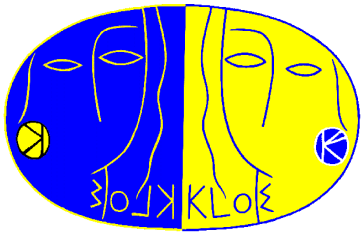
Offline Status Report



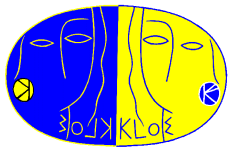
M. Moulson, 14 November 2003

Summary presentation for KLOE General Meeting

Outline:

- *Status of MC production*
- *Future MC production requests*
- *Computing resources*
- *Offline planning for 2004 run*

MC production status



Production completed to date (~450 pb^{-1} 2001-2002 data)

Program	Events (10^6)	Time (CPU days)	Size (TB)
all_phys	255	1100 (375 ms/ev)	6.9
ppgphok3	36	110 (264 ms/ev)	0.8
neu_kaon	410	1800 (375 ms/ev)	11.0

- Numerous upgrades to simulation
- Time-variable conditions
- Accidental activity from data
- MCDST's

Access to MC DST's

New DB2 view (logger.dtr_mcs_data) and KID protocol (dbmcdst)

To see available files using kls:

```
kls mcdst "run_nr between 19000 and 21000 and
dtr_stream_code='mrn' and
mc_mccard_code='all_phys'"
```

To access files from within ANALYSIS_CONTROL:

```
input url "dbmcdst:run_nr between 19000 and 21000
and dtr_stream_code='mrn' and
mc_mccard_code='all_phys'"
```

Further examples available at:

http://www.lnf.infn.it/kloe/private/mc/mcfaq.txt

MC production: known problems

Problems concerning generation:

- ϕ cross section constant for all_phys 2001
- a_0 produced 10× more frequently than natural
- $K_S K_L$ resonance shape for f_0/a_0
- $\omega\pi^0$ cross section "resonant" for all_phys 2002
- π and ν reversed in $K_S \to \pi e \nu$ generator
- bug in $\eta \to \pi^+\pi^-\gamma$ generator

Problems concerning reconstruction:

- order of hot/dead reversed in DCDELETE
- error in *s*-*t* relations used for DC reconstruction

Problems concerning DST's

- FILFO/par=1 used when streaming rad DST's

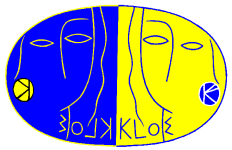
Known problems: a warning

If you are aware of any further problems with the MC production that affects your analysis:

- Please communicate it A.S.A.P.,
- in writing (by e-mail),
- to me and Caterina Bloise,
- if appropriate, with solution described

If I/we have not received a ***written communication,*** the problem is not “known” (and may not be fixed!)

MC: DC reconstruction issues



DCDELETE parameters mixed up in MC reconstruction!

Instead of: `INSERT DROP_DEAD TSKT DROP_HOT`

we had: `INSERT DROP_HOT TSKT DROP_DEAD`

DC trigger efficiency not correctly simulated

Reconstruction OK: both hot & dead dropped anyway

DCDELETE now simulates DC hardware efficiencies

Error in *s*-*t* relations used for DC reconstruction!

Realistic wire sags in MC → new *s*-*t* relations for reconstruction

s-*t* relations obtained from “calibration” using MC cosmic rays

Calibration job used DC stereo angles for data instead of MC

Large residuals, decreased software efficiencies

Highly advisable to re-reconstruct MC samples A.S.A.P.

MC: simulation of DC HW efficiencies

A. Antonelli, S. Dell'Agnello, M. Moulson

HW effficiencies in data < 99% on 1st 5 layers

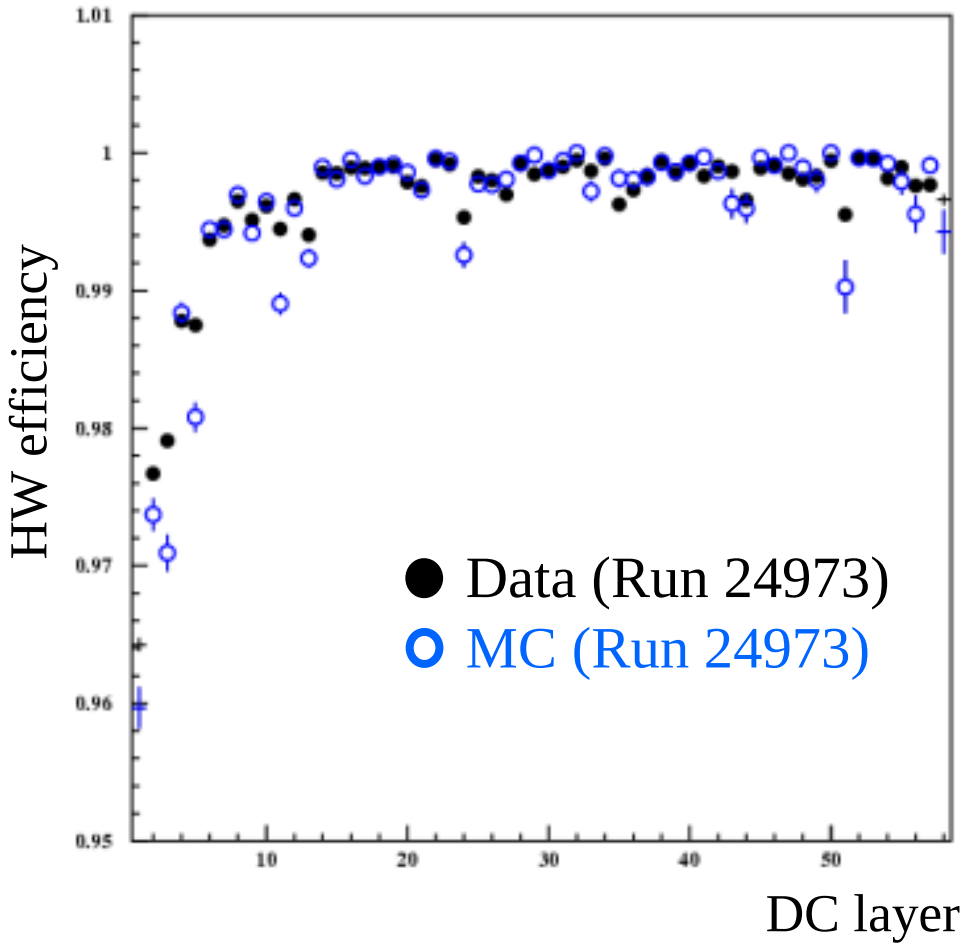
Important for analyses that:

- cut on position of 1st hit:
  $K_S \to \pi^+\pi^-/\pi^0\pi^0$
- use tracks concentrated near origin
  $K^\pm$ analyses

Now simulated by DCDELETE using physmon efficiencies for each run

How well are SW efficiencies reproduced?

HW efficiency vs. DC layer

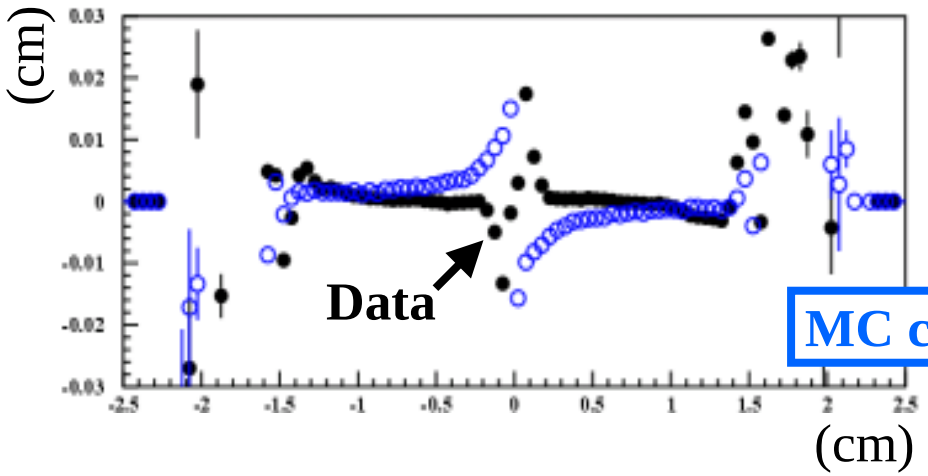


MC: s-t relations for reconstruction

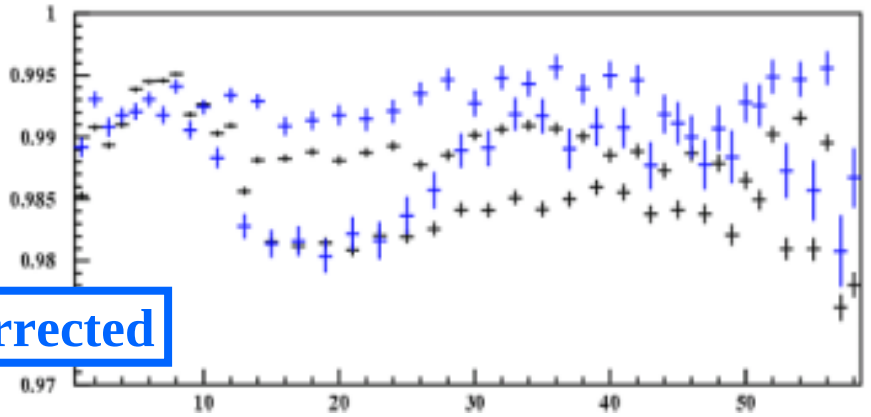
A. Antonelli, S. Dell'Agnello

physmon diagnostics: Run 24793 + simulation, large cells

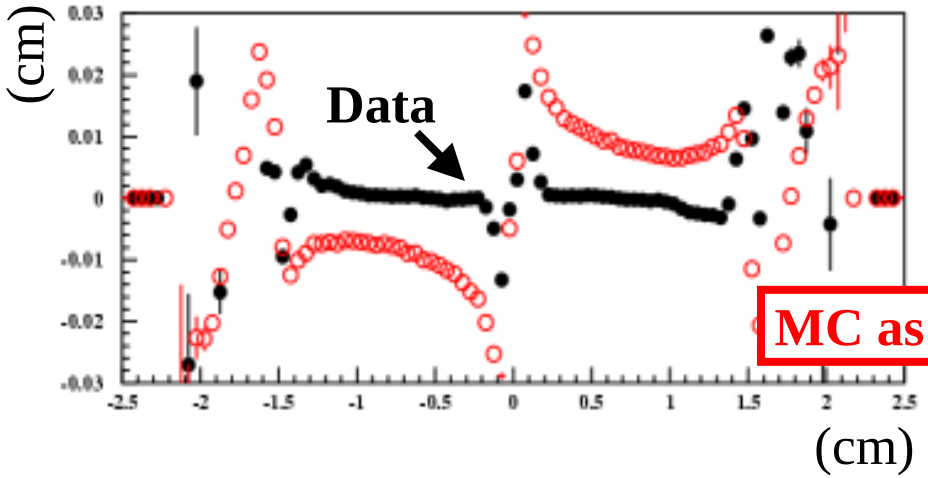
Mean residual vs. impact parameter

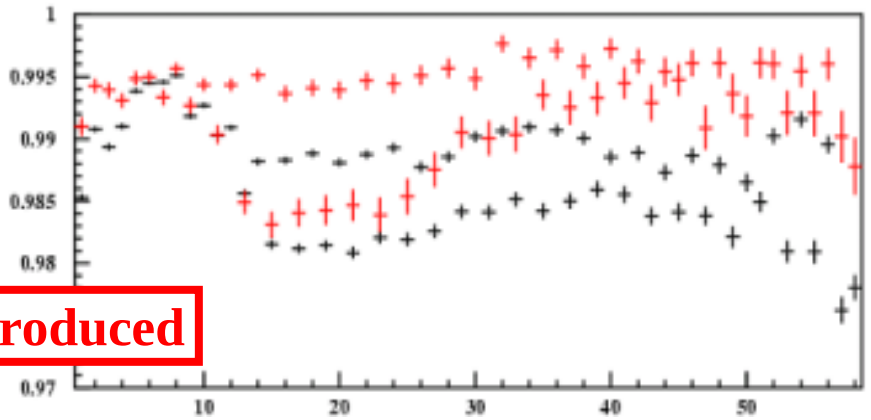


Software efficiency vs. layer number



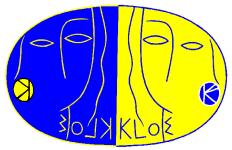
MC corrected





MC as produced

MC: tuning of EmC response to π and μ



C. Bloise, T. Spadaro, M. Testa

Problem: π, μ deposit 20% more energy in MC than in data

Observed *e.g.*, using E/p distribution for π from $\phi \rightarrow \pi^+\pi^-\pi^0$ events

Solution: Tweak scintillator response in MC

- π, μ energy loss from restricted Landau distribution + δ-ray generation (E_{thresh} = 10 keV)
- δ-rays not explicitly simulated ⇒ do not contribute to signal

  Assumed to be lost to absorber

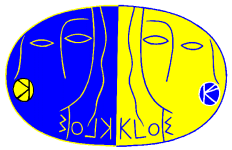
Gives good MC-data agreement for peak of *E*/*p* distribution

Excess as *E*/*p* → 0 for MC relative to data

π clusters in MC have more fragments than in data? Under study…

Expect these modifications to be available in 1-2 weeks

Further MC requests



Minimum request for MC production via new mechanism is ~30M events (or $\sigma \times$ LSF = 60 nb)

1. **All radiative ϕ decays plus $\phi \to \rho^0\pi^0$; $\rho^0 \to \pi^0\gamma$ and $e^+e^- \to \omega\pi^0$**
   450 pb^{-1} @ LSF = 5, σ = 49 nb ⇒ 110M evts
   Makes fixing radiative MCDST's less of a priority
   Satisfies: Radiative group

2. **Radiative bhabha events: $E_\gamma > 100$ MeV, $20° < \theta < 160°$**
   450 pb^{-1} @ LSF = 0.1, σ = 720 nb ⇒ 32M evts
   Satisfies: Gatti, Spadaro (tracking efficiency)

3. **All ϕ decays at high stat. for runs off peak:**
   a) $\sqrt{s}$ = 1017 MeV: 7.8 pb^{-1} @ LSF = 2, σ = 1305 nb ⇒ 20M evts
   b) $\sqrt{s}$ = 1022 MeV: 6.1 pb^{-1} @ LSF = 2, σ = 1488 nb ⇒ 19M evts
   c) $\sqrt{s}$ = 1019 MeV: 5.8 pb^{-1} @ LSF = 1, σ = 3055 nb ⇒ 18M evts
   Satisfies: Bini (15% $\rho\pi$ for Dalitz plot), ϕ cross section analysts

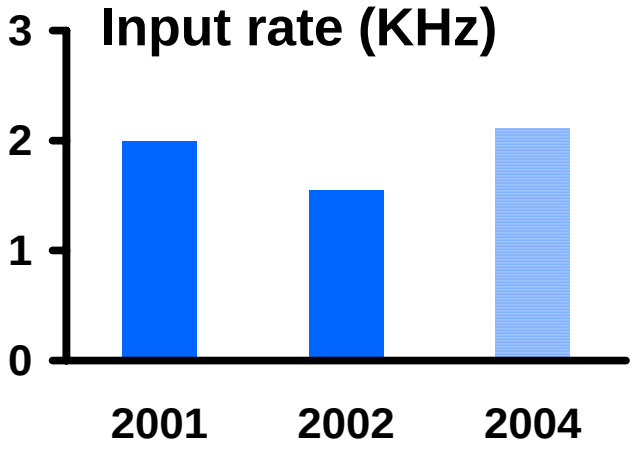
Each program ~7 days elapsed

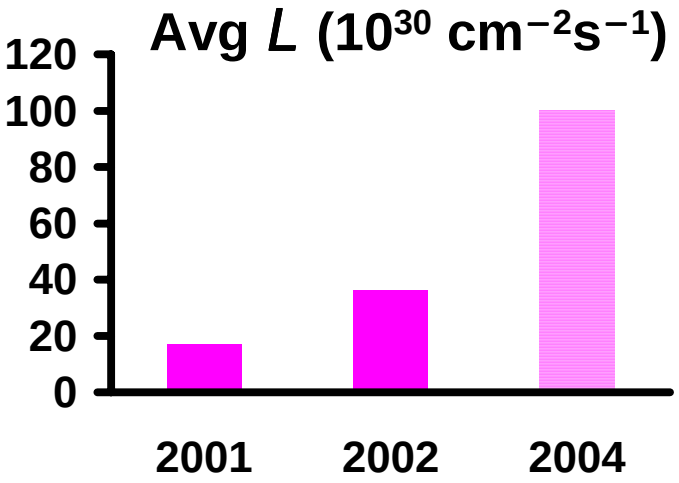
Reprise: Next steps for MC production

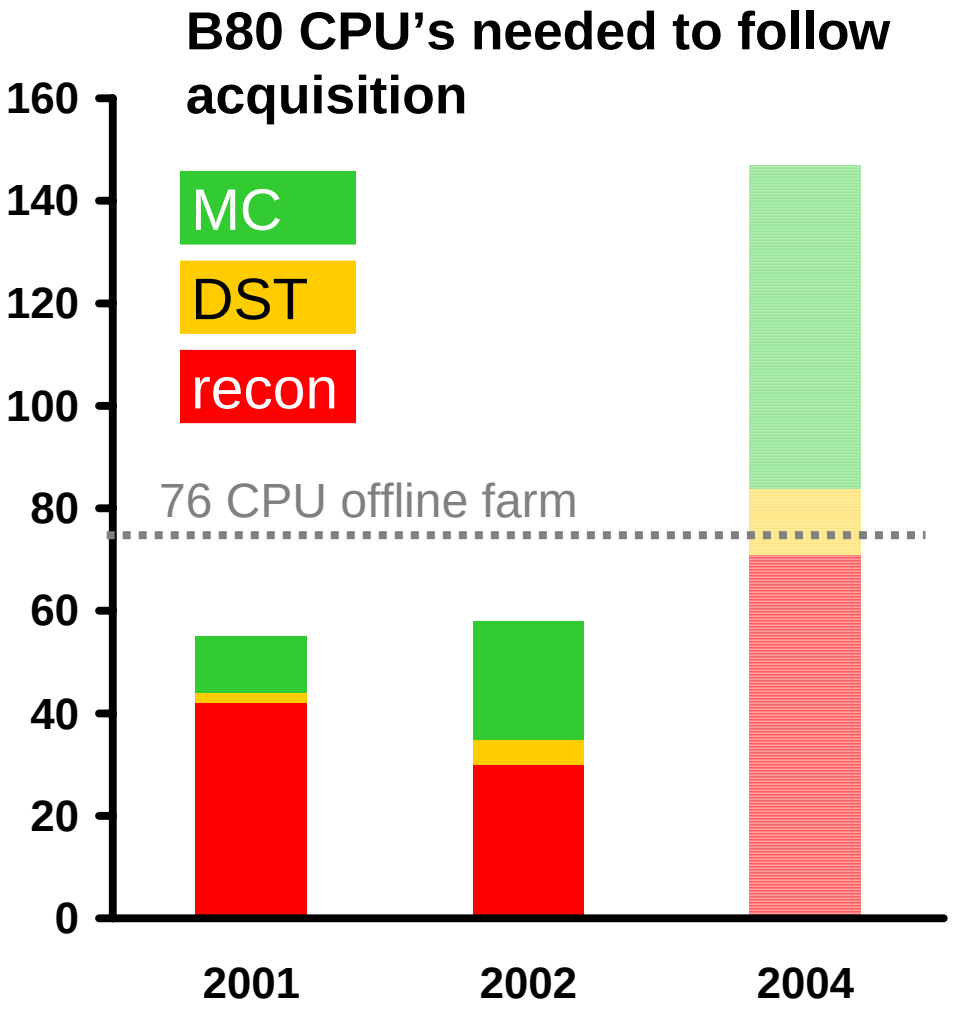
In rough order of suggested priority: to be discussed
Execution times are indicative only

1. Re-reconstruct 5-10 pb^{-1} (LSF = 1:5) of all_phys for tests **1 day**
2. Re-reconstruct all MC samples
   a) all_phys **9 days**
   b) ppgphok3 (use PHOKHARA-3b and redo generation?) **7 days**
   c) neu_kaon **14 days**
3. New MC production: radiative ϕ decays **7 days**
4. New MC production: $e^+e^-\gamma$, E_γ > 100 MeV **7 days**
5. New MC production: all_phys off peak **7 days**
6. Redo all_phys **20 days**
7. Redo neu_kaon **30 days**

CPU power requirements for 2004







Mass storage requirements

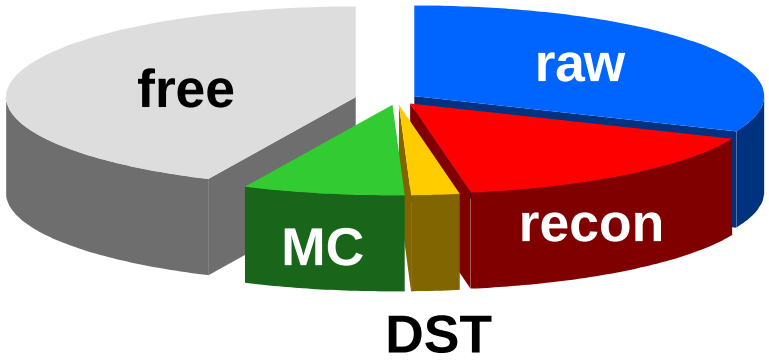
Installed hardware:

5.7 TB recall disk cache (DST)

IBM 3494 tape library

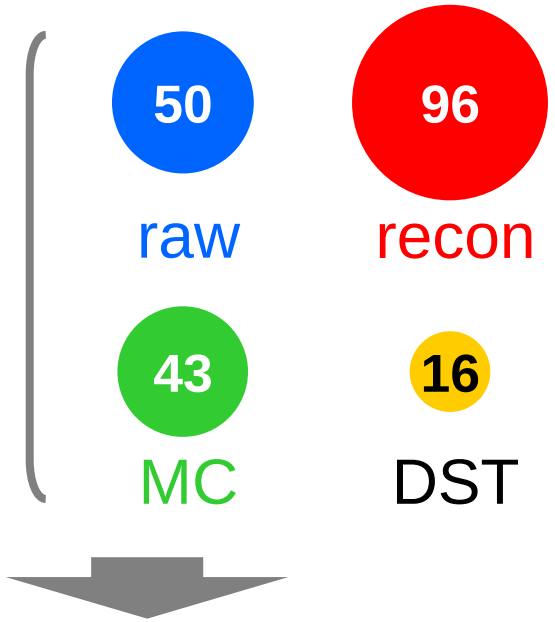
- 12 Magstar 3590 drives (14 MB/s)
- 60 GB/cartridge
- 5200 cassettes, 2 accessors
- Tivoli Storage Manager

Max. capacity: 312 TB
In use: 185 TB



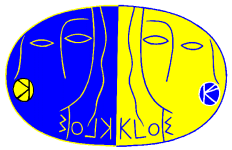
Predicted needs for 2004:

2 fb^{-1} at L = 1×10^{32} cm^{-2}s^{-1}



410 TB tape storage
16 TB DST disk cache (50%)

Upgrades for 2004



Additional servers for offline farm: ~80 B80 equivalents

10 IBM p630 servers: 4 × 1.45 GHz POWER4+

Additional disk space: ~20 TB for DST cache and AFS cell

Status: Approved by Direttivo, not yet ordered
Delivery 30 days after order goes out

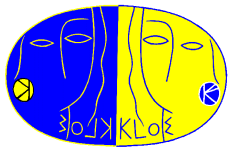
Additional IBM 3494 tape library: 300 TB

Magstar 3592 drives: 300 GB/cartridge, 40 MB/s
Initially 1000 cartridges with space for 3600 (1080 TB)
2 accessors, 6 drives, remotely accessed via FC/SAN interface

Status: Gara not yet open
6 months or more to delivery (personal estimate)

What if new library unavailable for part or all of 2004?

Status of reconstruction executable



How true is the claim that all DBV > 12 good for analysis?
If we re-reconstruct data now, what changes do we pick up?

DBV-13

- Fixed coordinates of PCA to origin in DTFS banks
  Fixed in DST's for DBV-12, does not affect streaming
- Magnetic field rescaling
  Magnet current stable to within < 0.1 A

DBV-14

- Bug fix for acollinearity cut in FILFO/VLAB
  0.1% effect on number of LAB/VLAB evts
- Bug fix for $K_S \rightarrow \pi^+\pi^-\pi^0$ tag
  Affects only one analysis

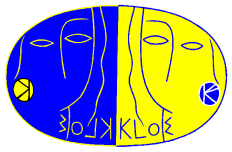
DBV-15

- Charged kaon retracking installed
  Affects only DST's

DBV-16

- Bug in *dE/dx* integration fixed in VTXFIN
  Appears to be most significant change

Status of reconstruction executable



Disclaimers:

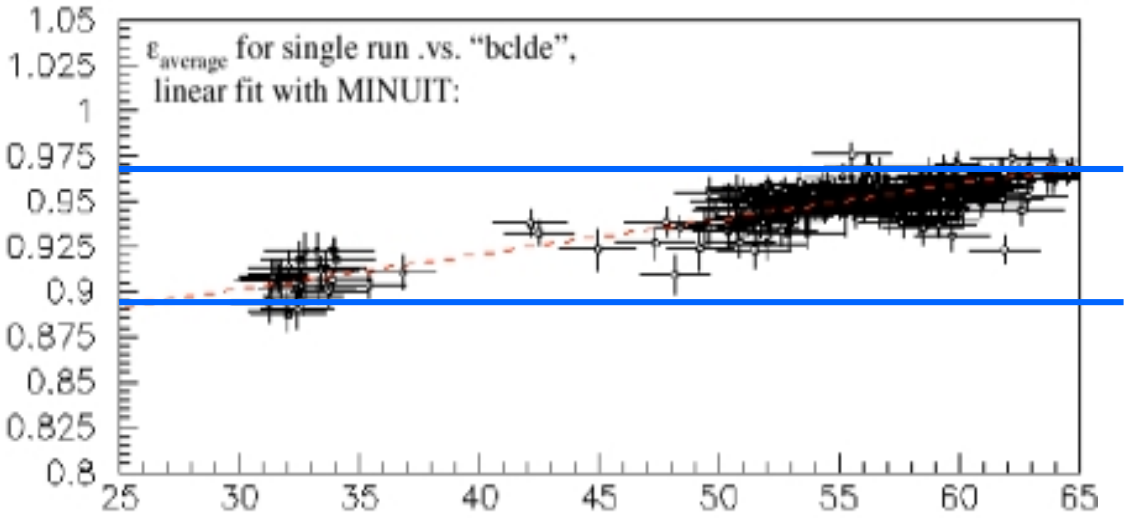
- This list assembled quickly, should be done more carefully
- Many small bug fixes (arrays out of bounds, etc.)
  Shouldn't have an effect on data quality, but who knows?
- Many changes that concern only MC reconstruction
  First good MC version is DBV-17

Conclusions:

- Most analyses can mix DBV-12 to DBV-15 with impunity
- Probably could mix some data reconstructed with current datarec version (DBV-17) as well
  Have to look at *dE*/*dx* in DBV-16
  DBV-18 will be significantly different (wire sags)

FILFO needs an overhaul

1. FILFO inefficiency is non-negligible and highly variable



Background cluster energy, MeV

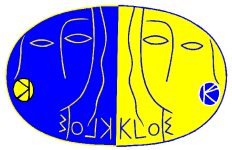
S. Müller
24-Jun-03 $\pi\pi\gamma$ group

6 % variation with background conditions

Reasons for this are understood, but fact remains…

2. FILFO cuts were tuned to eliminate as much background as possible from 1999 and 2000 data and have not seriously been looked at since beginning of 2001

FILFO is complicated!



S. Müller

Most $\pi\pi\gamma$ events are rejected by the BHABREJ-condition within FILFO:

BHABREJ = $N_{cl} \leq 7$ .and. (MIN .and. ENE .or. ANG)

This gives a correlation with accidental clusters

Not relevant for $\pi\pi\gamma$ analysis

MIN = $N_{cl} = 0$ .or. $N_{cl} = 1$ .and. $E_{cl} \leq 50$ MeV .or.

$N_{cl} > 1$ .and. < 4 clusters with $45° < \theta_{cl} < 135°$

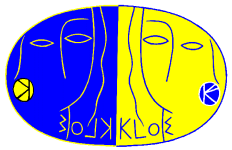
ENE = $N_{cl} < 7$ .and. { $N_{cl} < 2$ .or. [Δt(last – first cluster) $\leq$ 6-10 ns .and. EMAX $\geq$ 300-400 MeV] }

EMAX: Cluster with highest energy and Δt(cluster – first cluster) $\leq$ 3.5 ns and $60° < \theta < 120°$

FILFO ideas

1. **Get rid of it! Alas…**
2. **New downscale policy**
   - Like **afl** stream but downscaled fraction normally streamed
   - Convenience: vetoed events reconstructed, classified, and present in DST's
   - Caveat: these events must be removed from analysis sample, or at least handled differently from standpoint of efficiencies
   - Pilot study indicates 1/5 downscale would increase CPU 10% and require FILFO to be enforced in **bha** stream
3. **Initial work on FILFO: identify & eliminate most unstable cuts**
   - Can we get reasonable rejection and high efficiency by eliminating a few cuts?
4. **More comprehensive work on FILFO?**

Dropping BHABREJ from FILFO



S. Müller

Decreases FILFO inefficiency

2001 $\pi^+\pi^-\gamma$ analysis: **$1 - \varepsilon_{\mathrm{FILFO}} = 5\%$, decreased to < 1%**

Other channels not yet studied, effect expected to be beneficial

Increase in reconstructed volume

Up to 20% for **rad** stream (smallest datarec stream)

No effect on **bha** stream (largest datarec stream)

Overall increase in reconstructed volume

~10% for **early-2001** runs

~5% for **late-2001 and 2002** runs

Overall increase in CPU-time

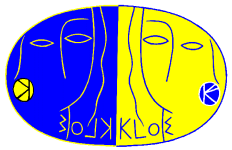
19% for **early-2001** runs

10-15% for **late-2001 and 2002** runs

Significant efficiency gain

Increases in CPU and volume probably tolerable

Reconstruction without FILFO



Request from $\pi\pi\gamma$ group: Reprocess 10% of 2002 data without FILFO

Useful to everyone for FILFO efficiency studies

Would allow more comprehensive studies of FILFO itself

Complications: Non-negligible effort/CPU consumption
Discard **bha** stream?
Define new streams for output files?

Alternative: Implement FILFO downscale and reconstruct (some) data

pros

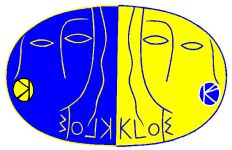
- More convenient for analysts
- Better coverage of run space
- Reprocessing all 2001-2002 data with streaming modifications frees ~30 TB of library space

cons

- Requires suspension of further work on reconstruction (or inclusion in MC reconstruction)
- Need to reconstruct quite a lot of the data set

We will produce a FILFO-study sample, but need to discuss details

Bhabha streaming proposal



Problem: bha files are:

- 60% of reconstructed output volume
- 30% of data written to tape in 2004
- 15% of space in tape library (30 TB)

} **Very large tape volume for low-interest sample**

Proposal:

Decrease E_{tot} cut: $E_{tot} > 250$ MeV (from 300 MeV)

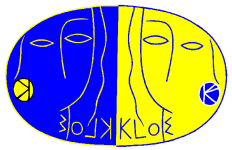
Downscale golden Bhabhas 1:5 *at least*

Create new tag for radiative Bhabhas for efficiency studies:

- 3 or more clusters with $\cos\theta_{ij} < 0.78$
- $-2 < t - R/c < 10$ ns
- $E_{cl} > 30$ MeV

Selects about 10% of current Bhabhas

Luminosity-scaling background trigger



LSB trigger automatically provides INSERT events for MC work:

- No time-intensive bggmaker pass
- No worries about background-dependent **bgg** selection efficiencies
- No messy attempts to isolate $\gamma\gamma$ clusters from background clusters

Technically feasible and tested:

- Combine normal trigger and TORTA pulser in freerun mode
- Implement the luminosity scaling by downscaling in software

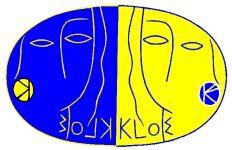
Example scheme:

- Currently collect **bgg** events with $\sigma \sim 40$ nb:

  Suppose we want **lsb** $\sigma = 100$ nb, assume $L = 2 \times 10^{32}$ cm^{-2} s^{-1}

  Need to collect **20 Hz** of truly uncorrelated triggers
- A/C module counts *e.g.* classified VLAB's from ECLS and streams next background event into **lsb** file (w/o reconstruction) every ~4 VLAB's
- Rest of uncorrelated events are discarded
- No uncorrelated events in streams

T3 policy and downscaled sample

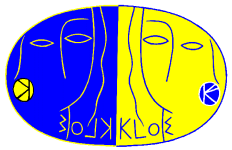


Foreword: T3 in 2002 data

- Trigger passes all events, only records cosmic-veto (CV) decision
- CV subject to T3, which decides to keep a certain fraction
- T3 decision determined by both EmC (fast clustering, cosmic-ray TOF) and DC (activity near IR based on clustering of DC hits)
- Overall CV inefficiency much reduced
- Not possible to estimate using single-track methods as in 2001 (EmC + DC decision)
- $1/(S = 64)$ of CV events retained by T3 for efficiency studies
- Must directly evaluate CV inefficiency using this sample

2004 data: Tune downscale factor for precision analyses

T3 downscale tuning



Precision on $(1-\varepsilon_{CV})$ at scaledown *S*:

$$1-\varepsilon = \frac{SN_{CV}}{N_{sig}} \qquad \delta\varepsilon = \sqrt{\frac{S\varepsilon(1-\varepsilon)}{N_{sig}}}$$

For 90% CL with $N_{CV} = 0$ use 2.3 cts.

?

Added rate = 1800/*S* Hz

Compare to total rate: 2100 Hz

***S* = 64 add 30 Hz: negligible**

***S* = 8 add 240 Hz = 10% of data**

Example 1: Worst case: $1-\varepsilon$ with T3 = 3.3%

$K_S \to \pi^0\pi^0$ **With *S* = 64, need 120 pb^{-1} for 0.1% error** in uncorrelated sample

K_L **crash** NB: Target for 2002 analysis is 0.1% *overall* systematic error

Example 2: $K_S \to \pi^+\pi^-$; $K_L \to 3\pi^0$: $N_{CV} = 0$, $N_{sig} = 6\times10^5$, $S = 64$: **$1-\varepsilon < 0.02\%$**

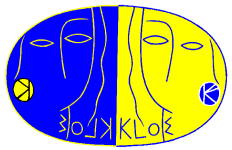
$K_S \to \pi^+\pi^-$; $K_L \to \gamma\gamma$: $N_{CV} = 0$, $N_{sig} = 1.8\times10^4$, $S = 64$: **$1-\varepsilon < 0.8\%$**

Example 3: $\sigma_{eff} = 100$ pb, $L = 2$ fb^{-1}, $N_{sig} = 2\times10^5$

$K_S \to \pi^+\pi^-$ Do not have ε: how efficient do we have to be to know it to 0.05%?

$K_L \to \pi^0\pi^0$ **With *S* = 64: $\delta\varepsilon = 0.05\%$, $1-\varepsilon$ must be less than 0.08%**

Other reconstruction issues for 2004



_dE/dx_ and DCDEDX A/C module:

- Include DCDEDX in datarec path before streaming
- New bank format for DST's

kpm stream:

- Eliminate some old streaming algorithms
- New streaming algorithm using cuts on *p* and *dE*/*dx* to classify on basis of a single track

ksl stream:

- Definitively eliminate KSTAG algorithm

rad stream:

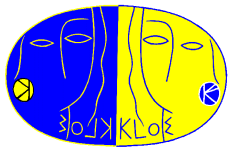
- Include downscaled sample with less stringent EVCL cuts in stream

ATFMOD:

- Include wire sags in track reconstruction

Any other issues?

Conclusions



First pass at large-scale MC production finished: ~1 year of effort!

- 700M evts produced, with upgrades, background, MC DST's
- Continuing effort: new upgrades ready to go
- Minor problems discovered: growing pains
- Lots of new production requests to fulfill

Tape storage for 2004 critical, will lead to painful choices

- CPU situation for 2004 looks good

Work planned for 2004:

- Modifications to FILFO, Bhabha streaming, physics streaming
- T3 scaledown to revisit; new background trigger (**bgg → lsb**)
- Other work on reconstruction: wire sags, dE/dx, etc.

Conclusions → Discussion

Much activity foreseen before data-taking restarts!

Important to discuss priorities:

- **Reconstruction w/o FILFO vs. MC production work**
- **Relative priorities of MC production tasks**
- **Re-reconstruct or re-generate MC samples?**
- **Reconstruct sample w/o FILFO or reprocess (some) data?**