

# Status of the ZEUS Offline Processing

and

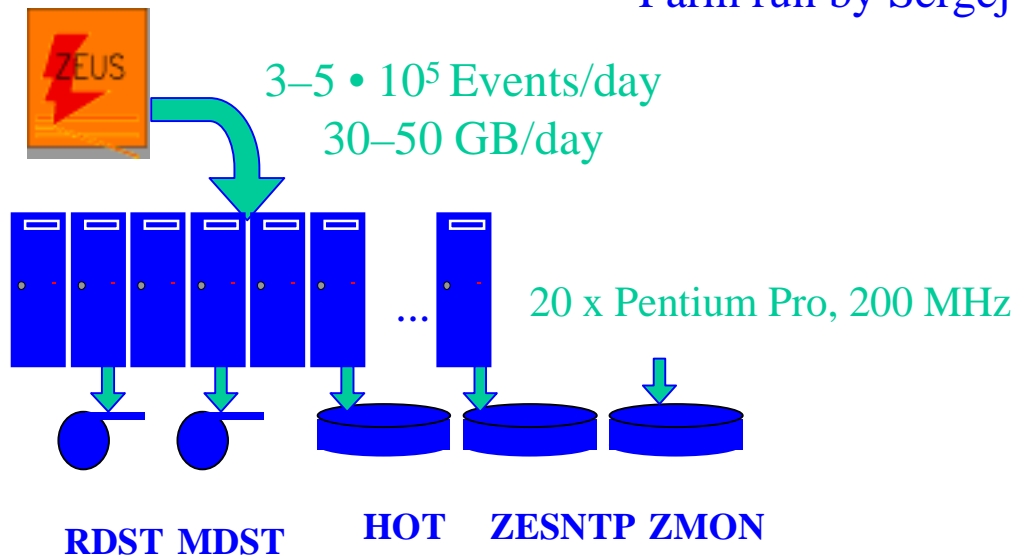
# Resource planning before & after shutdown

**Rainer Mankel**  
**DESY Hamburg**

**ZEUS Collaboration Meeting, DESY Hamburg**  
**Offline Parallel Session, 21-Feb-2000**

# Reconstruction Data Flow

Farm run by Sergej Zotkin



cpu time per event: ~ 4.2s (200MHz Pentium)  
program size: ~ 110 MB  
farm capacity (max): ~ 490 000 events/day

Datastreams:

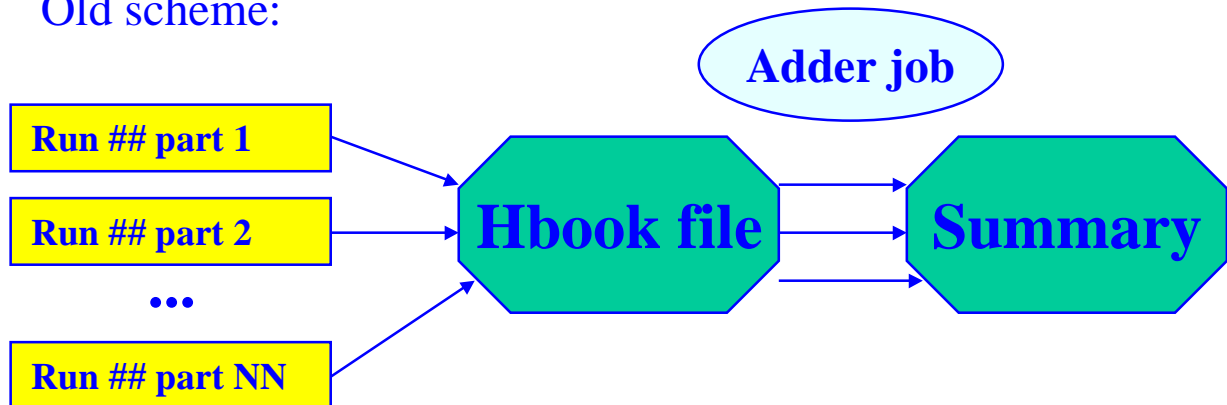
- RDST
- LOG
- HOTEVT
- MDST
- ZESNTP
- ZMON
- PMSTAT

# The Current Reconstruction Program

- Based on ZEUS software release v1999a.1 (appeared Jan-2000)
- Code changes:
  - trigger logic updates (o1recon)
  - mfrecon: fixes for miscabling problem
  - electron finders position/energy correction fixes
  - prrecon: presampler correction routines
  - other fixes/updates (vcrecon, fpreco, tfrecon)  
in total 11 packages modified
- Data selection changes:
  - DST bit introduced for cosmic ray events ( $\cong$  EXO19)
  - corresponding events added to hotline
  - some information for BMUON added to HOT events
- Changes in file environment
  - /zeus/pro/gaf resides now on AFS volume (Solaris, Linux, IRIX, OSF1)
  - accessible from all machines with AFS clients
  - simplified distribution
  - backward compatible
- Changes in the offline monitoring (ZMON)

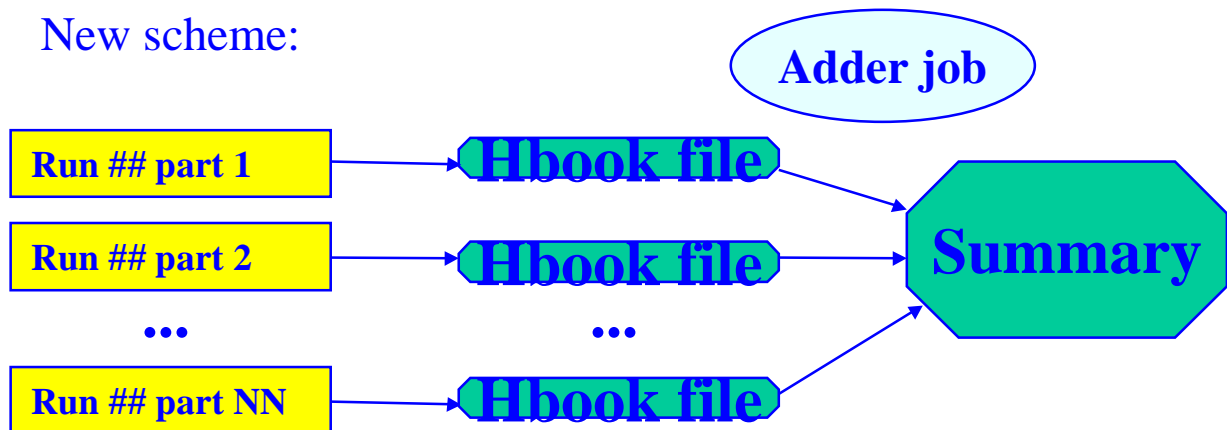
# Changes in Monitoring (ZMON)

- Old scheme:



- Needed file locking scheme, with occasional failures required manual merging

- New scheme:



- ZMON histogram files under
  - /shift/zarah1/data20/zmon/zmon00/

# Reconstruction History

- Test run reconstruction on 21-Jan-2000
  - found/fixed bug in computation of a **ZES variable**
  - some other small changes
- Reconstruction of current data started on 28-Jan-2000
  - some (rarely occurring) STOP conditions found & fixed (one fix underway)
  - some problems with corrupted luminosity GAFs
  - some problems with **ZES ntuples** found & fixed
  - some bad CAL GAFs (4 runs)
  - BPC timing currently not usable (south part dismantled)

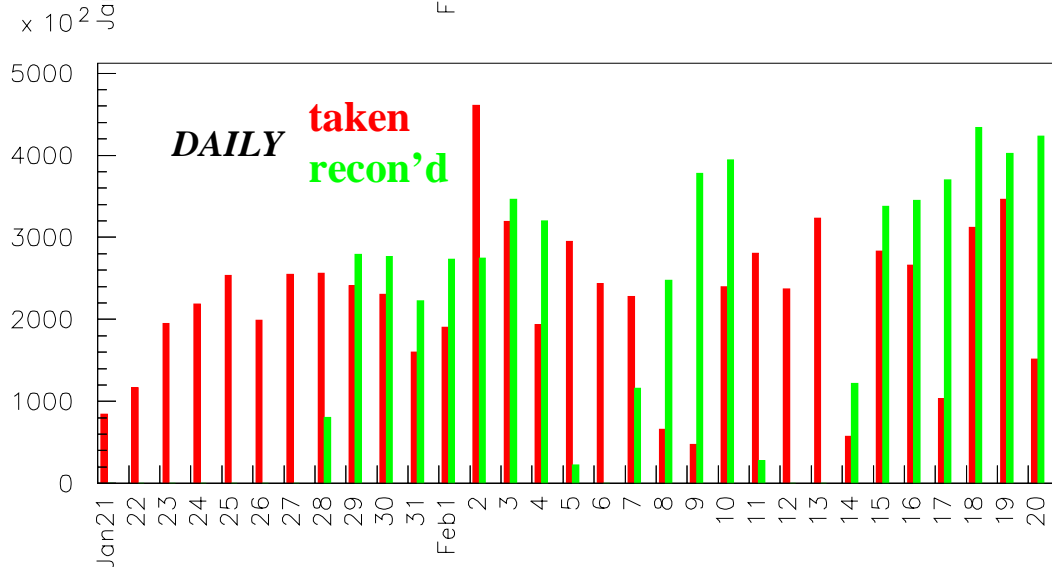
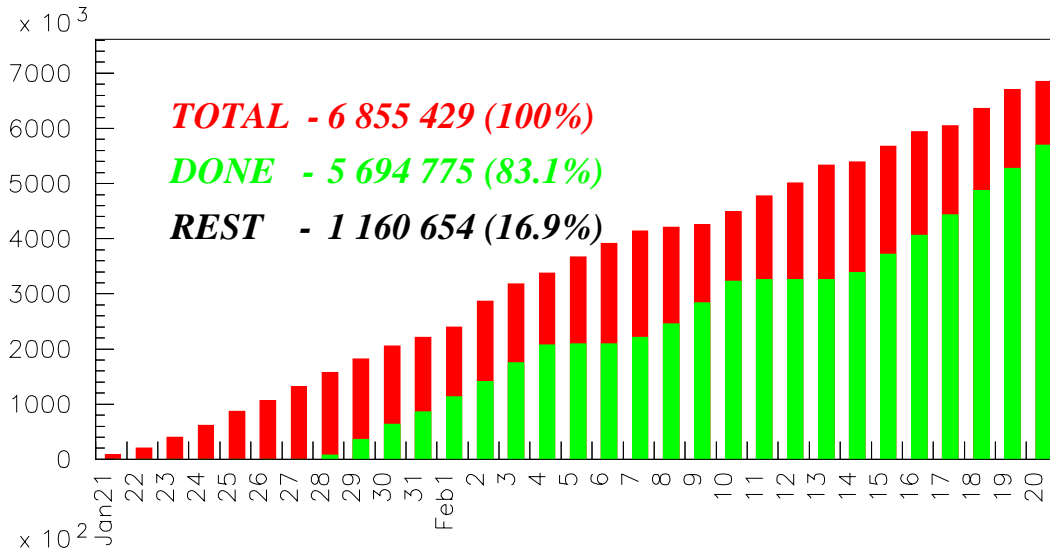
**reconstruction operation is smooth & stable**

Reconstruction not “online”, but reasonably close to data taking

Please use **every possibility to check** output & provide feedback

**Rainer.Mankel@desy.de**

# Current Status



- Good performance of HERA
- New delivered luminosity record ( $1.959 \cdot 10^{31}$ ) last week
- Reconstruction keeping up well with data taking
- Close enough to provide feedback

# ZEUS Computing Resources

## Situation & Goals

- Now:
  - 2000: Postponed shutdown  
8 months of data-taking instead of four  
requires more resources than originally planned for this year



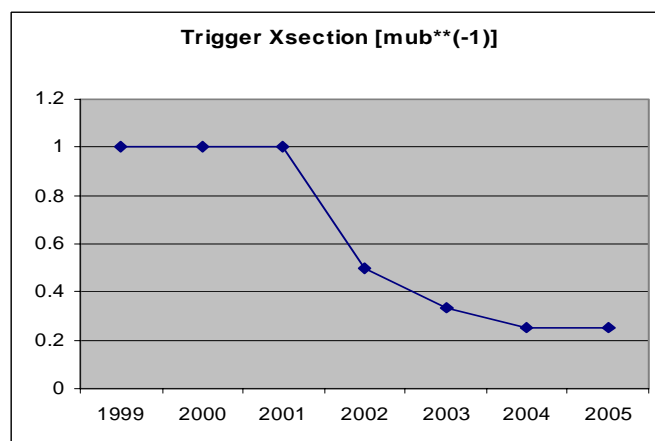
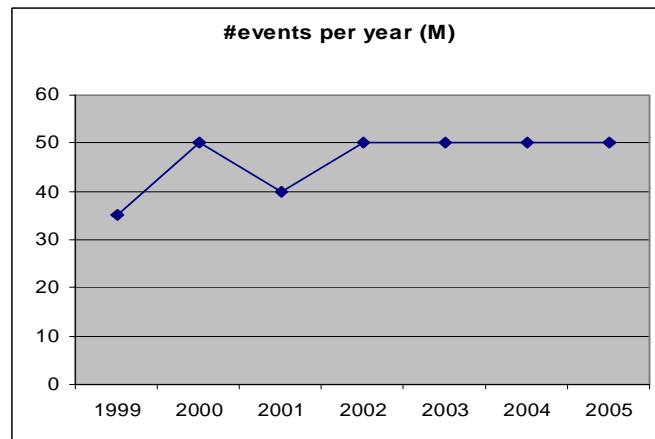
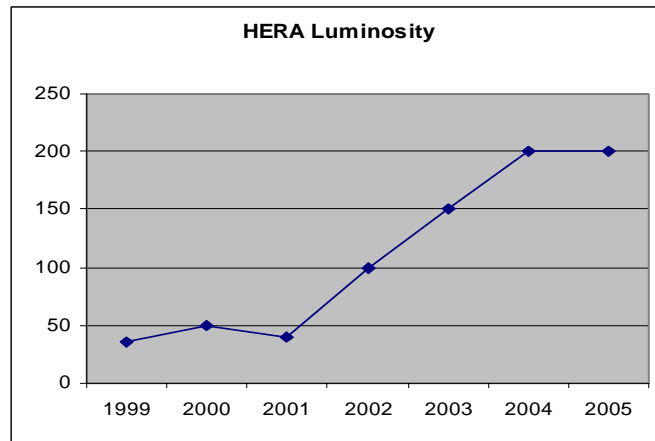
- Soon:
  - Shutdown: unique opportunity to bring our existing data (1996-2000) into consistent state
    - positions of calorimeters
    - full set of ZES variables for all events  
⇔ ORANGE
  - we should reprocess data from 1996-2000 during the shutdown

# ZEUS Computing Resources (cont'd)



- Mid-term:
  - 2001: **New detector components** (MVD + STT)
  - increase of data volume (RAW, R/MDST, MC)
  - additional computing in **reconstruction**, MVD/STT/global tracking
  - new reconstruction algorithms
    - time for SW development getting short
    - might be not fully optimized at the beginning
    - likely to require **extra output** in the beginning
      - diagnostic information
      - matching information
    - might require **extra reprocessing**

# Rate Assumptions



# MVD Offline Resource Estimates

thanks to : Uli Katz

**1 event: ~300 strips, 90 clusters, 12 tracks, 10 vertices + contingency for noise**

- RAW data: strips, clusters (from ADC's) 10 kB/evt
- R/MDST: clusters, tracks, pattern-rec info, extrapolations, vertices, relations 20 kB/evt + RAW (now)  
12 kB/evt + RAW (from 3rd year)
- CPU time per event 2x CTD today?  
10x CTD today?  
More??  
Much more??

**My working number: 4s/evt on PPro200**

- MC truth: impact points, strips, clusters 13 kB/evt
- extra space for global tracking?

# STT Offline Resource Estimates

thanks to : Ian Brock, Alexei Antonov

**1 event: ~200 fired straw r/o channels, 5%  
mean straw occupancy, 20 tracks, + noise**

- RAW data: hits, time information 30 kB/evt beginning, 3 kB/late later **10**
- R/MDST: hits, tracks 50 kB/evt + RAW (now) 6 kB/evt + RAW (later)

**My working number: 30 kB/evt first year  
9 kB/evt later**

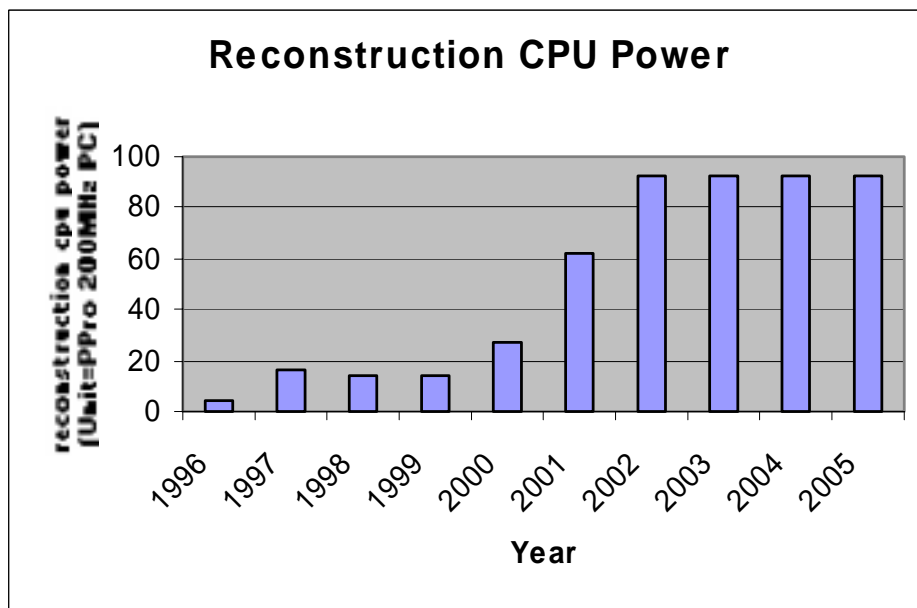
- CPU time per event 4s on SGI processor, ≈ 2.7s on PPro200

**My working number: 4s/evt on PPro200**

- MC truth: 20 kB/evt
- extra space for global tracking?

# Reconstruction Computing Power

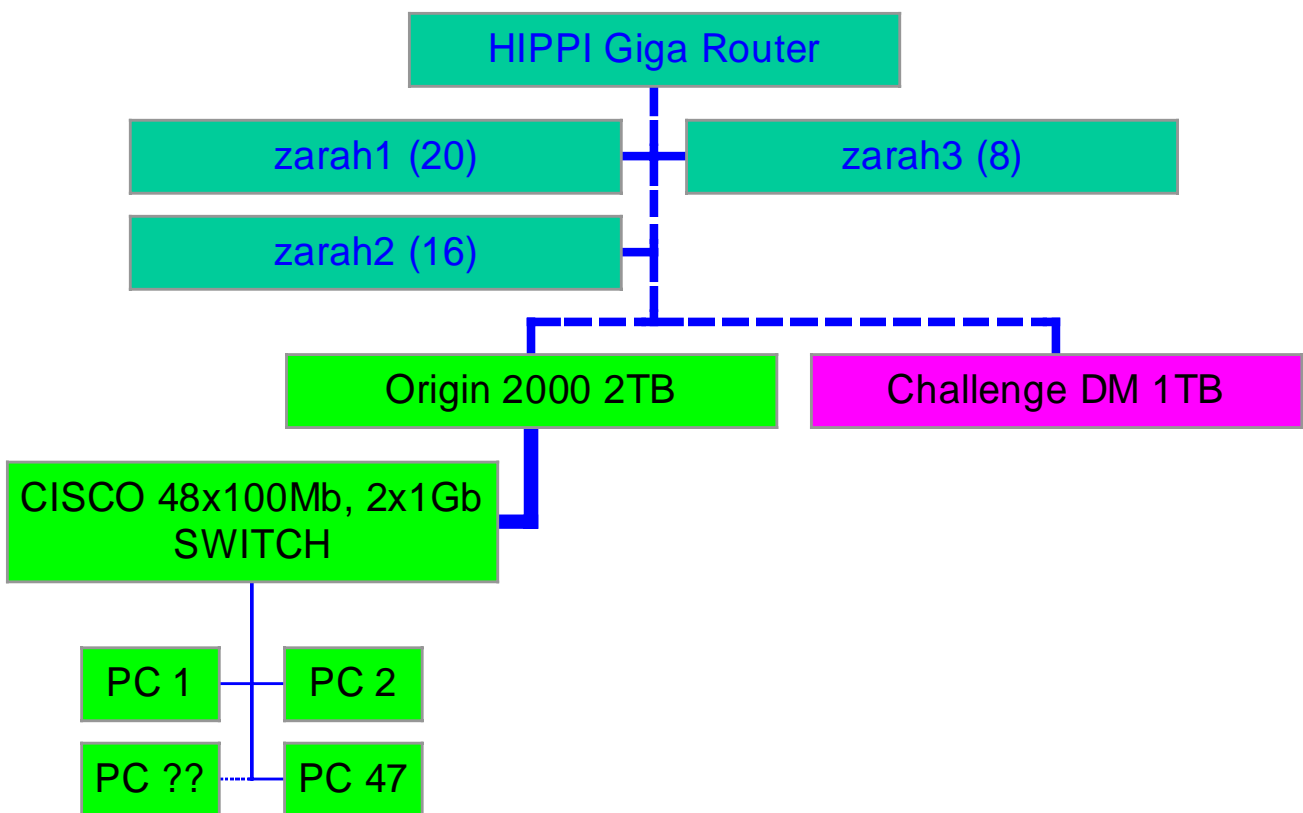
- No certain estimates on CPU time for **STT/MVD reconstruction** quite some margin
- Working hypothesis: CPU time per event **might triple**
- Model: enough power to **reconstruct** current year's data & **reprocess** last year's data in 5 months



- 2001 reconstruction requires update of a **factor of three**
- Shutdown: reprocess **150M events** of 1996-2000  
upgrade already this year, prior to shutdown  
2002: probably **another upgrade**, depending eg. on  
MVD/STT reconstruction needs

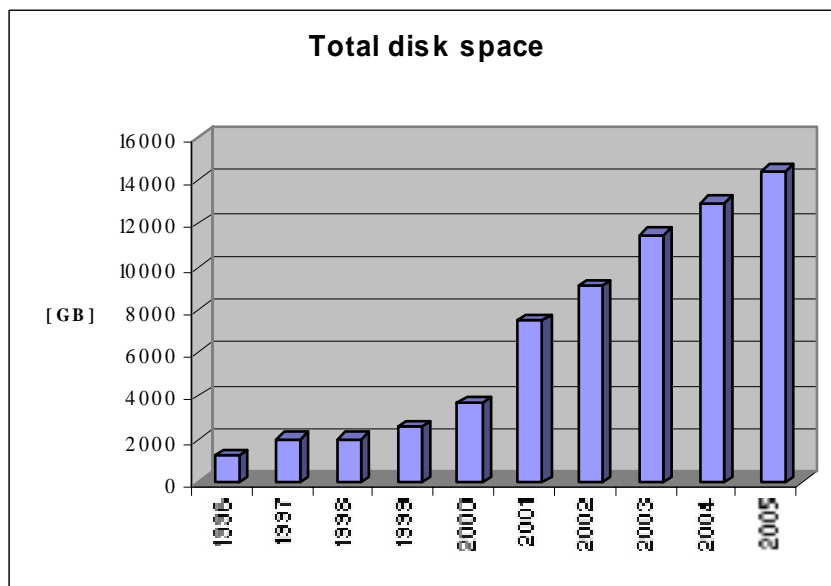
# Batch Analysis Resources

## ZARAH network environment



# ZARAH Disk Space

- This year: availability of some reprocessed data was delayed, mainly due to **disk space bottleneck**
  - 1 TB increase for 2000 already ordered
  - more is needed to accommodate this year's data
- Shutdown: need more workspace flexibility for **reprocessing**  
Another 2 TB this year



## ZARAH Disk Space (cont'd)

- 2001: General increase of event size, assume:
  - factor of 3 just for 2000
  - factor of 2 in the long runAdd 2 TB next year
- Another important long-term issue:
  - complexity of our system, number of components is ever increasing
  - will increasing maintenance effort kill us?
  - down-times of disks might slow down reconstruction & analysis

Improve reliability of disks

use RAID technology

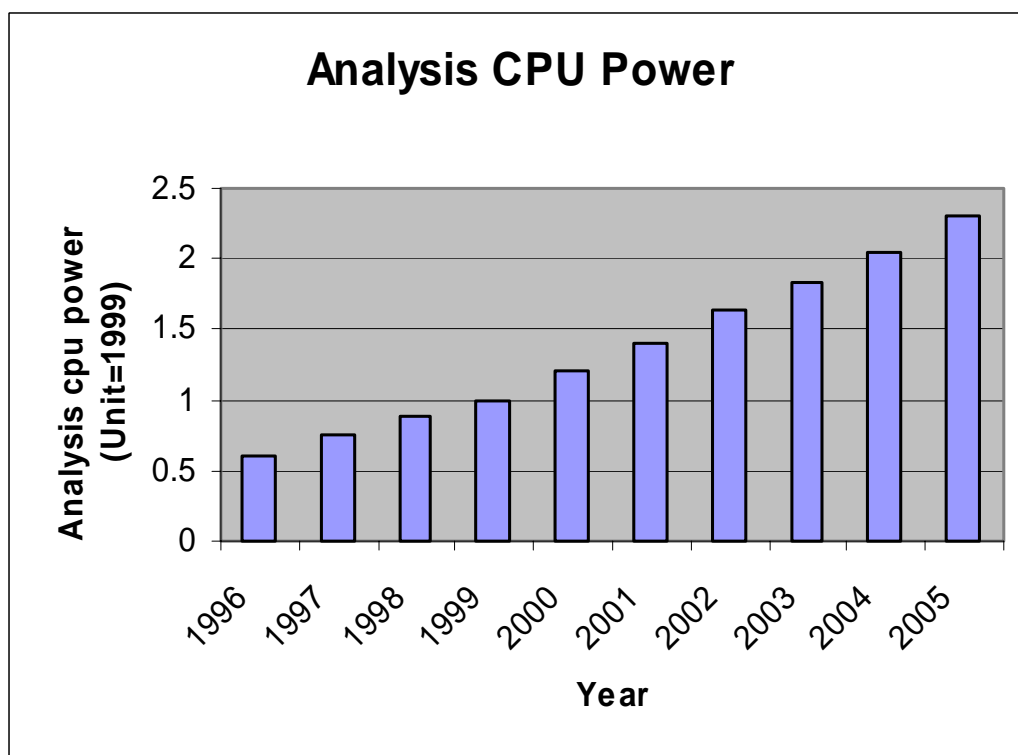
save (additional) personpower

continuous operation

Prepare for acquisition of another file server in 2002

# Batch Analysis Computing Power

- Batch queues grow too long in “prime times”
- More computing power needed
  - new **ZARAH PC farm** becomes available (currently under *beta test*)
  - exact gain to be quantified **empirically**
  - should be sufficient for this year
  - in mid-term future: SGI’s will run as long as they are functional & useful

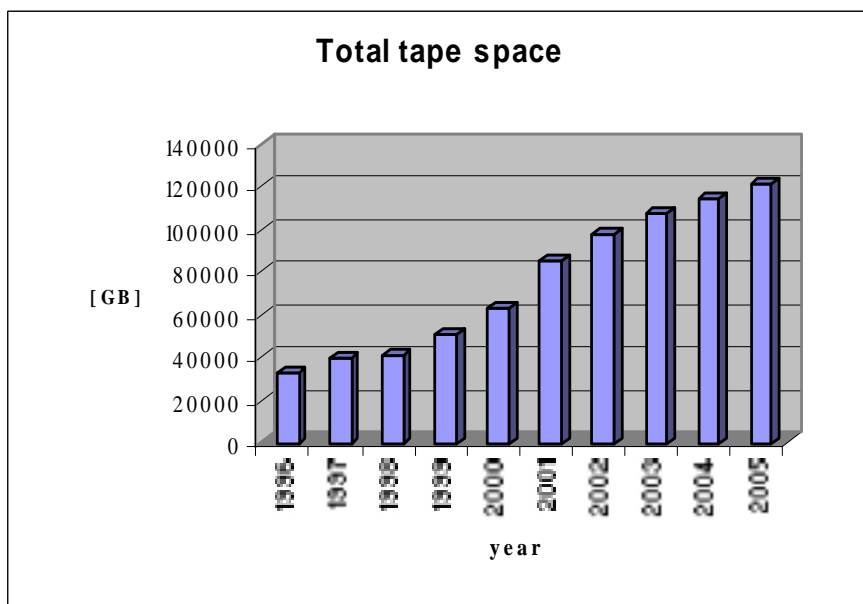
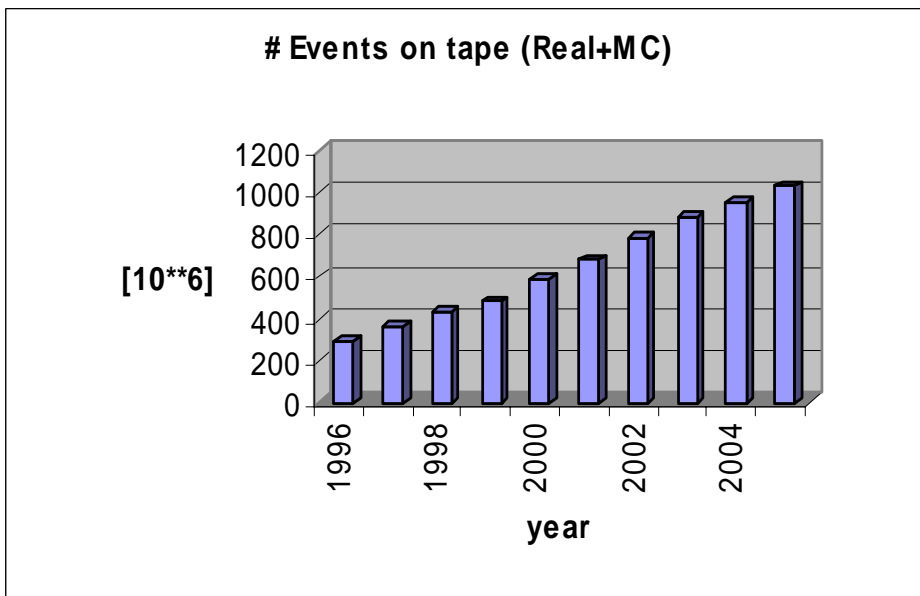


# Batch Analysis Computing Power (cont'd)

- Demands will grow with increasing amount of data
- More pressure towards “fast” production of results
- Extra analysis in connection with new detectors  
should prepare to purchase O(20 PC's) in 2001  
(assuming positive experience)  
need additional switch to connect PC's  
similar upgrade again in 2002, 2003

# Tape Space

- Governed by RAW, RDST, MC (+ MDST backup, Calibration)
- assume: keep RDST 1 year, delete RAW data after third year
- Typical increase  $O(15 \text{ TB/year})$  for next years



# Interactive Analysis

## Environment

- Reconstruction & batch analysis work on a network & file system architecture which is scalable & state of the art
- In contrast: environment for interactive data analysis
  - mostly n-tuples under PAW
  - very small AFS home directory + O(1.5GB/user) of “local disks”
  - local disks spread over many workgroup servers
  - unsatisfactory in terms of **maintenance, backup, scalability**
  - increasing number of “personal desktop workstations” (LINUX)
  - desktop systems are **poorly integrated**
  - WNT desktops hardly used for analysis
- Recent developments:
  - improved DESY network infrastructure (Gigabit ethernet backbone)
  - 100 Mbit network cards become standard on PCs
  - prices for disk space go down

# Interactive Analysis Environment (cont'd)

- Why not have a **centralized Large File Server** for interactive analysis?
  - **uniform access** from all platforms
  - lots of space (5 GB/user, 100 GB peak?)
  - flexible allocation of file space, independent of point where disk is mounted
  - eliminate border between **desktop PC & central workgroup server PC**
  - centralized maintenance & backup
  - if possible, integrated with AFS home directories

# Interactive Analysis Environment (cont'd)

- several models for DESY group under discussion

first experience from systematic tests with AFS

test system based on new network infrastructure being implemented

investigate performance with realistic analysis applications

bandwidth, AFS performance, benefits from caching

Decide on a solution soon

extendable to larger part of collaboration

# Summary

- Reconstruction of current data proceeding smoothly & in touch with data-taking
- Coming challenges for ZEUS offline computing:
  - Shutdown
    - Reprocessing project: bring 1996-2000 data in consistent shape
  - After shutdown
    - additional detector subsystems
    - increasing data volumes
    - maintenance of an ever-increasing system
- Detailed upgrade plans for the coming years
- Plans for centrally-maintained Large File Server