FERMILAB-PUB-89-278-PPD

Update on Central and Data Acquisition Computing at Fermilab

by Jeffrey A. Appel

The Tevatron era and its associated computing is now nearly five years old. Al Brenner wrote the last review of computing to appear in these pages near the beginning of this era (*Fermilab Report* issue of December 1984). A lot has changed in computing in these five years! This has been a period of exponential growth, technological innovation, and great tension. All these are the result of dramatically expanding computing needs and the resultant competition with other Laboratory priorities in a period of limited resources. October 1st, 1989, marked the beginning of the era of the Computing Division at Fermilab. It's appropriate, therefore, to stop and review how far we have come during these first five years of the Tevatron era.

The changes and growth in computing at Fermilab are driven by the physics research requirements which must be satisfied in order to take advantage of the Tevatron potential. The Collider Detector at Fermilab (CDF) is the most well known of the new class of experiments. However, it is only one of seven experiments which already count between 1500 and 3000 modules provided by the electronics pool. In each of these experiments, the purchase price of the pool electronics is between 1.5 and 3 million dollars. The increased size and complexity of the detectors has also led to an increase in the amount of data recorded. In the last fixed-target run, over 35,000 nine-track tapes of data were recorded, over 10,000 by one experiment alone. In the Collider run, CDF recorded almost 6000 tapes of data. These numbers all dwarf those of five years ago. It is no wonder that other changes in computing have been necessary.

Central and Distributed Computing

During the last two years, the Computing Department has implemented a three-pronged approach (Fig. 1, page 2) to the tremendous computing challenges of the experimental physics program. The three prongs include (1) the user-friendly, tools-rich VMS environment of the central VAX clusters and work stations; (2) eight production microprocessor farms for the running of stable, data reconstruction codes; and (3) the general purpose, fast turn-around,

---

The author, having just completed a term as Head of the Computing Department, has now returned to full-time research on the series of heavy-flavor physics experiments at the Tagged Photon Laboratory. He is spokesperson of E-769 and Co-Spokesperson of E-791.

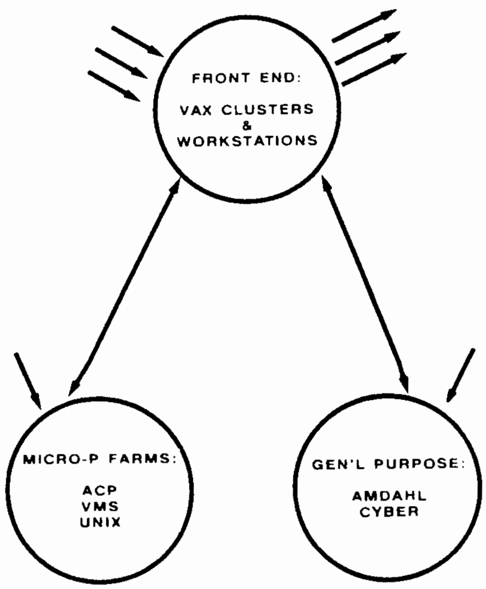


Fig. 1. Schematic representation of the three prongs of the Fermilab approach to scientific computing. Not fully shown are the interconnections and common file-serving components currently being implemented.

large-scale scientific computers, including the new Amdahl 5890 and outgoing CDC Cybers. Taken together, this triple platform provides over 1000 MIPS (millions of instructions per second) of computing power applied directly to physics research (Table I and Fig. 2). The increase in this power has been literally exponential, increasing a factor of two every six months during the most recent period (Figs. 3,4, page 4). Not only has the CPU power grown enormously, but the requisite numbers of disks and various types of tape drives and special input/output devices have grown, too. These must necessarily stay in balance with the CPU in order to make effective use of the tremendous raw computing power. Having the three platforms for computing has allowed these resources to be allocated in the most effective way, matched to the specific role of each of the platforms.

The microprocessor farms have been the major contributor to the growth of raw CPU power. During this period, the original Advanced Computer Program (ACP) system was turned into a computing center object. There are currently six ACP production systems in use. Seven different experiments have made use of these production systems, with as many as five experiments' programs operating at one time. This has required development systems, management software, and interfacing to the operators who mount the magnetic tapes of raw data. In addition, a great number of tools (e.g., librarian and linker routines) were written by Computing Department personnel. A beginning has been made on extending these capabilities to UNIX-based microprocessor farms in anticipation of using the newly available RISC (Reduced Instruction Set Computer) technology. One such farm, a commercially available system from Silicon Graphics, is in use by E-769 to help reconstruct the 10,000 nine-track tapes of raw data collected in the last fixed-target run.

The bulk of the new computing hardware and software resources were provided as part of the Central Computing Upgrade Project, first funded in FY86. However,

Table I. Scientific Computing at Fermilab
(Over 1000 MIPS of computing available)

A. LOAD:
700 tape mounts/day
3400 interactive terminal sessions/day
2000 batch jobs processed/day
500 users logged on in afternoon

B. CYBER COMPLEX:
approx. 30 MIPS (1-dual CPU 875)
36 gigabytes disk
20 tape drives

C. VAX COMPLEX:
approx. 65 MIPS (2 clusters)
65 gigabytes disk
20 tape drives

D. WORKSTATIONS AND LOCAL AREA VAX CLUSTERS:
Over 120 nodes, in 14 LAVC's
approx. 125 MIPS total
Additional 140 nodes in 12 LAVC's for mostly controls and engineering roles

E. MICROPROCESSOR FARM SYSTEMS:
6 ACP production systems
440 nodes total, approx. 312 MIPS
5 ACP development systems
46 nodes total, approx. 30 MIPS
CDF microVAX farm
22 nodes, approx. 60 MIPS
2 UNIX farms (physics + E-769)
approx. 320 MIPS

F. AMDAHL SYSTEM:
approx. 120 MIPS (5890/600E)
90 gigabytes disk
24 tape drives

G. Connected via hyperchannel, Interlink /DECnet intercomputer networks

H. CURRENT TAPE LIBRARY:
over 180,000 reels

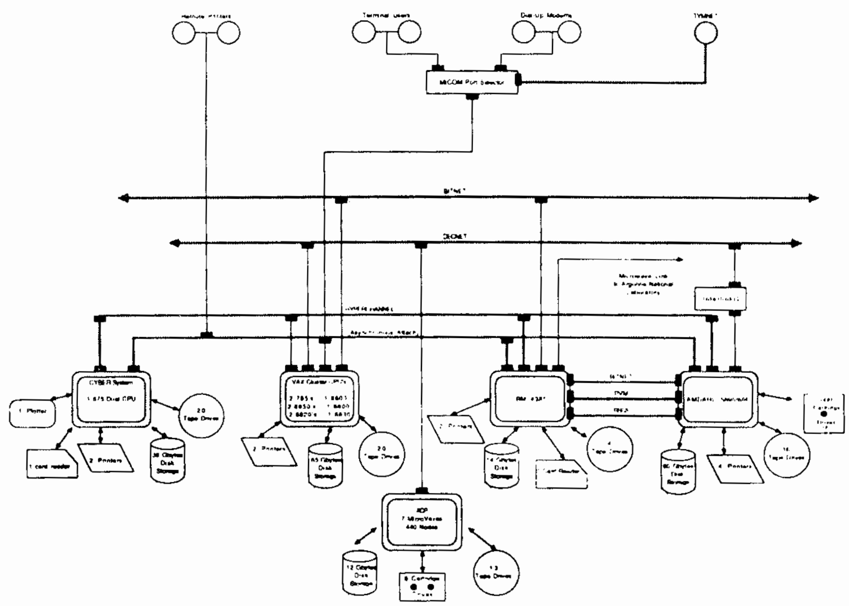


Fig. 2. Detailed schematic drawing of the computers in the Central Computing Facility at Fermilab.

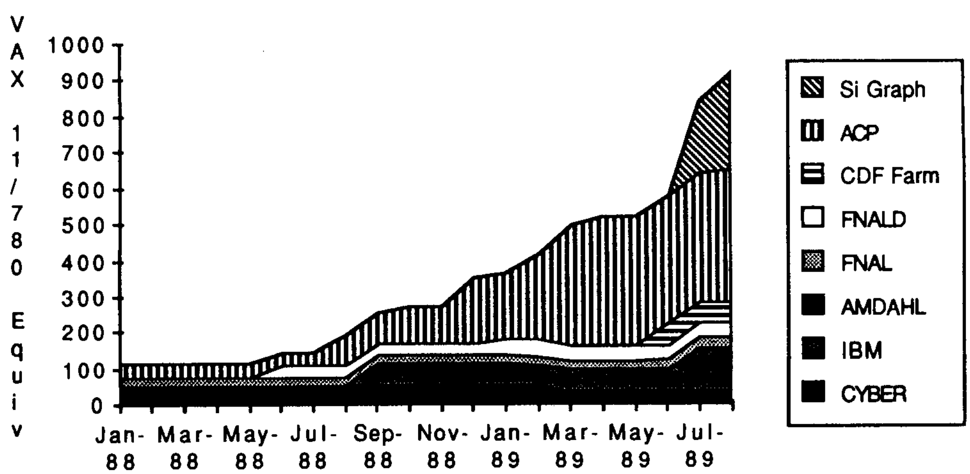


Fig. 3. Growth of the Fermilab CPU power over the last year and a half, showing the growth by system type.

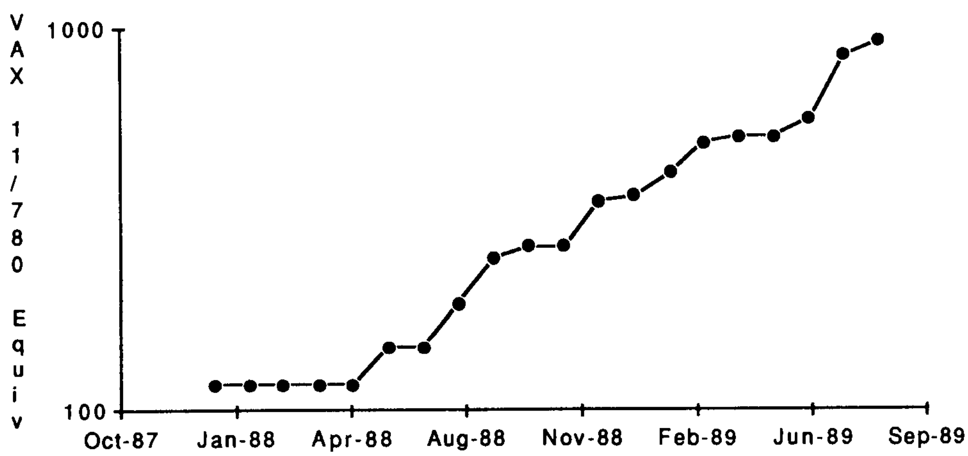
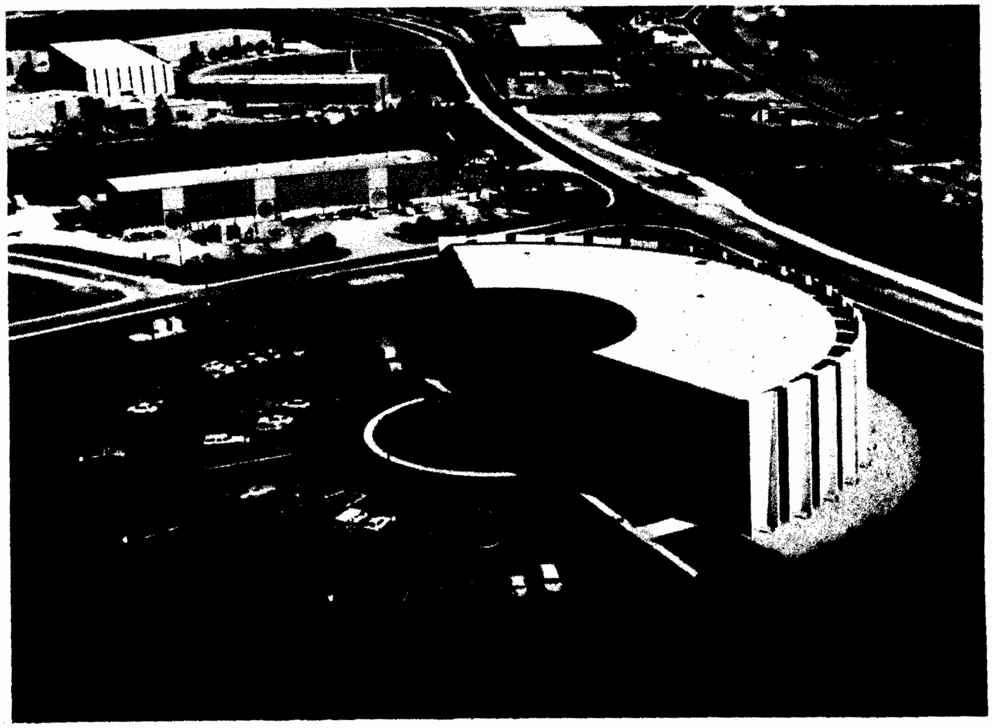


Fig. 4. Data from Fig. 3, plotted on a semi-logarithmic scale, showing the exponential growth of CPU power since May 1988.

a significant fraction of the resources have also come from the regular Computing Department equipment funds and from strong user groups. These user groups have wanted to increase the resources available to them at, and supported by, the Laboratory. In particular, the Japanese collaboration on CDF has added to the VAX cluster, many CDF institutions have provided nodes for a microVAX farm, and the INFN (Italy) has provided the base for one of the ACP farms.

In addition to its support of scientific computing at Fermilab, the Computing Department has continued to provide hardware and operating system support to the Business Section. This support of administrative needs is currently based on an IBM 4381-class machine. This administrative system has been upgraded several times in the last five years in support of the increased load of business activity at the Laboratory, the wider availability of data and reporting, and the development of wide-spread interactive data processing in this area.

In order to house all of the equipment associated with this tremendous growth, we have added the architecturally elegant new Feynman Computing Center (dedicated December 2, 1988, and shown in the photograph below) and additional operators and supervisors.



(Fermilab photograph 88-1033-5)

Aerial view of the Feynman Computing Center at Fermilab.

While the major part of our central computing resources are located in the Feynman Center, an increasing amount of computing resource is distributed elsewhere on the site. Not only is the hard-copy output from the computer available at many sites with traditional and laser printers, but the growth of local-area VAX clusters and workstations attached to the Center has been enormous (Fig. 5, page 6). In order to help support these local-area VAX clusters, the Computing Department has initiated an extensive program of help for the planning, installation, system management, maintenance, software distribution, and consulting with end users of these systems.

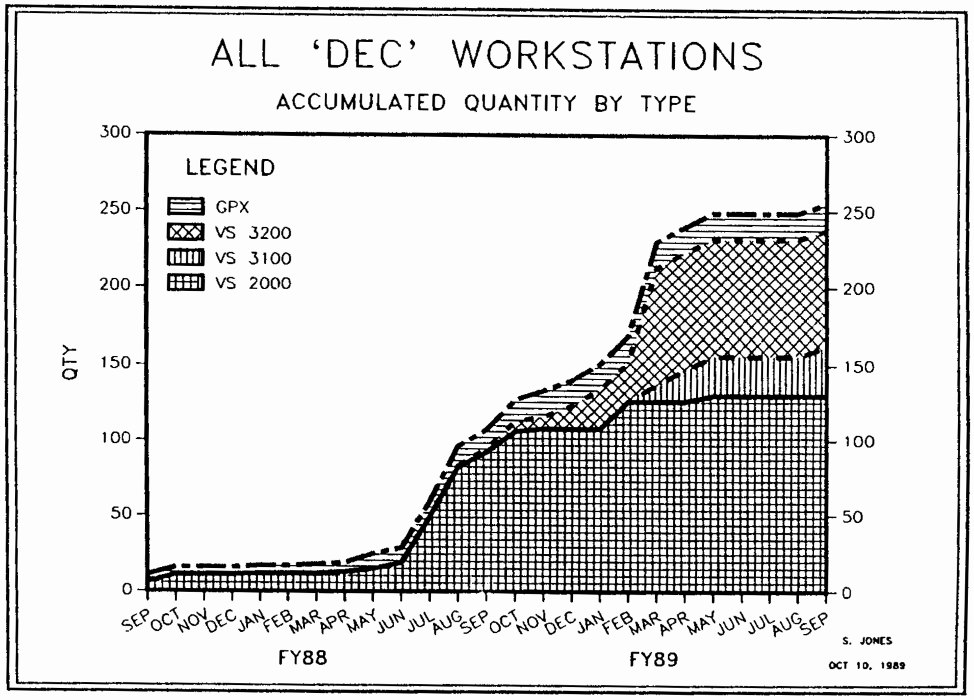


Fig. 5. Delivery dates of the most common workstations at Fermilab. The graph documents a very rapid growth in DEC workstations in FY 1988.

One of the less visible parts of distributed computing at Fermilab is that done by the engineering communities. The mechanical engineering community has coalesced around a common and unified approach, well positioned for centralized support. This has resulted in one of the major growth areas for the Computing Department. The support includes assistance with hardware and software selection, maintenance, and end-user consulting. When considering the use of workstations networked to data bases, file servers, and the central facility, the engineering computing is one of the most architecturally complex environments at the Laboratory. Yet it provides one of the most transparent environments to the end user.

In a sense, the most widely distributed computing organized by the Computing Department has been in support of several summer studies, most recently the Breckenridge Workshop on Physics at Fermilab in the 1990's. In conjunction with loans from several commercial vendors, the Department supplied a mini-Fermilab-computing environment, including access to the Fermilab site and wider HEPnet community. This was arranged with a minimum of funding, but provided a new high in capability for a remote high-energy physics location, the beautiful Rocky Mountains of Colorado.

Another innovation of this period has been the appearance of video technology for recording on 8-mm cassettes. The Department has taken a leadership role in developing applications of this technology. That role has included testing of the drives and various interfaces to computers, problem resolution in conjunction with the equipment vendors, and implementation of production copying facilities. At this point, video media are in use as input to microprocessor farm systems and as a distribution media for data summary information. Recently, the Department has been actively developing options for primary data recording from VME and VAXes at experiments and the capability of using these media on the VAX Cluster and Amdahl systems. Many problems remain to be resolved in this last area as the use of this capability is extended beyond the original disk backup function for which it was marketed. By the end of the last Collider run, CDF had started recording raw data on video media. It is anticipated that several experiments in the next fixed-target run will make use of the media as well. One experiment, E-769, has taken an approach which would have been unthinkable before the advent of this kind of cost effective data recording. Finally, we have just seen the last seven-track tape drives removed from the computing center. Can the last punched card be far behind?

Networking and Communications

The three-pronged approach to computing problems can only reach its full potential with the existence of a high throughput and transparent interconnection among the components. These are provided in the Fermilab system by DECnet and TCP/IP connections between the front ends and the microprocessor farms, and by Interlink hardware and software between the front end and the Amdahl. The Interlink product makes the Amdahl VM system appear as a DECnet node. Several experiments are using this feature in a fundamental way as they create their computing environment. Hyperchannel capability is also available among the larger systems.

Beyond the interconnected computers is a vast system allowing user access from next door and around the world. Fermilab has on site a segmented, but site-wide, Ethernet utilizing over 30 bridges. It's aggregate length is over 12 kilometers. One part of it is the beginning of a much higher bandwidth fiber optic capability. Users can connect to any of the major computers at Fermilab and to over 500 other devices on site.

Externally, the community is almost universally connected by DECnet and BITnet. These comprise what we have colloquially called HEPnet; that is, the network for high-energy physics. HEPnet has become one of the world's largest networks. Fermilab is an integral part of this. Fermilab maintains connections to Brazil, Canada, Europe (CERN in Switzerland and the INFN in Bologna, Italy), and via Lawrence Berkeley Laboratory to KEK in Japan. Much of this HEPnet backbone has been enhanced to 56 kilobaud and upgrades to T1 (1.544 Mbaud) will be in place this fall. All together, Fermilab has 39 direct DECnet links to universities and other laboratories, including the backbone. HEPnet now reaches 1500 systems in the U.S. alone. The Laboratory is also connected to the NSFnet, allowing the National Science Foundation-funded users a growing connectivity with increasing bandwidth. The Laboratory's role in networking has been recognized as a leading one, and the Department of Energy recently asked Fermilab to take on the overall management responsibility for HEPnet.

Data Acquisition Computing and Electronics

The move away from the old PDP-11-based Bison data acquisition systems began with the advent of the VAX-Online software environment. This widely used combination of standardized hardware and software incorporates the first distributed data acquisition system; multiple computers, each focusing on a part of the readout or monitoring of an experiment and interconnected for uniform system control and user interfacing. Like its predecessor, VAX-Online focuses on a single computer architecture. However, data acquisition needs have also grown rapidly and the VAX-Online system is evolving into a more universal and much more powerful PAN-DA system. Some claim the new name is meant to describe a sense of soft and fuzzy well being, but the name is also intended to convey the universality of the new approach (PAN DA). This approach opens the architecture to VME- and FASTBUS-based processors, as well as the more traditional VAX computers. It incorporates multiple levels of parallelism in readout and data collection. It anticipates UNIX as well as supporting VMS computers and the pSOS kernel operating systems run on Motorola 68K-series computers. A first implementation of the full PAN-DA system is scheduled for use in one experiment of the next fixed-target run. Many other experiments will use various components of the PAN-DA tools kit in support of their online system. Several experiments anticipate recording raw data on the new 8 mm video technology devices.

The wide variety of computer equipment in both the data acquisition and central computing systems have expanded the complexity of the maintenance job tremendously. While the largest computers are maintained by commercial vendors, the Fermilab maintenance groups have been enlarged and reorganized to handle everything from the more traditional mini-computer and its peripherals to distributed computer systems based in VME. The maintenance also extends to the hundreds of microprocessor farm nodes and system responsibility for the ACP farms in the Computing Center. Finally, the Department maintains 13 varieties of IBM and compatible PC's and Apple Macintosh computers, totaling over 1100 units! Over 540 repairs and upgrades were performed in the last year alone.

Two years ago microVAX 2000s and 3200s were only plans and subjects of interesting rumors. Today there are 130 2000s and nearly 100 3000-class workstations at the Laboratory. Exabyte 8-mm tape drives and Wren V disks were unknown two years ago. Today Fermilab provides repair service on about 150 of each of them. The approach to troubleshooting has also changed dramatically. Formerly, PDP-11 problems were investigated by chasing bits with octal switch panels. Now computers in the computers, using firmware-based tools, allow analysis of most problems from a terminal. Thirty percent of the Lab's PDP-11's have been retired. Versatec printers and 800-bpi tape drives are finally history. Where removable media hard disks were the primary program loading media at experiments two years ago, they have been largely replaced by electronic network distribution of software.

Front-end standardized electronics support has also grown tremendously to satisfy the needs of the Tevatron-era experiments. PREP (the Physics Research Equipment Pool) has grown to over 34,100 units with an almost $28 million purchase price, half of this in the last five years. The inclusion of processors and sophisticated controllers in many of the modern systems has added to the complexity of the distribution and maintenance jobs associated with this pool of modules. Continuous training and the use of more-powerful computerized test stands and a foray into the world of artificial intelligence have marked the efforts to keep up. The Department has added expert system technology in a prototype test stand for one of the more complicated FASTBUS modules, the Segment Interconnect.

Maintenance of the equipment in PREP is only one part of the repair effort provided by the Department. Over 53,000 instruments of 2500 different types are beneficiaries of this support. In the most recent year, over 7500 repairs were completed, with only 1-1/4 percent of these requiring return to the original manufacturer.

Personnel and Organization

It was evident rather early that all of these increasing technologies and support requirements would demand a larger staff and an organization that allowed people to focus their attention on the most crucial problems. Figure 6 shows the growth in each of the major parts of the Computing Department over the last five years. Overall, the Department has grown over 50 percent, and just before the reorganization into the Computing Division, it was recognized that the ongoing parts associated with the old Computing Department should grow by an additional 25 to 30 people minimum.

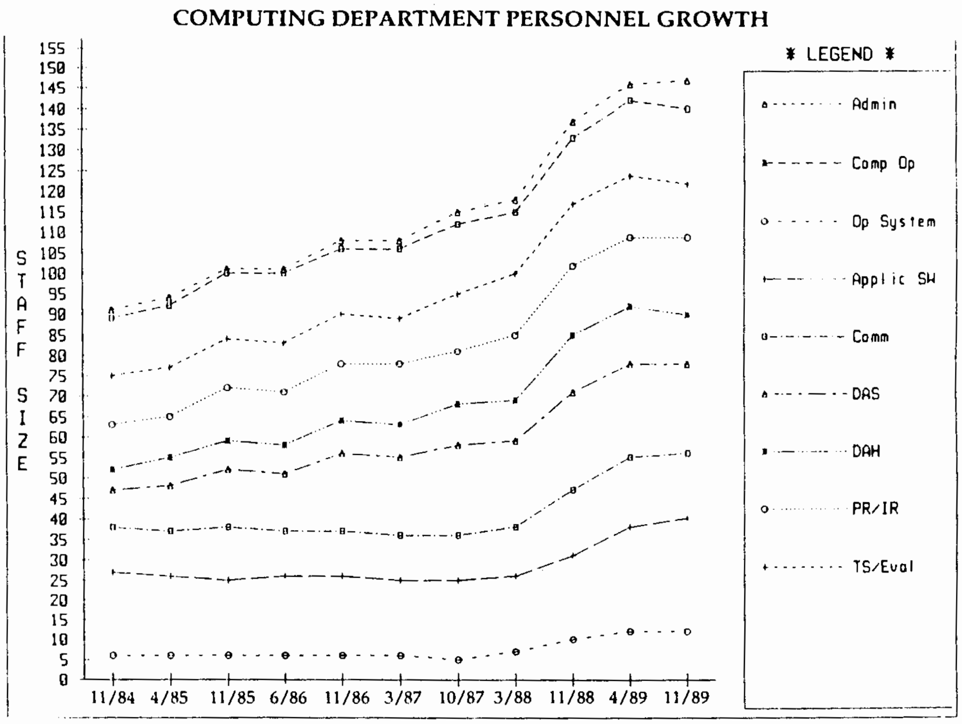


***Fig. 6.** Five-year growth of the size of the Computing Department by major activity area.*

The people in the Computing Department are only one of the human elements in the computing story. The end users are an equally crucial element. With this in mind, an aggressive program of consulting, training, and seminars has been an important part of the Department. The Department now publishes three newsletters and has organized a tutorial, seminar, or workshop at an average rate of one per week. Some of these training sessions have lasted as long as a week as well. These efforts are aimed at a wide audience, from the traditionally supported physics community to newer users from the office automation, engineering, and other communities. In addition, for the most rapidly changing areas, series of regular meetings with the most affected users have been set up. These include the ''Early Bird'' program for first use of the Amdahl system, meetings with engineering groups, and the newly instituted meeting to help users migrate off the Cyber systems and onto other available computing platforms at the Laboratory.

The number, size, and complexity of systems supported by the Computing Department has grown vastly faster than the number of people in the Department. Forging ahead has been possible only by the dedicated efforts and talents of the people who are so poorly represented by the statistics presented here. Based on the enthusiasm of these people and the demonstrated progress of the last five years, one can say that computing at Fermilab is poised for even greater success, success limited mostly by available funding and not by technological opportunity or talent.