ROUTING AND TRANSMITTAL SLIP Date 20 OCT 1986

TO: (Name, office symbol, room number, building, Agency/Post)	Initials	Date
1: EXA/DDA		20 OCT 1986
2.		
3.		
4.		
5. DDA REGISTRY		

Action		File		Note and Return	
Approval		For Clearance		Per Conversation	
As Requested		For Correction		Prepare Reply	
Circulate		For Your Information		See Me	
Comment		Investigate		Signature	
Coordination		Justify			

REMARKS

D/OIT RECEIVED A COPY.

STAT

DO NOT use this form as a RECORD of approvals, concurrences, disposals, clearances, and similar actions

FROM: (Name, org. symbol, Agency/Post)	Room No.—Bldg.
	Phone No.

5041-102

* U.S.G.P.O.: 1983-421-529/320

OPTIONAL FORM 41 (Rev. 7-76)
Prescribed by GSA
FPMR (41 CFR) 101-11.206

EXECUTIVE SECRETARIAT
ROUTING SLIP

TO:

		ACTION	INFO	DATE	INITIAL
1	DCI				
2	DDCI				
3	EXDIR		X		
4	D/ICS				
5	DDI		X		
6	DDA		X		
7	DDO				
8	DDS&T		X		
9	Chm/NIC				
10	GC				
11	IG				
12	Compt				
13	D/OLL				
14	D/PAO				
15	D/PERS				
16	VC/NIC				
17	D/OIT/DA		X		
18					
19					
20					
21					
22					

SUSPENSE _____ Date

Remarks

STAT

Executive Secretary
20 Oct 86
Date

3637 (10-81)

NATIONAL SCIENCE FOUNDATION
WASHINGTON, D.C. 20550

OFFICE OF THE
ASSISTANT DIRECTOR
FOR COMPUTER AND INFORMATION
SCIENCE AND ENGINEERING

October 10, 1986

Mr. William Casey, Director
Central Intelligence
Old Executive Office Building
Washington, D.C. 20506

Dear Mr.Casey:

Enclosed are a couple of papers that may be of interest to you:

(1) An article commemorating the 40th Anniversary of ENIAC...and the critical role of government in computing.

(2) A paper looking at the Scientific Computing Environment (i.e. supercomputers and alternatives for engineering and scientific use). The paper is to be used by us and our National Centers to clearly define various issues for resolution.

I hope they may be of interest to you. Please feel free to circulate, comment or feedback to me.

Sincerely,

Gordon Bell
Assistant Director/CISE

60-7

Enclosures

ENIAC as a Lesson in the Past, Present, and Future History of Government Sponsored Computing

C. Gordon Bell, Assistant Director
Computing and Information Sciences and Engineering Directorate
National Science Foundation
&
Gwendolyn Bell, Director
The Computer Museum

ENIAC established a critical role and pattern for government involvement in the advancement of computers and computation. It provides a number of important lessons today when, for the first time, we see increasing competition for the definition of the next computer generations.

Eckert and Mauchly, and the team of creative scientists and engineers at the University of Pennyslvannia needed a sponsor willing to take a risk. The U.S. Army Ballistic Missile Research Laboratory at Abardeen believed in research, risk-taking, and support. They needed the dramatic increase in computational power that ENIAC promised in order to carry out their calculations. Without the promise of a factor of 10,000 over the evolutionary relay technology, and the willingness of the creators and sponsors, the electronic computer as we know it, would have to have been invented in some other fashion. In contrast, calculators were being developed at Bell Labs, IBM, and Harvard utilizing evolutionary, conservative relay and electromechanical technology.

ENIAC was the first, operational, large scale, electronic digital computer, but first implemented as a calculator. It was subsequently used by Los Alamos for calculations on the first atomic bomb. Finally, ENIAC provided an environment where von Neumann authored the Edvac Draft Report describing the first stored program computer. (ENIAC later ran as a stored program computer, in contrast to the plugboard programming method it was designed to use.)

Eckert and Mauchly went on to form the Eckert-Mauchly Computer Company, as a Remington Rand Division, having been unable to get financing as a standalone company (Venture Capital hadn't been invented yet). They went on to produce the first commercial computer, Univac I, for commercial applications. IBM followed within a few years with the vacuum tube based, IBM 701, built as a scientific computer for defense applications.

ENIAC, with 18000 vacuum tubes was roughly 100 times larger than the largest system of the day (an electronic organ). It took the revolutionary environment of the university to even contemplate such a system. An engineer at Bell Labs or IBM would have been ridiculed for such thought, and indeed, much of the established, wartime engineering community was against building the ENIAC simply because it was so different, large and risky.

By 1960, the vacuum tube had been replaced by the semiconductor to build computer logic, based on the Bell Labs transistor. Today, we credit the exponent increase in semiconductor density, measured in transistors/chip, as the principle driving force for new computer generations (we're in the late fourth with very large scale integrated circuits with several hundred thousand transistors per chip). Similarly, exponental increases in magnetic storage densities play an equally important role in computer development, evolving from the early drum and IBM disk inventions. Thus, the computer revolution is NOT a revolution, but simply an evolution based on technology that evolves exponentially with time (the exponential rate is large). ENIAC, the stored program concept, and the transistor were the revolutions.

The exponential improvement creates three, new <u>evolutionary</u> computers:

- constant cost machines with exponential performance increases,
- larger machines, (supercomputers) at the limit of the new technology based on increased parallelism,
- new machine classes as the exponential increases in circuit densities permit new, lower cost computers (eg. the mini, micro, workstation, and personal computer).

The ENIAC-Eckert-Mauchly experience gives us a few principles that have been carried forth today, and show the principle strengths of the way computer have and are likely to continue to be discovered and evolve, provided the current research mechanisms remain in tact, while being significantly strengthened to deal with international competition:

- The risky, radical ideas come from the universities (and an occasional industrial lab) that create the revolutionary new tacts. Universities are seeking knowledge, and don't have the unlimited resources of corporations and hence must be clever. They may also be unconstrained and radical. The competition to seek knowledge and new ideas is rewarded and <u>required</u> by the publication and PhD production process of U.S. academic research.

• Government has been the main, critical supporter of radical new ideas, especially when coupled to the ability to solve war needs in an unconstrained fashion or to pursue new visions such as going to the moon. NASA's Apollo Computer used the first integrated circuits.

•Engineering in order to use is a critical part of research in computing. A "stretch" project such as ENIAC creates many, scientific and serendipitous results in additon to training scientists and engineers.

• Radical or even slightly eccentric ideas whether they be new machines or new machine classes (eg. minis, personal computers) will be rejected by established companies.

• New companies have to form to exploit new ideas. Many new companies follow suit to form an entire industry around the breakthrough.

• Significant ideas get adopted by the large, established companies in direct proportion to their perceived economic gain (or loss).

•Once established as a company and industry, evolution takes over to become the establishment, and ready to reject the next new idea.

•The role of the ultimate, critical user, such as Los Alamos, established a tradition of utilizing the largest computers for nuclear science and engineering. The Atomic Energy's National Labs (now Dept. of Energy) have served this role since ENIAC, encouraging the early IBM Stretch, Univac LARC, and Seymour Cray machines (i.e. supercomputers) from CDC and Cray Research.

As computing developed over the last four, post-ENIAC decades the pattern was reinforced with the Department of Defense, e.g. the Office of Naval Research, and later via the Advanced Research Projects Agency (now Defense Advanced Research Project Agency-DARPA), stimulating research, which ultimately turned into either new computer structures or key segments of the computing industry in a revolutionary fashion. The National Science Foundation with responsiblity for the health of science, engineering, and technology provided broad sponsorship of research and manpower training through the Computer Research program.

We now have enough post-WW II successes to know that the pattern established with ENIAC is not accidental, but continued to the present:

• Whirlwind developement at MIT yielding the core memory, interactive graphics (including CAM), fast real time command and control computing (archetype of DEC's first minicomputer), algegraic compiler, and air traffic control for defense

•Timesharing at MIT's Project MAC, as a precursor to wide scale timeshared computers and Bell Labs' Unix

• Graphics at Utah creating the algorithms and training the people to create the industry

• Arpanet for computer networks

•Understanding of LSI enabling the creation of VLSI computer aided design, including verification, simulation, and silicon compilation

•Artificial Intelligence in the form of vision and robotics

•Artificial Intelligence in the form of expert systems

• In addition, a serendipity process at research labs has created some of the first computer generated art and music, and new approaches to typesetting.

Today, with real competition from abroad, the process must be even more highly tuned and perfected. It is somewhat encouraging to see:

•DARPA's Strategic Computing Initiative building very high performance parallel processors, and NSF's Parallel Processing Program utilizing, understanding and building a new basis for computing.

•The National Science Foundation's Advanced Scientific Computing Program at five, national supercomputer centers involved in training, standards, networking, improved graphical access, parallel processing, and understanding new computers by utilizing them to solve problems.

•The SDI Network to provide a high performance "Global", Local Area Network (LAN) to connect the many disparate high speed LANs. NSFnet to connect the entire research community.

•Human interfaces utilizing speech and natural language input, with experiments in language translation.

Is the ENIAC (and transistor evolution model) complete enough to describe the entire history of computing? It doesn't explain the products such as PL/1 that come from the classical model where marketing departments talk to users, and then tell engineers what to build. Nor does it account for several important inventions at Cambridge (microprogramming), Manchester (virtual memory), and IBM (Fortran). It doesn't explain Cobol (the idea for the compiler coming from Univac and the standardizing effort coming from users, especially the U.S. Navy), or Visicalc (from the classical, creative American inventors which in this case were computer scientists trained in the research environments). On the other hand, it does explain quite a lot.

It is critical that we not only understand the roles of the various organizations, but are committed to keeping and improving the strong research base because for the first time, other countries are also committed to leading the next generation just as the U.S. has done in the past ones since ENIAC.

Preparing for Changing Scientific Computing Environments

Gordon Bell, Assistant Director
Computer and Information Science and Engineering
National Science Foundation
Washington, D.C. 20550
30 September 1986

Introduction

Recently, a hierarchy of scientific computers in three price ranges and computing styles have evolved with relatively the same performance/price and computational ability. The hierarchy includes: the supercomputer and large mainframe used as a regional or central computer costing between $10M-20M; the mini-supercomputer used alternatively as a central, departmental, or group computer costing around $500K; and a workstation/workstation cluster, used as a shared, departmental resource, as a single user system, and access to other machines in the hierarchy costing around $50K.

The comparable computational power of these new scientific computers raises various policy issues for NSF including the management of its Advanced Scientific Computing Program, the role of the five National Centers, and the way computation is supplied to the research community. Ideally, a user will utilize all forms of computation based on economics, networking, power, response time, and interaction (especially graphics) needs. This paper explores these parameters and outlines the policy implications required to provide the most productive environment for the research community.

The data for this analysis are key performance characteristics of a variety of scientific computers:

- number of processors, **#P.c**
- primary memory size in 64-bit Megawords, **M.p**, with virtual memory (shown as **.v**)
- secondary memory size in Megabytes, **M.s**
- speed measured in millions of floating point operations per second using Dongarra's Linpack benchmark for a 100x100 and 300x300 matrices, **Mflp**
- the price of the machine in millions of dollars, **$.M**
- the cost-effectiveness, i.e. performance per unit price for two sized matrices, **flp./$**
- introduction date, **Intr**
- stretch time versus Cray XMP single processor for a single job ()

Table of Computer Characteristics

System	#P.c	M.p	M.s	Mflp	$.M	flp./$	Intr	(Stretch) comments
Supercomputers								
Cray 416	4	16	9.6	**108.-480**	17	**6.4-28.2**	86	**(1)** 27/Pc, 8.5ns clock
Cray 48	4	8	9.6	**108.-480**	15	**7.2-32**	84	
ETA 10 (Est.)	8	288.v	9.6	**1040.-2K**	19.7	**52.8-107**	6/87	**(0.2)** 10.5ns, 7ns '88 = x1.5
Cray 2	4	256	9.6	**60.-372**	18.6	**3.2-20**	86	
Mainframe								
IBM 3090/400	4	16	60.	**48.-108**	9.8	**4.9-11**	9/86	**(2.25)** 128Mw=$2.2M
Mini-supercomputers								
Alliant F8	8	1.v	.4	**7.6-14**	.75	**10.1-18.7**	6/86	**(3.6)** with directives
Convex C1	1	1.v	.4	**2.9-14**	.4	**7.3-35**	1/85	**(9.3)**
SCS-40	1	2	.7	**7.3-26**	.65	**11.2-40**	7/86	**(3.7)** XMP compatible
Workstations								
Sun 3-200	1	1.v	-	**.47**	.04	**12.0**	9/86	**(57)**
Sun 3-200	1	1.v	.28	**.47**	.06	**8.2**		**(57)**
Sun 3-200	1	2.v	2	**.47**	.12	**3.9**	9/86	**(57)**
+3 diskless	4	8.v	2	**1.9**	.25	**7.5**		cluster of 4
Sun 87/B Joy	1	2.v	-	**1.5**	.02	**75**	87	**(18)**
Sun 88/B Joy	1	4.v	-	**4.0**	.03	**132**	88	**(6.75)**
Sun 89/B Joy	1	8.v	-	**10.0**	.04	**250**	89	**(2.7)**
Historical References								
Cray 1/S	1	1		**12.-66**	6	**2.-11**	75	**(2.3)**
VAX-11/780	1	.5v	.1	**.15**	.3	**2.**	4/78	**(180)**

Notes:

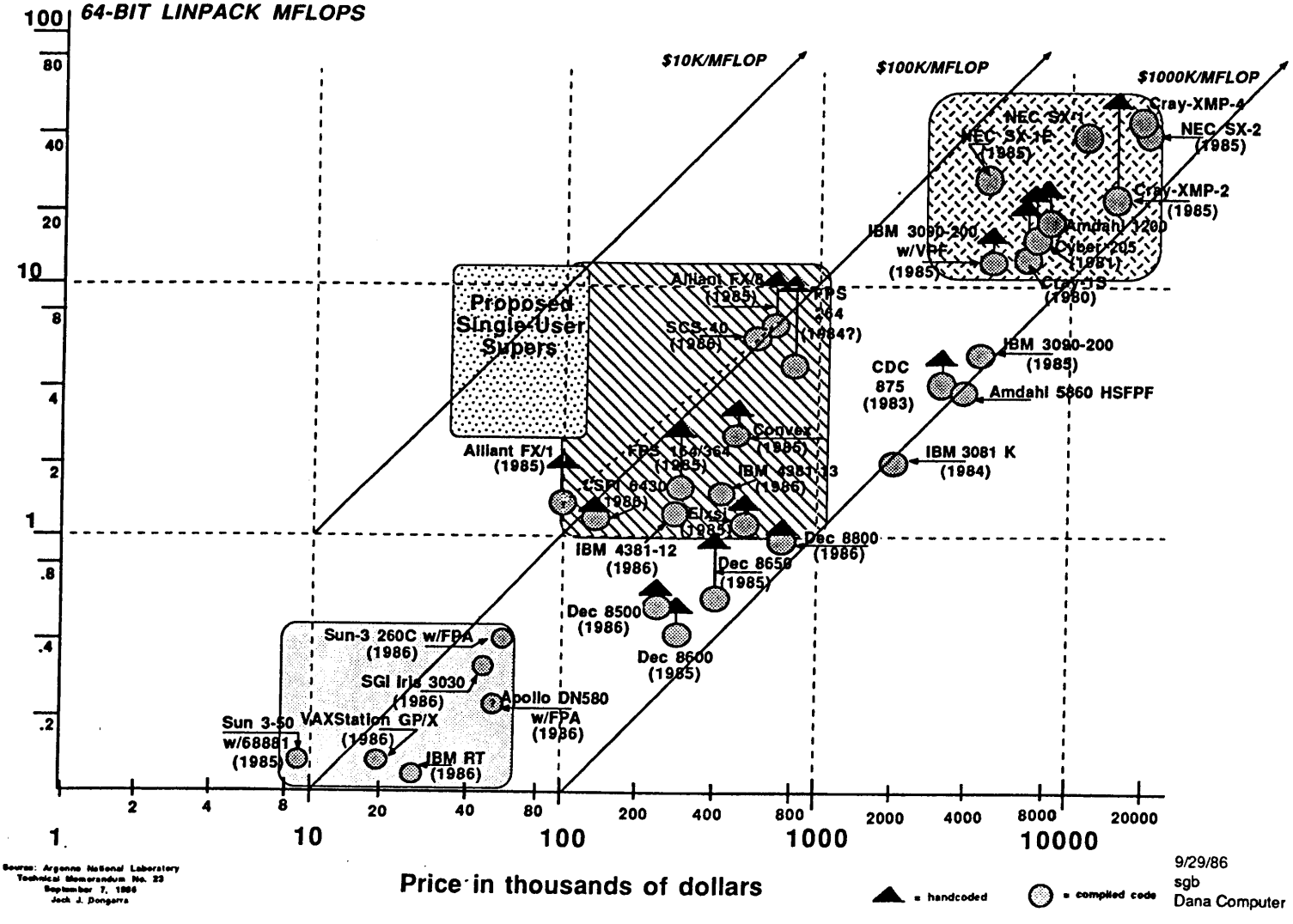
For the Crays and ETA-10, the performance is for 4 and 8 independent job streams. Linpack appears to be a good benchmark in that it correlates well with other scientific and engineering benchmarks, and with the average delivered power. As work is "tuned" for vector processing, the 300x300 matrix is a realistic target for typical applications.

A variety of different secondary memory configurations are given, including none for the 3090 and several Suns.

The fastest uniprocessor is the NEC SX-2 at 43-347 Mflops.

Super-minicomputers, and high-performance PC's are not included because they provide relatively poorer performance, and performance/price. For example, the PC AT/370 is a factor of 818 slower than a Cray and the cost to perform floating point operations is roughly double.

Comparative Linpack Performance



Source: Argonne National Laboratory
Technical Memorandum No. 23
September 7, 1986
Jack J. Dongarra

9/29/86
sgb
Dana Computer

Background

The current surge of interest in supercomputers becomes clear when we look at the evolution from the late 70's when the Cray 1 and VAX 780 were the standards for computation. The 780 entered the scientific and engineering community because it provided relatively the same price performance as a Cray 1, even though the performance differed by a factor of 80 (using Linpack as an indicator). A more reasonable estimate for the difference is more like a factor of 20-40. Those who bought VAXen observed that since the average user only got 1-2 hours of Cray time each week, (50-100 hours per year) they could get the same amount of computing done by letting a VAX grind 20-160 hours per week.

Over time, the Cray evolved; the XMP was speeded up by over a factor of two and built as a multiprocessor, which roughly trebled the performance/price. When the scientific community started utilizing Crays with improved compilers, they began to develop more effective algorithms for vectors that increased the effective power of the machines. The delay in getting a more cost-effective VAX (the 8600 was two years late), and the relatively high price of VAXen exacerbated the difference between the supercomputer, and the super-minicomputer (in essence a lower priced mainframe). The popularity of VAXen for more general computing also allowed the price to remain high, by giving it a market outside the research community. DEC, like IBM when it introduced a complete range of compatible computers, may have become less interested in and attentive to the research community. The Cray/VAX gap may have been a major motivation in the formation of the NSF Advanced Scientifc Computing Program.

In the early 80's Alliant, Convex, and Scientific Computer Systems formed to exploit the performance/price gap between the Cray XMP and VAX by utilizing vector data-types pioneered in the Cray 1. Thus, a new class of mini-supercomputers was formed, all of which have better performance/price than the Cray (almost a factor of 2 in the case of the new SCS-40).

By 1985, ten years after the Cray 1, IBM and Japanese manufacturers building IBM-compatible mainframes had added vectors and multi-processors to their machines.

Observations About the Computers From the Table

Three characteristics are important: the processing power in Megaflops; the cost-effectiveness in flops/$, and the stretch time versus a Cray. There are exceptional computers, when comparing the cost-effectiveness in each class: the (projected) ETA-10 (to be better by a factor of 8!), and the SCS-40 (better by almost a factor of 2). The SCS-40's virtue and principle flaw is Cray compatibility. Other mini-supers have virtual memory. A cluster of SUN workstations could provide up to a factor of 2 better performance/price, depending on the amount of secondary memory. The factor of 5 difference in the speed of the ETA-10 versus a Cray XMP should open up new problem solution domains. The ETA-10 uses large CMOS gate arrays on large, multilayer

printed circuit boards. This kind of fabrication provides a potential breakthrough in cost that is counter to the use of ECL to build supercomputers, large mainframes, and superminicomputers by Cray, DEC, IBM and the Japanese.Both the Cray 2 and ETA-10 have large memories that should open up new problem domains. All of the machines, except the Crays, have virtual memory. Because of the lack of paging, it may be difficult for multiple users with very large problems to effectively utilize the Cray 2. The use of large physical and virtual memories needs to be explored and understood.

While the table shows times for a floating-point intense program, Linpack, it is unclear how the machines perform under comparable workloads or whether they will actually be used in the same fashion. For example, a slower machine is likely to be used more interactively and results of the computation viewed constantly to avoid unnecessary work. Users of large batch machines may have to request more work and output because turn-around is longer. Scalar benchmarks aren't given, and most machines are used a significant amount of time either interactively or in scalar mode, both of which lower the performance and favor the 3090 (which outperforms the Crays in scalar mode), mini-supers, and workstations.

NEC's SX-2, not included in the Table, executes Linpack at about twice the performance of a single processor Cray XMP. The performance/price is unclear.

Many computers exhibit performance/price comparable to today's supercomputers. The Advanced Scientific Computing Program must understand the relative power and work capacity of all forms of computation and begin to develop ways to supply resources appropriate to user need and cost-effectiveness considerations.

Can Users Tolerate the Time Stretch/ Lower Cost Trade-off?

Can a user of a smaller computer, stand the lengthened turn-around time that comes with using a slower computer and stretching the computation time by factors of 4 to 10?. At present, only one or two users within our user community are receiving an hour of computer time per day. The mini-supercomputers, supplying the eqivalent of one hour of Cray time in 4-10 hours are competitive because the average turn-around for a one-hour job on a Cray can easily be this long. The typical turn-around for a 15 minute job is 2 hours (or factor of 8 stretch). The Sun Workstation might be used for longer computation provided the user "guides" the computation. The Sun's stretch factor is comparable to that experienced between the Cray and 780 during the late 70's. Alternatively, advances in partitioning programs for parallel processing make the cluster have the best performance/price if a job can be parallelized using a message-passing model of computation.

Based on the performance, and time allocations inherent in supercomputer use, a complete

hierarchy of computers will exist and is justified. Given that an individual user or project is likely to simultaneously access all levels of the hierarchy, a compatible (and most likely standardized) basic environment that can support user communities, who in turn have common applications environments, is essential.

Multiprocessors, Array processors and Multicomputers (e.g. Hypercubes) for Parallel Processing

A number of alternatives exist that may offer significant improvements in performance or performance/price. For example, a 64 computer NCUBE has been used to solve a problem that took twice as long on a single processor XMP. The improvement yielded almost an order of magnitude in cost. Given the decomposition for parallel procesing on the NCUBE, an XMP might be used to gain a 4 times speed-up; in fact, the XMP operating in this mode has computed Linpack at a rate of 713 Mflops which is 26 times the single processor rate. Likewise, array processors such as the FPS X64 have been lashed to minis and mainframes, yielding significant improvements in performance/price. None of these alternatives are explored.

Standardized parallel processing primitives in all programming languages based on a multi-process, message passing model of computation is needed for all structures. Programs used in this fashion will operate compatibly and identically across workstation clusters, multicomputers such as the hypercube, and shared-memory multiprocessors (e.g. Cray and ETA). Given the relatively constant performance/price and similar turn-around times for all of the computing alternatives, parallel processing becomes essential.

The Role of the Super Computer Centers

Historically, centers have existed for a variety of reasons including cost sharing, technology, performance, networking, user needs, local politics, government funding, etc. Clearly when hot ideas emerge and projects need ten to several hundred hours of supercomputer time that can't be supplied locally the centers are essential. The definition of the kinds of work that the centers will support is critical, given that computation can be done very effectively by local university centers, departments, projects, and individuals at workstations.

Our centers are critical to scientific and engineering computing for the research community. Today the centers train users about the parallelism inherent with vector data-types. They have the programs and staff to train the trainers and users rapidly, and to support large programs and datasets inherent in supercomputer use. Centers may be the best place to support certain large programs and databases for a given intellectual community; NCAR is an excellent example as it provides millions of lines of common programs and 17 terabits of common data for its community of atmospheric scientists to environmental engineers. Centers may also support common programs for communities of distributed users at mini-supers, super-minis, and workstations in order to

supply service when the distributed research requires significant computing power.

Large amounts of power (on the order of 1 hour per day) would be supplied to large projects that do not have machines, and to a community of student and casual users who access common programs and data. If the "average" project uses 1 hour per day or 350 hours per year, then a Cray XMP would support 24x4, or about 100 projects! Projects of this size would be, in effect, subsidized at about $100,000 with steady-state costs. It can, alternatively, service 640 users who use at most an hour a week, or 50 hours per year, providing them about a $15,000 subsidy. Finally, several thousand student and casual users who would use no more than 10 hours per year (a year on a PC/370) could be supported at negligible cost. Policy statements are needed which characterize useage across geography, user size, and discipline.

The centers have a lead role in supporting state-of-the-art computers of all types including supercomputers, mini-supercomputers, and larger scale experimental machines. The centers should be the beta test sites of all new systems, especially those which can not be easily purchased or supported by local researchers or departments. The centers must take the lead role in understanding benchmarks, workloads, and cost-effectiveness of all forms of computation.

Standards. The three alternative forms of computation that form the main line of computing all provide roughly the same computational service at comparable costs (not including the cost to the user). We must establish standards that make it equally easy for users to work at any of the places in a compatible fashion. In many cases, a user will use the super or mini-super or existing super-mini for calculations and the workstation to view results. Thus code will be run in a highly distributed fashion across different machines including new, and evolving UNIX-compatible PC's. Similarly, we should work toward establishing and supporting common programs and data across engineering and scientific disciplines so they may compute at any level of the hierarchy.

Conclusions

Computers now exist which allow various styles of computing ranging from regional supercomputers to personal workstations. All of the computers in the hierarchy will continue to exist and flourish because, with the exception of the ETA 10 to be delivered next year, all offer relatively the same cost and effectiveness.

Having the wide range of styles and locations demands attention to:

- •training, education and program support;
- •networks for intercommunication of programs, data, and terminal access;
- •benchmarks, workloads, accounting, and pricing i.e. understanding cost and effectivenss;
- •allocation of time across user communities by size, discipline, and geography;
- •standardized programming environments and graphics enabling effective use;
- •supporting specialized community programs (e.g. NASTRAN) and databases (e.g.NCAR);
- •specialized and alternative computers; and
- •standards, understanding and training for compatible, message-passing parallel processing.

With the center program entering phase II, attention and resources will have to be focused on these demands.