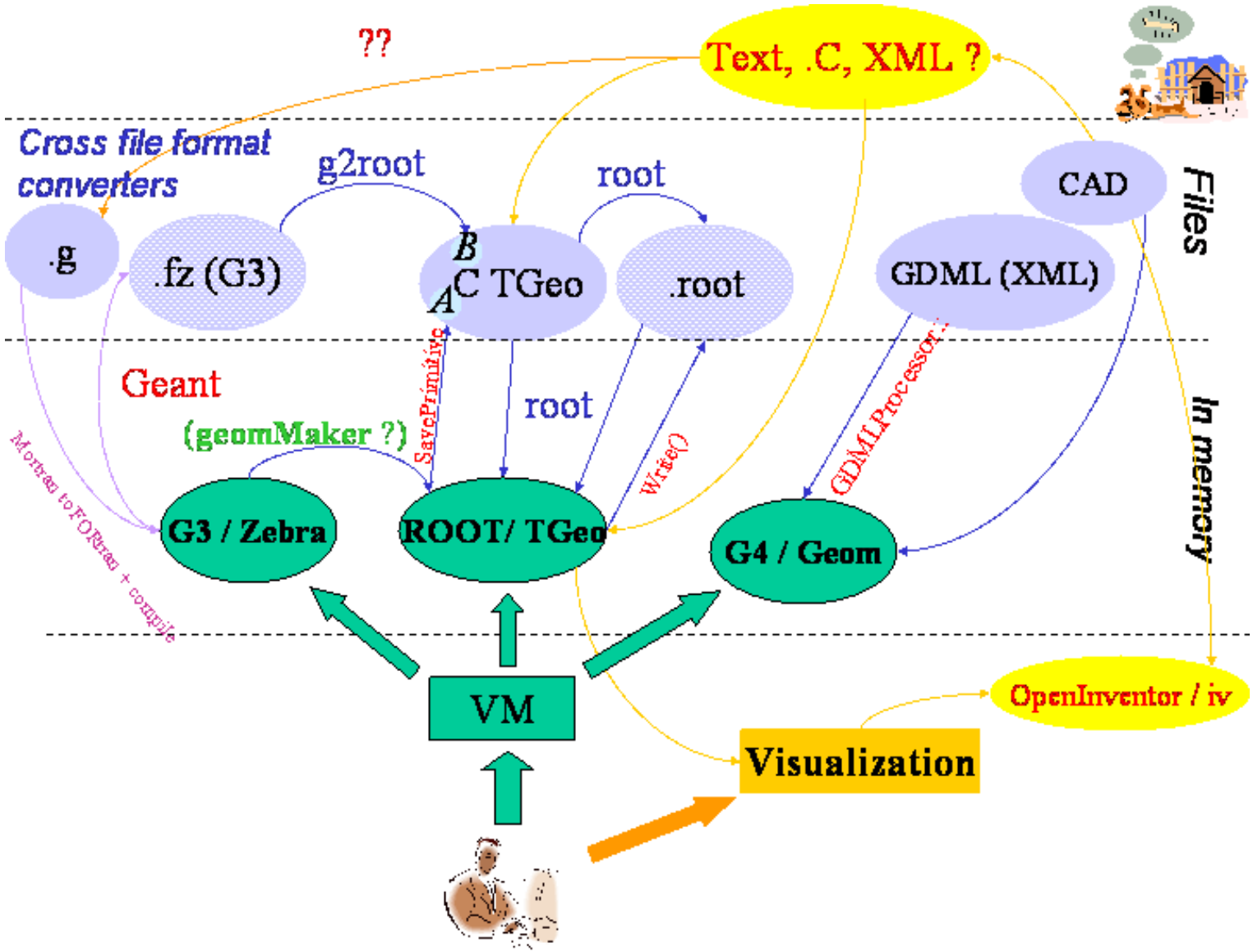
Infrastructure required components

- Changing database technology ?? Transparent / abstract interface to Db. **ODBC** ?? Data Access and Integration (**DAI**) ??
- Running on the Grid implies: reliable, traceable, consistent.**Log4j** !?
- **Integrated Catalog !! GridCollector !!** IO to revisit ??

Provide tools to developers (code longevity)

- Detector R&D, time frame & foggy Geant3 future DOES NOT allow to continue in the current direction (yet other abstract interfaces, from **Virtual Monte-Carlo** to detector description)

(*Yuri Fisyak / Maxim Potekhin*)


??
Text, .C, XML ?
Cross file format converters
g2root
root
CAD
Files
.g
.fz (G3)
B
C TGeo
A
.root
GDML (XML)
Geant
(geomMaker ?)
SavePrimitive
root
Write()
GDMLProcessor
In memory
Mortran to FORtran + compile
G3 / Zebra
ROOT/ TGeo
G4 / Geom
VM
OpenInventor / iv
Visualization

The future ...

Would be nice to be like this ...

Likely to be, from time to time, a bit like that ...

Now for the organizational changes ...

Computing Model change (the bitterness slide)

- C.f. talk at the Analysis Meeting *November 2003*
- Last collaboration meeting talk named "100 Million Events, what does this mean ??" *August 2003*
- ... and in 2002,
- "DAQ100 effects on Software & Computing" *November 2001*

(got 10 cents from John Harris)

Repeatedly asking for more money was a complete WASTE of time and efforts ...

That time has now come !!..

- **How do we analyze x5 (x10) more data with Moore's law expansion ??**
- **The question being : are we ready ??**

Jérôme Lauret, Collaboration Meeting, February 2004

Reminder (scale of the problem)

(*Jamie Dunlop*)

- We have 40 M minbias. Will have 50 M ?? 60 M ??
- We will have worth 30 M central ?? 40 M ??

Numbers do not need to be accurate to grasp the magnitude of the problem ... Below a 50/50 central/minbias, 75 M total estimate

Au+Au 200 (minbias)	35 M central	35 M minbias	Total
No DAQ100 (1 pass)	329 days	152 days	481 days
No DAQ100 (2 passes)	658 days	304 days	962 days
Assuming DAQ100 (1 pass)	246 days	115 days	361 days
Assuming DAQ100 (2 passes)	493 days	230 days	723 days

Just in case, that's ~ 1 year ...

Scale of the problem

Where are we now ??

- FY04 purchase allows for more resources (yesterday's news) – 361 days 1 pass becomes 280 days (beware, no expansion to user analysis farm within this figure / plan)
- ICC compiler, 30% gain in speed ?? ~ 200 days ...
- Fine, we can do 1 ½ passes in 1 year

Easy said, not easy go ... (1) untested
(2) can we afford +0 nodes to CAS ??
We may have a possible part-solution ...
Anything else ??

Large dataset consequences, solutions

Long range production have several organization consequences

- production team or more automation needed (Cataloging, Registration, error and disaster recovery, failure sorting)
- strong and TIMELY QA-ing necessary – Announcing the lack of signal XX or value YY after ½ way through a production will no longer be considered as funny (is actually NOT already)

In all cases, something needs to be done now while we (I?) can ..

Jérôme Lauret, Collaboration Meeting, February 2004

Large dataset consequences, solutions

Large experiment base their resource on an obscenely distributed model (we want to remain practical)
Grid technology helps "discovering" free resources

- Distributed local resource (disk space) is much cheaper
- We MUST (this is NOT a suggestion) move user analysis to a locally distributed model (no money for centralized storage in the first place and DOES NOT scale)
- We MUST try to run simulation on the Grid and gather experience later needed for Data Production on the Grid

Jérôme Lauret, Collaboration Meeting, February 2004

Large dataset & models (last word)

Ultimately, simulation, embedding, user analysis will need to migrate to the Grid or use Grid tools (Scheduler, GridCollector)

Other consequence : how do we do QA on the Grid ??
Maintain our code and libraries in sync across sites ??
package and distribute approved code ?? bundle analysis code ??

What is clear: BNL cannot remain the sole data center

- Old model – Tier0 center / Tier0 analysis
- Intermediate model – Tier0 center / Tier1 embedding (PDSF)
- STAR, the next generation : Tier0 / Tier1 / Tier2

Jérôme Lauret, Collaboration Meeting, February 2004