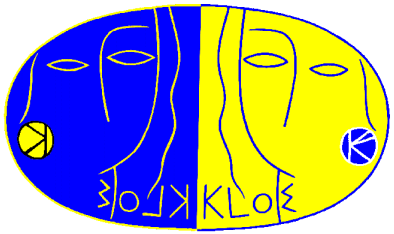
Large scale MC production

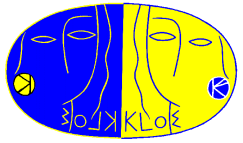


M. Moulson
Offline Discussion, 21 October 2002

Outline

- ***What to produce?***
- ***How to produce?***
- ***Proposal for MC DST's***
- ***Timescale and tasking***

What to produce?



My original proposal:

500 pb^{-1} of $K_S \rightarrow$ all, $K_L \rightarrow$ all $\approx$ 500M events

2—3 months if done efficiently on AIX farm

Useful for background studies for all neutral kaon analyses

Universally approved (?)

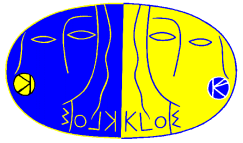
Only counter-proposal to require K_L decay before EmC

Should we consider being more ambitious?

$\phi \rightarrow$ all: 500 pb^{-1} $\approx$ 1.5 M events, 6—9 months

See how it goes, maybe try K^+K^- next, if there is demand

Main issues for production



To what extent do these issues need to be resolved before we start generating?

- Upgrades to geanfi/datarec
- Production of MC DST's
- Insertion of machine background
- Allocation of computing resources

Upgrades to MC

Are there upgrades planned that must be finished before production starts?

- Initial state radiation
- Wire sag in DC

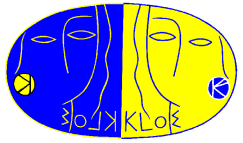
  Different on first ~10 layers. Easy to implement.

  Probably best not to include in reconstruction for consistency with 2001-2002 data
- Thickness of DC wall

  Missing 100 um of Al, 50 um too much CF (to be checked)

  Should be easy to correct
- ***Other issues?***

DST's for MC events: a proposal



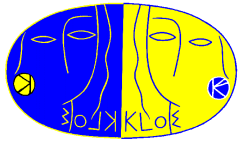
- Produced from **mcr** output
- Reconstruction bankset exactly identical to data DST's
- MC-truth bankset highly compressed

  All variables needed to fill PROD2NTU structures precalculated and stored

  Banks eliminated:

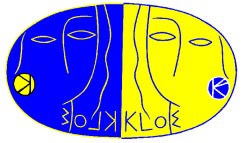
  - MC hit banks (**DHIT, CHIT**)
  - MC raw-bank equivalents (**MCEL, MDTC**)
  - MC link banks (**CEKA, DTHA**)

Bank list and data structures



Banks affected	Action	Code affected
LRID, EVCL	Retained	
PART, MATE, TMED	Dropped	
HEAD, KINE, VERT	Retained	
MDTC, DHIT, **DTHA, DTKA**	**MDHS:** hit summary **MTRK:** p at FH/LH, length, etc.	getdhitval tfmchit
BRIN, RUNG	Retained	
TCOE, T0MC	Retained	
MCEL, CHIT, CEKA	**MCLU:** E, t, x of clusters **MCHI:** only KINE index for hit cells	getclustru getcspsstru
CFHI	Retained	
QHIT, QIHI, *QCAE, QCKA*	Need proposal	???

Background insertion: status



Desirable to introduce background for large scale production

Probably will not be complete on timescale for starting generation but could become ready shortly afterwards

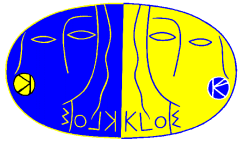
Begin generation/first reconstruction anyway?

Status: Code is under development

Background selection criteria:	*Miscetti*
Insertion module:	*Moulson, Patera*

Haven’t seriously begun to address logistics yet.

Background insertion: plans



Global sampling proposal:

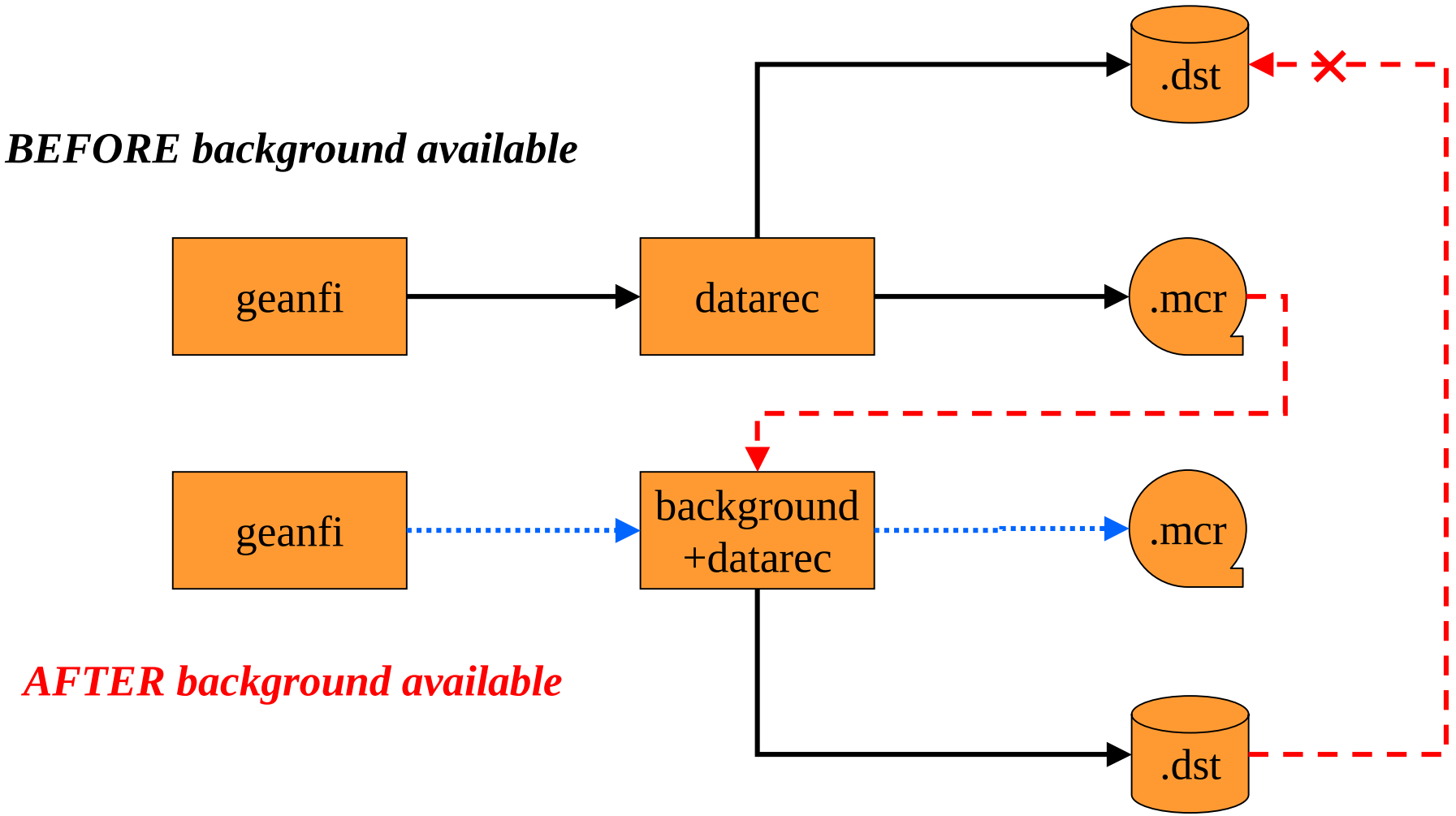
- Harvest background from all 2001-2002 data:
  Recognized $\gamma\gamma$ events have $\sigma \approx 120$ nb^{-1}
  ~50M $\gamma\gamma$ events in 450 pb^{-1}, ~100 GB volume
- Insert each background event in ~10 MC events
- If $\gamma\gamma$ selection efficiency not significantly dependent on background, obtain background profile that matches data

MC sample then refers to entire data set, not easy to subdivide

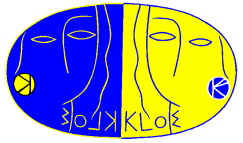
Can also subdivide into run groups. How big should they be?

- Larger means better averaging, simpler organization
- 2001, 2002 defined by separate MC run ranges (in any case)
- Groups should be of order ~10 pb^{-1} to avoid problems matching size of MC (35 nb^{-1}) and data (100—200 nb^{-1}) files

Possible scheme for production



Allocation of computing resources



Current CPU allocation:

AIX (92 CPU's):

Users:	**16**	(fibm11—fibm14)
DST's:	**8**	(fibm30, fibm32)
Reconstruction:	**68**	(all others except fibm16)

SunOS (40 CPU's, ~16 AIX equivalent)

Users:	**8**	(fsun01, fsun02)
MC production:	**32**	

Big tasks during coming months: MC production, **kpm** DST's

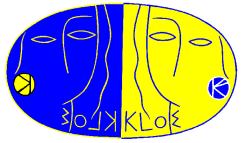
Need ~54 IBM CPU's to complete MC production in 60 days

Assumes 1M evt/day/(40 Sun CPU's), 1 IBM CPU = 2.5 Sun CPU's

Very desirable to produce **kpm** DST's on IBM nodes:

Also desirable to produce MC events on IBM nodes?

Task list



MC upgrades:	
ISR	
DC geometry	
MC DST's	
Bank-reduction code	Moulson
Database modifications	
Background insertion	
Event selection	Miscetti
Coding	Moulson, Patera, Miscetti
Testing	Others
Resource allocation	

Other issues? Volunteers?