

TWR DATA MANAGEMENT 2003 and UCI TWR WEBPAGE

ANDREA SILVESTRI
silvestri@HEP.ps.uci.edu
UNIVERSITY OF CALIFORNIA, IRVINE

- Raw TWR data of AMANDA 2003
- Set up of data Management at UCI
- Processing of TWR data
- Strategy of Filtering
- Conclusion
- Action Items

AMANDA Collaboration Meeting, Bartol
March 2004

RAW TWR DATA on SDLTs

- Currently all AMANDA 2003 TAPES at UCI
 - (Thank you: Christian S. and Matthias)
- Total # of TAPES: 107
 - 104 TAPES of 2003 = [TAPE_1 - TAPE_104]
 - 1 TAPE of 2002 = [TAPE_142]
 - 5 TAPES of 2004 = [TAPE_105 - TAPE_108]
 - 2 TAPES of 2003 = [TAPE_1 - TAPE_2] (at UW)
- ONLY 1 COPY of TWR data
 - Let's put them on Hard-drive
- 2 copies available of muon data
- High multiplicity data: merged f2k also available

TAPE CHARACTERISTICS

- TAPE_1 to TAPE_45 = data in software compression
(file.gz, file.bz2)
- TAPE_46 to TAPE_104 = data hardware compressed
- Typical Content of TAPE:
 - Muon data: ab_2003_267_7311_064.data.mu.to_tape_1
 - Daqlog: daqrun_2003_265_7309.log.mu.daq_log.to_tape_1
 - Rice: data-2003266042244-test_20.dat.rice.to_tape_1
 - Spase data: s2_2003_267_21a.spase.to_tape_1
 - SN data: sndata_0968_00.root.sn.to_tape_1
 - TWR data: twr_2003_267_7312_0129.dat.twr.to_tape_1
- UCI will take care only of TWR data processing
- Other data ~ 2TB (not on Hard-drive)

STRUCTURE OF THE FILE CONVENTION

- Canonical naming of file:
 - `twr_2003_267_7312_0129.dat.twr.to_tape_1`
- Def.:
 - `twr_YYYY_DDD_RRRR_FFFF.dat.twr.to_tape_1`
 - `YYYY` = Year
 - `DDD` = Day of year
 - `RRRR` = Run number
 - `FFFF` = Run File number
 - `dat.twr.to_tape_1` = File extension
- 330 DAYS = [35-365]
- 2914 RUNS = [4814-7728] (~ 1800 cal.runs)
1100 runs ~ 3 runs/day
- ~ 67300 FILES

EXTRACT DATA FROM SDLT

- Time requirements:
 - TAPE_1 to TAPE_45 S/W-compression:
 - To extract 1 TAPE ~ 95GB in 2.5 hr
 - If twr.gz 1GB/min ~ 2hr/TAPE (3.2GHz)
 - If twr.bz2 0.08GB/min ~ 1day/TAPE
 - 4.5hr/TAPE (.gz) 1day/TAPE (.bz2)
 - TAPE_46 to TAPE_104 H/W-compression
 - Extracting time=Reading time
 - 1 TAPE ~ 146GB [2-3]hr/TAPE
 - Optimistic: All TAPES 380 hr
 - All TAPES ~ 16 DAYS



TAPE AMOUNT OF DATA

- TAPE_37 = Typical TAPE of Software compression:

Files	#files	MB	GB	%
Muon	387	38	14.7	15.3
Daq-log	12	--	--	--
Rice	199	2	0.4	0.4
Spase	22	10	0.2	0.2
Sn	10	40	0.4	0.4
TWR	670	120	80.4	83.7

TAPE AMOUNT OF DATA

- TAPE_73 = Typical TAPE of Hardware compression:

Files	#files	MB	GB	%
Muon	356	52	18.5	12.6
Daq-log	6	--	--	--
Rice	62	8	0.5	0.4
Spase	19	26.5	0.5	0.4
Sn	13	39	0.5	0.4
TWR	630	200	126	86.2

TOTAL AMOUNT OF DATA 2003

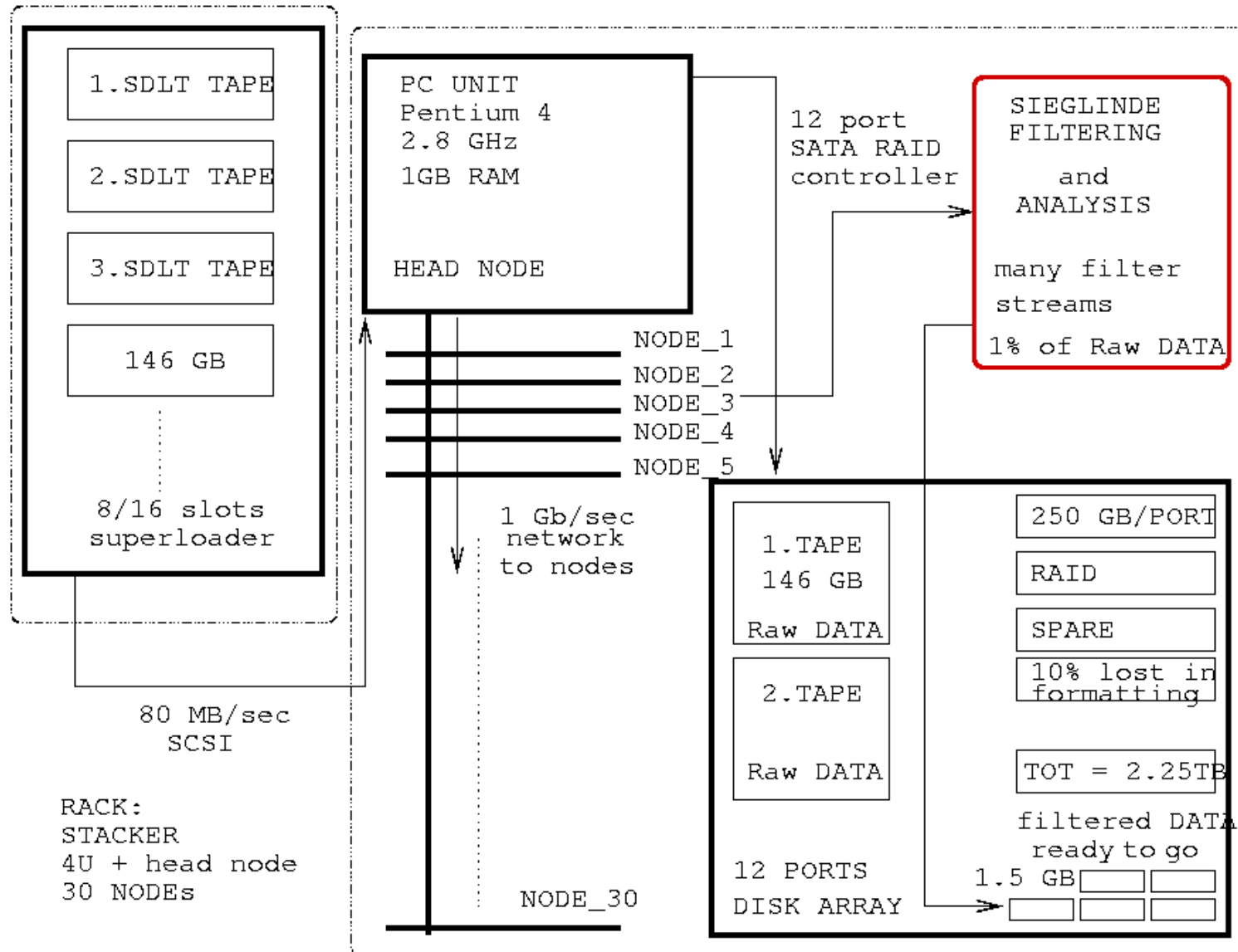
- TAPE_3 starts: Feb 04 2003 twr_2003_035_4814
- TAPE_104 ends: Dec 31 2003 twr_2003_365_7728

DATA	#files	SIZE	%	Summ.
Muon	38419	1.9TB	12.6	2.29TB
Daq-log	894	--	--	--
Rice	12613	25GB	0.16	26.6GB
Spase	1188	20BG	0.13	21GB
Sn	1217	15GB	0.1	16GB
TWR	67320	13TB	86.6	13.4TB

HOW TO MANAGE DATA?

- How to manage 13TB?
 - No way to store it on hard drive!
- Solution:
 - We need to filter it prior to store it
- New high-performance cluster facility at UCI
- Set up:
 - SDLT Stacker
 - U4 RAID + head node
 - 30 nodes (3.4 Ghz)

UCI CLUSTER FARM



HARDWARE COMPONENTS

- Performances: Fast, flexible, stable, lots of space
- NASA has exactly same cluster for HUBBLE-telescope data processing



4U head node



ParaSoft Corporation



1U node



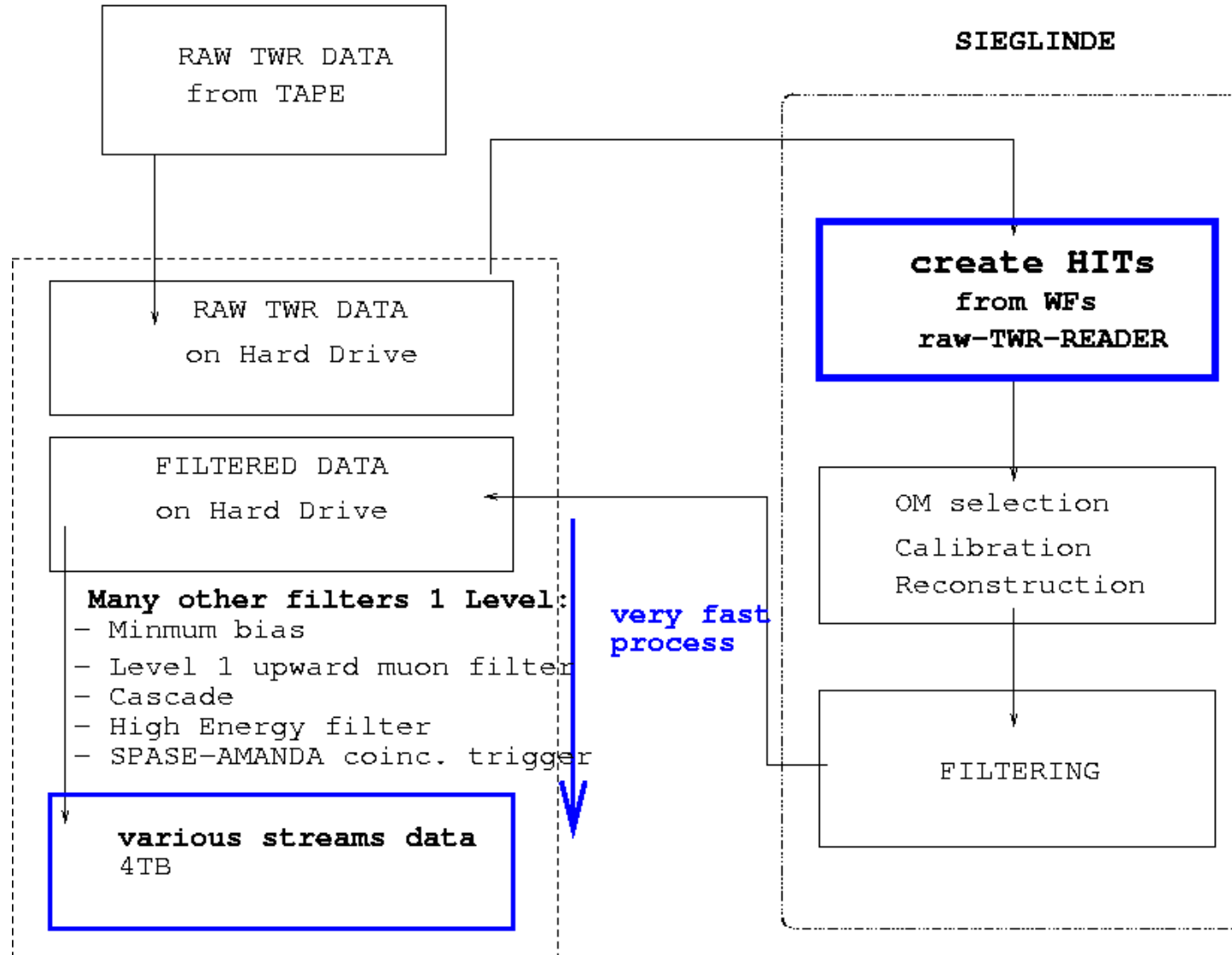
Quantum. stacker



Rack with 30 nodes

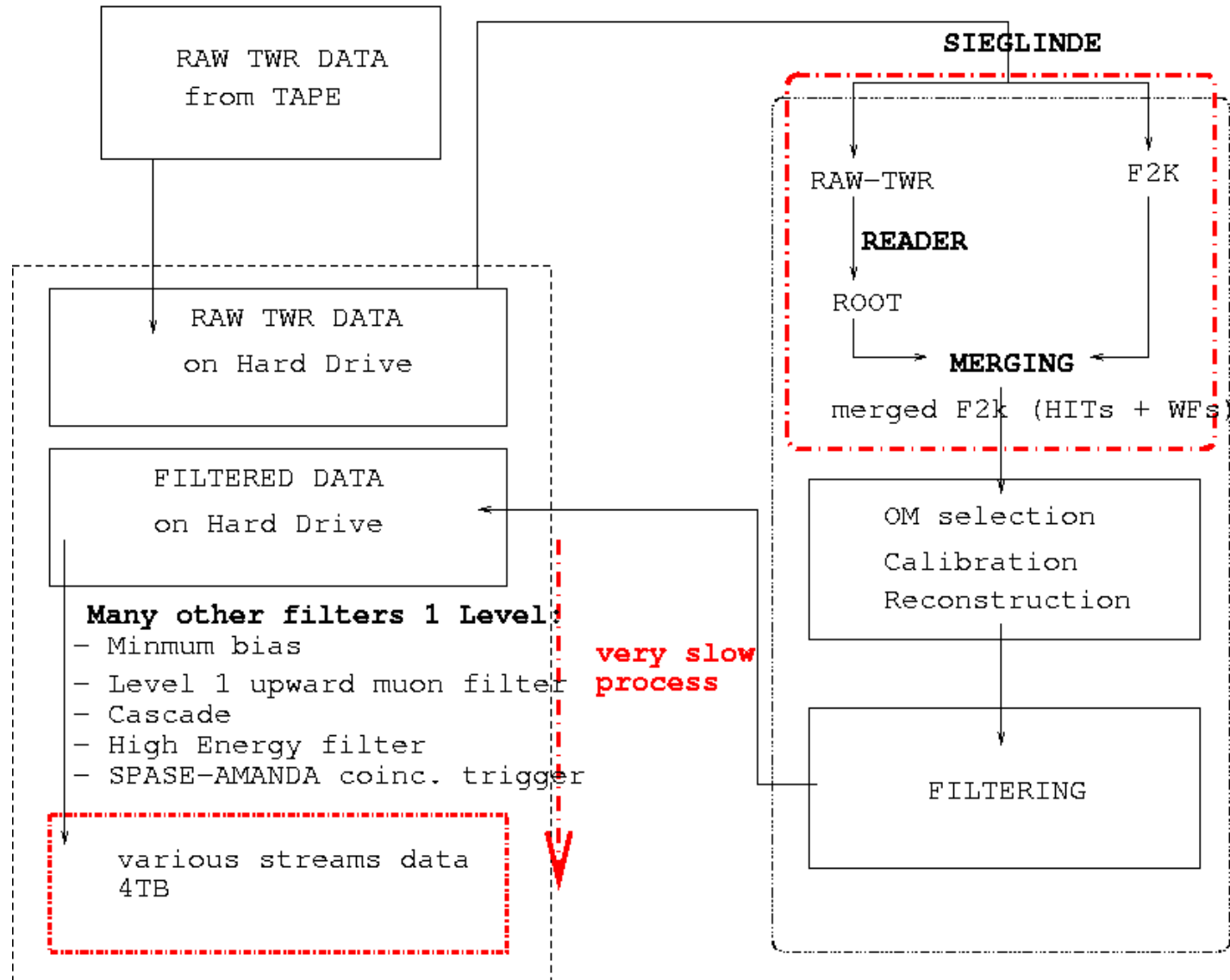
STRATEGY of FILTERING

- + No merging; + more info from WFs; + Sieglinde faster



ALTERNATIVE STRATEGY of FILTERING

→ - slow; - merging; - lose events;



UCI TWR-2003 WEB-PAGE

- amanda.uci.edu/twr_03_04.html
- Goal of the website:
 - ftp server will be available: transfer data via mirror
 - Updated status of TWR data processing
 - All filters, scripts used available
 - High mult. f2k events data 2003 (satellite)
 - Detailed explanation of processing done
 - TAPE, run, file, livetime statistics information
 - Link to MC mass production and Sieglinde websites
 - Plots: comparison 2003 filter data with mb MC.
- Entire filtered data set, depending on filter used:
 - Approx. 1% raw data = 130GB.
 - Not simple over network
- Solution: it fits on 1 SDLT TAPE
 - BUT, you will need a SDLT drive.

MC MASS PRODUCTION

- We do need of course a parallel MC mass production:
 - (?) Current status of simulation
 - (?) Thomas and Christin Burgess – Stockholm
 - (?) AMASIM ready to produce WFs to be piped into Sieglinde
 - (?) Hardware, CPU and Hard-drive resources
 - (?) MC Data dissemination
 - (?) Estimation of time needed
 - (?) Expected similar effort for MC mass production

EXPECTED TIME ESTIMATION for TWR 03

- With UCI compute farm of 30 nodes
- Extract data from SDLT: ~ 16 Days
- Time of processing for filters:
 - HE (use satellite DATA) available on the TWR WEB
 - CASC requires too much CPU
 - UPMU (2-fits) = 92 days (very conservative)
 - No SPASE-AMANDA coincidence data
- Of course with raw-TWR reader into Sieglinde
 - Much shorter CPU-time (a factor 10?)
- Estimated Livetime of ~ 200 days
 - $1.2 \cdot 10^9$ events (70Hz trigger-rate)

WHAT NEXT

- Set up of Hardware, compute farm at UCI
- Sieglinde installation and testing
- Set up of processing-filtering chain scripts
- Parallel MC mass production
- Upgrade and update of UCI TWR webpage
- Data available to entire collaboration, SDLT TAPES
- First TWR UHE neutrino analysis at UCI
- If OK offline processing and filtering
 - Online implementation at the Pole 04/05

MANPOWER AT UCI

- Andrea: (Manage of processing, Sieglinde testing + filter evaluation)
- Jiwoo: (Sieglinde testing + filter evaluation)
- Jed Rogge: (System Manager: Hardware Installation)
- Bayan: (Setup website interface, tape feeding, summary statistics)

CONCLUSION

- All TWR TAPES shipped at UCI
- Start of data processing TWR03
- High performance cluster farm at UCI (40k–50k)\$
- New implementation of Sieglinde
- UCI TWR webpage, and its continuous update
- MC effort in parallel
- Soon filtered data available
- If offline processing fine -> online at the Pole
- Action Items

ACTION ITEMS

- UCI data management: (Andrea, Jiwoo, Jed and Bayan)
- MC mass production:
- Sieglinde + TWR effort:
- Other items:
- Time schedule:
 - ◆
 - ◆
 - ◆
 - ◆