LA--12326-MS

DE92 019191

The State of Computing at Los Alamos National Laboratory FY 1991

I. A. Agins
C. Slocomb
M. Trainor
D. Land

Los Alamos National Laboratory
Los Alamos New Mexico 87545

Table of Contents

THE STATE OF COMPUTING AT LOS ALAMOS NATIONAL LABORATORY FY 1991

by

I.A. Agins, C. Slocomb, M. Trainor, and D. Land

ABSTRACT

This study is an effort to provide quantitative data concerning the state of computing at the Los Alamos National Laboratory as of the end of Fiscal Year 1991. It includes information pertaining to the Laboratory's computing equipment inventory, costs associated with the acquisition and support of the Laboratory's computing efforts during that fiscal year, and information related to the Laboratory's central and distributed computing and networking capabilities.

The bulk of the data was obtained from the Laboratory's central property and financial databases. Additional information was obtained from the Computing and Communications Division's personal computer and network support organizations.

1. Introduction

The effectiveness of using computers in scientific research at Los Alamos has been well established throughout the history of the Laboratory. Indeed, computational science has become a cost-effective complement to theoretical and experimental science.

Historically, the demand for computers has been most acute in the areas of nuclear weapon design and other programs relating to national security. This is still true, particularly in the face of additional programmatic requirements relating to economic competitiveness, energy, and health, safety, and environmental issues. There has also been, in recent years, growth in a number of other research areas requiring the use of computers. At Los Alamos, as the research problems have evolved and changed, so has the computer environment that exists to support them.

It is of vital importance, for Laboratory planning, to monitor the state of computing at the Laboratory. The *purpose* of this report is to provide base-line quantitative information about Laboratory resources currently being used for computing for this planning process. This report represents the Laboratory's first attempt at approximating the resources involved. The approximations will become more accurate in subsequent years, as the methodology used is refined. The report should prove useful not only to Laboratory management but also to users of computers within the Laboratory and external to the Laboratory.

This report provides a presentation of the data collected regarding computing resources. These data are presented as follows:

- o Section 2 provides information relating to the Laboratory's distributed computing.
- o Section 3 is a presentation of the Laboratory's central computing capabilities.
- o Section 4 presents an overview of networking activities.

- Section 5 is a breakdown of the Laboratory's external costs related to computing.

- Section 6 discusses the in-house personnel costs involved in the support of Laboratory organizations' computing efforts.

There is also an appendix detailing the data acquisition sources and methods employed for this study. The reader is encouraged to study this appendix in order to better understand these methods and the limitations imposed by the nature of the available data.

Some of the highlights of this study are as follows:

- There are approximately 17,600 total computers at Los Alamos, of which 15,000 are personal computers.

- The popularity of scientific workstations is increasing, while the acquisition of minicomputers continues to decrease. There are now almost 1700 scientific workstations at the Lab. Of these, almost 1300 are Sun Microsystems units.

- The acquisition cost of the computing equipment in the Laboratory's property database is approximately \$348M. Approximately \$172M of that total is C-Division property. Twenty other divisions control at least \$3M each of property controlled computing equipment.

- The Laboratory spent over \$94M on computer-related procurements in FY 1991. Of this total, approximately \$22M was spent on personal-computer-related hardware and software.

- In-house personnel costs in FY 1991 for the support of the Laboratory's computing efforts were estimated at almost \$54M.

2. Distributed Capabilities

Over the past decade, there has been extensive growth in the populations and capabilities of small computers, primarily personal computers and scientific workstations. The extraordinary growth in the numbers of these machines is reflected in Figure 2.0.1. Although this chart shows the cumulative numbers of all computer systems at the Laboratory, the principal growth is in small distributed computer systems. One should note that these figures were derived from the Laboratory's property database at the end of FY 1991. Therefore they more properly reflect the acquisition dates of systems <u>currently</u> in the inventory, as opposed to a delineation of when all computers (including those no longer at the Laboratory) were acquired. Nevertheless, these figures are indicative of the extensive growth in the number of systems.

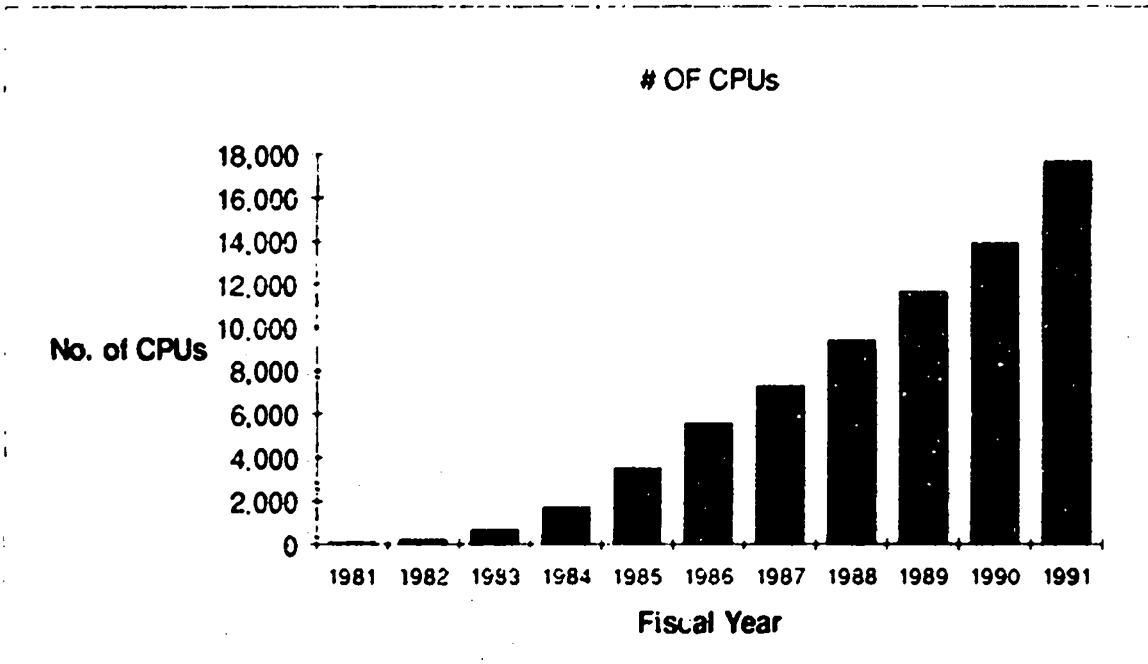

Figure 2.0.1. Number of computer systems at Los Alamos.

Technology developments have driven this growth in computer systems at the Laboratory. As mini- and microcomputers became more available, Laboratory staff found them both adequate and convenient for many applications. The small computer systems, from workstations to minisupercomputers, provide many functions that are not available to, not appropriate for, and not easily provided by central supercomputer systems. They have increased users' productivity because of the combination of a rich software environment with a good user interface. They are used for a wide variety of tasks including program development, calculations, data collection and analysis, and document and presentation preparation.

2.1. Distributed Processors

A minicomputer is defined here as a central machine built to be used by multiple users and less capable than a mainframe or supercomputer. They typically include a range of VAX machines, from the powerful 8000-class down to the much smaller microVAX.

Figure 2.1.1 shows the minicomputer acquisitions at the Laboratory. As you can see in the graph, there is a normal distribution with the peak being in the mid 1980s. The sudden rise in 1991 is an artificial rise in that it includes the new microVAXs (cost $5K–$8K). These mini VAXs can be seen as consistent with the migration trend towards more distributed computing.

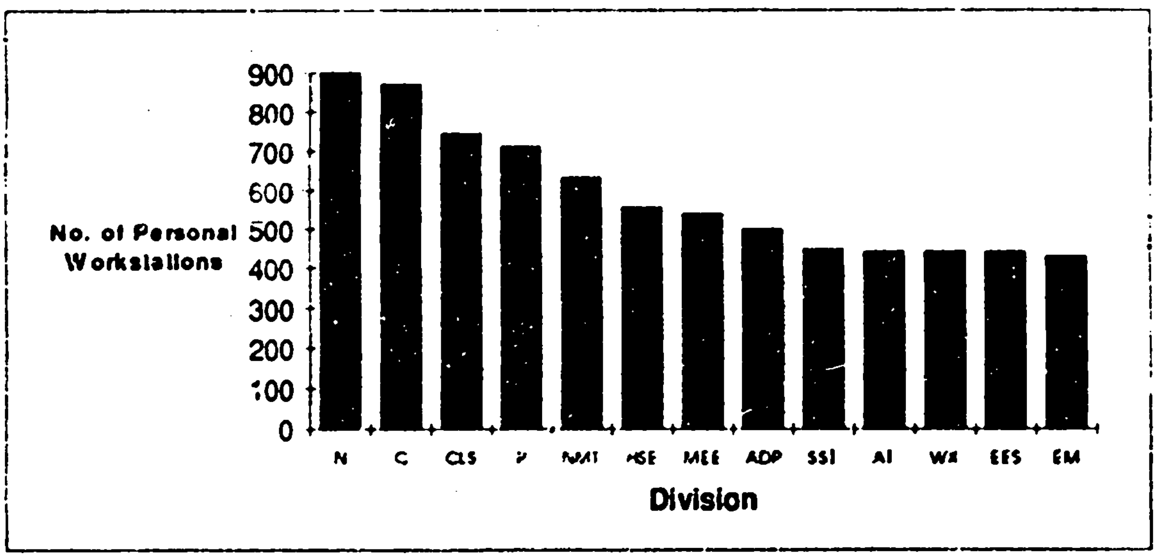

Figure 2.1.1. Minicomputer acquisitions.

There were a total of 555 minicomputers at the Lab at the end of FY 1991. Figure 2.1.2 depicts the divisions owning 10 or more of these systems. The large number of minicomputers in C Division can be attributed to their use as communications controllers within the Integrated Computing Network (ICN); others form the hardware platforms available for Laboratory-wide use.

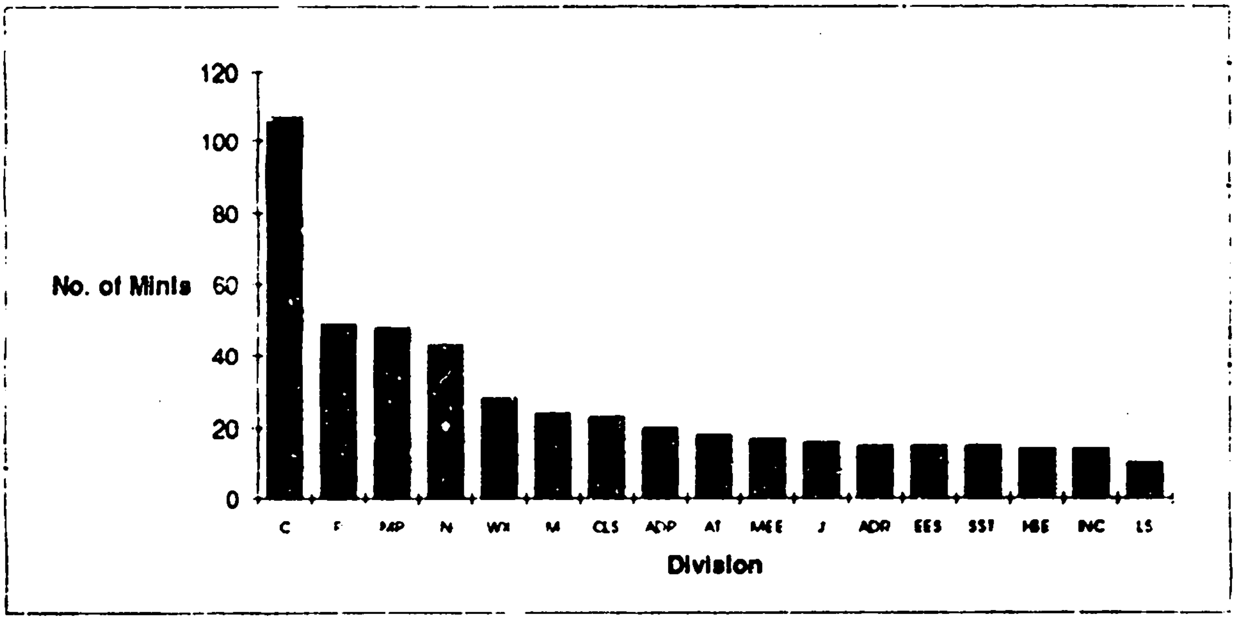

Figure 2.1.2. Divisions with 10 or more minicomputers.

Figure 2.1.3 shows the dollar value, by division, of the minicomputers for those divisions with over $1M in installed systems. Note that these figures only represent the acquisition cost of the central processing units (CPUs) and do not include the cost of the peripheral equipment such as disk and tape drives. Peripheral equipment costs can easily equal or exceed the cost of the CPUs.

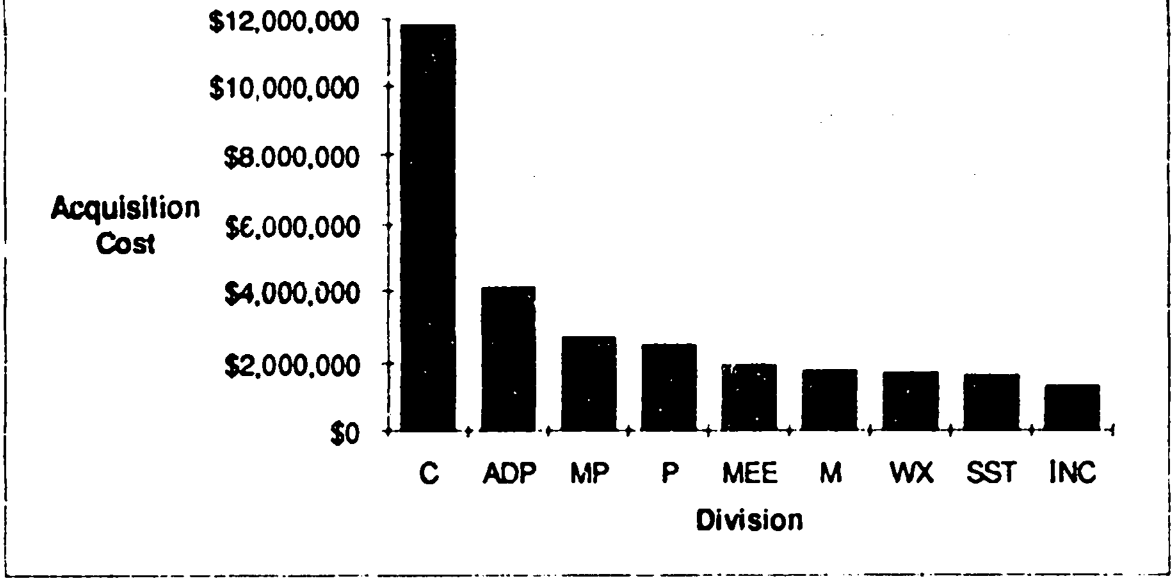

Figure 2.1.3. Acquisition cost of minicomputers by divisions with >$1M in CPUs.

It is interesting to note in Figure 2.1.3 that ADP Division ranks eighth in the number of minicomputers but ranks second in the cost of those systems. This reflects the need for that division to use large, robust systems for both the development and delivery of Laboratory-wide administrative applications. Conversely, P Division owns a large number of more modest configurations, indicative of the need to distribute smaller-scale computing over a larger number of decentralized applications.

2.2. Workstations

In 1982 and 1984 procurements for desktop workstations began at Los Alamos. The first of these procurements was for personal computers, the least expensive machines, and the next was for more-expensive and more-capable machines called "scientific workstations." These machines became very popular at the Laboratory because of their user-friendly software (spreadsheets, word processors, and drawing tools), their responsive computing capabilities, and their low cost. They have proved useful to a broad spectrum of the Laboratory population.

2.2.1. Personal Computers

Perhaps no other computing tool has found as wide a range of application as the personal computer. These relatively inexpensive systems are currently used for applications ranging from data acquisition and process control and monitoring to financial analysis, terminal emulation, and graphics preparation. The advent of the personal computer, in fact, was the primary cause of the demise of the shared-logic dedicated word processor at the Laboratory in the mid 1980s.

From the initial establishment of the IBM Personal Computer as the Laboratory standard in 1982, the number of personal computers has grown steadily to over 15,000. This number is almost twice the number of regular employees at the Laboratory, but includes personal computers for contract employees and for laboratory and off-site use. Figure 2.2.1.1 shows when the personal computers currently in the property database were acquired. As can be seen, the Apple Macintosh has grown in popularity over the last few years. This can be attributed to its icon-driven user interface, ease of use, and high-quality output capability. IBM systems and IBM clones together account for approximately two-thirds of the personal computer acquisitions because of the breadth of applications software available for them, the "standard" nature of their interface interconnects, and the large number of peripheral devices that are designed for easy attachment. Improvements to the user interface for these systems, as exemplified by Microsoft Windows, have improved their usability, thus placing them almost on par with the Macintosh interface.

Another interesting trend is an increase in IBM clone system purchases which can be ascribed to lower cost, improved performance, and acceptable reliability. Systems from over 200 different manufacturers are represented in the "Other" catagory in the figure below.

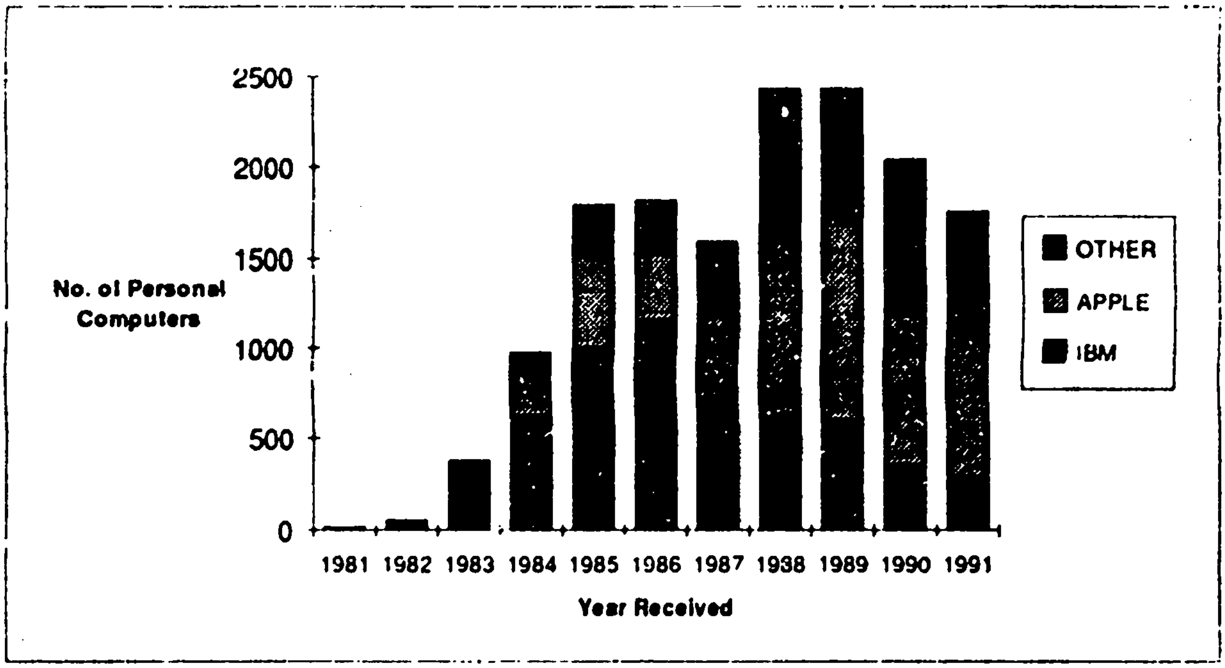

Figure 2.2.1.1. Personal computer growth (15,003 total units).

The decline in the total number of systems acquired from the high of 1988–1989 may be attributed to a combination of budgetary limitations, the reutilization and/or upgrading of older systems, and approaching saturation (i.e., any employee who needs one already has one).

Figure 2.2.1.2 shows the divisions owning at least 400 personal computers.

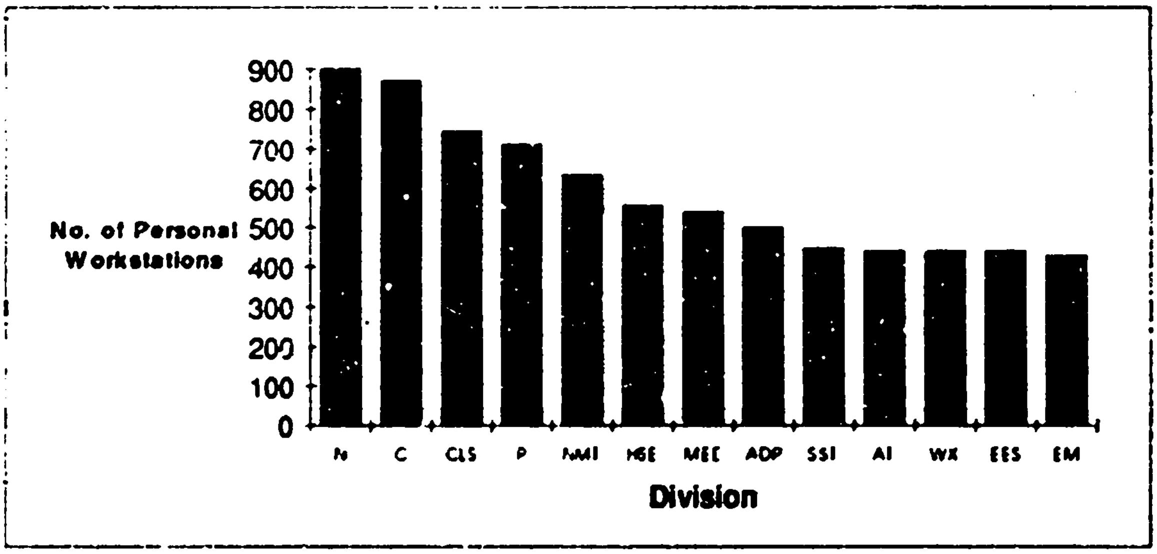

Figure 2.2.1.2 Divisions owning at least 400 personal computers

Figure 2.2.1.3 shows the acquisition value of the personal computers by division. These figures represent the value of the system units only. Unlike the distributed processors, however, these system units normally include disk drives. Monitors and other peripheral devices are not included in this figure.

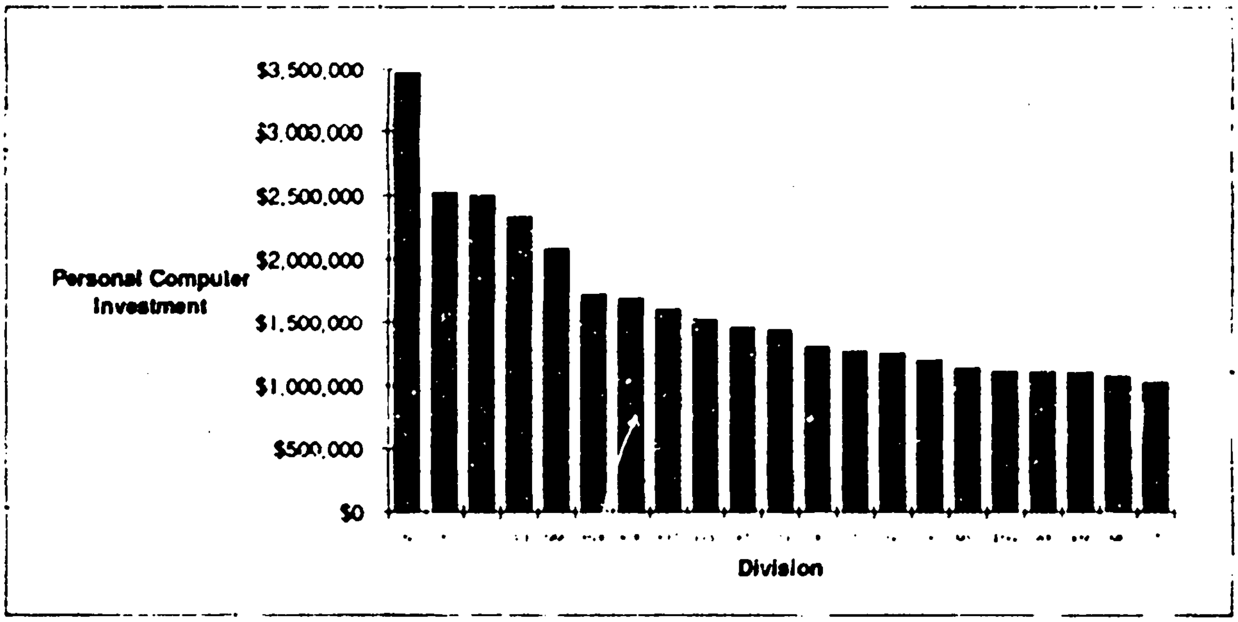

Figure 2.2.1.3. Divisions with at least $1M invested in personal computers.

Again, it is worthwhile to note that the division with the largest investment is not necessarily the division with the greatest number of units.

2.2.2. Scientific Workstations

Scientific workstations differ from personal computers in their greater capacity, capability, and availability of technical applications. They tend to operate under some variant of the UNIX operating system and generally operate as part of a local area network (LAN). Scientific workstations in this study are considered general-purpose in nature, as opposed to high-performance workstations that are optimized for a particular task (e.g., computer-aided design, artificial intelligence, graphics).

Note in Figure 2.2.2.1 that the number of scientific workstations has increased significantly in the past year. This rise in workstations is a result of the advancement of the technology as well as a result of the nature of funding. Along with the new workstations, however, comes the responsibility for systems management and maintenance. Thus, the need for centralized computing on the supercomputers is becoming more special purpose, for the very large grand-challenge type of problems.

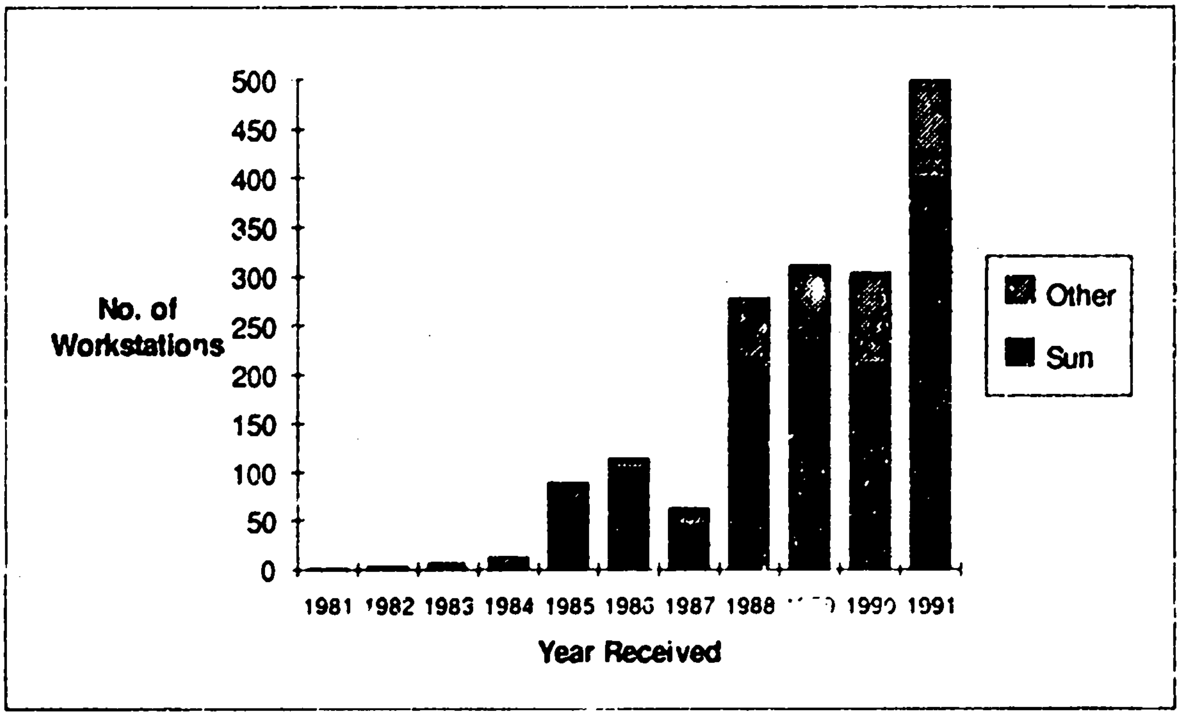

Figure 2.2.?.1. Scientific workstation growth (1,682 total units).

Figure 2.2.2.2 below shows those divisions with more than 50 Scientific Workstations.

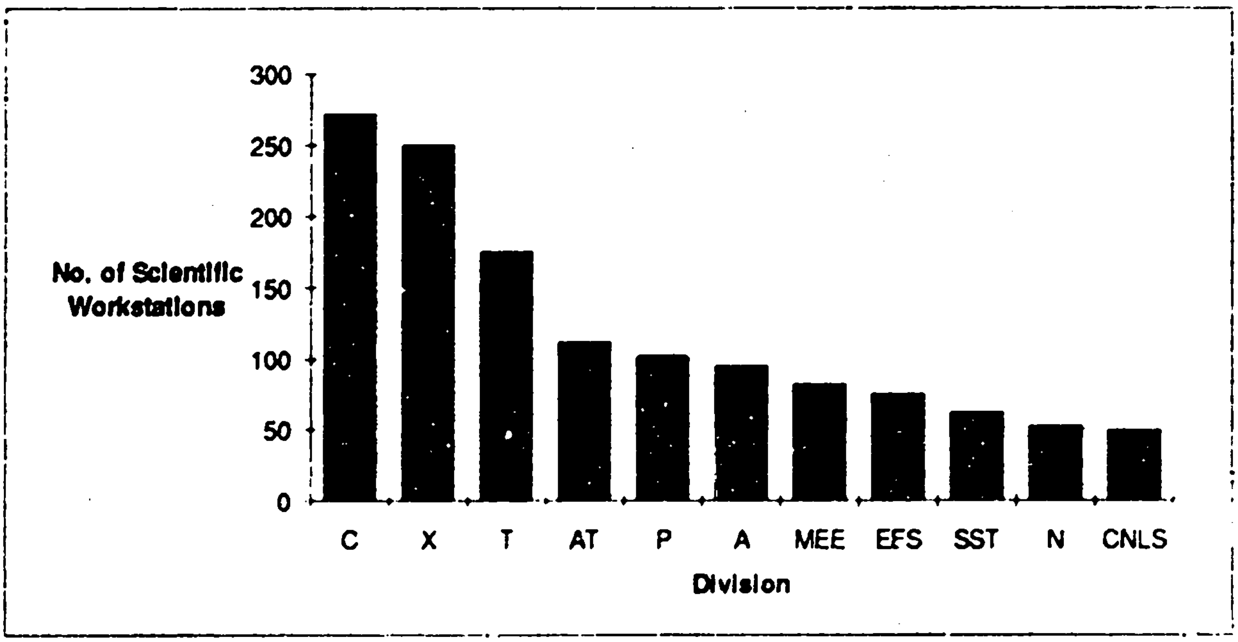

Figure 2.2.2.2. Divisions with > 50 Scientific Workstations.

Many of these divisions, such as X Division, also use substantial amounts of supercomputer cycles through the ICN; others, such as P, MEE, and SST, have large investments in minicomputers; and still others do not appear to be heavy users of either the central resources or distributed processors. It appears therefore, that for some, scientific workstations provide a complementary set of computing tools to those found on the other systems and are used in conjunction with them. For other organizations, the capability and capacity of these workstations is sufficient and cost-effective for their computing requirements.

Figure 2.2.2.3 shows those divisions with investments of at least $1M in scientific workstations. Again, these figures only include the cost of the system units. In many cases, these units have little or no disk capacity; they rely instead upon file servers connected via LANs.

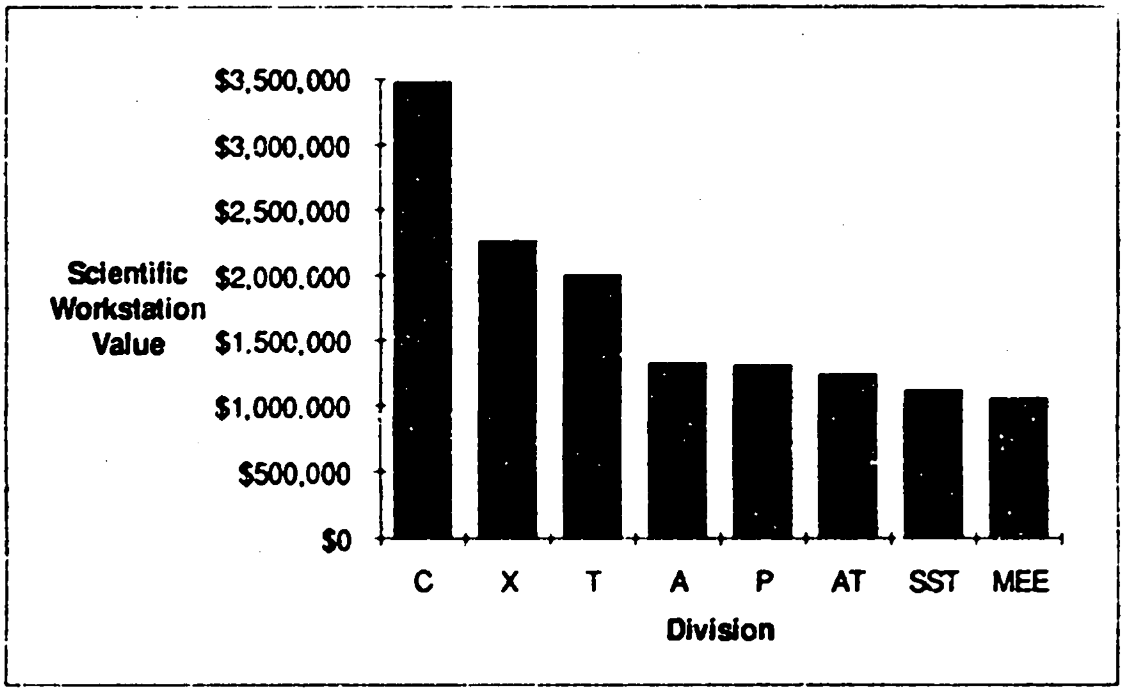

Figure 2.2.2.3. Divisions with at least $1M invested in scientific workstations. (Lab total for scientific workstation CPUs is $20.8M.)

Note that the most computationally demanding users tend to acquire high-end workstations, as do users who depend upon scientific workstations as their primary or exclusive tool.

3. Central Computing Capabilities

Los Alamos National Laboratory maintains the most powerful scientific computing center in the world with more than 103 CRAY-1 equivalents available to users. The center provides a range of services to clients throughout the nation via a national computing network.

The technology supporting supercomputing has grown rapidly. Over the past 40 years, some applications have had an increase in computing speed of 10 orders of magnitude. This increase in computing speed has come in almost equal parts from advances in computer hardware and from more sophisticated numerical algorithms. In the past few years, it has become increasingly apparent that new algorithms and massively parallel computer architectures offer the greatest potential for dramatic performance improvements. In addition, with increased speed has come the recognition of other significant requirements for large-scale computing: these include real-time visualization, which allows scientists to "see" physical phenomena and more intuitively understand the physics; high-speed networking, which provides convenient and powerful computing capability at the scientists' desks; and software that facilitates development of sophisticated scientific computer programs. Figure 3.1 depicts major supercomputer acquisitions since 1950.

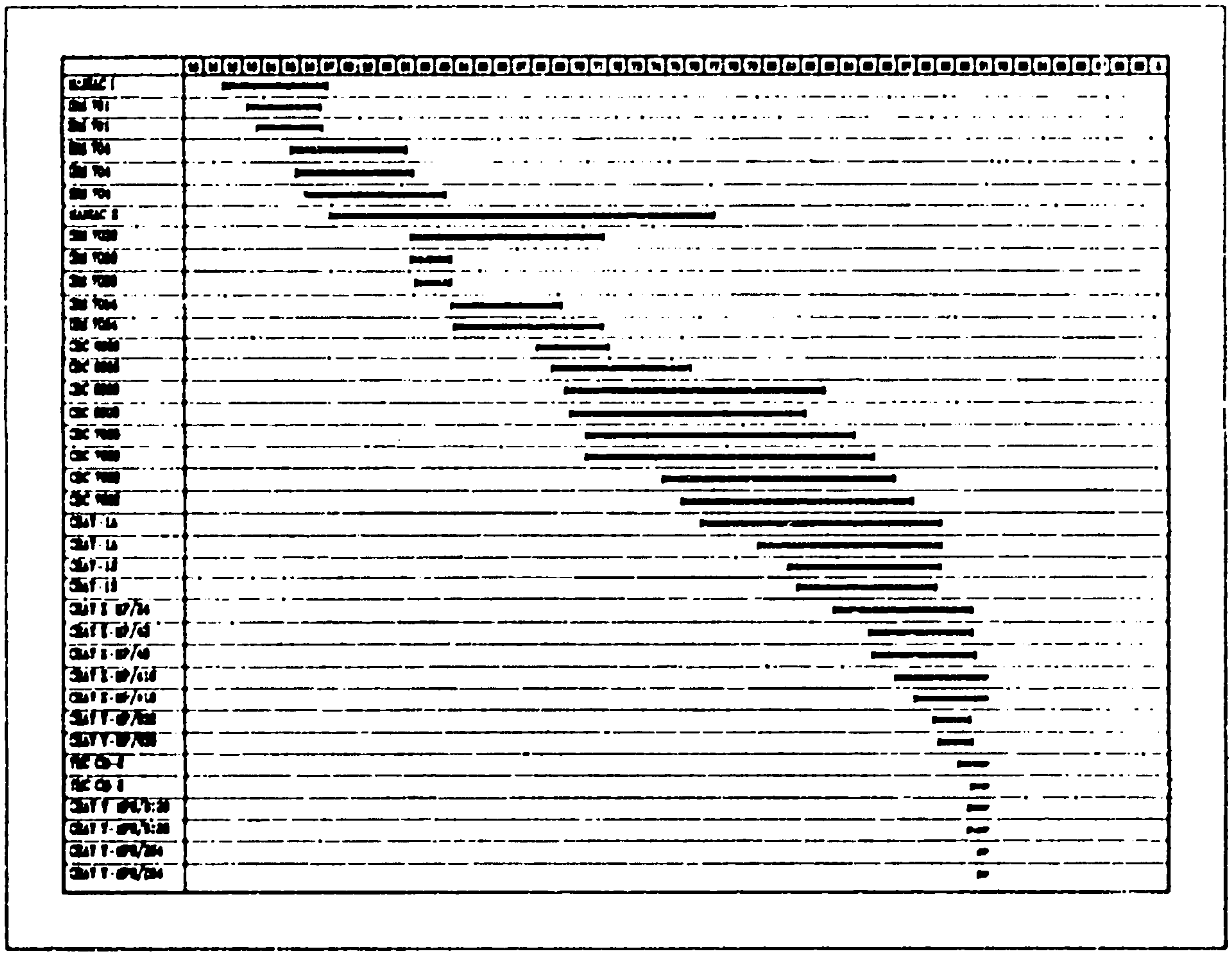

Figure 3.1. LANL Supercomputer acquisitions since 1950.

Increases in total computing power must be accompanied by corresponding increases in all the auxiliary capabilities that compose the Central Computing Facility (CCF) such as the Common File System (CFS, an archival file storage system), the Print and Graphics Express Station (PAGES, an output server), the Facility for Operations Control and User Statistics (FOCUS, a production control system for the supercomputers), and the network that interconnects central services to ensure there are no performance bottlenecks that restrict the overall capability.

3.1. Capacity and Usage

Figure 3.1.1 shows the growth of the Laboratory supercomputing capability in CRAY-1 equivalents.

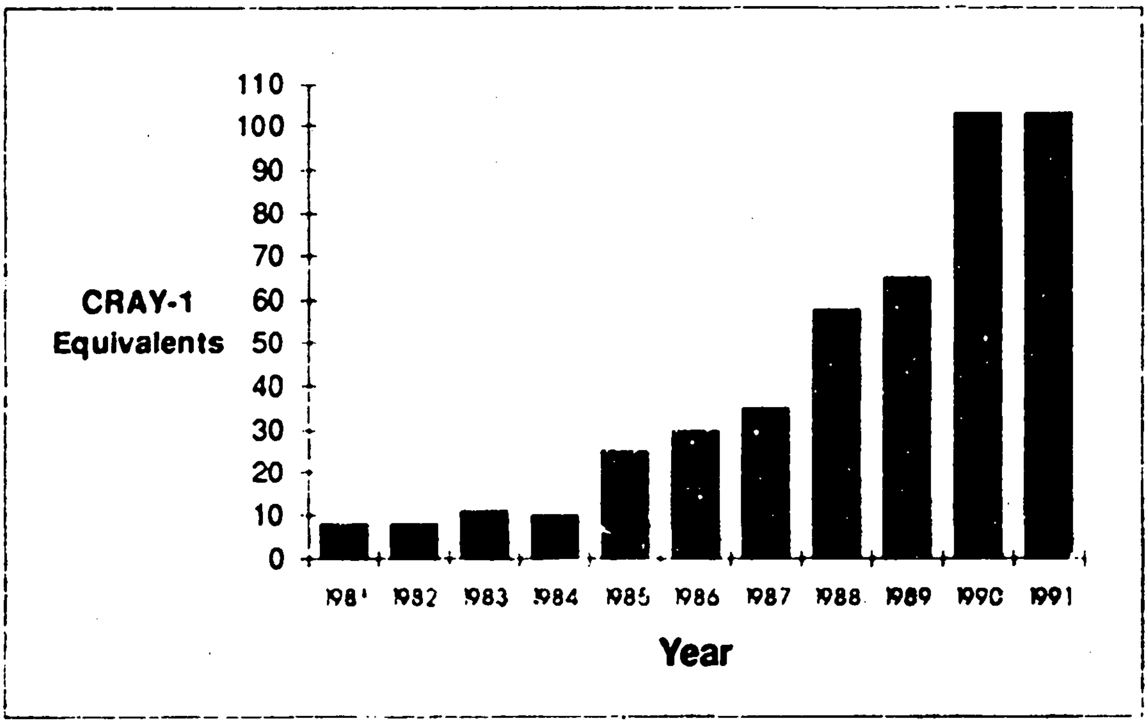

Figure 3.1.1. Los Alamos supercomputing capacity in CRAY-1 equivalents.

During the period since 1985 the computing capacity has increased by a factor of 3 and the load on the CFS and the network has increased slightly faster. This increase in the network traffic is shown in Figure 3.1.2.

In FY 1991, C Division devoted its available supercomputer line item funds to increase the memory and disk storage capacity of the two existing Thinking Machines Corporation model CM-2 computers. No new systems were acquired in that year, which accounts for the lack of significant increase in supercomputer capacity.

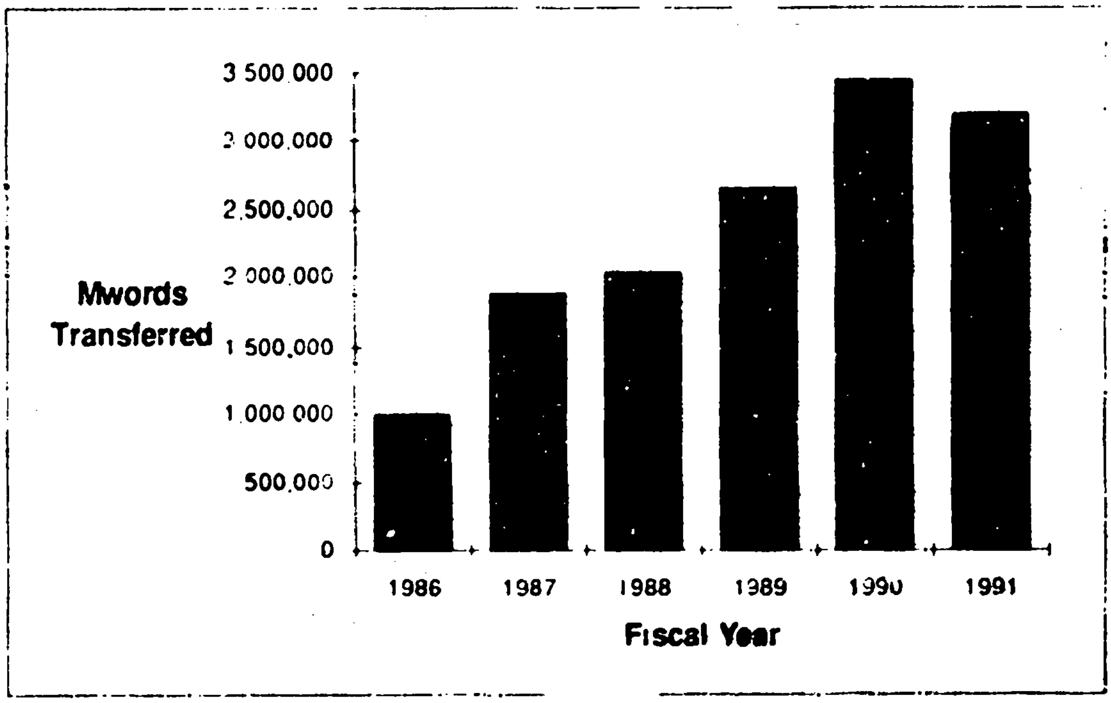

Figure 3.1.2. Network u[illegible] within the CCF.

This increase in network load reflects the increase in both the computing capacity and memory available on the supercomputers. In 1985 the maximum memory on the Crays was 4 Mwords (1 word = 8 bytes). Today, the smallest X-MP has 4 Mwords and the largest, an 8-processor Y-MP, has 128 Mwords. The Thinking Machines Corporation CM-2 system, with the largest memory configuration at the Laboratory, has 1 Gword. In addition to the effect on the ICN services, this memory increase has a profound impact on the kinds of problems that can be addressed by Laboratory scientists.

How is the central supercomputing capability used? Figure 3.1.3 displays the usage by large program affiliation.

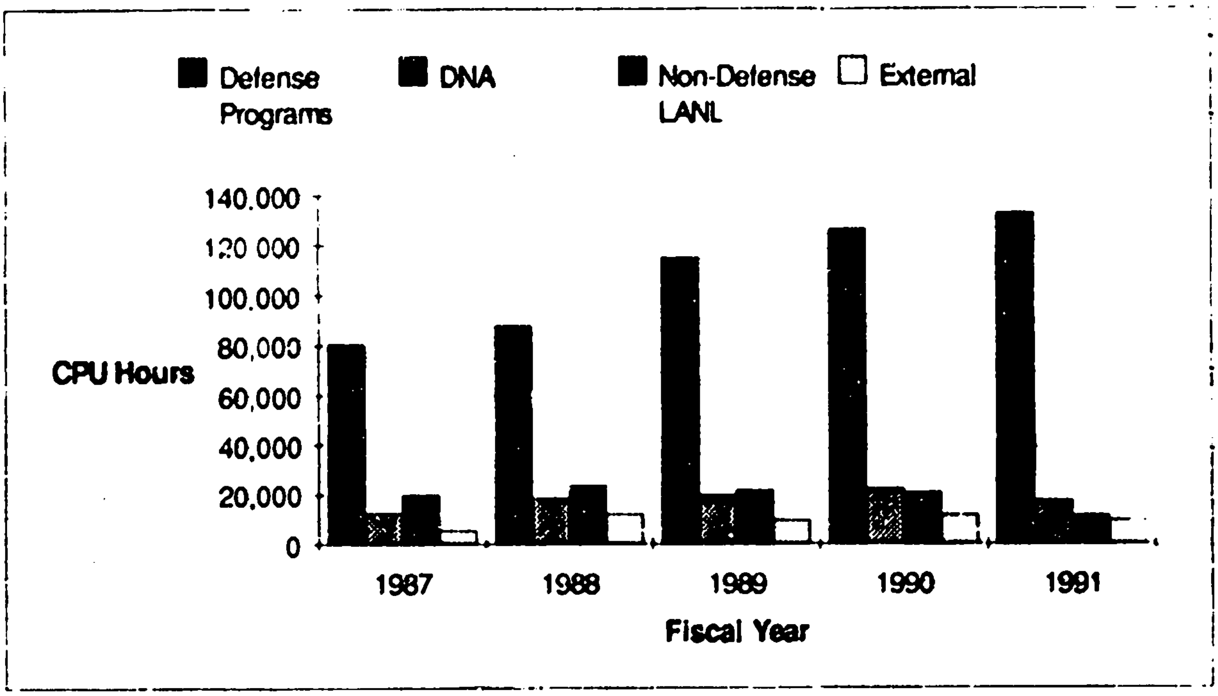

Figure 3.1.3. CCF supercomputer usage in central processor hours.

The nuclear weapons program, a subset of defense programs, continues to be the largest user of supercomputing at the Laboratory.

3.2. The Cost of C-Division Services

In FY 1991, $53M was spent for direct computing services from C Division. This money was used by C Division to support the Laboratory's CCF and [illegible]ratory-wide networking facilities. This figure has remained essentially unchanged since FY 1989.

In FY 1991, 56% of the computing revenue came from the weapons program and 17% came from the Defense Nuclear Agency (DNA) (see Figure 3.2.1). These figures are within a few percentage points of the FY 1989 and FY 1990 percentages. All but these two customers pay according to the services they use; the weapons program and DNA guarantee a level of support at the beginning of each year that covers their computing needs. Both DNA and the weapons program provide financial support for the acquisition of additional hardware. Much of the hardware acquired for the central facility has been funded directly by the weapons program. In FY 1988, DNA purchased a CRAY X-MP, which was installed in the central facility.

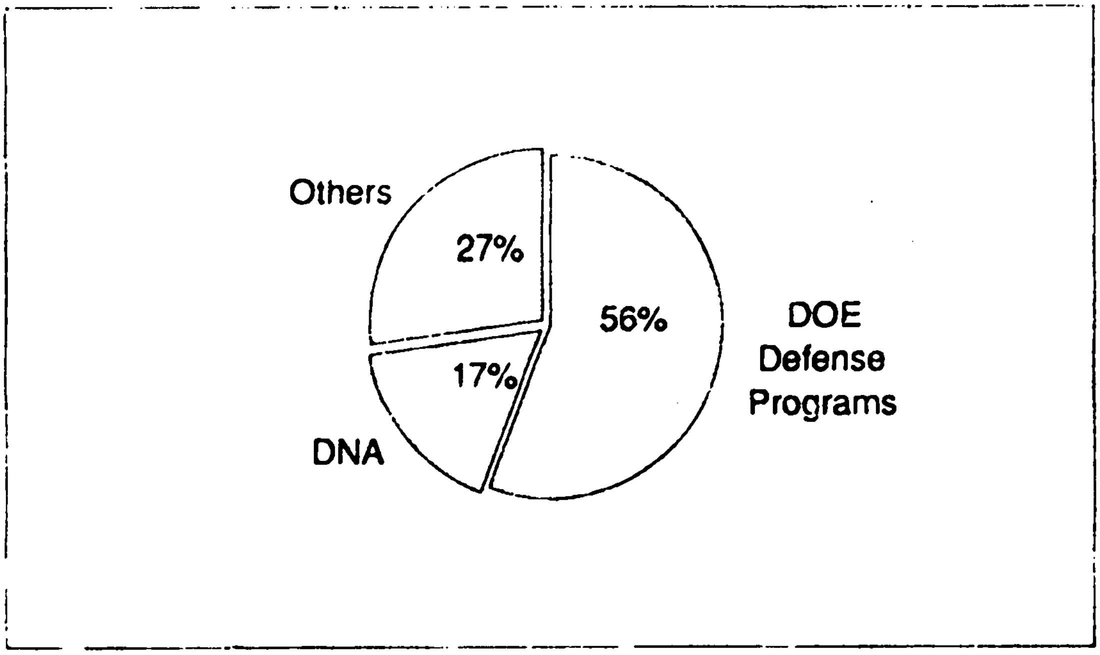

Figure 3.2.1. CCF recharge in FY 1991 was paid in the proportions shown.

The recharge system was originally developed to ensure that costs associated with the operation of the CCF were recovered. This system requires that users pay for central computing according to their use of the facility. The cost recovery in FY 1991 was split among the various resources as shown in Figure 3.2.2.

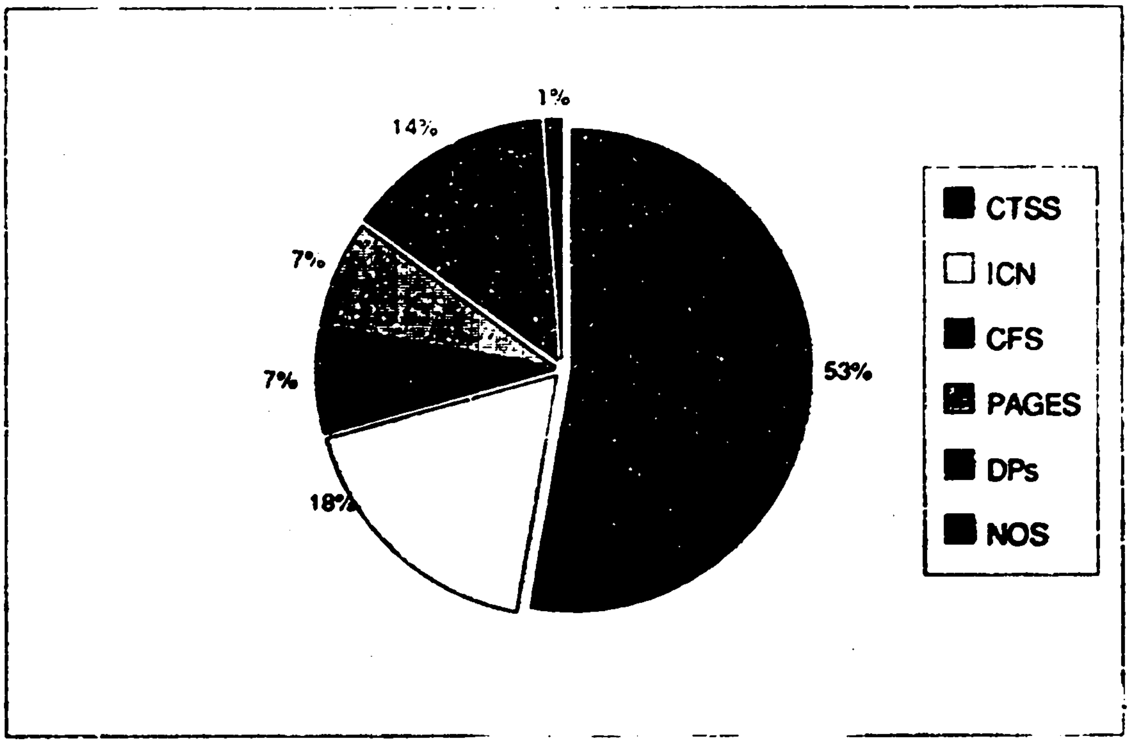

Figure 3.2.2. Cost recovery split among the resources.

The C-Division budget has remained relatively constant over the past several years, with computing expenses and revenues tracking closely since FY 1987 (see Table 3.2.1). With the changing technologies, however, there have been significant differences in how the budgeted monies have been allocated in the past few years. For example, the number of operators in the CCF has decreased because of improved automation. In turn, the budget for applications programming and network support has increased in response to the migration towards distributed processors.

Table 3.2.1. C-Division (ICN) Computing Expenses and Revenues

	FY 1987	FY 1988	FY 1989	FY 1990	FY 1991
FTEs	288	276	278	263	246
Expenses					
Salary	14.9	14.6	15.7	16.3	15.2
Burden	11.8	12.1	13.4	14.9	14.1
M&S*	11.9	11.2	10.6	9.5	8.7
Maintenance	8.4	8.1	7.6	7.2	7.1
Leases	10.5	12.0	10.7	8.2	9.7
Total Expenses	57.5	58.0	58.0	56.1	53.7
Revenue					
DOE Defense Programs	33.3	33.4	33.2	29.8	29.9
Other Internal	11.6	12.3	10.8	13.6	11.5
External	1.9	3.4	2.8	4.0	3.3
DNA	10.1	9.0	8.9	8.8	9.1
Total Revenue	56.9	58.1	55.7	56.2	53.7
Major Capital Equipment (DOE Nuclear Weapons Program)	13.7	14.0	13.9	12.4	9.4

*M&S is material and services.

4. Networks and Internetworks

Computer networks at Los Alamos are a valuable institutional resource. Networks make up the foundation of an important type of multidisciplinary collaboration capability at the Laboratory. Individuals in different disciplines and geographic locations, locally or around the world, can interactively collaborate in real time on projects, from small to grand challenges. Networking is also an essential part of high-performance computing.

This year the Los Alamos Network Operations Center (NOC) was formed to provide a Lab-wide focus on computer network infrastructure planning and operation. Prior to 1992 network growth (excluding the central resources) has been built based on programmatic or group needs. The establishment of the NOC was necessary to improve coordination of resource and coverage allocation.

Five internetworks, one each for processing open, administrative, DOE secure/unclassified, DOE secure/classified, and DoD secure/secret data, provide the network access to customers' workstations or distributed processors (DPs) (Figure 4.1.1). With these networks, a customer can talk to other colleagues or to resources at the central facility. The topology of each is an Ethernet backbone running at 10 million bits per second connecting customer LANs together using network routers to sort and separate the traffic. These routers make sure that the only traffic that traverses the customer LAN is traffic from or to that LAN. Different technologies are used to connect LANs to the backbone (Table 4.1.1).

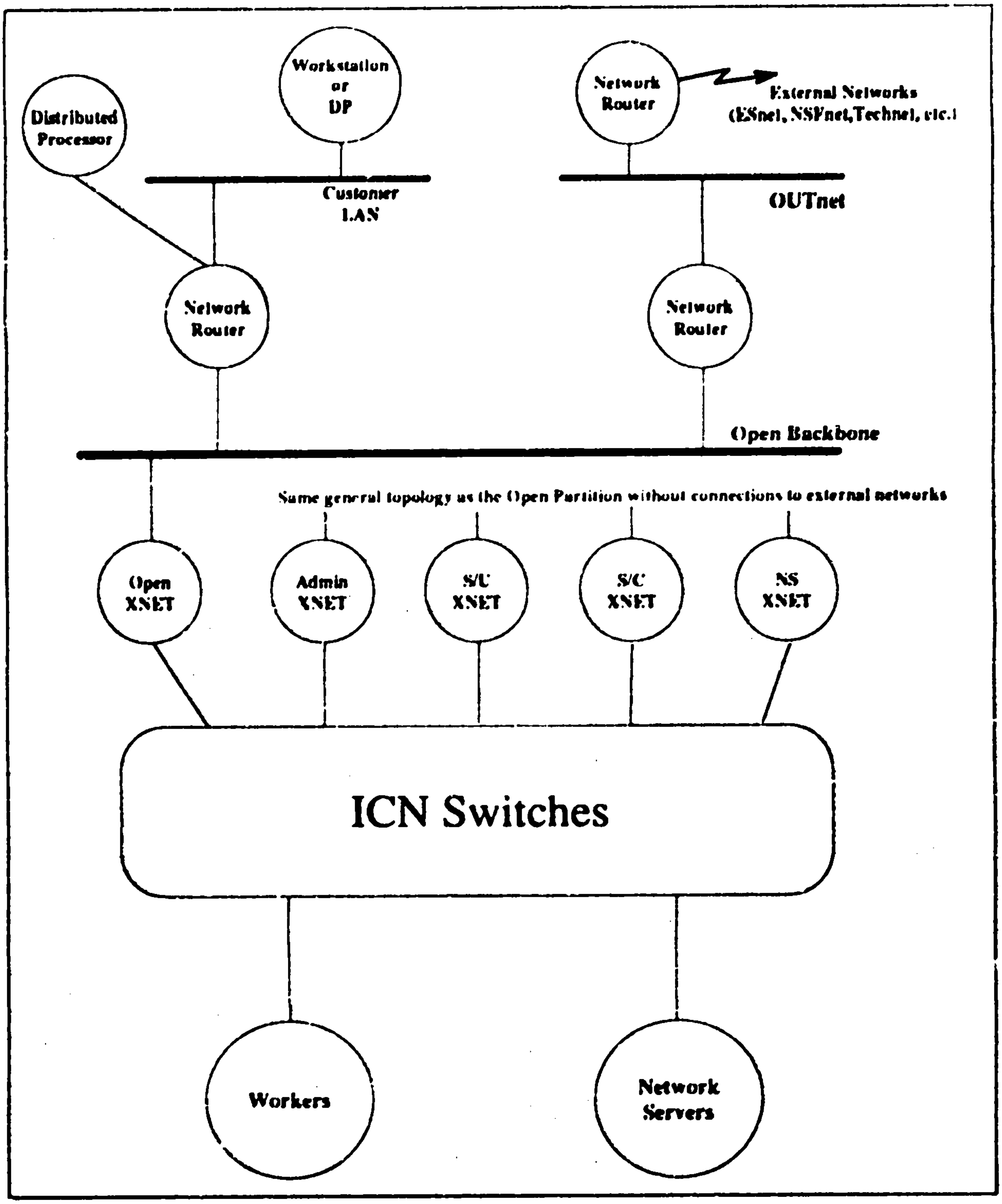

Figure 4.1.1. Internetwork Topology.

Table 4.1.1. LAN Interconnect Technology and Communications Speeds

Type	Speed
Serial	56K to 1.5M bps
"GS/7" over LABNET	5M bps
Ethernet via fiber	10M bps
Token Bus over LABNET	10M bps

These internetworks run both Transport Control Protocol/Internet Protocol (TCP/IP - an open protocol developed for the Defense Data Network and widely used on most non-proprietary networks), and DECnet (a proprietary protocol developed by Digital Equipment Corporation). They communicate with the central facility through gateways called XNETs that allow Simple Intermachine Protocol (SIMP) traffic to flow over TCP/IP and DECnet on the internets. The largest of these internets is the Open (unclassified) internet. It connects over 2,742 local nodes on 125 networks (Table 4.1.2). Estimates are based on informal use reports generated from host and network registration data. There are several LANs that are not connected to the Open internet and as such they are not tracked or reported.

Although these internetworks represent the major local area networking at the Laboratory, there is also a significant number of other network types, many of which are not connected to the backbones. These include Novell, Banyan, Apple, and token ring. There are currently no reliable figures for these; we are currently investigating the means to acquire the information.

Table 4.1.2. LANL LANs registered with the NOC

Division	IP Networks	IP Nodes
A	3	159
ADP	1	69
AT	4	217
C	30	614
CLS	1	56
CNLS	1	209
EES	8	111
ENG	1	13
HSE	2	63
IGP	1	9
INC	2	15
IS	1	156
LS	1	42
M	1	10
MEC	1	9
MP	3	11
MST	1	19
N	3	86
NMT	10	0 (moving to IP from Banyan)
P	4	227
SST	3	122
T	4	92
X	3	126
Other	3	(Subnets used in interconnections)

The open internet supports connections to state and international networks (Table 4.1.3) that provide Los Alamos with worldwide connectivity for sharing information This global connectivity also permits uncontrolled access to local systems. At Los Alamos, the NOC tests the vulnerability of our systems and notifies network managers of problems so they can be corrected. The NOC also maintains "filters" on the network routers to stop common attack mechanisms.

Table 4.1.3. Open Internet Connections to External Networks

Name	Sponsor	Connectivity	Speed of access
ESnet	DOE	Global	T1 (1.544 mbs)
NSFnet	NSF	Global	T1 (1.544 mbs)
NM Technet	State	Statewide	56 kbs

The Open internet is large and combines a diverse range of applications. The Open network also contains the greatest range of competence in the management of local LANs. Another of the NOC's functions is to act as a resource to help LAN managers. These needs are quite varied and include Appletalk and Novell network gateways. The statistics in the following table are collected by the NOC on packets that travel on the backbone (the central network segment that interconnects the network subnets). Some traffic was missed because LAN-only traffic and LAN-to-LAN traffic on the same router does not traverse the backbone. The DECnet statistics are somewhat incomplete due to a lack of tools and documentation that decode DECnet packets to determine what applications are running. Therefore, we assume 95% of the DECnet traffic is destined for central facility resources (because of the history of the distributed processor). *When the internets were first used, all the traffic was destined for the CCF. Now only about half of the non-DECnet traffic is for the central facility.* The rest of the traffic is for service from LAN to LAN or LAN to non-LANL destinations. Collecting statistics is a pilot project on the open backbone and has not been fully implemented for other backbones.

Volume (for comparison with other backbones):

Average usage: 24,810,782 packets/day, or 287 packets/second averaged over 24 hours. This represents an average load of over 10%. The absolute maximum throughput on Ethernet is 14,400 packets/second with a minimum 60-byte packet or about 570 packets/second with a maximum 1,500-byte packet. Under normal circumstances only about 80% of the above numbers can be achieved.

Average packet size: 148 bytes

Traffic destinations (on the backbone):

70% is LAN-to-LAN traffic

18% is LAN-to-CCF traffic

12% is LAN-to-outside-LANL traffic

Protocol (% of total traffic):

14.9% is DECnet

85.1% is IP

IP Breakdown:

92.9% is TCP, which provides reliable delivery of messages

6.1% is UDP, which provides only basic delivery of messages

0.8% is ICMP, which is used for some diagnostics such as reachability

0.2 is IGRP, which is a routing protocol used on the backbone

TCP Breakdown:

46% is communications with central services (CONNECT, CFS, MOVE, PAGES)

18% is interactive "terminal" communication between internet peers (telnet and rlogin)

18% is for remote command execution including remote shell (rsh)

12% is for remote file transfer using ftp

6% is X/Windows for remote windowing to workstations and UNICOS CRAY.

The Administrative backbone primarily carries traffic between ADP Division and the central facility and some traffic from other administrative customers. The

Administrative network has 39 registered DECnet nodes, 13 IP nodes, and 1 IP network. Access today is primarily via the asynchronous network via Micom switches.

The rest of the networks are shown below. These networks have traditionally not had the breadth of network services such as IP access to PAGES, email, and network news that the open network has.

The heaviest used of the remaining three backbones is the DOE secure/classified. It is also interesting to note that because there is no direct external connection or need-to-know router filtering on the classified and secret backbones, almost all traffic is destined for the central facility.

Secure/Unclassified Network Statistics:

TCP/IP networks:	12
nodes:	232
DECnet networks:	4
nodes:	61

IP breakdown by division:

Division	Networks	Nodes
C	7	163
EES	1	1
P	1	27
T	2	51
X	1	0

Volume (for comparison with other backbones):
Average usage: 189,760 packets/day, or 2 packets/second
Traffic destinations (on the backbone): 5% is LAN-to-LAN traffic 95% is LAN-to-CCF traffic
Protocol (% of total traffic): 89.9% is DECnet 10.1% is IP

DOE Secure/Classified Network Statistics:

TCP/IP networks:	10
nodes:	344
DECnet networks:	6
nodes:	106

IP breakdown by division:

Division	Networks	Nodes
A	1	26
C	4	15
IT	1	0
MEE	1	0
P	2	5
WX	1	26
X	11	272

Volume (for comparison with other backbones):

Average usage: 666,640 packets/day, or 8 packets/second

Traffic destinations (on the backbone):

2% is LAN-to-LAN traffic

98% is LAN-to-CCF traffic

Protocol (% of total traffic):

10.9% is DECnet

89.1% is IP

DoD Secure/Secret Network Statistics:

TCP/IP & DECnet networks: 7

nodes: 17 DECnet

13 TCP/IP

(all but two of these networks are external to LANL)

Volume (for comparison with other backbones):

Average usage: 50,540 packets/day, or <1 packet/second

Traffic destinations (on the backbone):

0% is LAN-to-LAN traffic

100% is LAN-to-CCF traffic

Protocol (% of total traffic):

42.5% is DECnet

57.5% is IP

5. The Cost of Computing

This section describes spending on both central and distributed computing resources

5.1. Laboratory Expenditures for Hardware, Software, and Support

During FY 1991, the Laboratory spent slightly more than $94M on computing-related equipment and services. Figure 5.1.1 categorizes these FY 1991 expenditures. This compares with approximately $106M in FY 1990 and $109M in FY 1989.

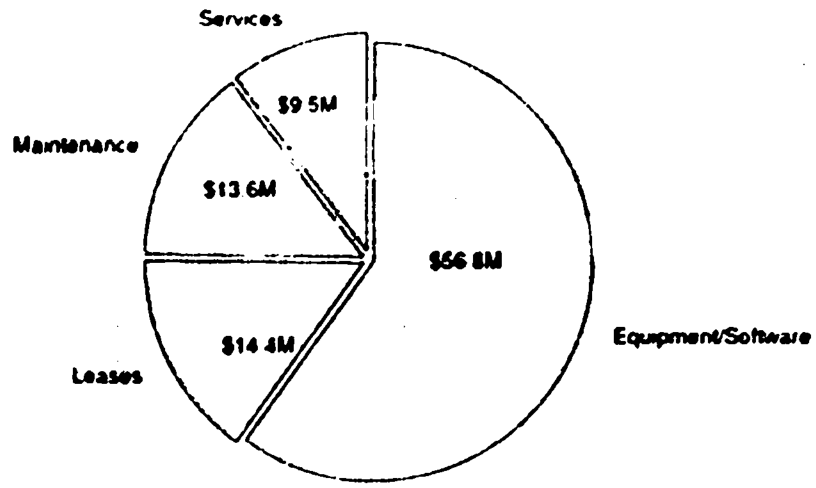

Figure 5.1.1. FY 1991 computing expenditures totaled $94.3M.

Of the $56.8M spent on direct purchases of equipment and software, $[illegible]4M was capital money, $3.4M was money sent to the Laboratory by external agencies (reimbursable), and the remaining $29.4M was charged to operating expense. Services expenditures include software support contracts and computer time acquisitions from commercial companies.

Of the $94.3M, C Division spent $41.6M in support of its activities, largely in support of the supercomputers. The remaining $52.7M was spent by the other divisions in the Laboratory for their computing requirements outside of C Division. Figure 5.1.2 compares these expenditures by category inside and outside C Division.

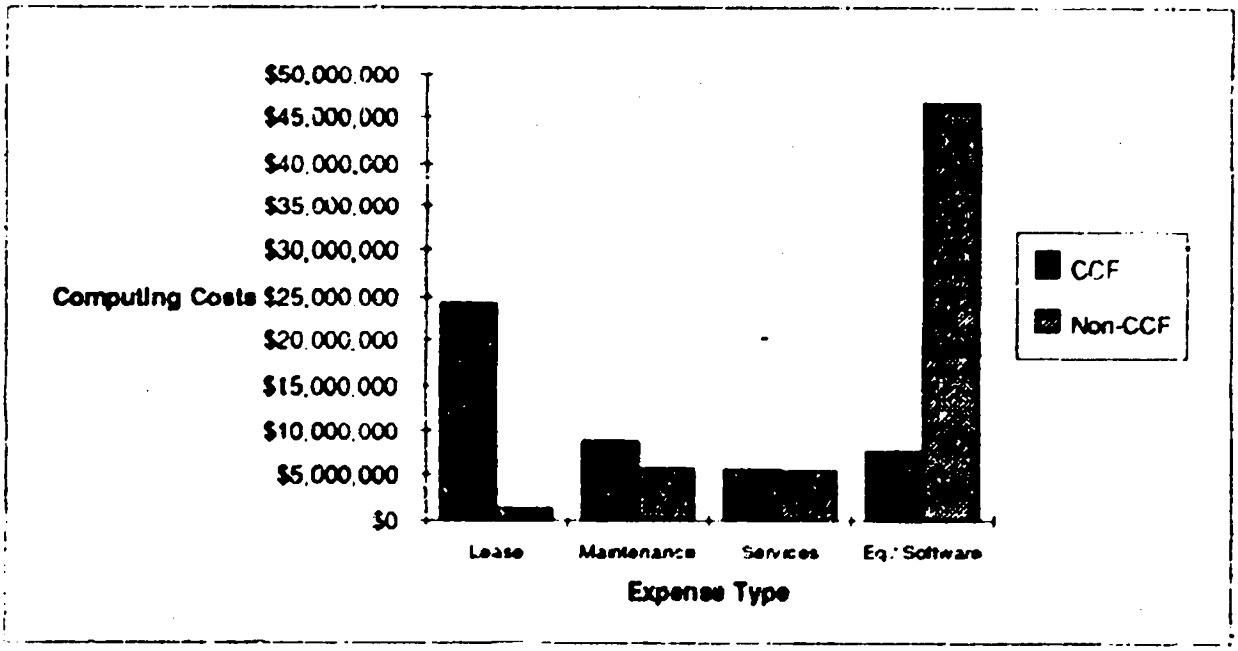

Figure 5.1.2. Computing costs inside vs outside of C Division.

Figure 5.1.2 shows that there are significant expenditures on direct purchases of computing equipment and software throughout the Laboratory. Virtually every organizational entity in the Laboratory had computing-related expenses during FY 1991. An examination of those divisions that spent more than $1M revealed 14 divisions (excluding C Division) spending a total of $24M (Figure 5.1.3).

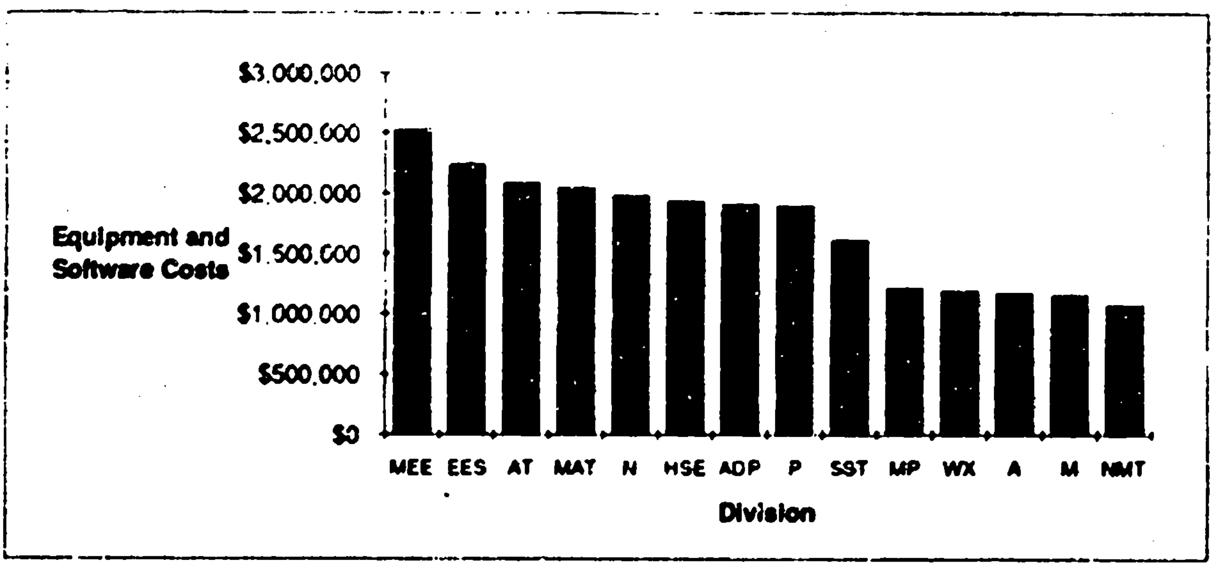

Figure 5.1.3. Equipment and software expenses in FY 1991 for those divisions that spent more than $1M (C Division's expenses, not shown above, were $15.97M).

5.2. The Cost of Services

Figure 5.2.1 represents the cost of various computer-related services, excluding hardware maintenance services, by division for those divisions that spent $100K or more. These include such services as data preparation, program development, graphics services, and commercial timesharing.

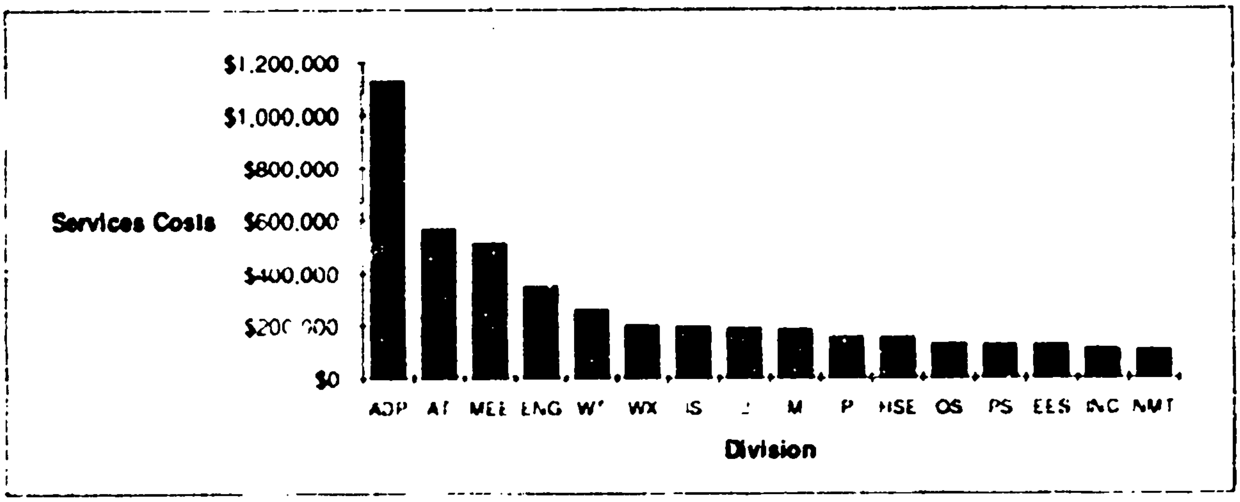

Figure 5.2.1. Expenditure for computer-related services for those divisions that spent more than $100K (C-Division expenditures, not shown here, were $3.93M).

5.3. The Cost of Maintenance

Figure 5.3.1 depicts the cost of computer-related maintenance services for divisions that spent over $100K.

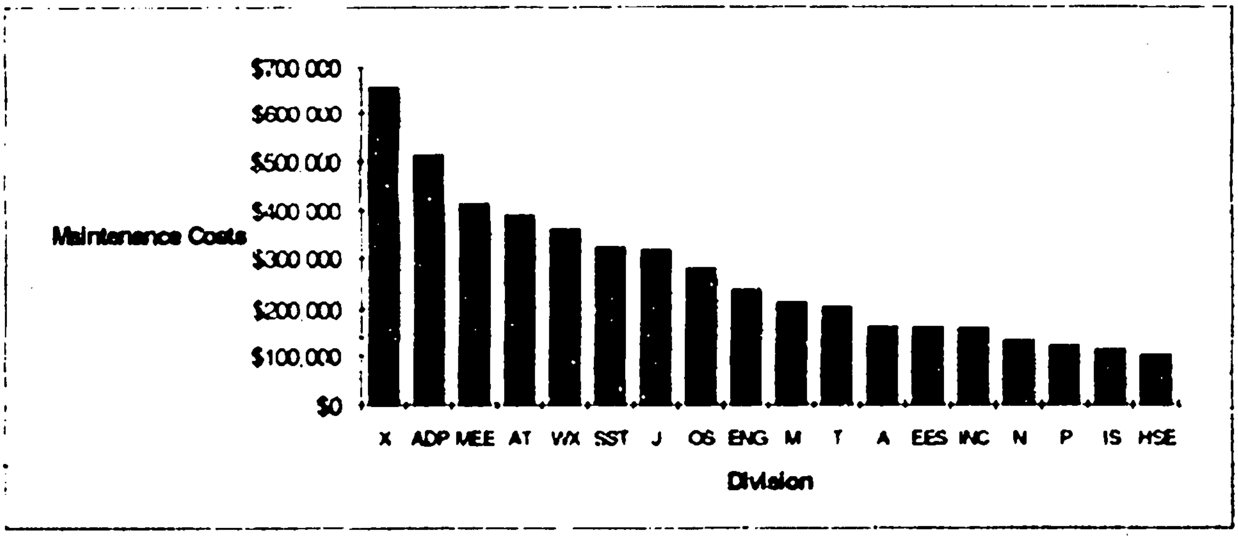

Figure 5.3.1. Divisions spending over $100K for maintenance services (C-Division expenditures, not shown here, were $7.88M).

5.4. The Cost of Leases

Figure 5.4.1 shows the divisions incurring mainframe computer-related lease costs during FY 1991.

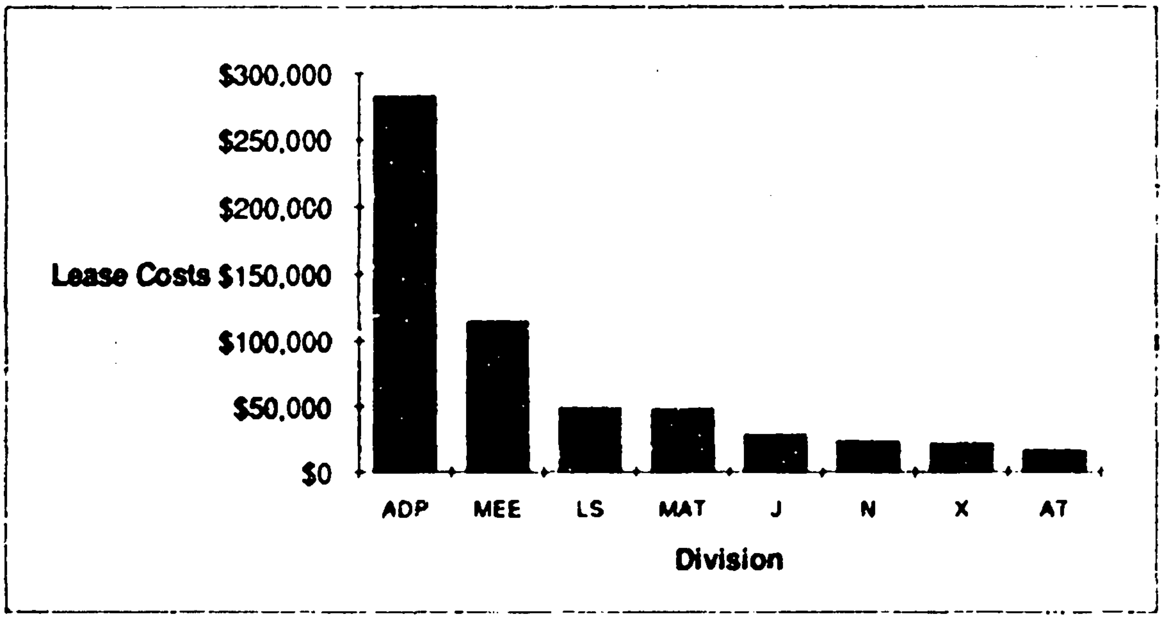

Figure 5.4.1. Divisions incurring lease costs in FY 1991 (leasing costs during FY 1991 for C Division were $13.77M).

Leasing is considered a viable alternative only under certain conditions. These include a short-term requirement in which purchase is not necessary; instances in which the technology is changing rapidly and obsolescence would occur before the useful life of the equipment has been exhausted; and when capital equipment funds are unavailable for the purchase of an expensive item of equipment that is programmatically critical.

For these and other financial reasons, leasing has traditionally been discouraged at the Laboratory for anything less than a major item of equipment, such as supercomputers.

6. In-House Personnel Costs

By far, the most difficult figures to obtain for this study are those for in-house personnel costs related to the support of the Laboratory's computing efforts. The information in Table 6.1 is derived from the Laboratory's input to the DOE Information Resources Management (IRM) Long-Range Plan. These figures only include those personnel who *support* the computing effort, and do not include personnel who only *use* computers.

Table 6.1. In-House Personnel Costs

Software Development	$37,008,000
ADPE Operations	6,721,000
ADPE Maintenance	2,331,000
ADP Studies and Other	6,753,000
Software Studies and Other	1,096,000
TOTAL	$53,909,000

The definitions for these categories are as follows:

Software Development Personnel costs relating to the planning, design, development, testing, conversion, and maintenance of software used at the site.

ADPE Operations Personnel costs relating to the operation and use of Automatic Data Processing Equipment (ADPE) for production work.

ADPE Maintenance Personnel costs relating to the in-house maintenance of ADPE.

ADP Studies and Other Personnel costs to manage or perform services and tasks including hardware feasibility studies; ADPE acquisition, selection, and use; internal ADPE consulting services; training of in-house personnel in the use of ADPE; contingency planning and security; and other ADPE-related personnel costs not reported in other categories.

Software Studies and Other Personnel costs to manage or perform services and tasks including software feasibility studies; software acquisition, selection, and use; internal software consulting services; training of in-house personnel in the use of software; and other software-related personnel costs not reported in other categories.

Because no consistent, quantifiable data exist, these figures necessarily are best management estimates. Nonetheless, it is apparent that the Laboratory expenditures for in-house support of computing represents a significant personnel and dollar investment.

Appendix: Notes on Data Gathering Techniques and Limitations

The information in this report was obtained from two primary sources: the Laboratory's PROPMAN property database and the DDA accounting system. The following describes the methods used to extract the data and the limitations with regard to completeness and accuracy.

General

In both cases, the data were extracted and downloaded to tables within a Microrim, Inc. R:Base for DOS database. Once loaded, columns were added for division identifier (e.g., 80 for C Division) and a division name. The identifier and name were derived from the field or fields in the original extracts that contained cost center information.

PROPMAN Data

PROPMAN is the Laboratory's property database. It resides on a Control Data Corporation (CDC) CYBER system, designated Machine N. PROPMAN is in the form of a System 2000 database.

It would have been ideal if one could have keyed on a single field to extract all ADPE; however no such field exists. It was therefore necessary to construct a query using the available fields to attempt to extract as many of the records of interest as possible.

That query requested the information on records that were coded as BSA "5" or BSA "E" (ADP capital equipment purchased with Lab capital or reimbursable funds), or records coded as "R" or "Z" (ADP) orders, or orders beginning with "PIS" or "MR" (in-house PC Store orders), or orders for which the description was "Computer Personal."

This returned approximately 38,000 records of property-numbered items. The download of this information required two passes. This was due to the limitations imposed by System 2000 on the total number of fields and/or the total number of characters that a "list" command can return. The first pass returned the property number, description, acquisition document number, th month and year received, the group, cost, and technical area location. The second pass (using the same query) returned the property number (a second time), model, manufacturer, building, and room number for the equipment.

Each of these passes was loaded into its own R:Base table. A relational "union" operation was then performed to create a new table with a single record containing all fields from the two source tables by using the property number as the common field.

After this "master table" was built, a tally operation was performed to generate a unique listing of descriptions. Because some non-ADPE is purchased under ADP-related procurement activities, not all items downloaded were ADP in nature. These included television sets, safes, generators, etc. An operation was performed to eliminate the non-ADPE, resulting in the database shrinking to approximately 35,700 items.

Because there is no data dictionary support in System 2000, there are no standard descriptions. This necessitated an extensive cleanup of that field to create such standard descriptions. Every effort was made to determine the nature of the equipment and establish standard nomenclature by checking the manufacturer, model number, and cost fields and consulting the manufacturers' catalogs and other vendor material.

Even after these attempts, some level of incompleteness exists. The major category consists of those ADP items of equipment that are not in the PROPMAN database at all. The property management regulations exempt the majority of low-value equipment from property control. This includes most of the computer display monitors, pointing devices such as mice, keyboards, and some other peripheral equipment.

Similarly, software is not property controlled; no figures are therefore available for the nature, volume, or dollar value of this indispensable commodity.

A further complication exists in determining a standard nomenclature for equipment. At what point is a personal computer a workstation, or a workstation a minicomputer? In that same vein, some workstations can be configured as LAN file servers, and are therefore distinguished more by use than hardware characteristics. In these cases, "best estimate" was the guide.

DDA Data

The financial figures used in this study were downloaded from the Laboratory's DDA accounting system. In all cases, these figures represent contract obligations (the value of the contract over its term), as opposed to costs (checks written) for FY 1991. This distinction, in one sense, is important because obligations can (and frequently do) span fiscal year boundaries. In another sense, however, the difference between obligation and cost loses importance because it "washes" across fiscal years. Because this study is intended to be published annually, it was felt that the DDA data were as good a source as Commitment System costing information.

The DDA data are initially in the form of an rdb database located on a MAT-controlled DEC VAX system. A query was performed to extract all records in which the order type was "M" (maintenance), "R," or "Z" (ADP-related), or the buying group was MAT-6 and the order type was "L" (lease) or "X" (blanket orders).

For each record, the purchase order number, vendor name, cost center, program code, parenthetical (for capital equipment items), expense type, total order amount, group, description, and order date was downloaded. This resulted in an R:Base table of approximately 14,000 records.