GEM TN-91-000004

GEM Computing Meeting SSCL

July 19, 1991

Abstract:

Transparencies and Minutes of First Computing Meeting held at SSCL on July 19, 1991

7/25/91-KMcF

Minutes of First meeting of Computing Group -- July 19, 1991 at SSCL

Please send any corrections to K. McFarlane (MCFARLAN@SSCVX1) (Some transparencies are available which give more details.)

GEM participants present (apologies for any omissions): N. Akchurin (UIowa), K. Arisaka (UCLA), M. Atiya (BNL), T. Bolton (Columbia-Nevis), J. Branson (UCSD), L. Cormell (SSCL), O. Fackler (LLNL), G. Fanourakis (Rochester), P. Haridas (MIT), H. Kasha (Yale), D. Marlow (Princeton), K. McFarlane (SSCL/Temple), M. Mohammadi (SUNY Stony Brook), B. Moore (Umiss), H. Newman (CAlTech), J. Reidy (UMiss), J. Ritchie (UT Austin), L. Roberts (SSCL), T. Wenaus (LLNL), G. Yost (SSCL), R. Zhu (MIT)

Guests: B. Courtney (IBM Houston), R. Blanken (IBM Houston), C. Kung (UT Arlington) H. Johnstad (SSCL), M. Botlo (SSCL), S. Frederikson (SSCL)

1. K. McFarlane: Outline of the scope of the computing effort for GEM: This includes on-line, communications and off-line computing and interactions with other areas including Level 1 and 2 triggers. A possible organization for a computing group, including an on-site group was presented. The tasks for this group and its evolution over the period 1991 to 1994. Beginning with 2-3 people in the near term, it should evolve to 20 FTE's by 1994. It was noted that SDC currently has several SSCL people working on SDC-specific projects.

2. D. Marlow: Physics simulation for L1 trigger. Results from basic simulation of Higgs search, using BaF2 EM calorimeter. This GEANT-based simulation has been used as the starting point for other groups.

3. H. Newman: SSC Computing profile; Architecture and Dominant tasks. A profile of off-line computing evolution for GEM was presented:
   1991-1992 : Workstation farms. Individual simulations, test beam data analysis, limited data storage, 10-200 kB/s I/O; some 1 MB/s (Ethernet, Exabyte; FDDI)
   1993-1994: Fast backplane. Simulations focussed on detector design, realistic performance estimates. Managed access to 04.-1 TB disk files, tape files of several TB. Database buildup and distribution. Secure access and update. First realistic subdetector reconstructions.
   1994-1996 Centralized computing. Early version of 'final' simulation code, with data structures and links to database. Buildup of reconstruction code, global reconstruction. Large scale storage of simulated data, large files of recon. data-summary events (10-20 TB). Summary of projections by L*. Summary of L3 experience and plans. Summary of need for, and implementations of, high performance computing architectures. In particular, H. Newman emphasised the need for main- frame-like file serving architectures and very tight coupling between high-performance workstations and a file-server with the performance currently achieved by mainframes.

4. L. Cormell: SSCL Approach to support of simulation efforts. SSCL organization as it affects computing. New computer division. PDSF Phase I (in use). Phase II (planned for Jan.1992). Here there

was a discussion of the I/O limitations of the Phase II configuration. It was emphasised that PDSF was available and not yet fully used. Cormell emphasised the need for a contact person.

5. H. Johnstad: SSCL Physics software support.
Organization of software support, current and projected size of group. Relationship with CERN: SSCL is working closely with CERN, being a source of updates to CERN, as well as a distribution center to U.S. (or SSC-related) institutions for software, manuals and updates. The group maintains current versions of CERN software on various platforms (most Unix machines and VAX/VMS). This software can be obtained from the support group. Email requests to: CSL@SSCVX1 (SSCVX1::CSL or csl@sscvx1.ssc.gov).

6. S. Frederikson: SDC Computing
Summary of computing section of SDC proposal. Description of the SDC 'SHELL' program. It was pointed out that L3 and other efforts had similar systems; this will be reported on at the next meeting.

7. C. Kung (UT Arlington). The computing science perspective.
Outline of possible future evolution of computing, from a top-level view. Division of tasks between physicists, professional computer scientists and industry-created applications. UT Arlington has interest in, and capability for, working with detector collaborations. Students, post-docs and faculty are available. There was discussion of the role of comp. science people in HEP, where the software is highly evolved and not easy to depart from.

8. R. Courtney (IBM): IBM Houston efforts in SDC.
(Reporting for A. Elam, the IBM effort leader.) IBM is establishing a 'HEPLAB' in Houston with a mix of machines and a network configuration much like one proposed for L3. This includes Intel i860 machines, disk array, Ampex high-speed tapes system, etc., linked by Ultranet, FDDI and a HIPPI switch. It is doing this work in cooperation with SDC. IBM is providing the capital funding for the lab. and at least 4 FTEs. SDC is giving a token amount of funding. IBM is building up some experience in HEP software and a library of relevant documents. GEM should tap in to this effort.

9. M. Botlo. Simulation of front-end, L1, L2 and L3 input.
This is work on simulating data flow in the data-acquisition using MODSIM (an object-oriented language for discrete simulations) and graphical user interfaces. Questions like the best L1 gating strategy and the best depth and sequence of buffering for data are being explored.

10. H. Newman. High performance computing projects at CERN/L3
The SHIFT project was described. A variety of workstations and mainframes are linked by an Ultranet hub and software has been written to manage the system in a machine independent way. The need for high-performance networking was emphasised, along with the opinion that it is not too soon to start planning for the time when simulations will be dealing with large amounts of data and need better performance than can be given by workstations using Ethernet or FDDI. Also, better recording technology will be necessary.

11. A discussion followed of a number of topics:
a) The overlap with the trigger and data-acquisition group. This would certainly include the L3 area, and L1 and L2 in the area of managing configuration files and integrating L1, L2 and L3 simulations into a full detector simulation. A result of this discussion was to ensure that the two groups would coordinate their meetings to avoid conflicts on common issues.
b) Languages One v. two or more?
c) Code development and maintenance. PATCHY will be needed to use the CERN software. But for other software, there are several choices. The lab. will have CMZ available, and SCCS is one component of Unix software. This issue will be discussed at the next meeting. In the meantime, K. McF will investigate SCCS.
d) A draft outline for the GEM proposal was circulated. A section for the LOI was discussed.
e) Forms for signing up for SSCVX1 and PDSF accounts were made available. If in doubt fax them to K. McFarlane at 214-708-0006.

11. The time for the next meeting was set to be Sept. 4th 1991, the day before the next Council meeting. The agenda will include: a) LoI section b) language c) Code development/maintenance environment, tools. d) Input to SSCL on PDSF.

COMPUTING FOR GEM

K. McFarlane

July 18, 1991
+19

1. Computing Tasks

- Support preparation of LoI
- Support preparation of proposal
  - section on plans, schedules, costs
- Design and construction of detector
- Run detector
- Make data accessible
- Obtain physics results

Report of the SSC Computer Planning committee - L. Price, ed.
-- (inc. H. Newman) -- Jan 1990
Report of the Task Force on Computing for the SSC
-- M. Gilchriese, ed. Dec. 1988

2A

Near term Activities

- Preparation of LoI
  - simulations largely ad hoc ; approaches used do not commit collaboration
  - brief section on plans and needs, (for proposal preparation), organization
- Preparation of Proposal
  - computing done with eye to longer-term solutions
  - Detailed section on approaches to simulation, on- and off-line computing,
- Formation of working groups
- Central support group
- Working group meeting 7/19/91

(agenda)

Difficulty of projecting needs, costs:

Computing Projections in L*

- Much less definite than for apparatus *
  - costs and needs only estimated to factor of 10
- Need to make better estimate of costs and schedules and understand contingencies
- Areas of uncertainty
  - Simulation demands
  - Detector data rate
  - L3 farm requirements
  - Networking demands
  - Computer performance and cost evolution
- Research needed
  - Data storage and access
  - L3 farm
  - Networking
  - Software development
  - Parallel/distributed computing

* H. Newman comment: was $12M, descoped to $10M so actually 2 sig. figs.

Computing →

Total Cost L* (k$)
89747.0
166419.1
59578.5
71257.7
52548.6
27687.8
10000.0
477238.7
98792.1
166419.1
59578.5
71257.7
52548.6
27687.8
10000.0
486283.8

Computing Functions

- Communications
- Simulation
  - Physics
  - Detector response
  - Trigger
- On-line
  - L1, L2, L3
  - Simulation of on-line data flow/filtering
- Off-line
  - Staged processing (Pass 1,...)
  - Selection, analysis, visualization

(contd.)

Computing Functions (contd)

- Data management
  - Physics data
  - calibrations
- Document management
- Engineering
  - Mechanical
  - Electronic
  - Information systems

In all these areas, there are choices that need to be made.

People

300-500 at more than 50 institutions:

U.S.
India
Germany
Romania
China
USSR
Japan
Brazil

- Need communications
- Distributed computing resources
- Agreed standards
- Training

- ♦ Form linked working groups
- ♦ Central (?) support group

Communications

- Email -- 1 kBs^{-1}
- News/bulletin board -- 1 kBs^{-1}
- Enhanced mail for work groups -- 10 kBs^{-1}
- Code, document transfer -- 10 - 100 kBs^{-1}
  (10MB in 20 mins)
  Video conferencing
- Remote graphics -- 1 MBs^{-1}
- Data transfer -- 10-100 MBs^{-1}

Ideally, the combination of computing resources (CPU power, data storage, communications) should be such that all users are 'local' in the computing space. All the data, all the power is as accessible to a (distant) user as it is to a local user.

Computing Needs -- Simulations

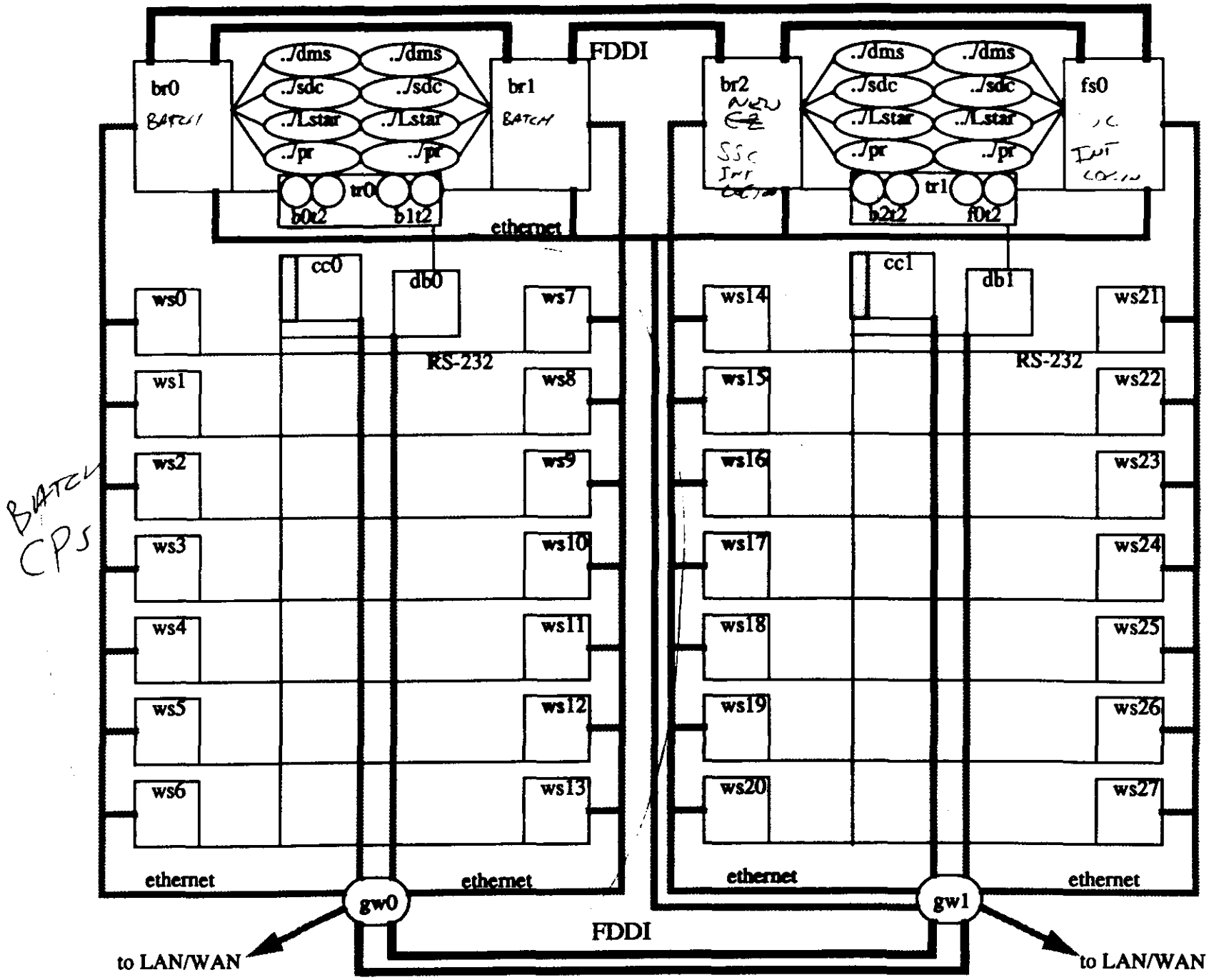
Use SSCUPs -- SSC Units of Processing power. 1 SSCUP = 0.9 VUP = 1 MVUP, on event generators.

- 10^5 parameterized events at 400 SSCUP-sec per event. One-day turnaround means 400 SSCUPs.
- 10^4 full simulation events at 20,000 SSCUP-sec per event (L3: 6 mins on VAX8650 x 10). One-day turnaround means 2,000 SSCUPs.
- 10^6 parameterized events at 400 SSCUP-sec per event. Three-day turnaround with 2,000 SSCUPs.

Workstations now are 20-30 SSCUPs.
SSCL has 1,000 SSCUPs in PDSF,
3,000 by Jan. 1992

This does not include pileup, full shower for all events, overlap of events.....
Not all power available for one job.

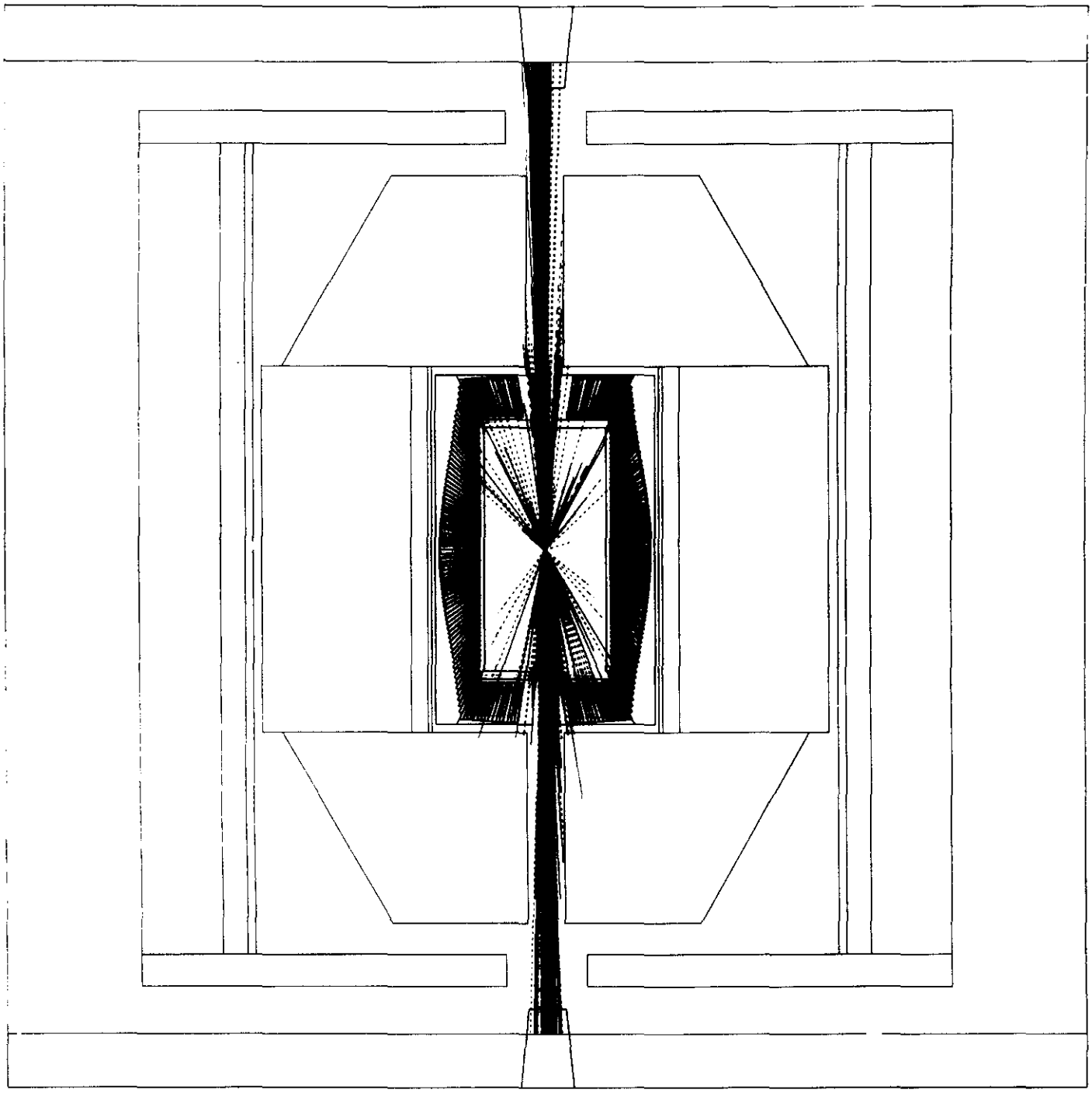
PDSF Reconfiguration for SYSDEV M.D. Marquez 7/8/91



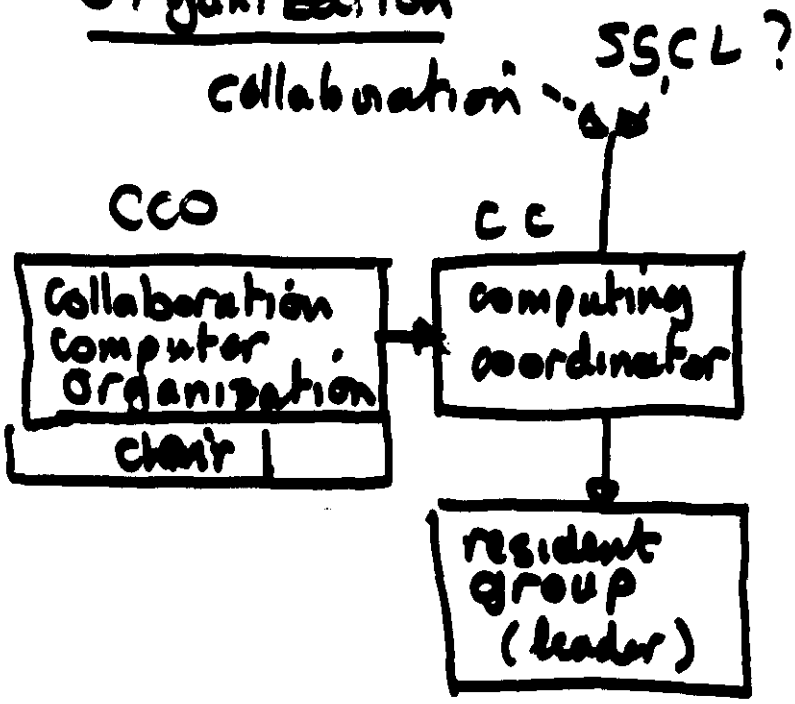
PDSF - exists, in use

10

Min bias event in "L*"
on PDSF



Organization



Resident group

1991 Coordinator, 2? programmers, secy visitor

1992 Add people in data-base, off line on line, communications (8 FTE?)

1993 20 FTE's ?

Functions of resident group

All in consultation with CCO

- Implement collaboration policies
- Coordinate effort with SSCL, SDC
- Coordinate research efforts
- Monitor the (distributed) collaboration effort as to cost and schedule
- Develop, maintain 'environment'
- Maintain and distribute code
- Provide consultation, training, organize workshops
- Maintain local facilities for collaboration
- Evaluate Hardware
  Applications
- Maintain data base
- Networking for collaboration
- Develop code
- Support LoI, proposal preparation.

The ratings on various benchmarks are given in the following table.

Table 1: Ratings on various benchmarks of several computer systems.

System	MIPS	DP Mflops	SPECmarks	std. SSCUPs	opt. SSCUPs
DEC VAX-11/780	1.0	0.14	1.0	1.1	–
DEC VAX 6420 (1 CPU)	7.5	1.18	7	7.0	8.2
SGI 4D/380 (1 CPU)	30	5	18.5	13.5	21.6
SGI 4D/35S	33	6	23.0	15.3	–
IBM RS6000/530	33	11.3	28.6	14.3	21.3
IBM RS6000/520	29.5	8.5	22.0	11.9	–
Sun Sparc 2	28.5	4.2	21.0	11.8	16.3
DECstation 3100	14	1.6	10.8	6.4	10.0
DECstation 5000	24	3.7	18.5	10.5	–
Apollo DN10000 (1 CPU)	22	6	18.6	11.6	17.4
HP Apollo 9000/720	57	17	55.5	22.7	29.4 *
HP Apollo 9000/730	76	22	72.2	–	–

As can be seen the performance on physics codes, given the restrictions, is not accurately predicted by the other quantities across the architectures.

* ≈ $0.4k / SSCUPs

Benchmarks from SSCL-37 + supplement

Table 5: CERN benchmark results for various machines (from CERN)

default

Computer	CERN units
Cray YMP 6ns	11
HP9000/720	10.5
DEC9000	9.1
IBM3090/400E	6.6
SGI 4D/310	6.2
IBM RS6000/530	5.9
HP/Apollo 10000	4.9
IBM RS6000/320	4.7
DEC 5000/320	4.4
SGI 4D/25	3.5
DEC 3100	2.8

Aut

SSCUPS based on:

ISAJET	OPT	STATIC	OPT
PYTHIA	NOOPT	AUTO – default	OPT
JETSET	"	"	"

geometric mean.

Table 5: CERN benchmark results for various machines (from CERN)

Computer	CERN units
Cray YMP 6ns	11
HP9000/720	10.5
DEC9000	9.1
IBM3090/400E	6.6
SGI 4D/310	6.2
IBM RS6000/530	5.9
HP/Apollo 10000	4.9
IBM RS6000/320	4.7
DEC 5000/320	4.4
SGI 4D/25	3.5
DEC 3100	2.8

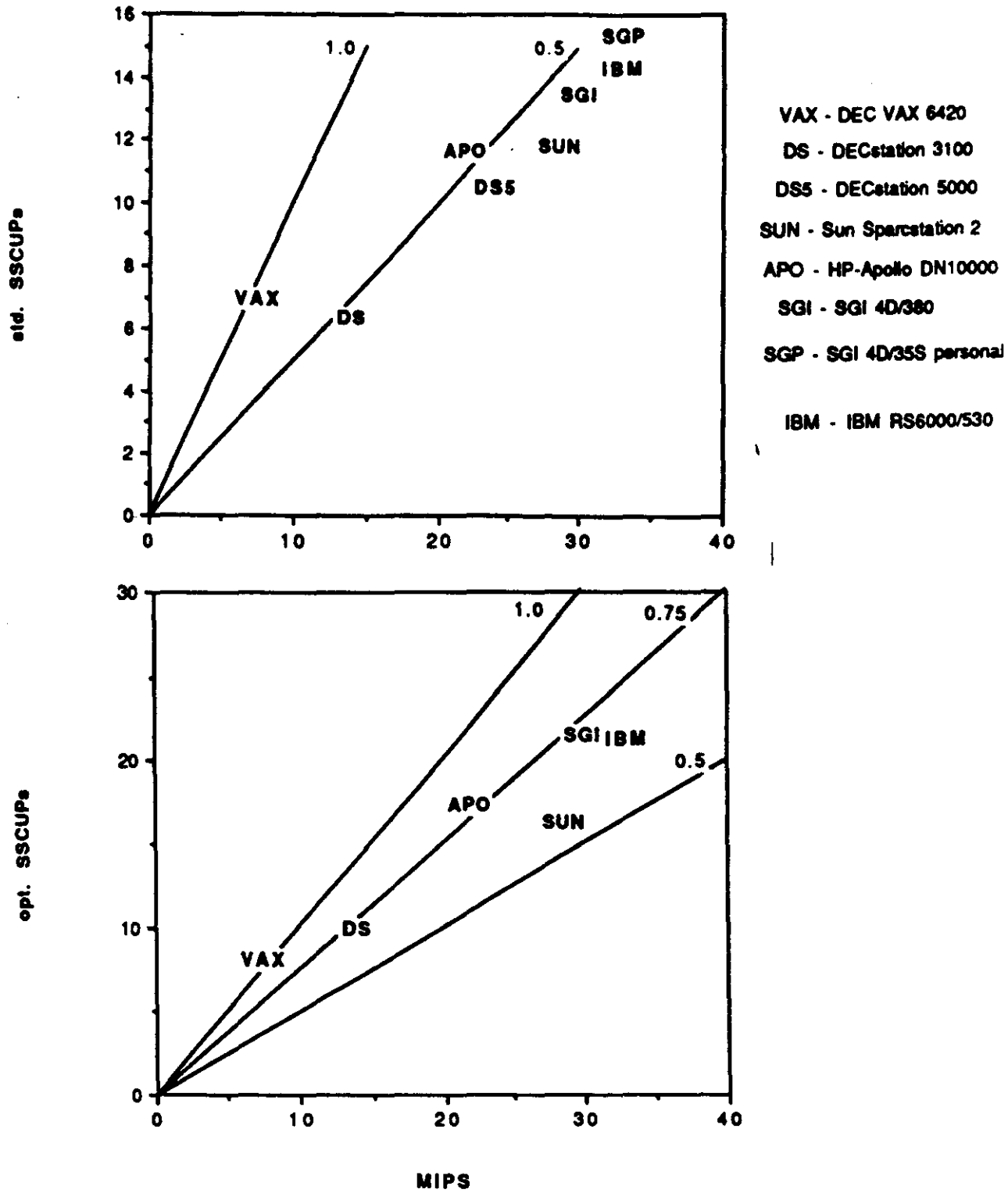


Figure 1: Rating in std. (a) and opt. (b) SSCUPs for several systems, as a function of MIPS. Systems are identified in the key. Lines of labelled slope are to guide the eye.

H. Newman

SSC Computing Profile

Architecture & Dominant Tasks

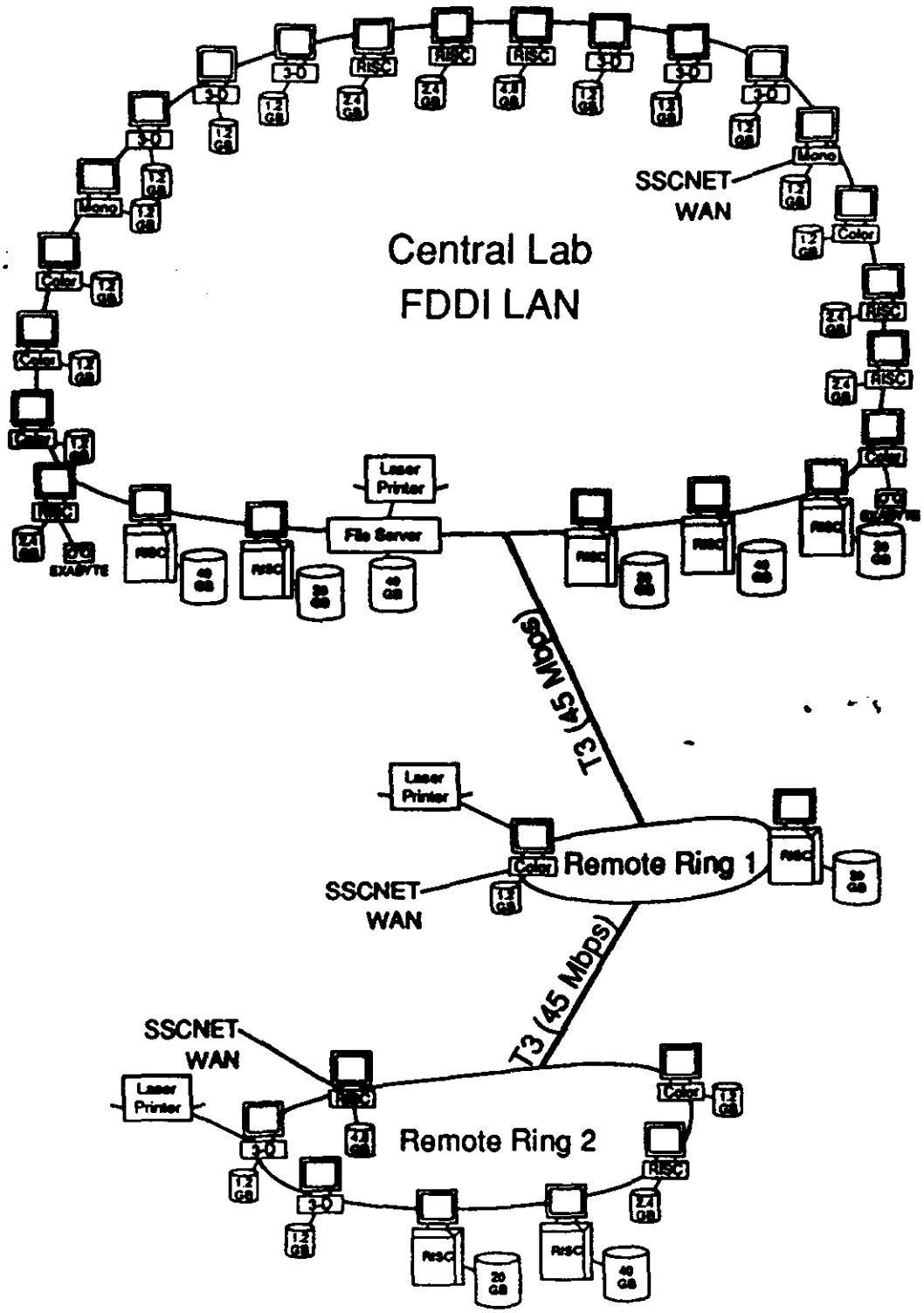
1. 1991-1992 Workstation Farms
   - Individual Simulations; Test Beam Analysis
   - Limited Data Storage
   - 10-200 kByte/sec I/O; some 1 MB/sec (Ethernet; Exabyte; some FDDI)

2. 1993-1994 Fast Backplane
   - Simulations focused on Detector Design; Realistic Performance Estimates
   - Managed Access to 0.4 → 1 TB Disk Files; → Several TB Tape Files...(?)
   - Database: Buildup, Distribution, Secure Access & Update
   - **First Realistic Subdetector Reconstructions**

3. 1995-1996 **Central?**
   Early Version of 'Final' Simulation Code
   - With Datastructures & Links to Database
   - Buildup of Reconstruction Code:
     - Global Reconstruction

   ⇒ Large Scale Storage of Simulated Data + Large Files of Reconstructed Data-Summary Events: 10-20 TB?

SSC Lab Workstation Environment
(Ca. 1991 - 1992)



→ $\mathcal{L}^*$ COMPUTING NEEDS ESTIMATE

DETECTOR SIMULATION & ANALYSIS

	3/91	12/91	12/92
INTERACTIVE WS	(250) 600	30 × 20 VUPS = 600 VUPS	1000 VUPS
MEMORY/CPU	(8) 16	16→32 (INT. WS) 32 (BATCH)	32 ≥ 32
BATCH RANCH CPU (VUPS)	(250) 480	1500	2500
USER DISK (GB)	(25) 42	60	100
EVENT DISK (GB)	(25) 40	60	250
BULK STORAGE (TB)	(.75) .25	1	6-10

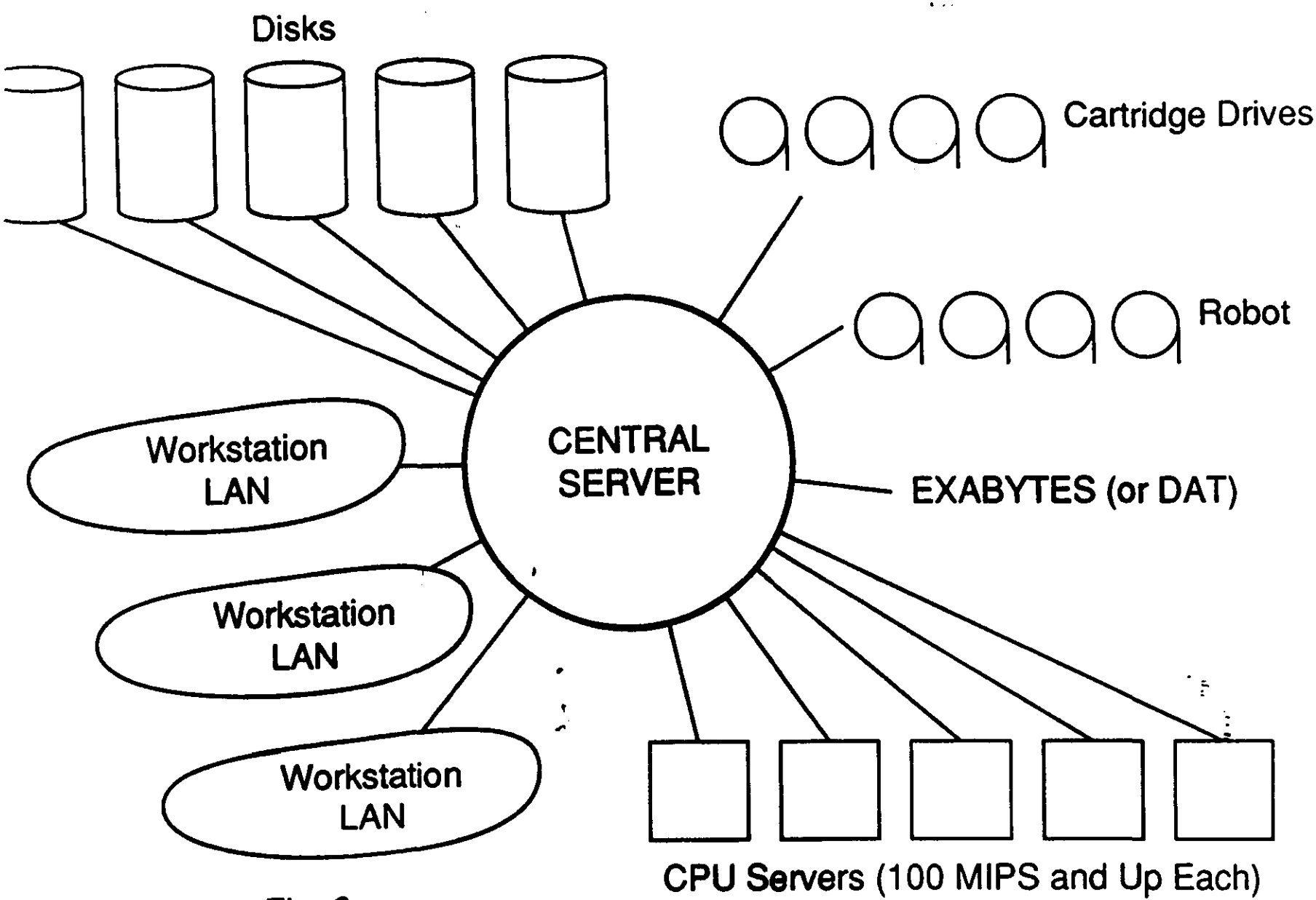
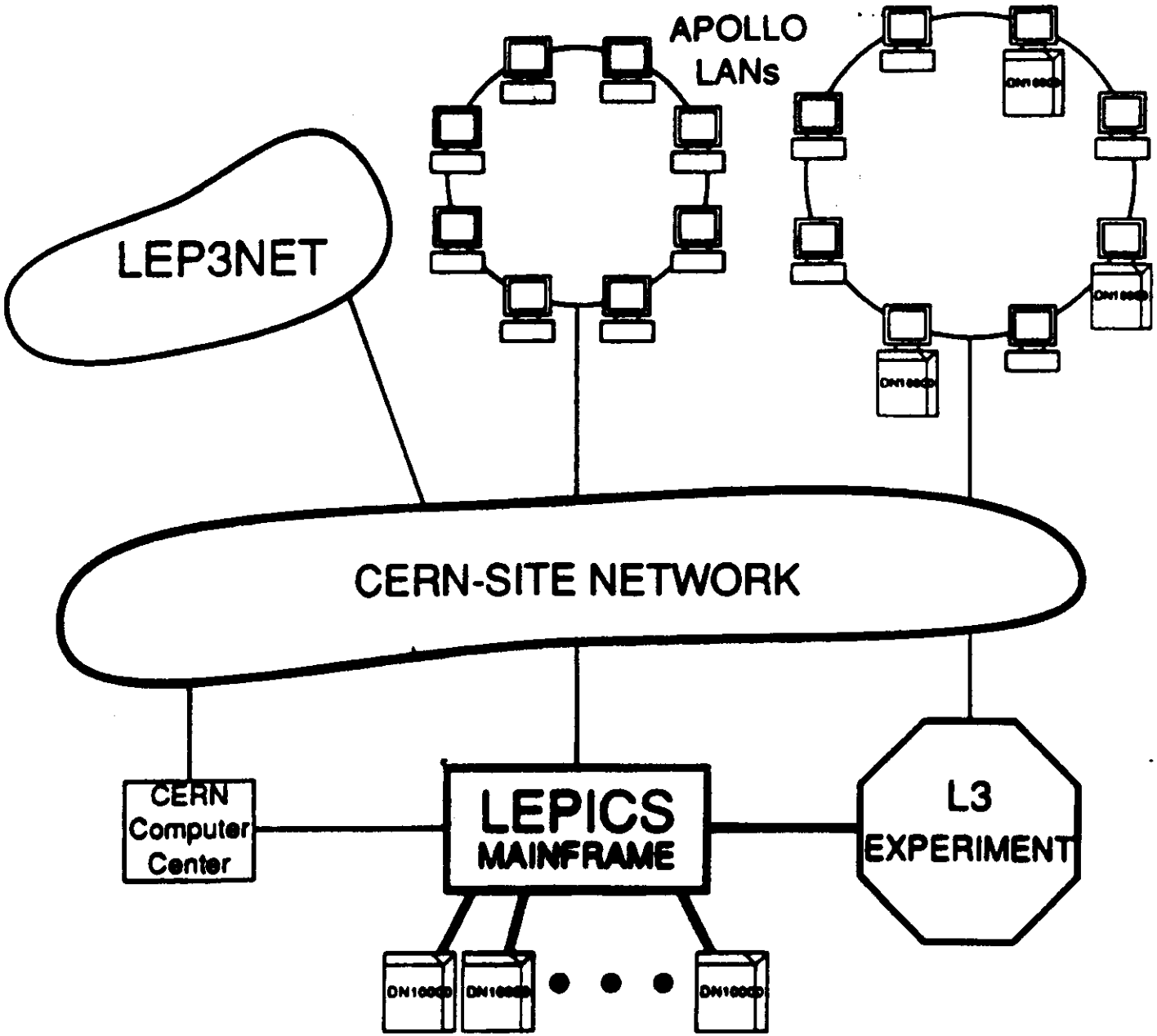

Disks
Cartridge Drives
Robot
CENTRAL SERVER
EXABYTES (or DAT)
Workstation LAN
Workstation LAN
Workstation LAN
CPU Servers (100 MIPS and Up Each)

Fig. 2

Summary of Requirements of the L3 Data Analysis Task

- An integrated local and wide area computing system.
- Everybody should have good access to 40 Terabytes of data.
- 500 to 1000 MIPS are needed, much of it with high-speed access to the data.
- The user interface (especially graphics) should be the best offered by today's workstations.



L3 Offline Computing at CERN

L3 Apollo Network at CERN

- ~~100~~ ≈120 nodes:

  22 DN10000s (36 CPUs).

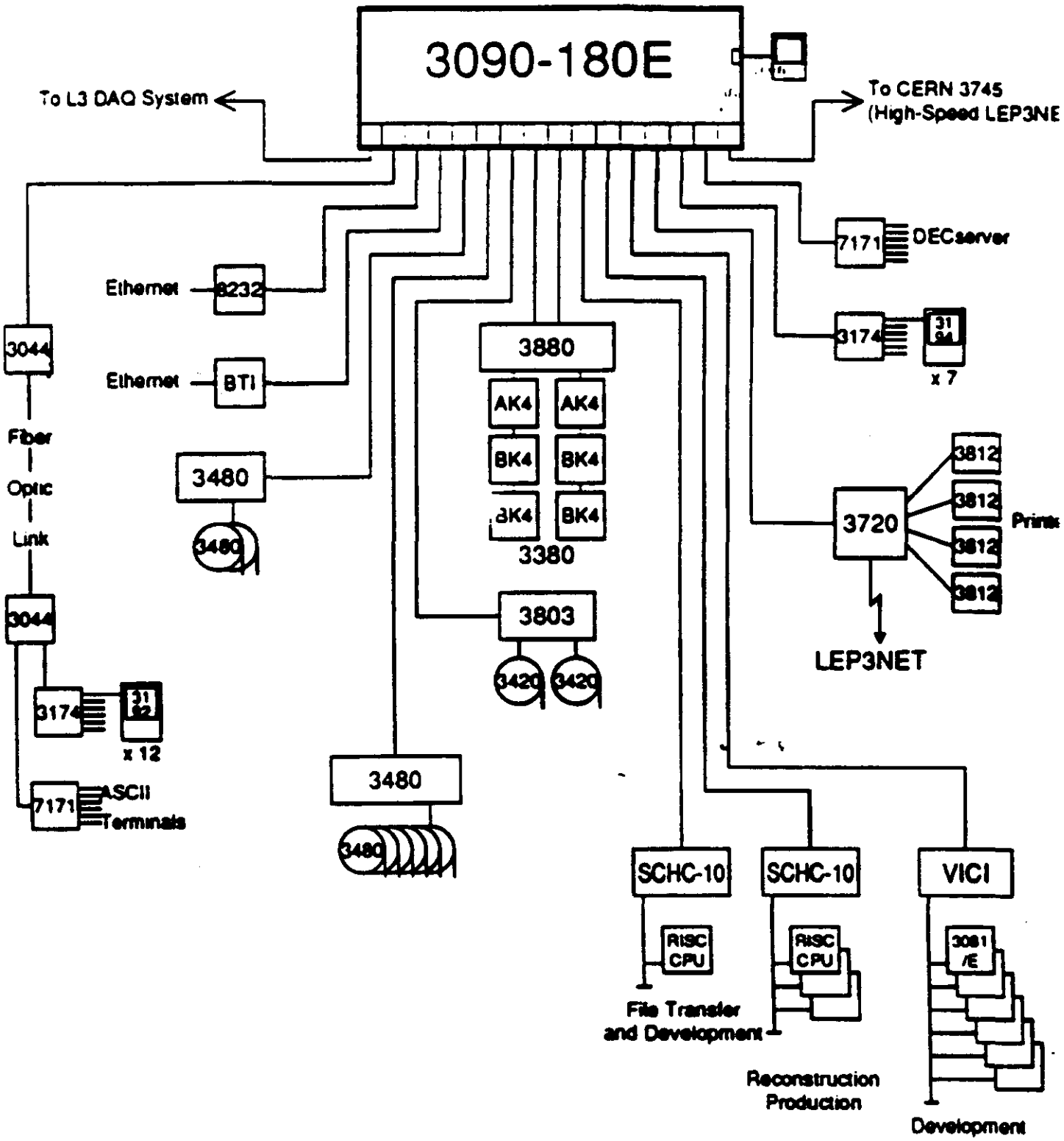
- Over 60 GB disk.

- Optimum topology: one ring.

  (Each user needs access to data stored on all other nodes in the ring.)

- Massive (factor > 100) insufficiency of Domain ring for file access.

- FDDI also expected to be insufficient.



LEPICS Configuration

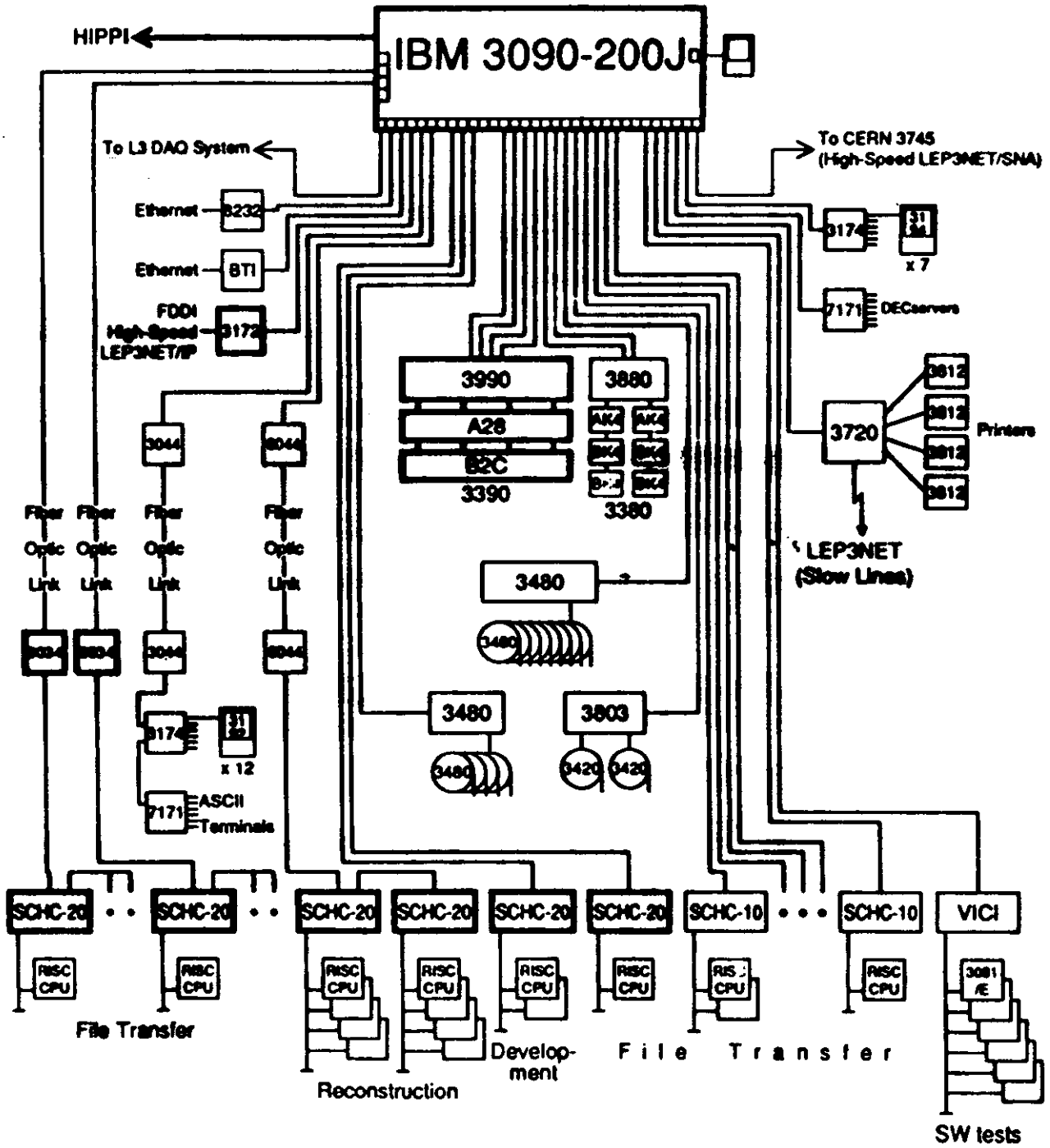
LEPICS

- Upgrade to 3090-200J scheduled for July 15, 1991 (Dependent on contract signature by MIT.)

- Projects with IBM and HP bring:
  - Financial and technical benefits for L3,
  - Technical advances for future HEP data analysis.

- Upgraded system is well matched to the needs of the 1991 data analysis.

- Data volumes in 1992 onwards are expected to require additional peripherals and attached processors.


HIPPI
IBM 3090-200J
To L3 DAQ System
To CERN 3745
(High-Speed LEP3NET/SNA)
Ethernet
Ethernet
BTI
FDDI
High-Speed
LEP3NET/IP
3174
x 7
7171
DECservers
3990
3880
A26
B2C
3390
3380
3720
Printers
LEP3NET
(Slow Lines)
Fiber Optic Link
3480
3803
x 12
ASCII
Terminals
SCHC-20
SCHC-10
VICI
RISC CPU
File Transfer
Reconstruction
Develop-
ment
File Transfer
SW tests

Statistics Display for Node 3A67A //13dsp10k2

Process Name	0% CPU Percentage 100%
Null Process	0.0%
Null Process	0.0%
Null Process	0.0%
Null Process	0.0%
OS processes	4.1%
init	0.0%
(receape)	90.9%
(receape)	89.9%
(receape)	90.0%
(receape)	91.0%
(sil3v160a.exe)	10.7%
(sh)	0.0%
(sh)	0.0%
(sh)	0.0%
(nqsdaemon)	0.0%
(sil3v160a.exe)	11.9%
(sh)	0.0%
(sh)	0.0%
(sh)	0.0%
(nqsdaemon)	0.0%
(p3_daemon)	0.0%
server_process_m	0.0%
(stell_fxfer)	0.0%
(netdaemon)	0.0%
(nqsdaemon)	0.0%
(logdaemon)	0.0%
mbx_helper	0.0%
(writed)	0.0%
netmon	0.0%
(cron)	0.0%
llbd	0.0%
snmpd	0.0%
(timed)	0.0%
(ruptimed)	0.0%

L3 Off-line Computing at CERN
Investment and Running Costs

INVESTMENT: Hardware

Workstations	$ 3.2 M
Mainframe (CPU/Memory/Channels)	$ 1.4 M
Mainframe Peripherals	$ 1.2 M
Miscellaneous	$ 0.2 M

INVESTMENT: Manpower

Workstation-Mainframe Communications	3-4 man years
Parallel Processing	3-4 man years
Data Management	2-3 man years

Table 2.1. Summary of requirements for physics and detector simulation studies.

Job (Run) Type	CPU Requirement (SSCUPS-seconds/event)	Events/study	Total Time (SSCUPS-seconds)	Event Length (kbytes)	Total Output /study (Gbytes)
Test and Debug	50-100	10-20	1000-2000	300-500	5-10
Event Generation	5-10	10^5-10^6	10^6-10^7	10-30	1-10
Minimal Simulation	50-100	10^5-10^6	10^7-10^8	300-500	100-500
Full Simulation	10^3-5×10^4	10^4-10^5	10^8-10^9	300-500	10-50
Data Reduction and Analysis	5-50	10^5-10^6	10^6-10^7	300-500	50-1000†

†Input data rate.

Table 2.2. Model of Peak User Requirements.

Activity	Job Requests (Simultaneous jobs)	CPU Requirements (SSCUPS) per job	I/O Traffic (MB/sec) total*	Storage Requirements (GB/day) total*
Misc. Interactive	50†	10-15	1-5	1-10
Compile and Debug (mostly interactive)	50†	10-15	5-10	10-20
Event generation (batch)	4	200	1-5	50-100
Minimal Simulation (batch)	10	200	10-20	200-500
Full Simulation (batch)	4	400	1-2	50-100
Data Reduction and Analysis (interactive + batch)	20	20	50-100††	—
Backup	1	10	10	300-700
TOTALS interactive	100	1000-1500	10-50	10-20
batch	50	4000	50-150	1000

† Simultaneous interactive users

†† Input from storage

* Integrated over all jobs in the row category

L3 Off-line Computing at CERN
Investment and Running Costs

RUNNING COST: Maintenance/Licences

Workstations	$ 250 K/year
Mainframe	$ 310 K/year

RUNNING COST: Manpower

	'Professionals'	'Helpers'
Mainframe Systems Programmers	2	2
Workstation System Managers	2-3	2
Network Specialists	2	2
L3 Applications Support	2	2

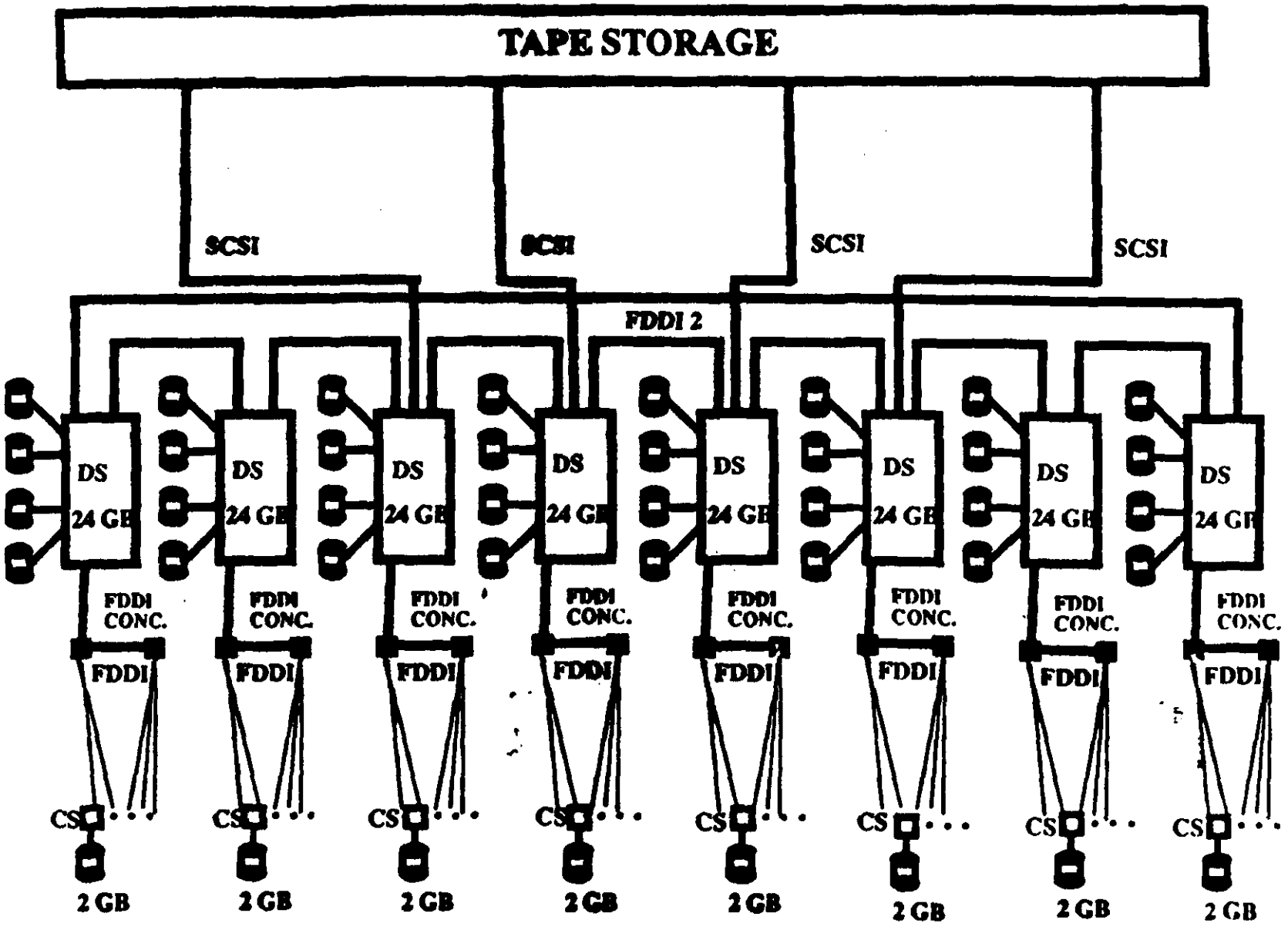


Figure 2.1. Functional diagram (block diagram) of the simulation facility.

Joint Projects

PHASE I

L3/IBM	**Parallel Processing**
L3/Apollo	**DN10000 to IBM Channel Interface**

IBM Channel - DN10000 Link Performance (using SCHC-10)

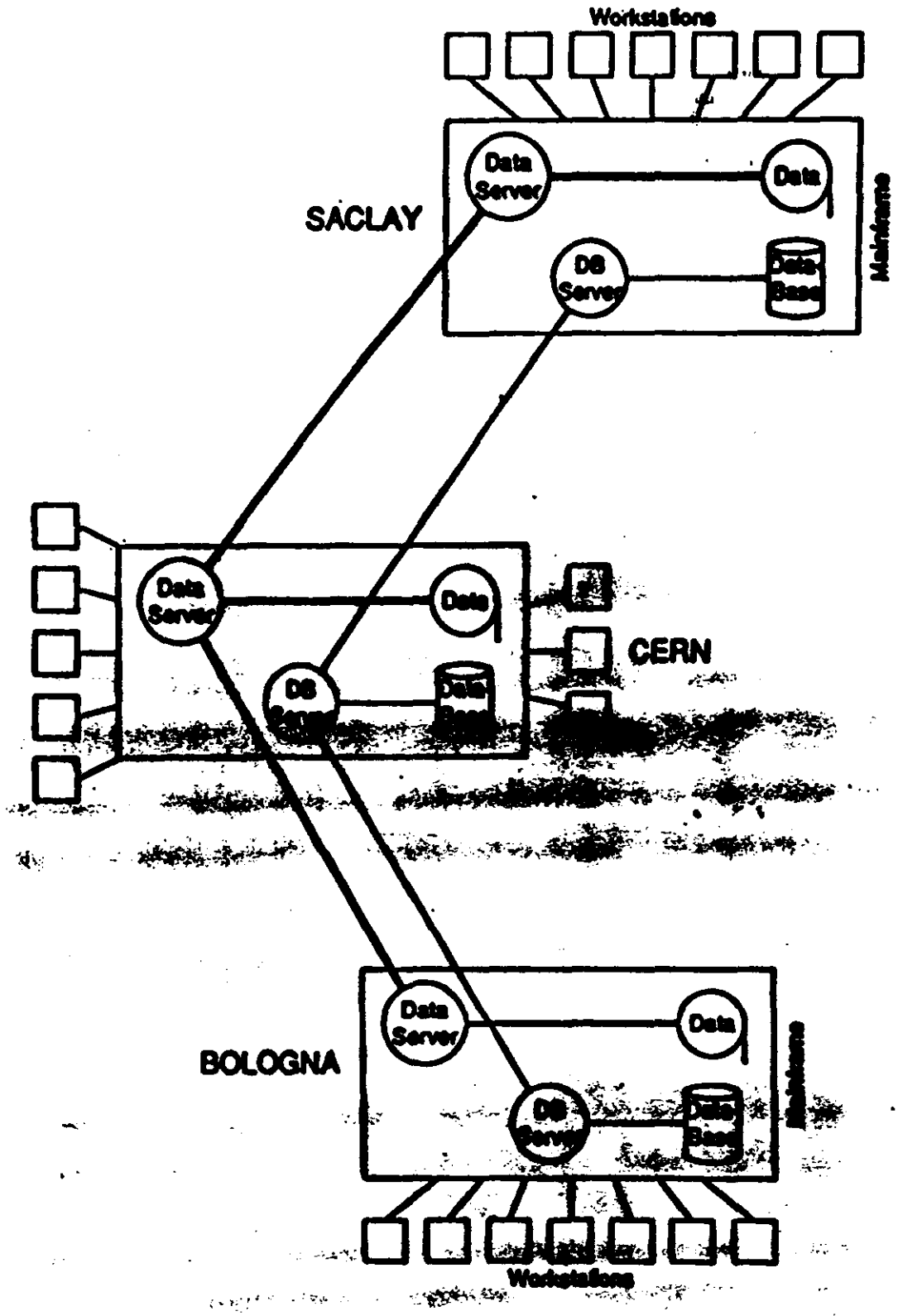
File Transfer	up to 1.1 MB/s
IBM Memory - DN10000 Memory	1.7 MB/s
IBM Memory - VME Memory	1.9 MB/s
VME Memory - DN10000 Memory	3.3 MB/s

IBM 3480 Tape Cartridges to be Written in 1990

Raw Data	11,000
Reconstructed Data	11,000
Simulated Data	6,000
Reconstructed Simulated Data	4,000
Selected Events (Physics Datasets)	6,000
TOTAL CARTRIDGES for 1990	38,000

IBM 3480 Tape Cartridges Written Each Year

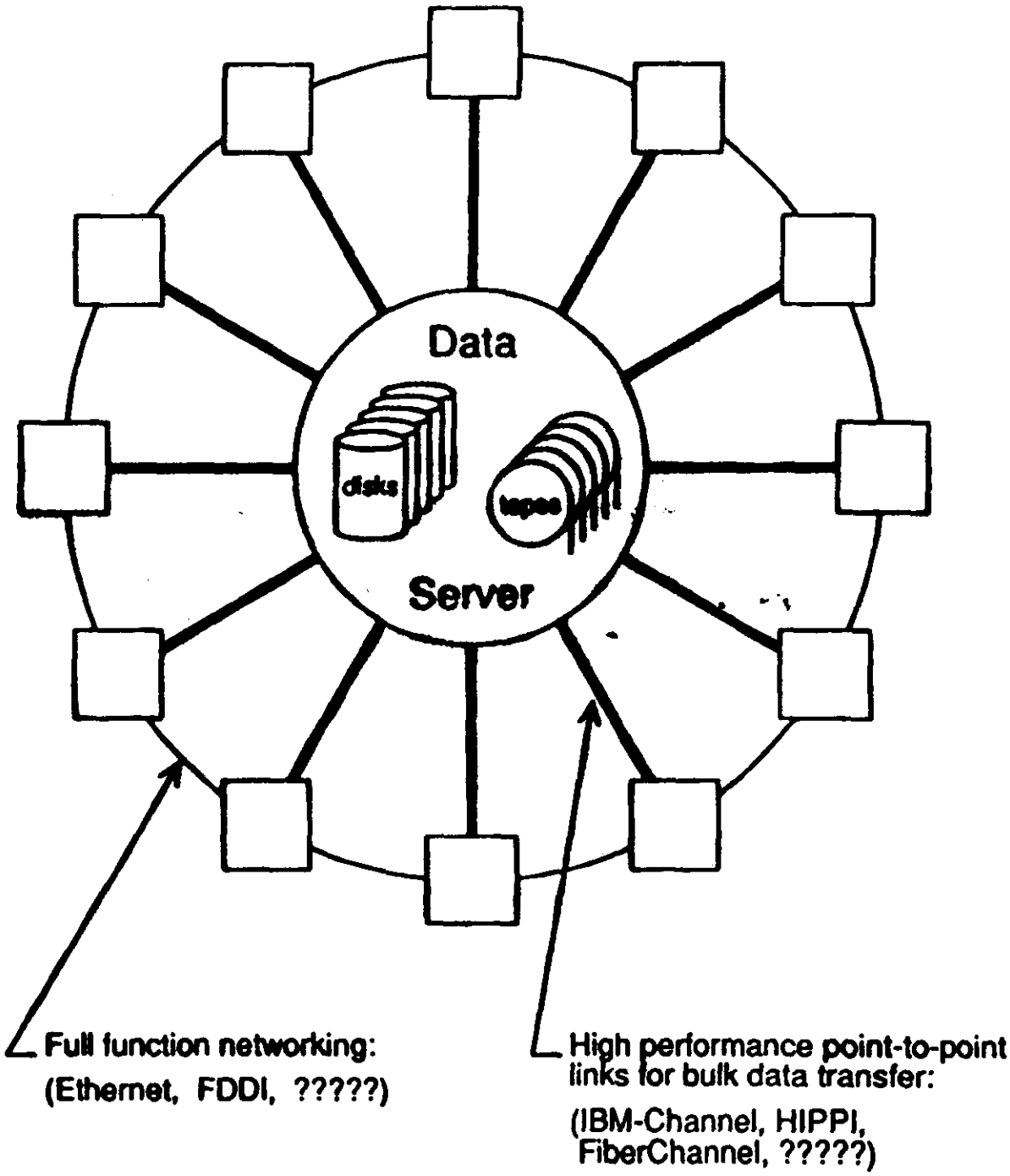
1989	4,500
1990 (Est.)	38,000
1991 (Est.)	60,000
1992 (Est.)	90,000
1993 (Est.)	100,000
TOTAL CARTRIDGES BY 1993	300,000



Example of a distributed data management and database management system for an HEP experiment.

A Model for HEP Computing

Workstations/CPU Servers and Data Server



R. P. Mount, January 1991

CERN/L3 - HP FDDI Project

Nov. 1990 HP-funded CERN Associate appointed. (Hans Lubbers)

Dec. 1990 Participation (equipment/manpower) in beta test of X-Bus boards.

Jul. 1991 Delivery of 15 X-Bus boards.

Attempt to set up intensive use of Apollo Domain over FDDI within one building.

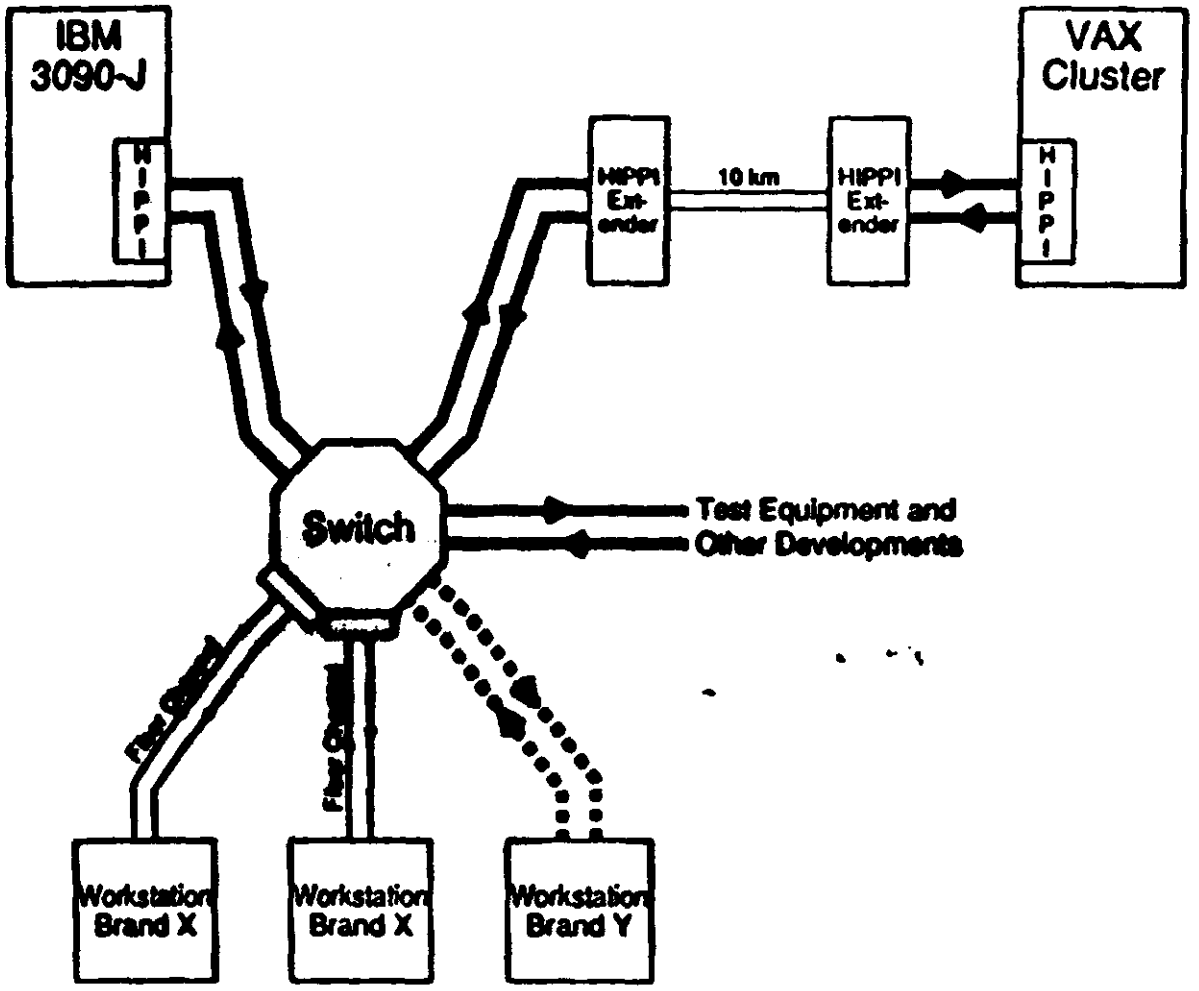
Sep. 1991 Extend FDDI ring to Building 513 (1.5 km away) for access to LEPICS

Install FDDI-FDDI IP-router for access to CERNVM IBM-3090 system.

Nov. 1991 Extend FDDI ring to:

Building 188 (2 km)
Point 2 (10 km)
Building 892 (7 km)

R. P. Mount, January 1991

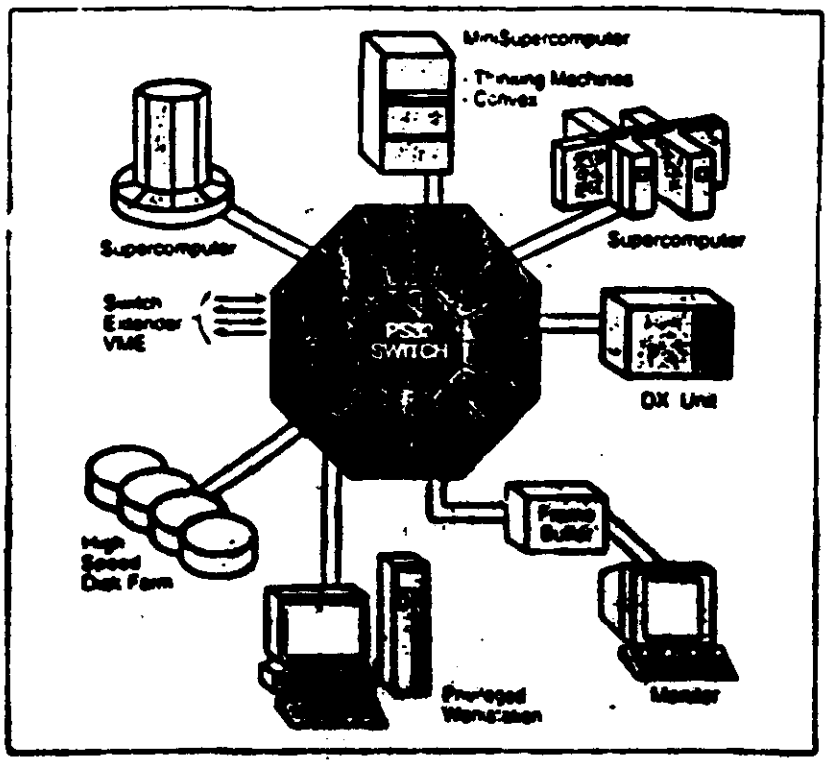


//13cit3/localuser/mount/lepics2/figs /hippi switch fig 2

PS32 HIPPI Switch

Enables Full-Bandwidth Switching of Multiple HIPPI Channels

- Complies with the HIPPI ANSI standard.
- Operates with vendor-provided device drivers which conform to the ANSI standard.
- Has an aggregate rate of up to 25.6 Gigabits per second and data latency of less than 240 nanoseconds.
- Provides up to 32 simultaneous simplex connections or 16 simultaneous full duplex connections.
- Can be connected to a DX unit, providing HIPPI users with access to FDDI, HYPERchannel, Ethernet, etc.
- Expands from 8 x 8 (8 input and 8 output ports) up to 32 x 32 in increments of 2 x 2.
- Has a camp-on feature to avoid rejection of connection requests.
- Runs self-test diagnostics at power-on or at the operator's discretion.
- Has built-in capacity for collecting operational information.



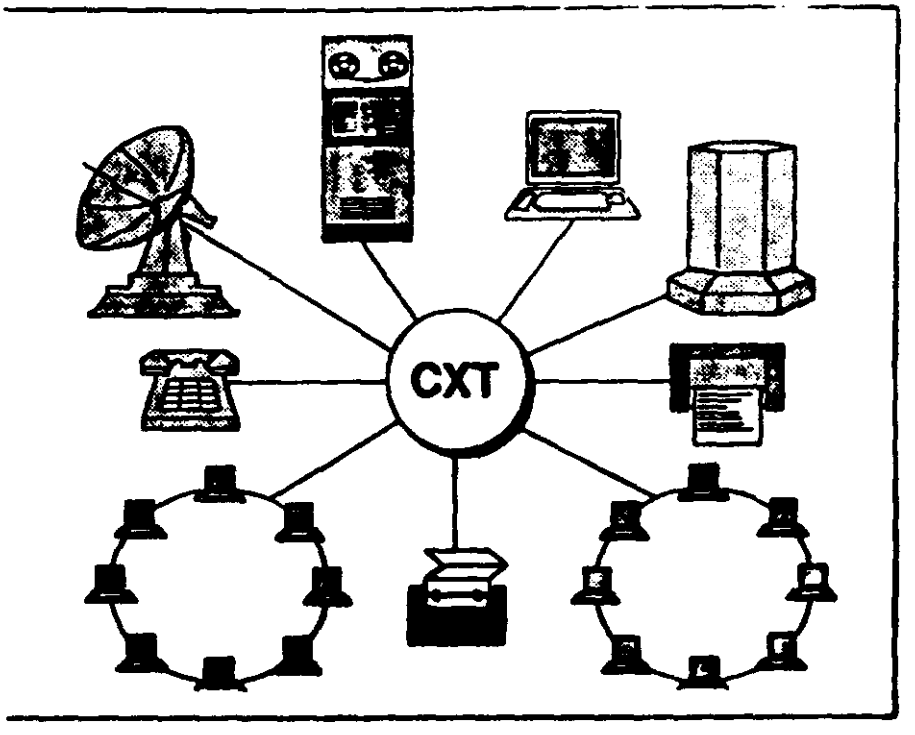
Overview

The PS32 HIPPI Switch is a high-speed switch that can be used to establish a network of channels that conform to the ANSI High Performance Parallel Interface (HIPPI) standard (X3T9.3). The switch has up to 32 HIPPI input and 32 HIPPI output ports to which the channels are attached. Any input port can be connected through the switch to any available output port. A HIPPI connection between an input and an output port forms a simplex (unidirectional) path. Up to 32 simplex paths can be established through the PS32 Switch, and all the paths can be used simultaneously.

Two simplex paths through the PS32 Switch can be used to form a full duplex (bidirectional) path. Each full duplex path requires two input and two output HIPPI ports. Up to 16 full duplex paths can be established through the PS32 Switch, and all the paths can be used simultaneously. The PS32 Switch will initially support only the single-wide (32 data bits) HIPPI channels.

ANCOR Communications, Inc.

The CXT®/1000 System



The CXT-1000: An Interconnectivity Breakthrough

Features

- 1.0625 Gbaud/channel data rate
- Meets the ANSI X3T9.3 Fiber Channel Standard
- Supports connectionless service
- Supports connection service
- Unlimited expandability without performance degradation
- Distributed architecture
- Nonblocking operation
- No single point of failure
- Dynamic circuit selection
- Individual circuit security
- Automatic power-up table construction
- Modular approach for easy service
- Remote diagnostic capability
- Modules replaceable under power
- Standard ST fiber-optic connectors

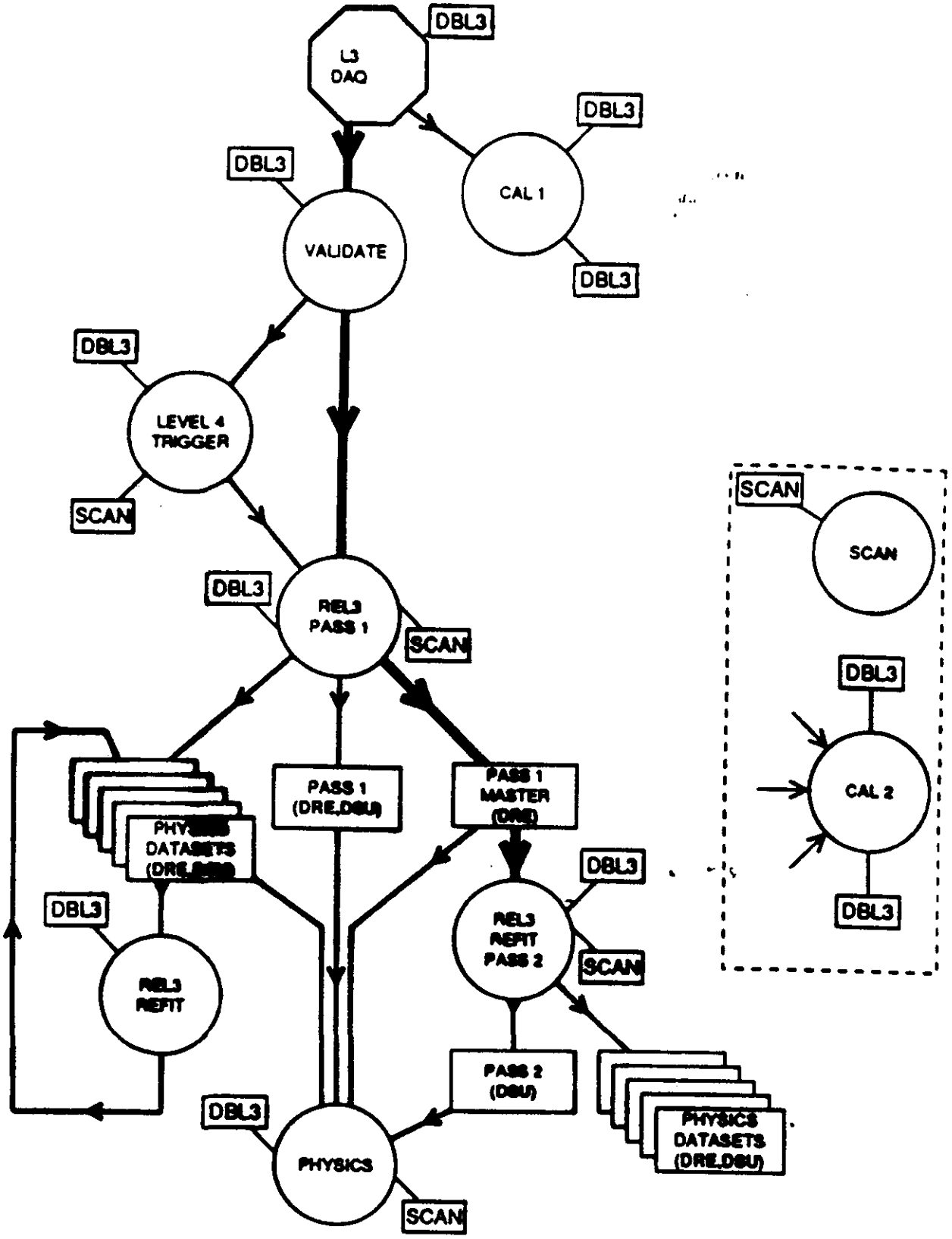
Overview

The Ancor CXT/1000 establishes a new era in high-speed electronic communication. It incorporates breakthrough space-switching technology that permits the two-dimensional interconnection of almost any type of electronic device in a single integrated communications network.

The CXT/1000 is the culmination of more than a decade of extensive network and space-switch research. It is an ANSI-standard-based connection/connectionless network, and delivers a new level of interconnectivity for any communications environment.

Ancor's advanced space-switching technology enables a CXT/1000 system to expand infinitely *without performance degradation*. The unique Ancor approach overcomes the fundamental limitation of other data communication networks, which is a fixed bandwidth that must be shared by all nodes.

The CXT/1000 system provides a bandwidth that increases in direct proportion to the number of nodes on the network. The aggregate bandwidth of a CXT system is equal to the number of nodes multiplied by the individual channel bandwidth. Because the CXT/1000 features a 1.0625 Gigabaud/ channel data rate, the aggregate system bandwidth can easily exceed a terabaud.



L3 Data Analysis Strategy

SDC Collaboration

OVERVIEW of SDC COMPUTING

Soren G. Frederiksen
SSC Laboratory

Outline

Online

- Overview
- Functional Requirements
- Relationships with other subsystems
- Online Computing Architecture
- Design philosophies and methodologies
- Online Computing Components
- Schedule
- Needed R&D

OUTLINE(cont)

Offline

- Summary
- Computing Model and Operational Concepts
- Offline Computing Requirements.
- Communications
- Software
- Hardware Design
- Organization
- Computing for SDC outside SSCL
- R&D required
- Technology Forecasts

OUTLINE(cont)

Schedule

Cost estimates

Upgrades and scalability

URGENT TOPICS

Software Development
- Code Maintenance *
- Languages
- Software Engineering
- Kernel Specification *

Online Requirements
- Architecture

Offline Requirements *
- Architecture
  - distributed versus centralized

Software Architecture

Standards

Code Maintenance:

Presently researching 3 packages

CMZ

SCCS

Codebase

CMZ and SCCS will be used to manage the SDC SHELL program, and then evaluated.

Codebase will also be studied, but no plans at present to use it with SHELL.

Kernel Specification

Kernel program for analysis

Written by computer specialist

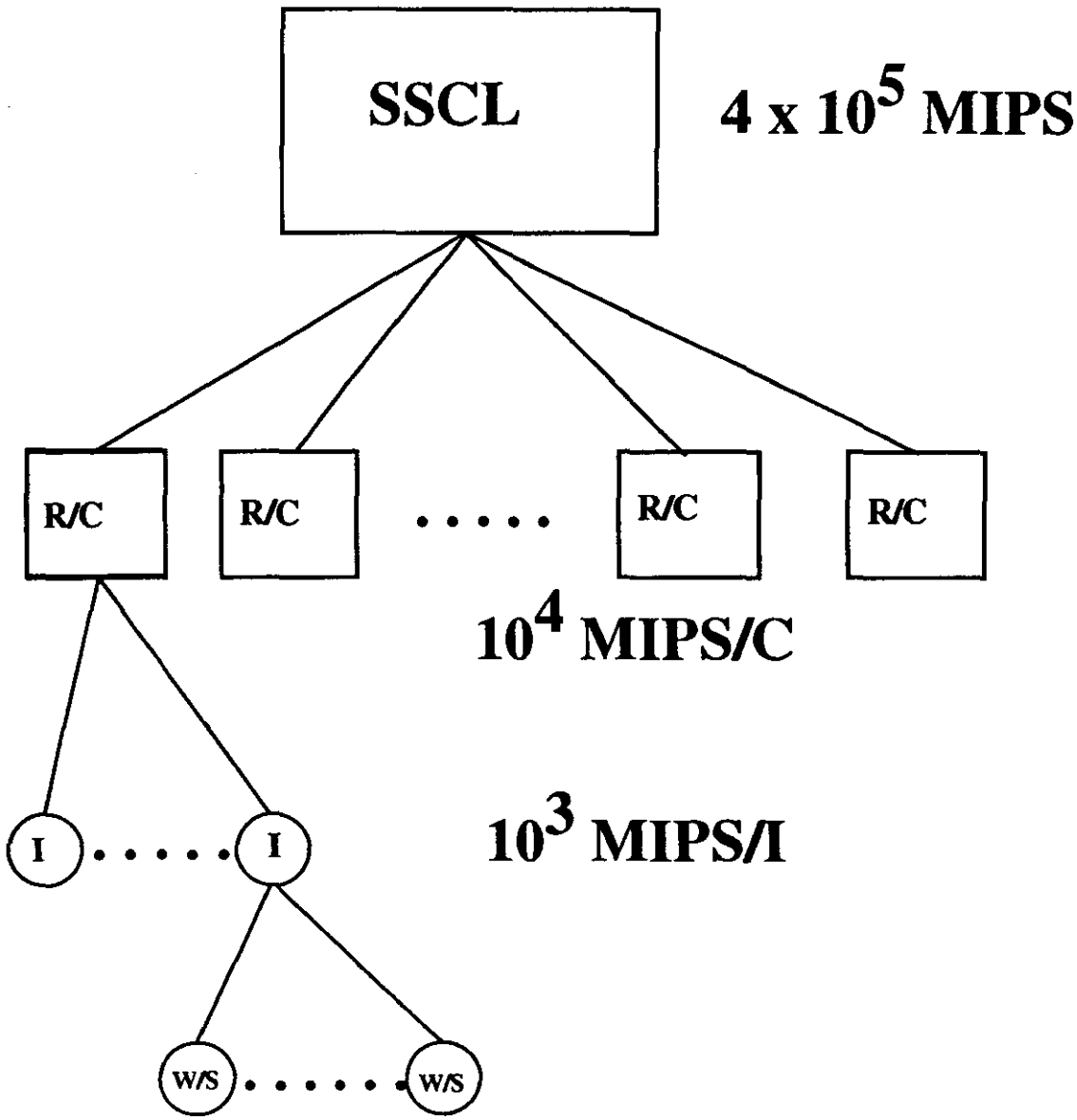
Conform to industy standards

User packages plug into Kernel, and interface only with the Kernel.

- No comunication between packages.

Offline Requirements:

Computing Model



Assumptions:

100 Hz rate into Offline

1000 MIPS-second/event for reconstruction

1 MByte event size.

10% reduced sample.

Implies:

10^5 MIPS for Reconstruction

10^5 MIPS for 2nd pass

2×10^5 MIPS for Monte Carlo

10^{15} bytes raw data/year

Architecture Studies:

Estimating event sizes based on physics processes.

Studying CDF requirements for event sizes, and computing needs.

Using P-NUT to study architecture and check of network/disk/comunication requirements.

Estimate/guess locations of regional centers.

Conclusion

SDC Computing on Track.

Always try to use industry standards whenever possible.

Do not get tied to one architecture.

Must be flexible enough to change with the changing technology.
(required because of long lead time)

Must start worrying about computing problems *now*, before wrong solution gets ingrained into the system.

SDC Collaboration

THE SDC SHELL PROGRAM

Soren G. Frederiksen
SSC Laboratory

SHELL is still in development stages.
Version 1.0 release 1 month ago.
Version 2.0 to be released next week.

SHELL behaves like a software bus

- Can insert different packages corresponding to different detector components.

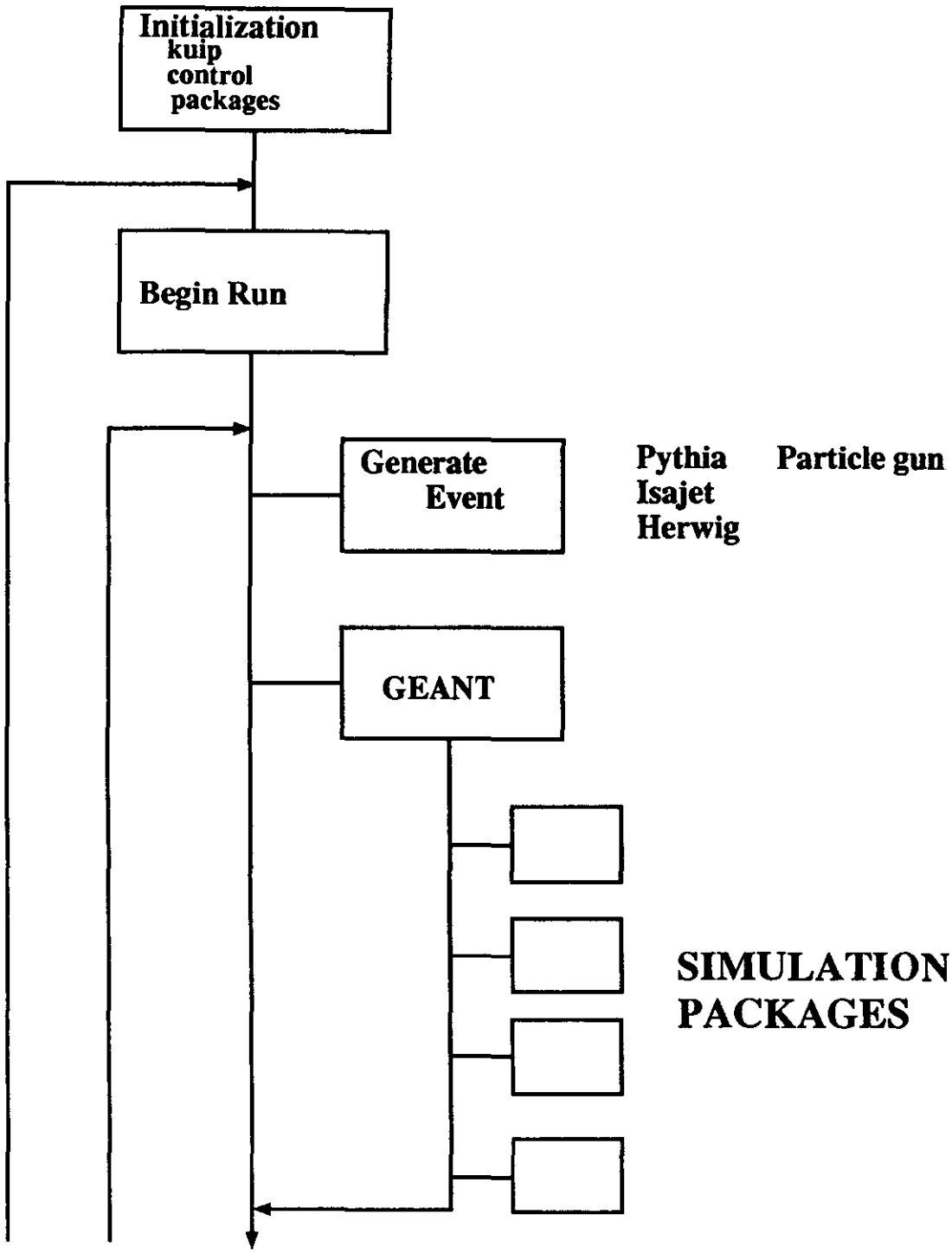
- All communication between packages is through ZEBRA banks.

SHELL maintenance is done using PATCHY

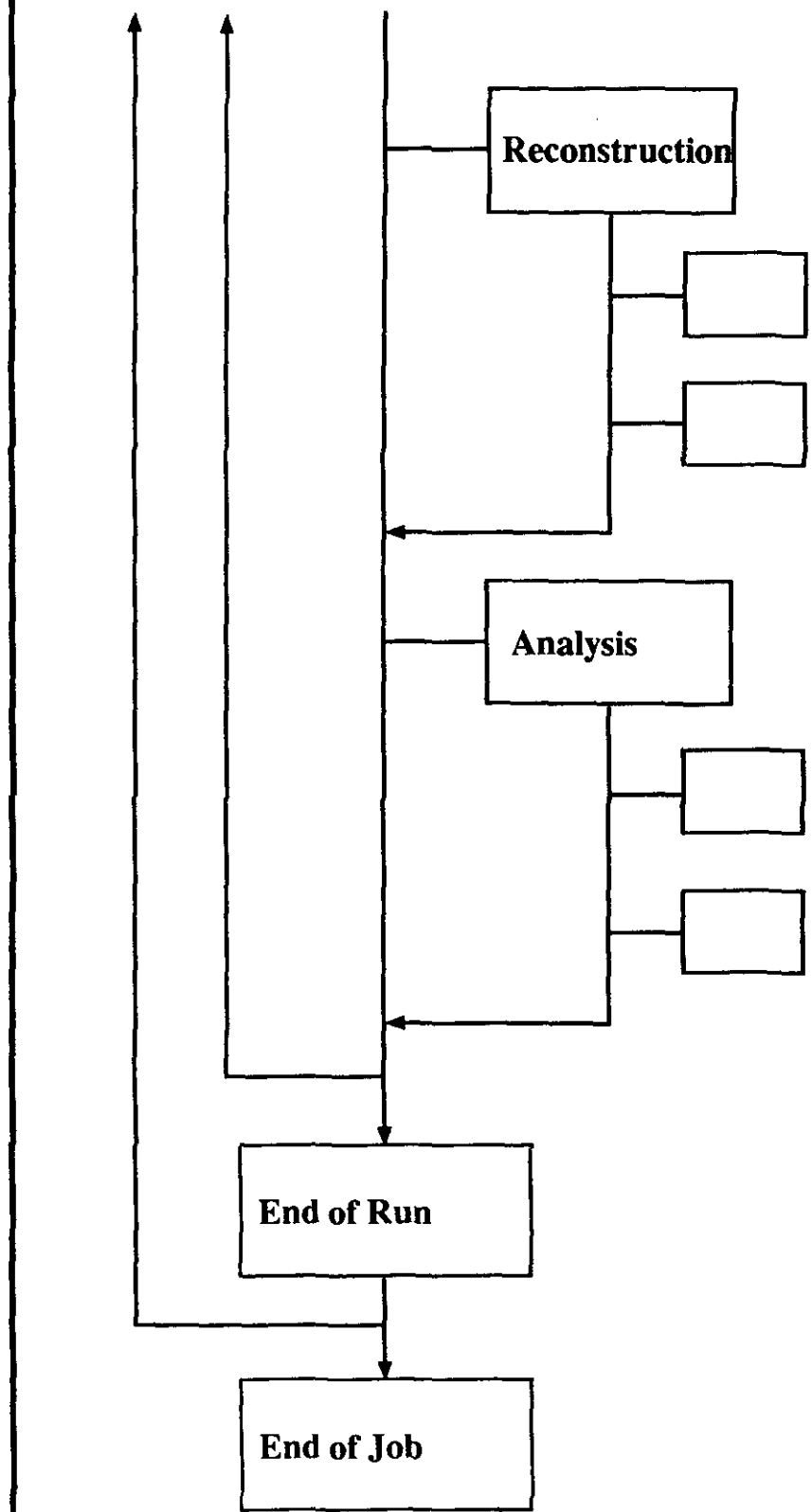
SHELL control is interactive through ASCII files

SHELL is written in standard Fortran-77, with very few (if any) extentions.
-Enables us to run on any platform

SHELL OUTLINE



SDC Collaboration



CURRENT PACKAGES

SIMULATION PACKAGES

AN	Calorimetry
CF	Fast calorimeter shower
CL	Liquid argon calorimeter
FC	Forward Calorimeter
MU	Muon
RT	Ray tracing package
SF	Sci-fi tracking
ST	Silicon tracking
WF	Hybrid tracking
WR	Straw tube tracking

RECONSTRUCTION PACKAGES

TS	Silicon Reconstruction

ANALYSIS PACKAGES

AH	Analysis of hybrids
AT	Analysis of Tracks

RUNNING SHELL

Making the executable:

Modify patchy cradle (SDCSIM.CRA)
Include packages of interest.

Run PATCHY and produce fortran program.

Compile and link program.

Runing the executable:

Modify control files

SDCSIM.CTL	- General control
	- Turn on/off packages
SGINI.CTL	- Simulation control
SGGEO.CTL	- Geometry control
SGEVT.CTL	- Event control
SDEVENT.CTL	- Flow control

Run program
Can also control flow of program interactively using KUIP commands.
Can analyze data interactively.

SDCSIM

Works on Apollo, DECstation, IBM RS6000, SGI, SUN, and VMS; possibly others.

Constantly evolving and being improved.

Presently being maintained using PATCHY, but will soon be maintained using CMZ and/or SCCS.

Simple installation procedure (written by SGF).

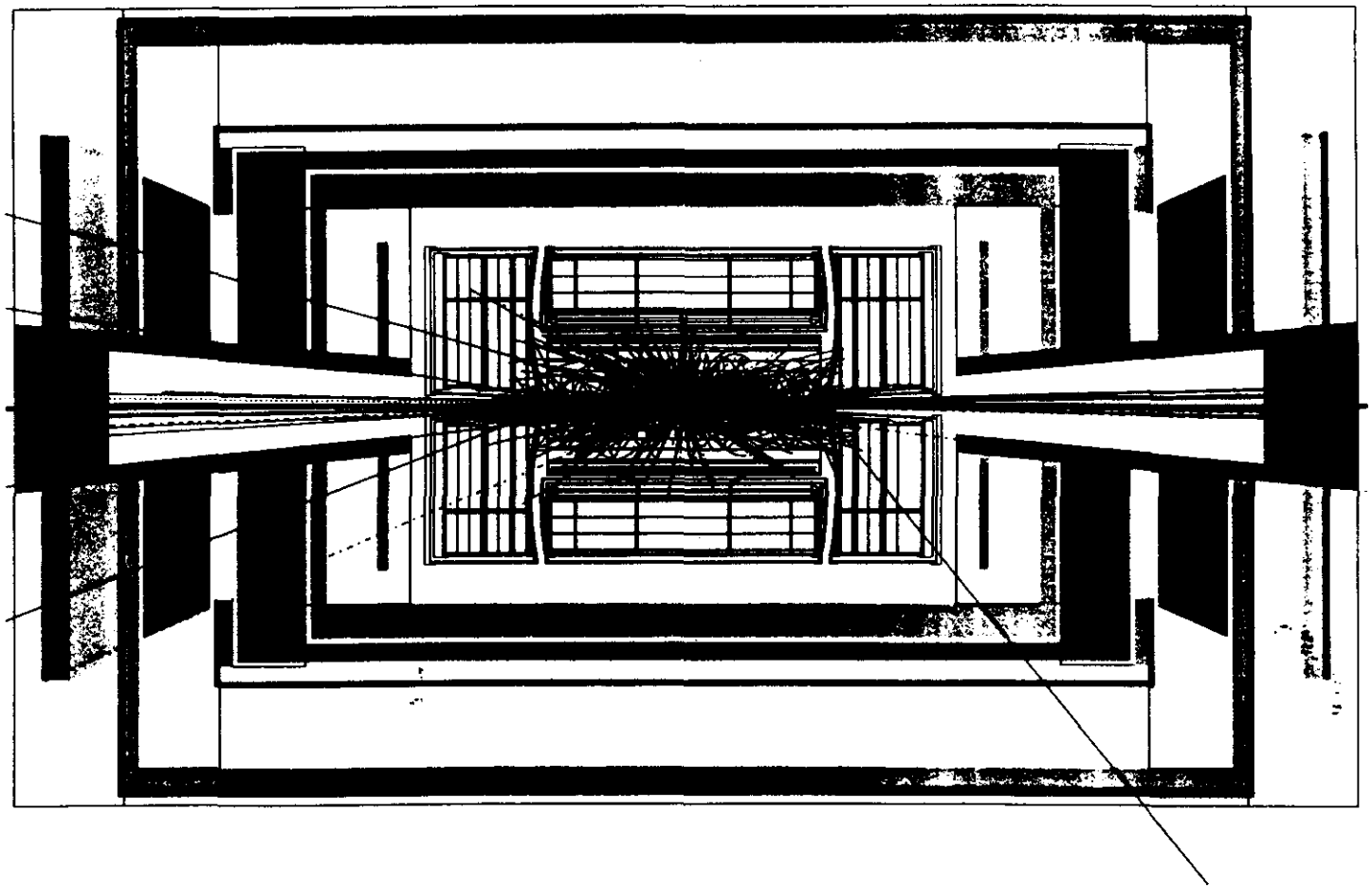
Version 2.0 is in the process of being released.
A new release implies a new library.
PATCHY correction cradles will be distributed between releases to correct serious bugs.

Interested people can get source from:

SSCVX1::USER5:[SDCSHELL.V01.SRC]
.LIB]
.DOC]
.MGR]

pdsf:/sdcshell/v01/src
/lib
/doc
/mgr

SDCSIM



SDC Computing and Analysis Software

SSCL Dallas

July 18, 1991

A.J. Elam
W.J. Courtney

IBM Federal Sector Division
Houston, Texas
(713) 282-8776

AELAM@HOUVMSCC.VNET.IBM.COM

IBM Dallas
Dallas, Texas
(214) 280-4765

bcourtn@dalvmic1.vnet.ibm.com

Contents

1991 IBM SSC/SDC Activities Update 1

Letter of Commitment 4/17/91 5

IBM Study Plan/Options 6

1991 SDC Study Tasks Schedule 8

IBM Library of HEP Documents 9

Independent Research & Development Plans 10

Near Term Plans/Activities 13

Additional Information 14

1991 IBM SSC/SDC Activities Update

Time	Activity
Jan 9-11	SDC Meeting - Dallas
Jan 10	IBM Briefing to SDC Computing on Houston Capabilities • FSD Houston Overview • EOSDIS Overview • OBS Software Quality • DSH-MSSWG Activities
Jan 24	FSD All Site Coordination Telecon
Feb 1	SSC/SDC PAR - Funding Issue
Feb 7-9	IBM Houston Hosted SDC Computing - Presentations and Working Sessions
Feb 12-13	FMI - DOE Conference - Crystal City
Feb 14-15	SSCL Planning Session - Bethesda
Feb 21	FSD All Sites Coordination Telecon
Feb 22	IBM Technical Visit to SSCL - Alan Fry. • Alan - Intro of HEP computing • Soren - PAW and GEANT Tools • Lee - ZEBRA Data Management
Feb 28	GBSI Study Plan Review - Funding Issue
Mar 1	IBM Technical Visit to Rice - David Adams • PAW and GEANT demos • Tools and Dev Env discussion
Mar 1	DOE Review with Gerry Ebker by J. Brunner • SSCL and SDC Highlighted • Funding not discussed
Mar 3	SDC Study Plans and Status Memo (w/ Funding Issue) (Elam)
Mar 6	IBM Technical Visit to Argonne National Lab - Larry Price • Larry Price - Introduction and Overview, Host • Bob Blair - CDF S/w Environment • Ed May - Fixed Target Experiments (Soudan) • T.D. Lee - Low Temperature Superconductivity Theory • Various demonstrations of tools

Time	Activity
Mar 7-8	IBM Technical Visit to Fermi National Accelerator Lab - Irwin Gaines • Irwin Gaines - FNAL Overview and Introduction • D0 Detector Visit • D0 Computing Working Group Session • ACPMAPS Demo and Technical Discussions • E791 - Fixed Target Visit • CDF Detector Visit • G.P. Yeh - Demo of tools/discussions • Jack Pfister - Fermi Organization and Tour • Eric Wickland - CDF Distributed Computing Environment • Fermi Computing Center Tour
Mar 12	SDC Offline Computing Funding Issue Memo (Elam)
Mar 12-15	HEP Conference - Japan (Merritt Jones)
Mar 13-15	SSC Industry Show in Atlanta • FSD Manassas, Boulder, Owego and Houston - S&TC Double Booth • Lack of Washington Participation Alarming • Program and Project Management • Industry Exchange (Martin, LeCroy, AT&T)
Mar 19	IBM Observations to SDC Computing Working Group at ANL
Mar 20-22	SDC Meeting at Argonne National Lab • Trilling - SDC Status • Gilchriese - Technical Status • Fred Gilman - SSC, LSTAR and Funding Status • Computing Technical Presentations • Banquet - Dinner with DOE representative
Mar 25-26	Present SSC/SDC Activities to IBM FSD Technical Advisory Board
Mar 27	IBM Houston GBSI Independent Research & Development Review
Mar 28	FSD All Sites Coordination Telecon
Apr 5	IBM Program Assessment Review - Funding Issue
Apr 10	Fermi Joint Study Meeting, Chicago Branch Office
Apr 17	IBM Letter of Commitment for SDC Study
Apr 22-30	IBM SDC Requirements Review • Computing Section Document Creation • Process Diagram • System/Software Bus Diagram • System Management Architecture Diagram

Time	Activity
Apr 30	IBM HEP Info Exchange (Hou/Kingston)
May 1	IBM Accelerator Tour - Fishkill • Advanced Lithography Facility (ALF) • LinAc, Electron Storage Ring, X-ray beam lines, photolithographic process lab, power and control system • 700 Mev • Lithography capability leading to 0.1 micron technologies
May 2-4	SDC Computing Working Group Meeting Yorktown • Parallel Processing Research – Massively Parallel (Vulcan) – Distributed Processing Clusters – Victor Demo – GF11 Demo • High Performance Networking (NSFNet, MetaRing, MetaNet) • Graphics and Visualization Research • Data Compression Research • Software Development Productivity • SDC CWG Session
May 16	IBM HEP Strategy Meeting, Dallas ESC • Houston, Bethesda (FSD) • Dallas, Kingston (ESC,S&TC) • FNAL, SSCL, SLAC, LBL (Accounts) • AWD (Austin Product Lab)
May 22	Zitel Disk Array Technical Presentation
May 30-31	SDC Tracking/Simulations Meeting, SSCL
June 3	IBM FSD Exec Call on DOE/SSCL (Mr. Macina)
June 4	IBM/SDC SOW sent from IBM to LBL
June 5-6	Technical Conference Calls • LBL - Chris Day • IU - Fred Luehring
June 10-14	Physics Computing '91 Conference
June 14	Loral Neural Network Demonstration, Houston
June 17-18	IBM/SDC Review of Computing Section Document/Information Model
June 19	Exec Call on Dr. Trilling (SDC/LBL)
June 19	Exec Call on Dr. Schwitters (SSCL)
June 20-22	SDC Collaboration Meeting , SSCL
July 10-11	SDC Computer Working Group Meeting , SSCL
July 16-19	IBM Technical Visit to LBL

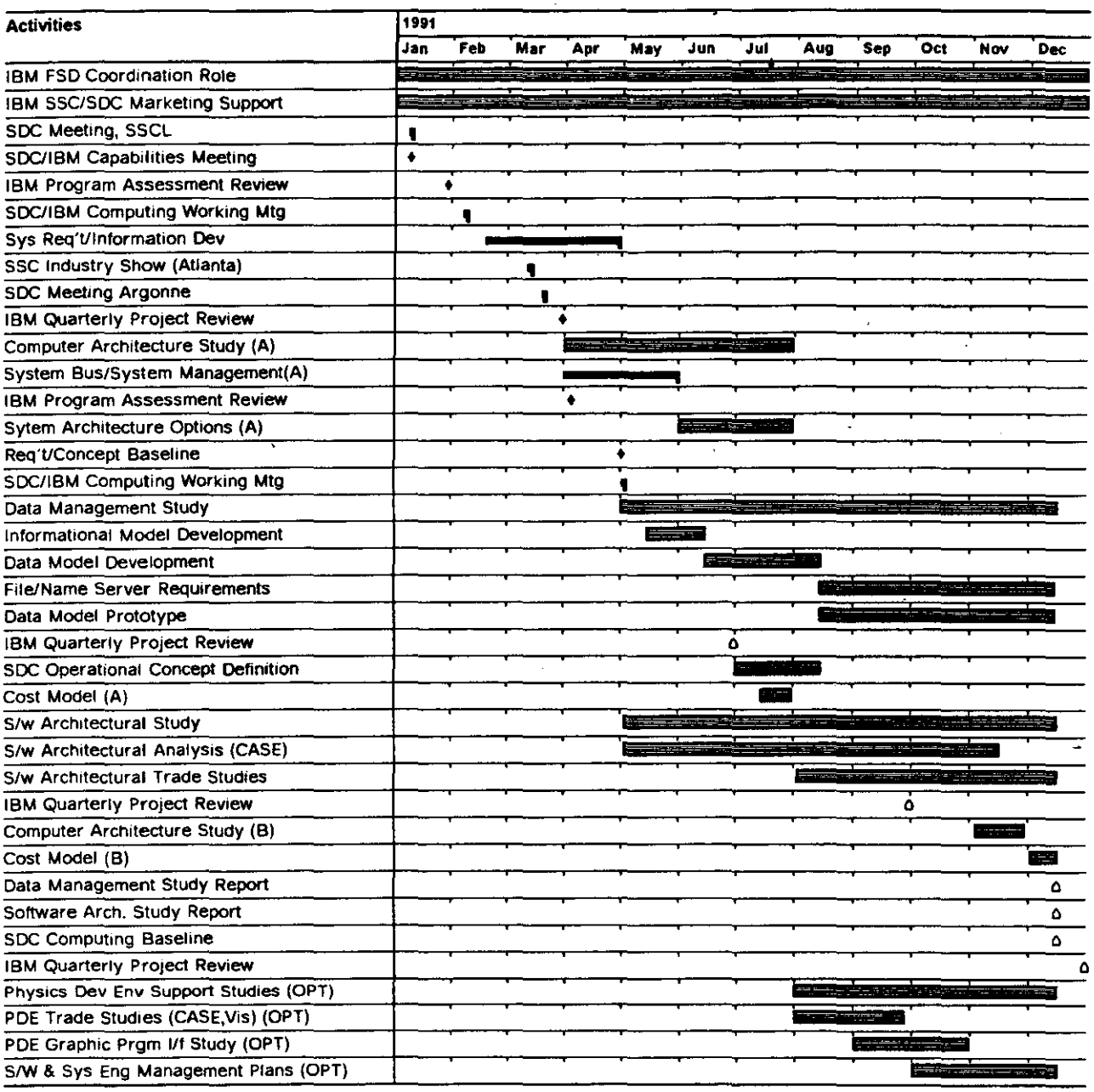
IBM Houston Team

Name	Responsibility	Contact
Tony Elam	Program Manager, System Engineering	(713) 282-8776 AELAM
Glen Kubena	Software Architecture, System Engineering	(713) 282-KUBENA
Denis Dorris	Data Management, Data Modeling	(713) 282-8543 DORRIS
Dr. Jim West	Physics Development Environment, Simulations	(713) 282-8722 WEST
Bob Ennis	System Requirements, Ops Concepts	(713) 335-6534 BOBENNIS
Note: IBMers Internet at HOUVMSCC.VNET.IBM.COM		

Additional IBM Technical Support Available

Name	Expertise	Contact
Dr. Bill Courtney, Dallas	Physicist, IBM System Engineer	(214) 280-4765, BCOURTN @ DALVMIC1
Dr. Jeff Durachta, Kingston	Physicist, Science & Technical Computing	(914) 385-5056, JWD @ KGNVMA
Dr. Steve Hicks, Dallas	Physics, Engineering Support Center	(214) 406-5524, HICKS @ DFWVM03
Dr. Yoram Ofek, Yorktown	High Performance Communications	(914) 784-7085, OFEK @ YKTVMH2
Dr. Ephraim Feig, Yorktown	Mathematics, Access to Data	(914) 945-3507, FEIG @ YKTVMZ
Note: IBMers Internet try at userid@nodeid.VNET.IBM.COM		

1991 SDC Study Tasks Schedule



Letter of Commitment 4/17/91

Distribution: Gil Gilchriese, LBL;
Larry Price, ANL;
Drew Baden, University of Maryland

Subject: IBM Initial Commitment To SDC

I am pleased to inform you that IBM is committed to support the Solenoidal Detector Collaboration. IBM Houston will support the SDC's Computing and Analysis Software Subcommittee with technical studies in the following areas of prime importance to the SDC: Requirements Analysis and Operation Concept Development; SDC Data Management; SDC System Architecture; SDC Software Architecture; and SDC Costing. As planned, IBM will work closely with key SDC representatives throughout the study period to focus applicable results to support the SDC Design Proposal. IBM's level of effort commitment to these study tasks is approximately four dedicated engineers for the remainder of the year.

In addition, IBM Houston plans to attend and support all relevant SDC technical meetings throughout the year. We will also continue to make available key research and appropriate development engineering skills as required to address specific questions and challenges during the study.

I will continue to work aggressively to seek additional approval for IBM involvement in other study topics such as the Physics Development Environment and "proof of concept" prototyping. I am optimistic that such support will be available in the future but unable to commit to these tasks at this time.

I will present my new support plan, tasks description and working schedule at the SDC Computing Working Group Meeting at IBM Yorktown on May 2.

If you have any questions or concerns, do not hesitate to call or e-mail. I look forward to our challenging work together.

Sincerely,

Anthony J. Elam
IBM Houston SSC/SDC Program Manager

IBM Study Plan/Options

Level of Effort Activity - Committed

Task	SDC Lead	IBM Lead	IBM Support	Effort
Sys Req/Ops Con	Price, Day, Cormell	Ennis	Dorris, West, Poole	6MM
Data Management	Grossman, Baden	Dorris	Buck	10MM
Computer Architecture	Gaines, Scipioni	Elam	Poole, Coyne	4MM
Software Architecture	Day, Fry	Kubena	McCain, McLellan	8MM
Total (w/ cost estimates)				28MM
Note: SDC Reps from ANL, LBL, SSC, FNAL, U of Md, U of Ill				

Level of Effort Activity - Uncommitted

Task	SDC Lead	IBM Lead	IBM Support	Effort
Physics Dev Environment	* * TBD * *	West	Graf, Everson, McCain	10MM
Cost Modeling	* * TBD * *	* * TBD * *	Keller, Rone	3MM
S/w & S.E. Mgmt Plans	* * TBD * *	Kubena		2MM
Total				15MM

Additional funds for travel and overhead approved.

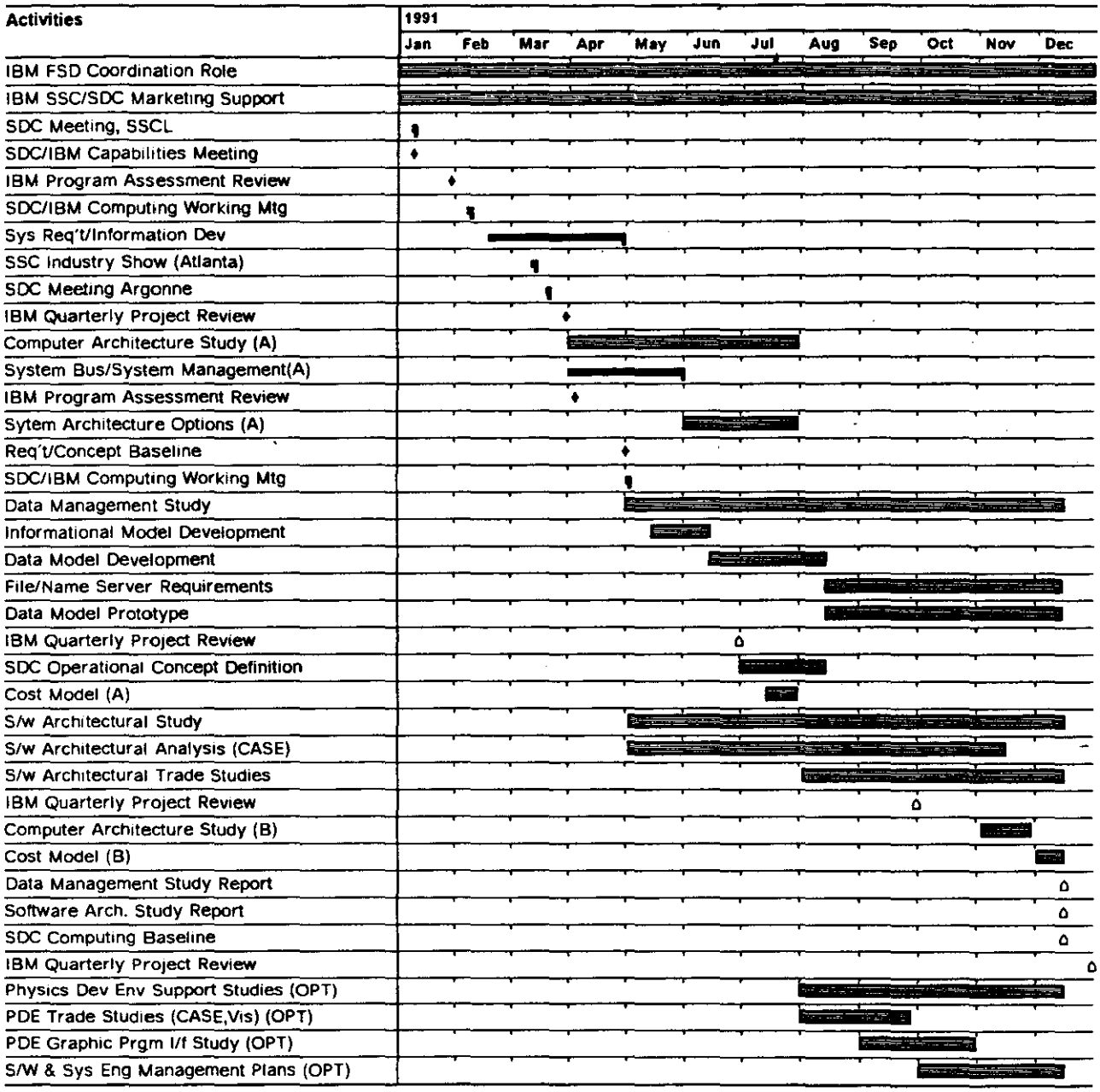
IBM Houston Team

Name	Responsibility	Contact
Tony Elam	Program Manager, System Engineering	(713) 282-8776 AELAM
Glen Kubena	Software Architecture, System Engineering	(713) 282-KUBENA
Denis Dorris	Data Management, Data Modeling	(713) 282-8543 DORRIS
Dr. Jim West	Physics Development Environment, Simulations	(713) 282-8722 WEST
Bob Ennis	System Requirements, Ops Concepts	(713) 335-6534 BOBENNIS
Note: IBMers Internet at HOUVMSCC.VNET.IBM.COM		

Additional IBM Technical Support Available

Name	Expertise	Contact
Dr. Bill Courtney, Dallas	Physicist, IBM System Engineer	(214) 280-4765, BCOURTN @ DALVMIC1
Dr. Jeff Durachta, Kingston	Physicist, Science & Technical Computing	(914) 385-5056, JWD @ KGNVMA
Dr. Steve Hicks, Dallas	Physics, Engineering Support Center	(214) 406-5524, HICKS @ DFWVM03
Dr. Yoram Ofek, Yorktown	High Performance Communications	(914) 784-7085, OFEK @ YKTVMH2
Dr. Ephraim Feig, Yorktown	Mathematics, Access to Data	(914) 945-3507, FEIG @ YKTVMZ
Note: IBMers Internet try at userid@nodeid.VNET.IBM.COM		

1991 SDC Study Tasks Schedule



IBM Library of HEP Documents

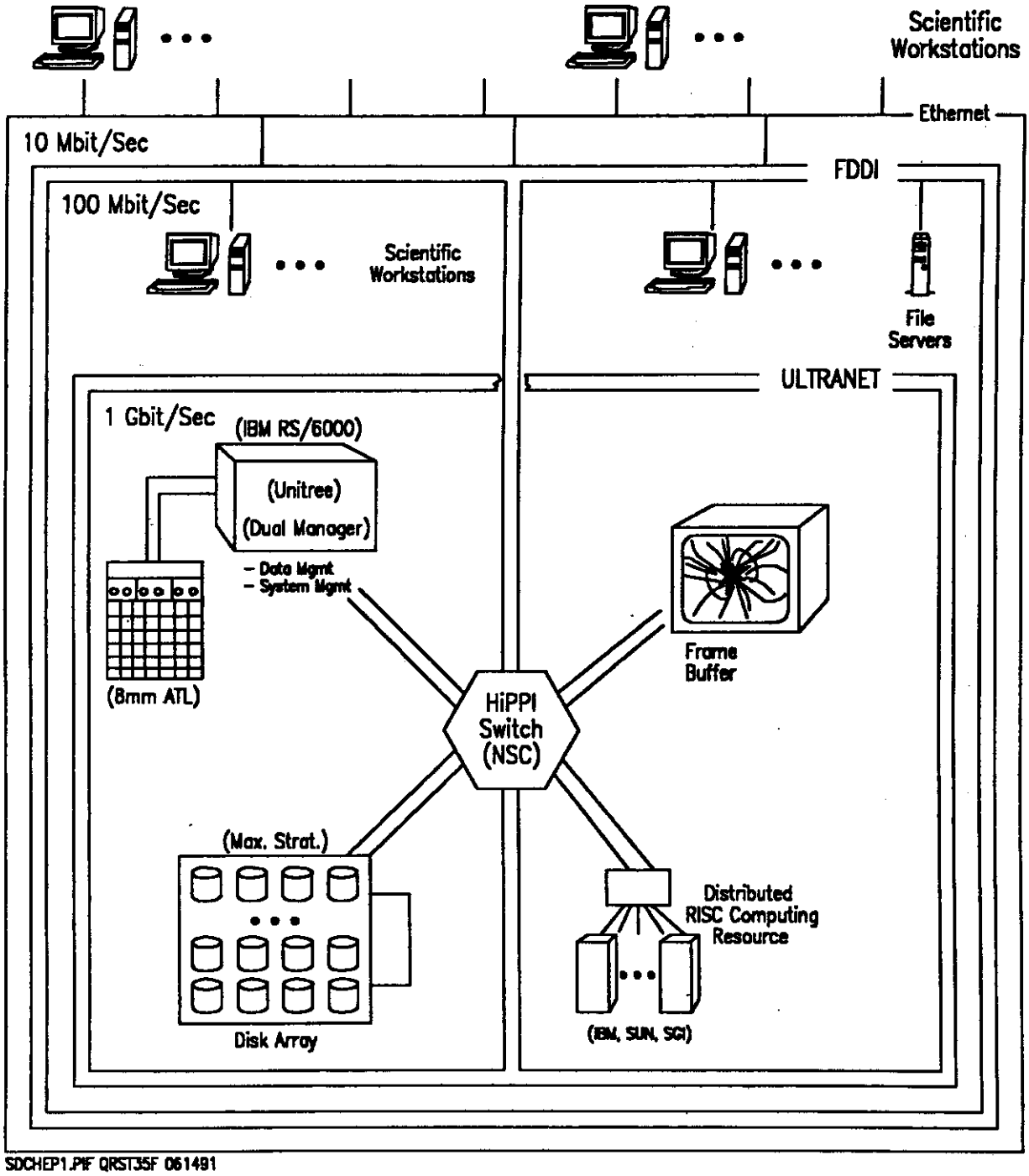
- 112 Documents in Database
- Name, Date, Author, Categories, Abstracts
- Available for review from Tony Elam

Code	Category
Adm	Administrative Information
Ar	Architecture
Dm	Data Management
Exp	Experiment/Physics Specific
Gen	General Interest
Geo	Geometry Specifics
Hw	Hardware Specific
IBM	IBM Internal
Net	Networking, Communications
On	On-line
Off	Off-line
Org	Organizational Information
Pr	Processing
Sim	Simulation
Sw	Software Specific
Use	User Interface, GUI

Independent Research & Development Plans

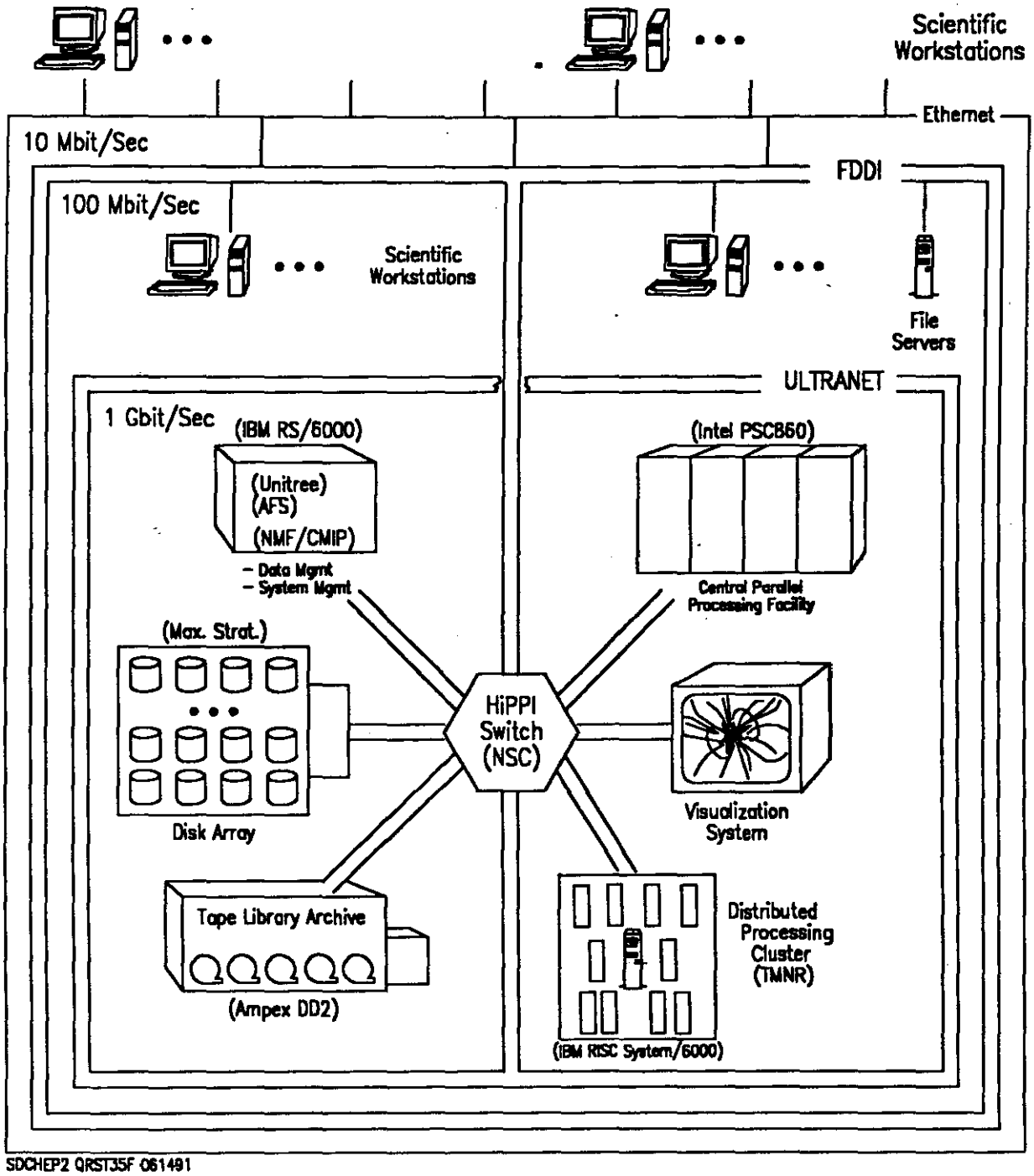
- Additional IRAD Committed for 1991
- Build a HEP Proof of Concept Lab (see charts)
- Obtain DOE/SSC/FNAL Evaluators
- Support a major FNAL Experiment in 1992
- Gain Experience and Knowledge in HEP codes
  - SDC simulation codes
  - Event reconstruction and classification codes
  - PAW and analysis codes
  - Investigate visualization tool sets for physics
- Make available the IBM HEP Lab to SDC collaborators
  - Distributed RISC Workstations
  - ES/9000 VM/AIX with Vector
  - Visualization tool sets

HEP LAB Stage 1



SDCHEP1.PIF QRST35F 061491

HEP LAB Stage 2



SDCHEP2 QRST35F 061491

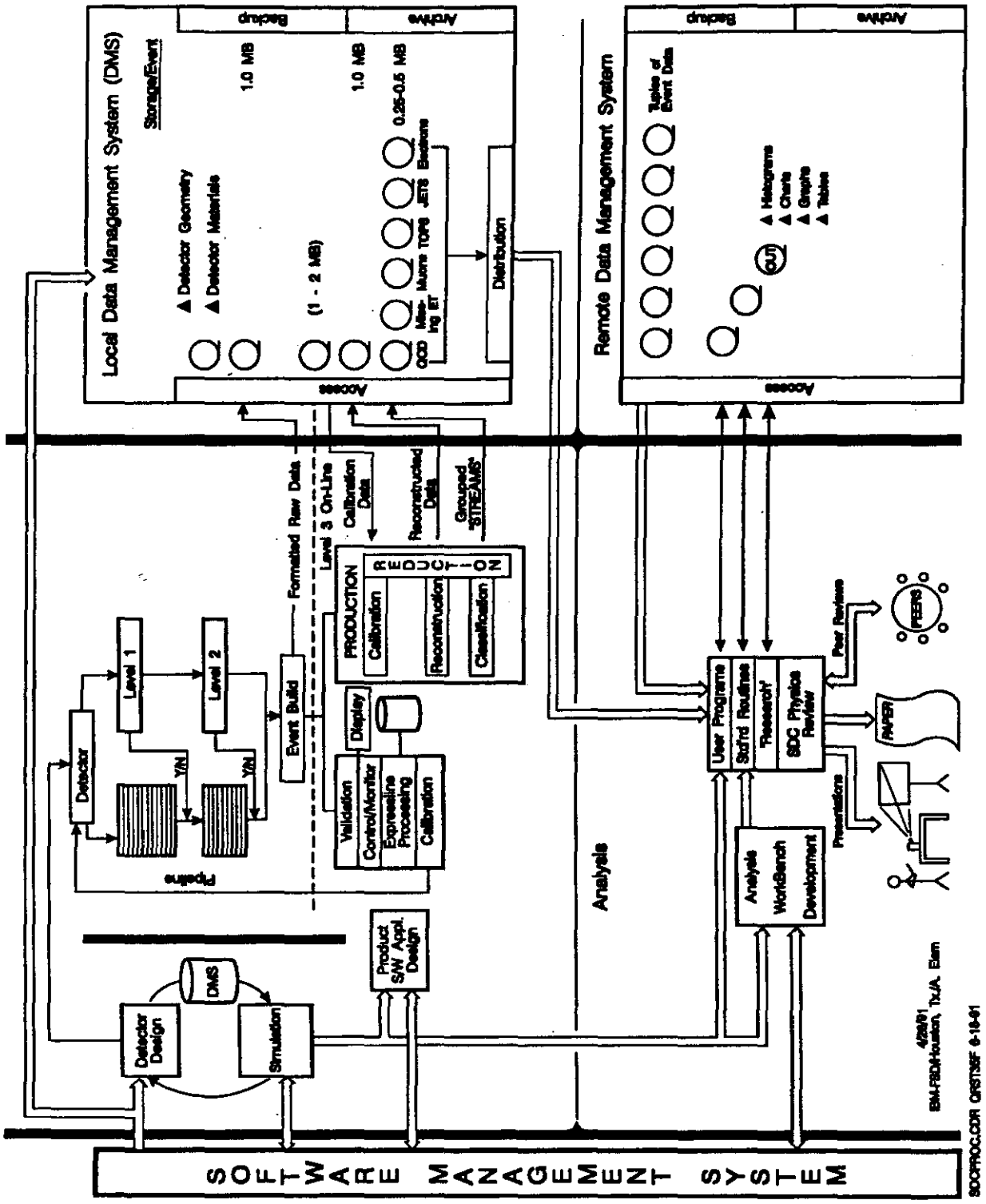
Near Term Plans/Activities

- MOU between IBM Houston/SDC
- IBM Houston role in support of CDF
- IBM Houston role in support of other HEP activities

Time	Activity
June	Establish "baseline" + options for costing
TBD	IBM Host SDC CWG at Almaden
July	• PAC Status Review • Draft Mgmt and Budget Plan to SSCL
Aug 5-10	SDC Meeting at LBL
Sept 26-28	SDC Meeting at ORNL
Oct	IBM Final IR&D and SDC Study plans for 1992
Nov	SDC Baseline Review
Jan '92	SDC Technical Proposal Draft

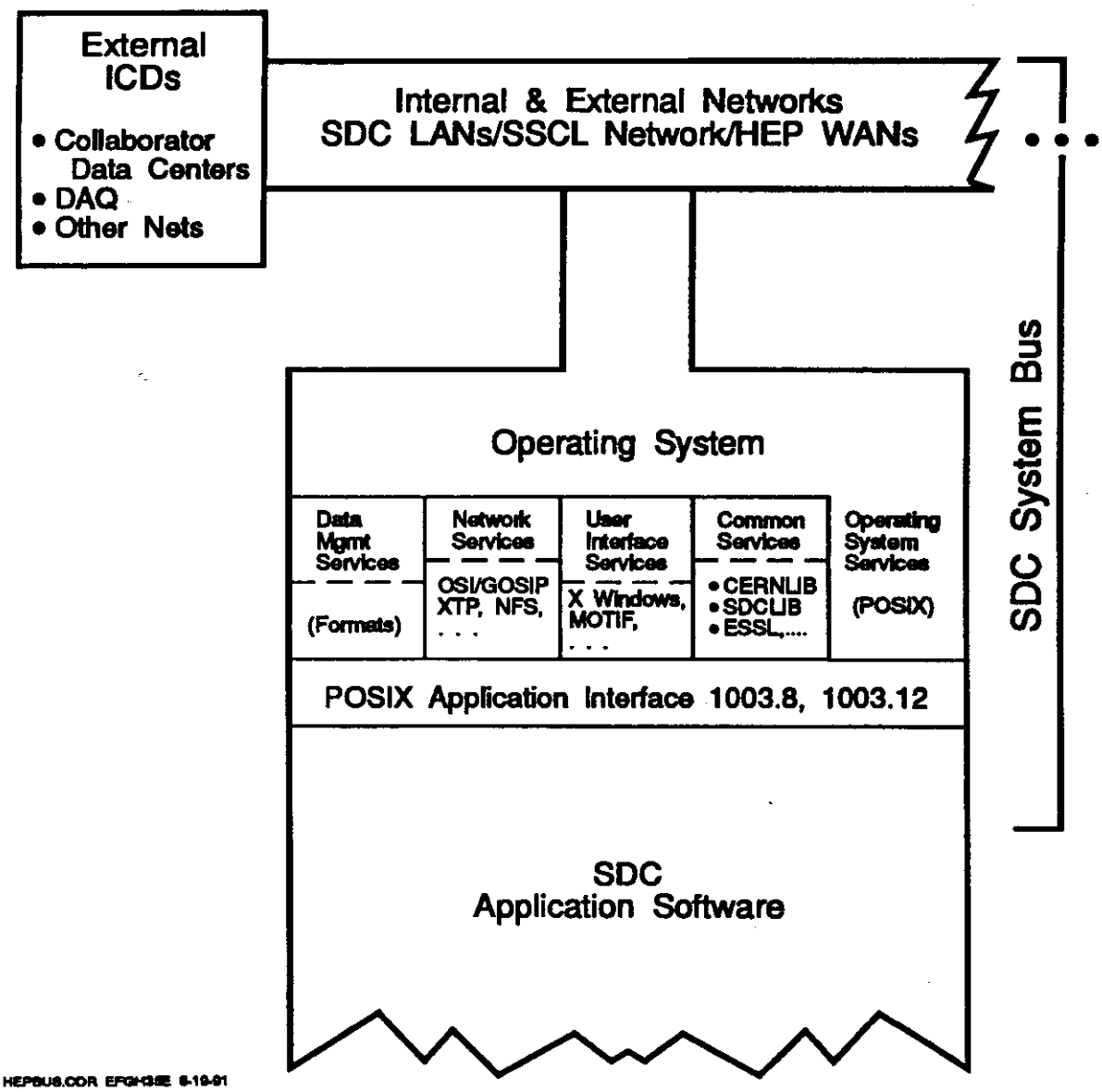
Additional Information

- SDC Computing Section Document
- SDC Process Diagram
- SDC Software/System Bus Diagram
- OSI Systems Management Architecture Model Diagram
- SDC Information Model
- System Engineering Process


Local Data Management System (DMS)
Storage/Event
Backup
Archive
Access
Detector Geometry
Detector Materials
1.0 MB
1.0 MB
0.25-0.5 MB
(1 - 2 MB)
QCD
Mis-ing ET
Muons
TOPS
JETS
Electrons
Distribution
Remote Data Management System
Tuples of Event Data
Histograms
Charts
Graphs
Tables
OUT
Formatted Raw Data
Level 3 On-Line
Calibration Data
Reconstructed Data
Grouped "STREAMS"
PRODUCTION
Calibration
Reconstruction
Classification
REDUCTION
Level 1
Level 2
Detector
Y/N
Event Build
Display
Validation
Control/Monitor
Expressline Processing
Calibration
Pipeline
Product S/W Appl. Design
DMS
Detector Design
Simulation
Analysis
User Programs
Std'rd Routines
"Research"
SDC Physics Review
Peer Reviews
PEERS
PAPER
Presentations
Analysis
WorkBench
Development
SOFTWARE MANAGEMENT SYSTEM
4/26/91
IBM-FSD/Houston, Tx./A. Elam
SDCPROC.CDR QRST35F 6-18-91

STANDARDS Structure the SDC DATA SYSTEM

- Improved Interoperability
- Reduced Development Cost
- Obsolescence Avoidance
- Network Wide Information Services



HEPBUS.CDR EFGH3IE 6-19-91

OSI Systems Management

Architectual Model

Management Function

FAULTS

Alarm Reporting
X.733;ISO/IEC 10164.4

SECURITY

Alarm Reporting
X.736;ISO/IEC 10164.7

Audit Trail
X.740;ISO/IEC 10164.8

Access Control Objects/Attributes
X.74X;ISO/IEC 10164.X

ACCOUNTING

Accounting Meter
ISO/IEC 10164.10

PERFORMANCE

Workload Monitoring
X.739;ISO/IEC 10164.11

Measurement Summarization
X.738;ISO/IEC 10164.M

USER INTERFACE

Event Reporting
X.734;ISO/IEC 10164.5

Log Control
X.735;ISO/IEC 10164.6

CONFIGURATION MANAGEMENT/CONTROL

Object Management

Attributes for Representing Relationships

State Management
X.731;ISO/IEC 10164.2

X.756;ISO/IEC 10164.3

X.730;ISO/IEC 10164.4

CMIP Platform

CMISE

Mgmt/Info Service Elements
ISO/IEC DP 9595.1

CMIP

Mgmt/Info Protocol
ISO/IEC DP 9595.2

Figure IV-9

WHC16.COR UW0036A

SHIFT

The Scalable Heterogeneous Integrated Facility for HEP Computing

Jean-Philippe Baud Julian Bunn Fabrizio Cane David Foster *
Frédéric Hemmer Erik Jagel Gordon Lee Les Robertson Ben Segal
Antoine Trannoy Igor Zacharov

CERN - European Laboratory for Particle Physics
1211 Genève 23, Switzerland
2 March 1991

H. Newman

Abstract

***shift*, the Scalable Heterogeneous Integrated computing FaciliTy, aims to demonstrate that commercially available workstations and associated disk and tape storage products can cheaply and reliably be networked together to provide an integrated computing service, comparable in quality and availability with traditional mainframe services, but suitable for installation in user-operated environments, and smoothly scalable to provide very large computing facilities.**

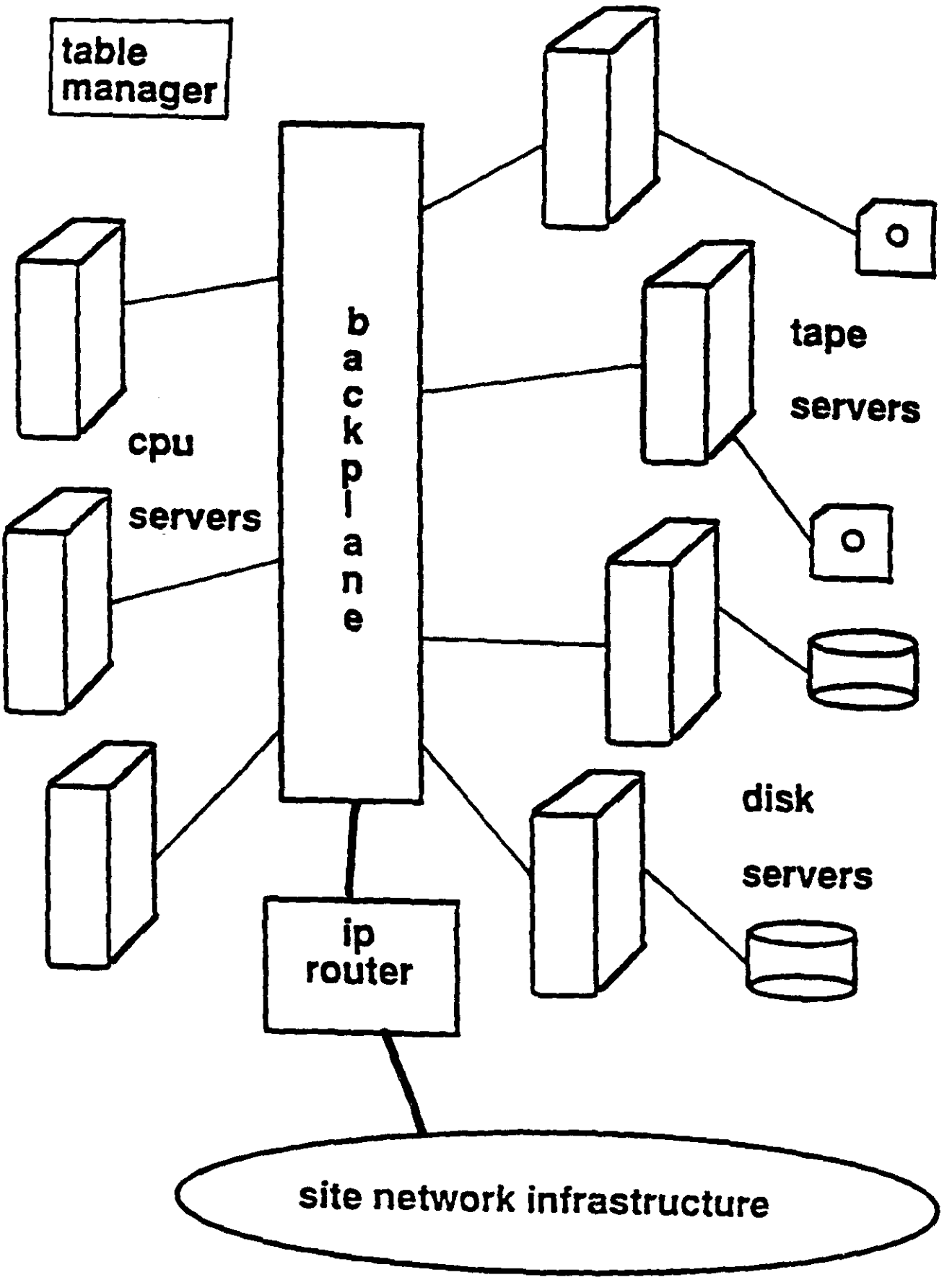
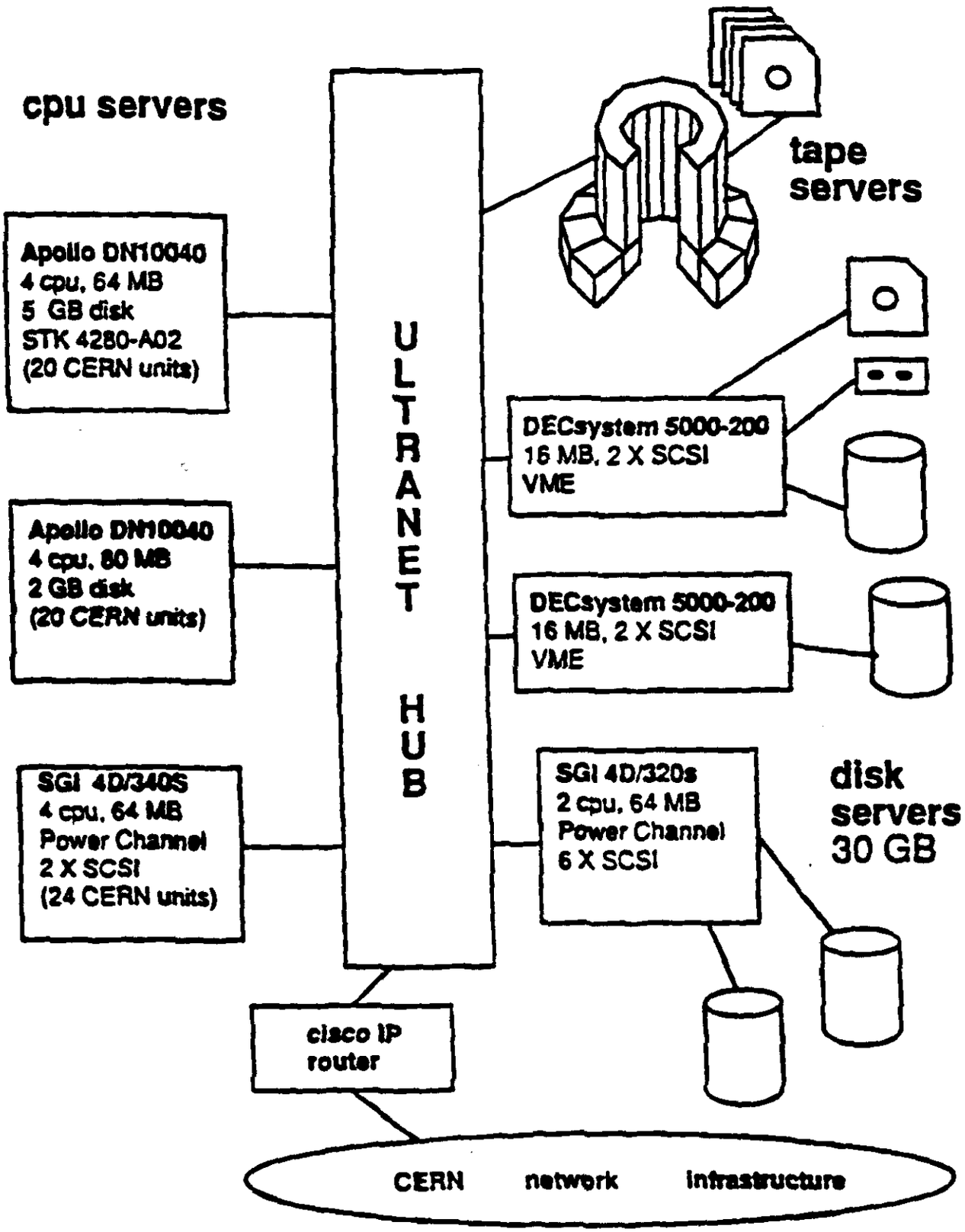


table manager
backplane
cpu servers
tape servers
disk servers
ip router
site network infrastructure


cpu servers
Apollo DN10040
4 cpu, 64 MB
5 GB disk
STK 4280-A02
(20 CERN units)
Apollo DN10040
4 cpu, 80 MB
2 GB disk
(20 CERN units)
SGI 4D/340S
4 cpu, 64 MB
Power Channel
2 X SCSI
(24 CERN units)
ULTRANET HUB
tape servers
DECsystem 5000-200
16 MB, 2 X SCSI
VME
DECsystem 5000-200
16 MB, 2 X SCSI
VME
SGI 4D/320s
2 cpu, 64 MB
Power Channel
6 X SCSI
disk servers
30 GB
cisco IP
router
CERN network infrastructure

3.2 Software Components

Disk Pool Manager The disk pool manager allocates files to pools, creates files and associated directories, locates existing files, deletes files on request and performs garbage collection on temporary pools. The user interface to the disk pool manager is through the *sfget*, *sfrm*, and *sfsh* commands.

An important component of the disk pool manager is the *table manager*, which maintains static and dynamic configuration information. It is responsible for locating *shift* files, making disk pool selection decisions and serialising access to the data when necessary. The data used by the table manager can be recreated in the event of a serious failure of the node on which it executes.

Tape Copy Scheduler The tape copy scheduler organises the copying of data between *shift* disk files and magnetic tapes. On request from a user through a *tpread* or *tpwrite* command, it selects an appropriate tape server depending on the device type required, location of the tape and current tape activity. It then initiates a tape copy using the *cptpdsk* or *cpdsktp* program on the tape server node. The tape copy scheduler informs the user when the operation is complete, queues concurrent requests for the same tape, and deals with error recovery.

Remote File Access System The Remote File Access System (RFAS) provides the most efficient way of accessing files stored on *shift* disk servers from compute and tape servers. These files may also be accessed using NFS [7], but the RFAS will take account of the network characteristics and the mode of use of the file to minimise overheads and maximise throughput.

Each *shift* node runs a special daemon, the RFAS server, which implements file access requests from other *shift* nodes. Normally this is done using TCP/IP protocols over the backplane.

The user interface to the RFAS is defined for the C and Fortran language bindings.

Remote Tape Copy The *cptpdsk* and *cpdsktp* commands copy data from tape to disk or disk to tape. They execute on the computer to which the tape unit is attached, but can access a local or remote disk file, normally using the Remote File Access System.

Tape Allocation and Control This component, called the *tape daemon*, provides the multi-user tape support which is usually missing on UNIX systems. This includes tape allocation, tape mounting, and label checking.

Distributed Operator Interface This provides a single operator interface for the *shift* configuration. In particular it is used by the tape daemon to request tape mounts, and deal with tape label conflicts.

Network Queueing System The Network Queueing System, NQS, is a distributed batch scheduler originally developed for NASA and which is now in the public domain. NQS servers execute on each of the *shift* compute servers, on a number of other UNIX services at CERN and, optionally, on end-user workstations running UNIX. Each server performs three functions: interfacing with users logged in to the local machine for job submission, query and control: communication with other NQS servers over a TCP/IP network; scheduling of job execution on the local machine.

NQS is an important facility used by *shift*, although it is entirely independent of it. CERN has extended NQS to cater for centralised batch queues which feed clusters of compute servers.

Accounting Facilities A number of utilities are available to provide resource consumption reports and controls.

Table 2: Local disk - Remote disk reading
using UltraNet and the Remote File Access System
Performance and CPU Cost Comparison

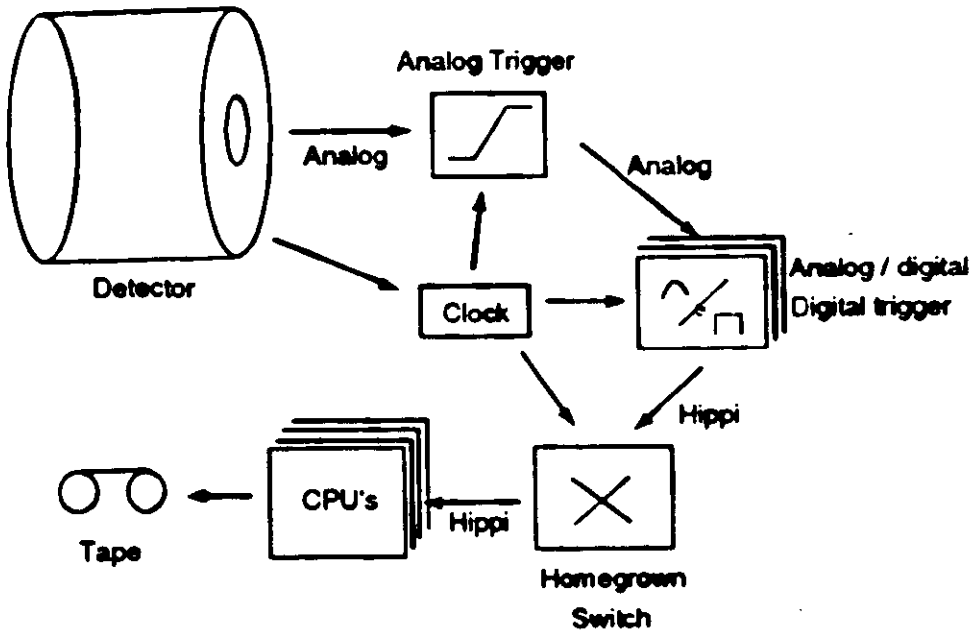
client	disk server	Application blocksize KBytes	Data rate MBytes/sec		CPU cost of client MBytes/CPU-sec	
			local	remote	local	remote
SUN 3-260	SGI 320S	20	0.5	0.6	1.4	1.7
		32	0.5	0.8	1.4	2.0
		200	0.5	0.8	1.4	5.3
SGI 340S	SGI 320S	20	1.5	1.2	4.8	4.8
		32	1.6	1.5	5.5	7.3
		200	1.8	1.4	5.2	23.0

Table 1: UltraNet - FDDI : Performance and CPU Cost Comparison

source	sink	Application blocksize KBytes	Data rate MBytes/sec	CPU cost of sink MBytes/CPU-sec
UltraNet				
SGI 320S	SGI 340S	20	4.4	15
		200	10.5	75
		2000	13.0	130
SGI 340S	Cray X-MP/48 (low-speed channel)	20	3.1	80
		200	5.5	550
		2000	6.0	600
SGI 340S	SUN 3-260	20	1.9	3.8
		200	3.2	6.4
		2000	3.4	6.8
FDDI				
DEC 5000-200	SGI 340S	20	1.5	3.3
DEC 5000-200	DEC 5000-200	20	1.8	2.4
Apollo DN10040	Apollo DN10040	20	0.75	1.0
SGI 340S	SUN 3-260	20	0.6	1.0

On-line computing

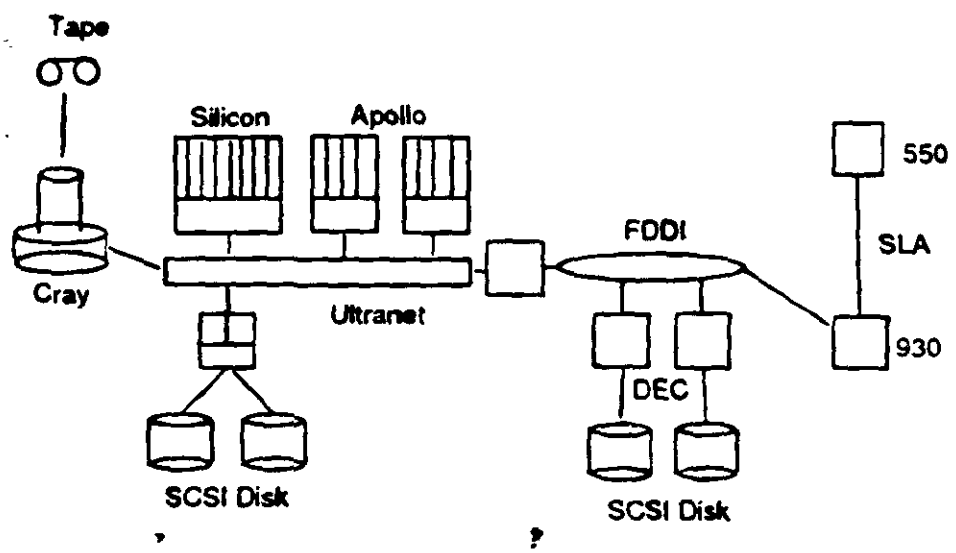
- LHC[1] is coming in 1996
- An event every 15 nano-seconds
- Megabytes of data per event
- Upcoming project
  - NA48 (Experiment in 1992/3)
  - Data coming in at high speed (50MB/s)
  - Will be processed by up to 16 CPU's in parallel
  - Based on Hippi and / or 1/4 speed FCS



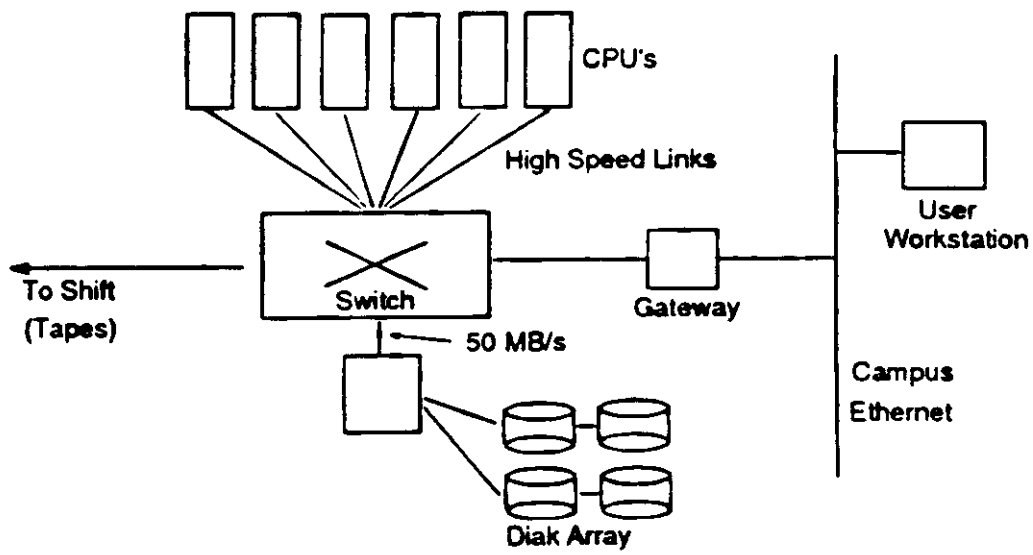
1 The next generation collider

Farms

- Price / performance is the driving force
- Projects at CERN
  - SHIFT, started in 1990, very successful
    - Heterogeneous: SG, Apollo, Dec
    - Provides similar throughput as mainframes



- Simulation Farm
  - 50 HP 'Snakes' (Joint study with HP)
- Visualization and Analysis farm
  - Analyze big datasets (> 100MB) interactively
  - Dataset on one single disk (-> Disk array)



CPU Requirements

- Applications are very Integer-bound
  - Integer SPECmarks reflects application performance
    FP and Total SPECmarks lead to excessive expectations
  - 'Mega-IF's per second' instead of MFLOPS
    Max. 15% floating point content
- I/O is not always a bottleneck
  - The '50 Snakes' will be linked by ethernet !
- "Ideal CPU":
  - Single board
  - Rack with 64+ boards
  - Fast Integer CPU (Fast clock or Multiple parallel integer units)
  - Memory (32M++)
  - Fast I/O connection (FCS ?)
  - Some integrated interfaces (SCSI, Serial)

DRAFT - For Discussion Only
Ken McFarlane 7/18/91

0. Computing

0.1 Overview of Tasks

0.1.1 On-line

0.1.2 Communications

0.1.3 Off-line

0.1.4 Auxiliary Tasks

0.2 Organization

0.2.1 Collaboration Computing Organization

0.2.2 Resident Group

0.2.3 Coordination with SSCL

0.2.4 Coordination with SDC

0.2.5 Institutional Responsibilities

0.3 Computing Model

0.3.1 Distributed Participation

0.3.2 Central Coordination and Support

0.3.3 Central Data Storage

0.3.4 Central Facility

0.3.5 Regional Centers

0.3.6 Local Institutions

0.3.7 Institutions outside U.S.

0.3.8 Relation to SSC Computing Division

0.3.9 Relation to SDC Computing

0.4 Design and Implementation Methodologies

0.4.1 Standards

0.4.2 Hardware Simulation

0.4.3 Software Engineering

0.4.4 Professionals and Professionalism

0.4.5 Training

0.5 Cost and Schedule Approach

1. On-line Computing

1.1 Overview of Subsystems

1.1.1 Front-end

1.1.2 L1, L2

1.1.3 L3

1.1.4 Control

1.1.5 Network

1.1.6 Other Related Systems (Simulation, Offline)

1.2 Requirements

1.2.1 Configuration (Front-end, L1, L2, ...)

1.2.2 Calibration

1.2.3 Control

1.2.4 Acquiring, Processing and Recording Data

1.3 Quality Assurance

1.3.1 Performance Monitoring

1.3.2 Alarm, Error Reporting

1.3.3 Reliability

1.3.4 Security and Safety

1.4 Physics Analysis

1.5 Software Tools and Methodologies

1.6 Design (Block Diagrams)

1.7 Implementation

1.8 Beam Test Support

1.9 R&D

1.10 Cost and Schedule

2. Communications

2.1 Overview

2.2 Requirements

2.2.1 Command and Control (On-site)

2.2.2 Data Movement (On-site)

2.2.3 Off-site WAN

2.2.4 Regional Network

2.2.5 Dial-up Capability

2.2.6 Teleconferencing, E-mail, News

2.3 Supported Systems e.g., E-mail, news: PC on up

2.4 Needed R&D

2.5 Cost and Schedule

3. Off-line Computing

3.1 Overview

3.1.1 Central Facility

3.1.2 Regional Center

3.1.3 Local Institutions

3.2 Requirements Summary

3.2.1 Data Storage and Access

3.2.2 Processing (CPU, Memory)

3.2.3 Analysis Workstation

3.2.4 Analysis Software

3.2.5 Availability

3.2.6 Software Development

3.2.7 Support

3.3 Hardware and Software Standards

3.4 Data Storage Requirements

3.4.1 Rapid-Access (Magnetic or Optical Disk) permanent

3.4.2 Rapid-Access Temporary (Staged)

3.4.3 Slow-Access (Tape Robots)

3.5 Data Access Methods

3.5.1 Sequential

3.5.2 Keyed

3.5.3 Partial

3.6 CPU Requirements

3.6.1 CPU and FPU Power

3.6.2 Memory and Swap Space

3.6.3 CPU/FPU Architecture, Precision

3.6.4 Distributed/Parallel Computing

3.7 Analysis Workstation

3.7.1 Graphics Performance

3.7.2 CPU Performance

3.7.3 Network Performance

3.8 Availability

3.8.1 24-hour Partial Results

3.8.2 Time Scale for Full Results

3.8.3 Distribution of Selected Events to Regional Center

3.9 Software Requirements

3.9.1 Portability

3.9.2 Distributed Execution/Parallel Execution

3.9.3 Execution Modularity (Software Bus')

3.9.4 Graphic Analysis Tools

3.9.4.1 Graphic Event Editing, Selection

3.9.4.2 Interactive Control of Execution

3.9.4.3 Histogramming, fitting

3.9.4.4 Interface to Hard Copy

a. Compound Documents

b. Video Conferencing

c. Recording in CD, ...

3.9.5 Self-Documenting

3.9.6 Self-Journaling

3.9.7 On-Line Help

3.10 Software Development and Management

3.10.1 Overview (the problem!)

3.10.2 Commercial Software

3.10.3 Centrally Developed (CERN, SSCL)

3.10.4 Software Engineering

3.10.4.1 Methods

3.10.4.2 Productivity and Reliability Gains

3.10.4.3 Tools

3.10.5 Development Environment

3.10.5.1 Standards

3.10.5.2 Code Management

3.10.5.3 Verification

3.10.5.4 Languages

3.10.5.5 Documentation

3.10.5.6 Debugging

3.10.6 Maintenance

3.10.6.1 Librarian

3.10.6.2 Versions

3.10.6.3 Distribution

3.10.6.4 Updates and Corrections

3.11 Human Interface

3.12 Engineering

3.12.1 Engineering-to-Physics (CAD to GEANT)

3.12.2 Data Acquisition Simulation

3.12.3 Networked System Simulations

3.13 R&D

3.14 Cost and Schedule

4. Miscellaneous

4.1 Document Preparation and Maintenance

4.2 Test Beam Data Acquisition Support

4.3 Standards Group Participation

4.4 Technology Forecasts

4.5 Range of Systems Supported

4.5.1 Upgrades and Scalability

4.5.2 Parallel

4.6 Benchmarking