

STNTUPLing June'2001 data

- **Introduction**

- Data taking, detector/offline commissioning in progress
- Have lots of data, what is the most efficient way of looking at them?
- Want disk-resident datasets, even though we are short of disk space
- Need highly compressed data format with convenient access
- Want to use the same datasets and tools for commissioning now and for analysis after that

- **STNTUPLE**

- comprehensive multi-branch ROOT ntuple+ (micro-DST):
- <http://www-cdf.fnal.gov/upgrades/computing/projects/Stntuple/Stntuple.html>
- Data organized in blocks, each block corresponds to a branch
- Portability: compiles with GCC (x 3-4 speedup on Linux)
- to access all the objects stored in the STNTUPLE one needs a shared library of about 6MBytes (Linux)
- Assume that preferred analysis mode relies on executing the scripts, compiled on the fly with ROOT script compiler
- Many examples of how to write the analysis scripts
- lots of nice utilities
- The goal for analysis datasets: event size of about 5-10KBytes, mostly reconstructed objects
- In early days store RAW data as well (already appreciated by the PLUG, XFT, CLC and CES groups during the commissioning period)

- **Reconstructed data blocks**
 - Header block
 - Jet block (cone 0.4, provision for “fat” jets)
 - Cluster block (CES, CPR, PES clusters) – Ray Culbertson
 - Met block – Chicago group
 - Electron block
 - Muon block: muons, CMU/CMP/CMX stubs (recent additions by Anyes Taffard from Liverpool)
 - Tau block:
 - Photon block – Ray Calbertson
 - Vertex block – Hartmut Stadie
 - Track block: track parameters, error matrix, hit content
 - Trigger block: TSID, TFRD, TL1D
 - XFT: (Richard Hughes)
 - SVX data block: Aaron Domingues started from SVX hits (a hit is a cluster plus global space coordinates)
 - SVT: Subir Sarkar (Geneva)/Un-ki Yang(Chicago)

- **MC data blocks**
 - GENP block
 - OBSP block (OBSP+OBSV)
- **RAW data blocks**
 - Cal data block
 - CES, CPR, PES data blocks
 - PES data block
 - CMU, CMP, CMX data blocks
 - CLC data block
- **Most important missing pieces of the data**
 - COT hit block (Evelyn Thomson (OSU) is interested)
 - Connection between MC particles, tracks and hits

- **STNTUPLE datasets for June**

- learning how to deal with the terabytes of data in quasi-real time
- Production: few days behind, use L3 output (available next day)
- People involved: Beate Heinemann (for QCD group) and I, starting from the next week: U.of Chicago group is going to join
- Trigger tables are changing, this affects the what you do
- **Location:**
FCDFSGI2:/cdf/data05/s0/commissioning/stntuple/36x36
- **documentation:** <http://www-cdf.fnal.gov/upgrades/computing/projects/Stntuple/Stntuple.html>
- **muon skim:** CMU wedge (≥ 3 hits) + COT track $P_t > 2.5$ GeV in $\Delta(\phi) < 0.5$
- **Jet skim:** ≥ 1 central ($|\eta| < 1$) jet with $E_t > 20$
- **EM object skim:**
 - ≥ 1 Em object $|\eta| < 3$ with $E_t > 10$ GeV or
 - ≥ 2 EM objects with $E_t > 5$ GeV (need to double check)
- **Conversion sample**
- **Data size** (run 118783 ,100000 events)
 - Muons: 12.5K events
 - EM objects: 7K events
 - Jets: 4.5 K events
- Total: 800 MBytes (about 33 Kbytes per event on disk, ROOT compression is about 3 for these events, so 100 K Bytes of data per event)
- **Even now can keep good data on disk long enough!**

- **Split vs non-split mode:**
 - Split: HBOOK column-wise type, can “click” on leaves to plot histograms
 - Non-split: 12 branches, each branch written in non-split mode. Can tune I/O routines to improve I/O performance
 - performance tests for CalDataBlock (array of CalTower’s)
 - Split mode, no compression: **4.2 Mbytes/sec**
 - Non-split mode, no compression: **13.6 Mbytes/sec**
- **Stream A means physics today!**
 - **Starting from yesterday processing of the stream A is a priority**
- **Next steps:**
 - **Will be adding missing pieces** (more trigger info, for example)
 - **Will do the best to keep up with the data taking.**
 - A job reading the tape via DH system currently finishes successfully 1 time out of 5, so it takes a few days to process a single run...
 - STNTUPLE datasets are small, can allow more datasets:
 - MET dataset,
 - Dilepton dataset
 - **It is a lot of work, help is welcome!**