Mock Data Challenge Production

Robert Hatcher
Fermilab Computing Division

Week-In-The-Woods Mtg. June 2004

Outline

- Who
- What
- Where
- How
- When
- Why

Who: a collaborative effort

- Robert Hatcher: GMINOS, config scripts, FNAL generation
- Mark Messier: GNUMI flux, oscillation code
- Howie Rubin: reco batch farm management
- Alex Sousa: production script, Tufts generation
- Nathaniel Tagg: PhotonTransport/DetSim
- Jon Urheim: IU production, coordination
- special thanks:
  - Anatael Cabrera: for bugging Nathaniel about oddities
  - Mike Kordosky: GCALOR testing
  - Niki Saoulidou: CPU cycles

What: Data Sets

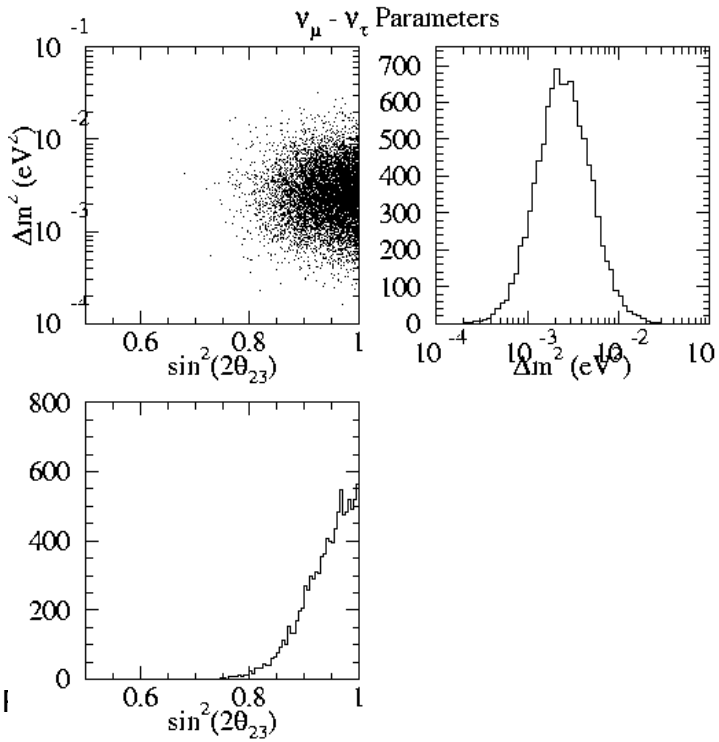
- Challenge set
  - oscillations (far), modified flux, modified Neugen3
  - truth information removed (from public access)
  - FarDet: 7.4e20 pots (3yr x 62%, 1.9s rep @ 2.4e13 pots/spill)
  - NearDet:
    - overlaid, 300[-54,-5] files, 550 snarls @ 2.46e13 pots/spill
    - horn off – coming soon...
- Monte Carlo set
  - truth information retained
  - FarDet: each file represents 6.5e20 pots, mix of NC+CC
    - nominal beam: 20 [-0,-1] files
    - $\nu_x \rightarrow \nu_e$: 9 [-0,-1] files (intrinsic ν_e is unmodified, additional NC's)
    - $\nu_x \rightarrow \nu_\tau$: 40 [-1,-5] files (additional NC's)
  - NearDet:
    - overlaid, 300[-77,-2] files, 550 snarls @ 2.39e13 pots/spill
    - single events – coming soon...

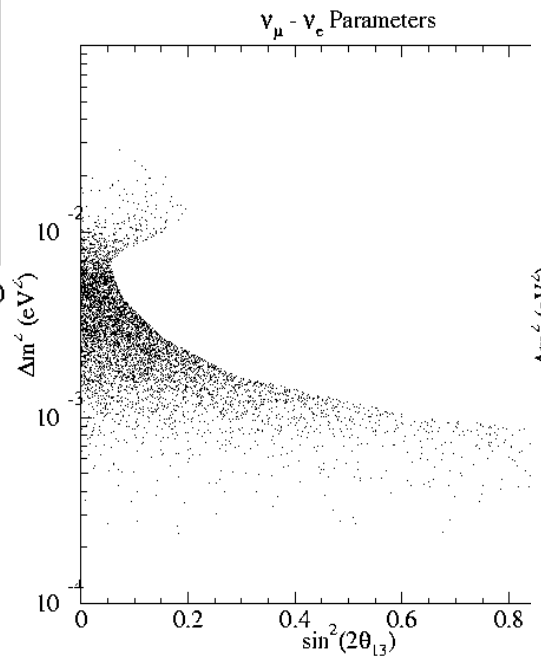
What: Generation Specifics

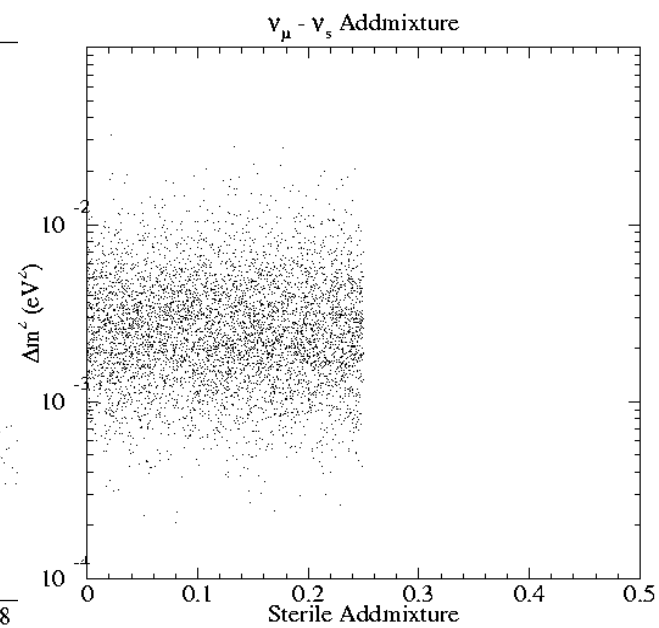
- Flux:
  - directly from GNUMI v16 column-wise n-tuples (horn off is v15)
  - re-weight GFLUKA to FLUKA (WFLUK.F)
  - re-weight energy/probability over face of flux window
  - challenge set has hadronic production warping
- Neugen3 mods for challenge set:
  - Resonance mass (`gv_ma_res`): 1.032 ± 4%
  - QE mass (`gv_ma_qel`): 1.032 ± 4%
  - DIS fudge factor (`gv_mdc_fudge`): 1.0 ± 3%
- GMINOS:
  - see Cambridge talk...
- GEANT:
  - GCALOR for hadronic model
  - 10 keV cutoff
- Oscillation parameters:
  - wouldn't you like to know...

Oscillations (ala Mark Messier)

- 100% chance of $\nu_\mu \rightarrow \nu_\tau$ oscillations
  - $\log_{10}(\Delta m^2_{23})$ gaussian(SK,1.5*1σ_{SK})
  - $\sin^2(2\theta_{23})$ gaussian(SK,1.5*1σ_{SK}) restricted to physical region
- 75% chance of $\nu_\mu \rightarrow \nu_e$
  - parameters uniformly up to CHOOZ 95% CL limit
- 75% chance of sterile oscillation
  - parameters uniform up to SK 95% CL limit
- No CPT violation, decoherence, Lorentz violation, nu decay, LSND, etc







Where: enstore tape

- dCache access (also via PNFS on minos1)
  - **dcap://fndca.fnal.gov:24125/pnfs/fnal.gov/usr/minos/mcout_data/**
  - + *production release* (currently **R1.7/**)
  - + one of: **far/**, **near/** (MC) or **fmock/**, **nmock/** (“Data”)
  - + one of:
    - **cand_data** – files w/ raw + candidate records
      - MC files also have SimSnarl (truth) records
    - **sntp_data** – “Sue style” n-tuple files w/ NtpSR, Ntp3D TTrees
      - MC files also have truth TTrees
    - **snts_data** – like sntp_data but w/out strip info

Where: file naming convention

- Base file name:
  - **f2410***srrr*_**0000** – FarDet MC nominal beam
  - **f2411***srrr*_**0000** – FarDet MC ν_e
  - **f2413***srrr*_**0000** – FarDet MC ν_τ
  - **n1430***srrr*_**0000** – NearDet MC overlaid beam
  - **n1410***srrr*_**0000** – NearDet MC single-event beam
  - **N1130***srrr*_**0000** – NearDet "Data"
  - **F2110***srrr*_**0000** – FarDet "Data"
  - **N1114***srrr*_**0000** – NearDet horn-off "Data"
- *srrr*
  - *rrr*: random # sequence
  - *s:* split file (Far=0, Near Overlay=1:3)
- Extension .*type*.*release*.**root**
  - *type*: **cand**, **sntp**, or **sntp**
  - *release*: **R1.7** (currently)

How: Two Steps

- GMINOS production
  - NearDet more complicated than FarDet
    - generate files of rock events – very time intensive
      - need not be independent; reused these events
    - generate file of individual detector events
    - overlay and mix in rock events to form snarl
    - split into more manageable files
  - Distributed over various sites:
    - IU: some NearDet detector files + overlay mixing
    - Tufts: the majority of the NearDet + overlay mixing
    - FNAL: rock files, FarDet files, smattering of NearDet
  - flux re-weighting slows generation down
- PhotonTransport + DetSim + Reconstruction
  - processing was done on the FNAL batch farm

When

- Lots of files are available **right now**
  - processed with R1.7
- More to come as CPU cycles allow
  - MC Near with no overlay – RealSoonNow
  - Horn off data – possibly problematic...
  - More NearDet overlaid MC + “Data”
  - Need to improve speed by:
    - profiling code to see where time is spent
    - fix dithering problem which results in lost files
- Reprocess with R1.8 when tagged

Why

- Test generation mechanics
  - pointed out some of the difficulties
- Test PhotonTransport/DetSim
  - pointed out some of the difficulties
- Test reconstruction
  - pointed out some of the difficulties
- Test file handling procedures
- Test analysis groups' procedures
  - ...we await results...

Rough Time Estimates

- FarDet example file:
  - 3.8e21 POTs = 36538 events = 41.9 GHz*hrs
  - this is ~ size of MC data set
  - estimate ~13 days to reconstruct (@30sec/evt)
- NearDet example file:
  - 1000 spills (8m,20s beam time) at 4e13 POTs/spill assembled from
    - 42016 events in the NearDet (40.6 GHz*hrs)
    - 20877 events generated in the surrounding rock that left energy in a scintillator strip (135.1 GHz*hrs)
    - 9.6 GHz*hrs overlay procedure (and redigitization)
  - a target cylinder R<50cm, 20<ipln<59 is ~0.65% of total mass means that only 1 in 3.8 spills (at this higher POTs/spill) has an event in this volume; even worse if one requires 25cm radius
- Near challenge, MC sets
  - ~1.4e7 total events or ~5.75e5 spills (13.3days beam time)
  - ~113.4 GHz*days + ~377.5 GHz*days = 16.4 GHz*months each
  - tricky file management issues due to overlaying
  - ? estimate of reco time