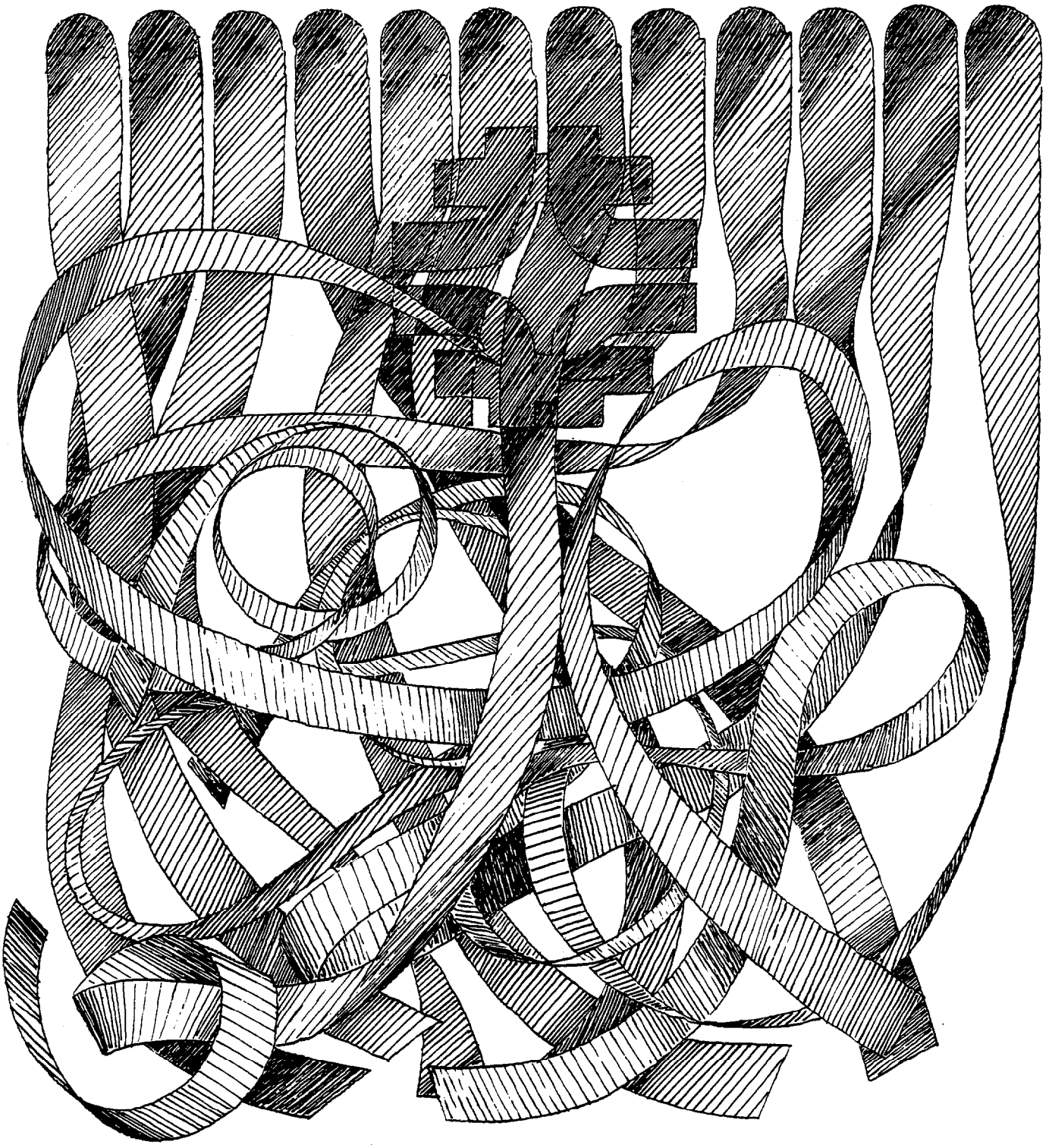
Fermilab

TM-1230
0062.000

REPORT OF AD HOC COMMITTEE ON FUTURE COMPUTING NEEDS FOR FERMILAB

Future Computing Needs for Fermilab



Report of the Ballam Committee **December 1983**

Foreword

A series of happenings conspired to convince me that a thorough, new look at the computer problem was overdue.

There was a growing realization that the TeV II computing load had been underestimated. There was proof that the large hadron colliding detector has the capability of saturating even so awesome a facility as that at CERN. UA1,2 with relatively modest runs at less than 10^{29} average luminosity were reported to be computer-limited. A new component of "engineering computation" within which we include CAD/CAM, accelerator orbit studies, magnetic field design and theoretical physics, represents a potent consumer of Cyber hours.

We recognize that by 1986, there will be a crisis of computer utilization and our best efforts will be required to address this problem so as to avoid diminishing the scientific potential of the Tevatron.

Joe Ballam's committee was first convened on May 31. Their report is an important document and will have a major impact on the Laboratory's program over the next five years.

Leon M. Lederman

Table of Contents

Section	Title	Page
I.	Introduction and Recommendations	1
II.	Estimate of Needed Capacity	9
III.	High Energy Physics Tasks and the Computing Environment	13
IV.	Experiment Area Computers	24
V.	Networking	27
VI.	Special Data Processors	31
VII.	Personal Work Stations	38

Appendix	Title	Page
A.	Committee Membership and Record of Meetings and Documents	43
B.	Fermilab Computing Department Strategy for the Mid-80'a	45
C.	Survey Results	53
D.	UA1 From a Computer and Software Viewpoint	64
E.	CDF Computer Requirements	82
F.	Suggested Topics for Work Projects	91
G.	Mainframe Developments	93
H.	Accelerator, Engineering and Administrative Computing Needs	97

I. INTRODUCTION

This committee was formed at the request of the Director of Fermi National Accelerator Laboratory as an ad-hoc group to review its computing needs for the next five to ten years and to make recommendations to him as to the means of satisfying them.

The timing of such a review is appropriate because of the imminence of the Tevatron I and II accelerator programs which should come into full operation over the next few years. The increased computing demands expected to be placed on the laboratory from experiments planned around these facilities have created an urgent need for an upgrading of the laboratory's computing facilities.

Up to the present, the laboratory's needs have been well served, given the the budgetary restraints imposed by the priorities of machine construction, by the gradual improvement of the central computer along with user and laboratory supplied mini-computers. The user community has expressed its satisfaction that the successful physics done on the 400 GeV proton accelerator was, in no small measure, supported by the laboratory's Computing Department and its facilities.

However, as shown in Section II of this report, the experiments preparing to use these new accelerators are expected to cause qualitative changes in the laboratory's computer problems. These are predicted to appear as early as 1984 and grow rapidly by the first half of 1986 when collider experiments are scheduled to begin.

These detectors will certainly need much more capacity than is presently installed, and if predictions are even partially correct, more than the additional capacity scheduled for 1984. The proponents of the Collider Detector Facility (which is well along in design and construction) claim that they will generate 10^{12} bytes per year that would require the equivalent of 50 VAX 780's or three times the present CYBER capacity. While the exact configuration of the second collider detector is not yet determined it should not be far behind in computing

needs. Those claims are basically supported by the experience of the detectors currently installed and running in the $p\bar{p}$ collider at CERN and the chances are good that when these collider experiments begin at the Fermilab accelerator, the full capacity of the detectors will be needed from the very beginning. Therefore one should expect a full blast of the generated data to hit the computers almost at turn-on and increase in intensity as the luminosity grows.

In addition some new fixed target experiments running at high luminosity at 1 TeV will also generate huge amounts of data.

The central computing facilities of the laboratory should be upgraded in "capability" as well as "capacity". Increased capacity means more CPU cycles, tape drives, physical memory, disk space etc. Increased capability implies a greater number of special input and output devices that can be supported, number of large jobs that can be run in parallel, variety of the types of work that can be done, such as interactive graphics, improved quality of interactive computing etc.

The recommendation of the 1983 HEPAP Sub Panel on New Facilities to begin the design of the next generation accelerator as a multi-TEV pp collider will very likely make new demands on the accelerator physics and engineering staff of Fermilab which will in turn increase the demand for computer capability over the next five years.

Faced with the magnitude and variety of these new computing demands the committee felt that a new era was on the horizon for Fermilab computing and that immediate steps be taken to meet this challenge.

Future major experiments will consist of very large collaborations, international in scope. They will depend on a large variety of computers, microprocessors, emulators, preprocessors, etc. Both people and machines will have to be connected easily to each other. Access from the world outside Fermilab to its computers has to be fast and simple. Physicists will have to interact with distributed and central computers on a sophisticated level.

Almost every scheduled and proposed fixed target experiment designed around the TeV II will need augmented capacity at the experiment itself. This is

so, not only because some of them will operate at higher trigger rates, and have increased the number of detector channels, but also because second generation experiments usually, by their nature, require a higher level of sophistication in the data analysis. In addition, these computers may be used for offline analysis and thus lessen the amount of needed increased central processing capacity.

The experiments at the CERN Collider, (UA1 and UA2) have amply demonstrated the usefulness of three dimensional graphics. Even the "scanning" of events by physicists has been greatly aided by these devices. Fermilab will certainly have to enter this field as well – and should take advantage of access to the most recent developments.

Personal work stations with a wide range of capabilities in graphics sophistication and computing power are beginning to have a substantial impact on the entire field. Concentrated study of the stations as they would function as stand-alone systems and as highly independent "terminals" attached to large computers should begin soon, and should continue over the long term. As the computing power of desk-top stations expands, and physicists get the opportunity to exercise their imagination as to their uses, it is likely that the use of these devices will revolutionize computing throughout HEP. Substantial portions of the physicist's data analysis, calculating and document preparation (paper writing) tasks could be entirely removed from the province of centralized computing.

There is also the question of networking. High speed networking is needed at the laboratory in order to connect computers and transfer large amounts of data between them. Such a network connecting most Fermilab's VAX's to each other and each to the mainframe will be needed. Access by users at home institutions to almost all Fermilab computers – or as a minimum to the mainframe and the VAX's, used on their own experiment, should be made available.

All of the above items are more fully discusssed in further sections of this report.

The Fermilab Computing Department has outlined its future planning in a document entitled "Strategy for the Mid-80's" which is attached as Appendix B

and represents the response of the laboratory's staff to these impending fundamental changes.

It is probably true as stated in Appendix B that there is no single main processor, now available from the industry, that could handle the total generated output of the two detectors at the TEV I if a luminosity of 10^{30}sec^{-1} is achieved early on. The results of the survey tabulated in Section II show a requirement for the equivalent of 24 CYBER 175's by 1987 of which 16 will be used in collider experiments. Projections of availability of main processors at that time as discussed in Appendix G indicate that it is unlikely that a working machine, including the reliable software needed to support it, would be on the market soon enough to do the job. The best estimate that committee members could make, given the financial planning of the laboratory for main processors, together with a reasonably conservative estimate of the gradual decrease of the cost per unit of computing, was that by mid-1986 one could have a central processor that would handle the equivalent of 12 CYBER 175's or roughly half of the stated requirements. The remainder of the task will in all probability have to be performed using many emulators, programmable microprocessors, and super mini computers. These estimates have already assumed that the initial trigger rate of many thousands per second at the collider have already been reduced to less than 10 per second by hardware triggers and sophisticated preprocessors.

However, the experience at PEP, PETRA, CESR and the CERN $p\bar{p}$ collider shows that raw number crunching power is not enough. The programs written for those collider detectors are enormous and need access to very large memories. This is also true for the simulations (Monte Carlo, etc.) that must be done in order to debug and fine-tune the programs as well as for calculation based on actual data. Some idea of the magnitude of this effort is given in Appendix D entitled "The UA1 from a Computer and Software Point of View.".

A cost effective way to gain access to large memory address space is through a computer that has virtual memory management in conjunction with a large physical memory. This allows the computer to run programs whose memory requirements are larger than the physical memory and to run multiple programs

concurrently whose sum of memory space exceeds the physical memory space without the need of special programming.

The committee feels that the survey of future needs shows that a significant increase in the central processor capacity and capability should be in hand some time in 1986 – a year earlier than current planning calls for. Fermilab does not have the luxury of more than just enough time in which to decide on the nature of this next acquisition. Once this decision is taken, it takes a year or so of effort to prepare for the smooth introduction of the new system, a period of time which includes an elaborate education of the user community. Therefore there is at most one year in which to prepare the specifications for any new system which therefore has to be based on what is now available or in the very least what has already been announced. The laboratory should begin now to prepare specifications for this next stage.

The use of super-mini computers such as the VAX -780 is now widespread at many laboratories and certainly at most university HEP groups. Although the VAX lacks capacity compared to most large computers, it is strong in capability. That is, it has a fully interactive operating system for running and debugging programs from the users terminals, large memory address space, networking, etc. These capabilities enhance productivity as users of the VAX and other systems with the same capabilities will attest. We support the Fermilab efforts to provide every reasonably large experiment at the TEV II with a VAX or equivalent as part of its data gathering system. This VAX should remain in place for use of the experimenters throughout the experimental program and certainly between data taking periods. These VAXs should be networked together and with the central computer facility for purposes of program sharing, uploading and downloading run constants, and even ad hoc resource sharing.

As mentioned above, the large $p\bar{p}$ cross-section coupled with reasonable luminosity will produce a physics event rate so large that it will not be possible to record all events on tape for future processing. Furthermore, the experience on the CERN $p\bar{p}$ collider (see Appendix D) has shown that recording at rates above a few Hz is not feasible. In addition, because of the multiplicity and complexity

of the individual events, the offline analysis is expected to overwhelm even the future central processor capacity of the laboratory (see Section II).

To solve the online problem, data will need to be filtered in real time. The first stage is a reasonably sophisticated hardware trigger which eliminates unwanted events in the most transparent manner possible. The second stage uses programmable processors which then select classes of interesting events according to criteria which can be changed from time to time during the course of an experiment. The UA1 detector at CERN uses a llel set of 168/E emulators which makes choices based on algorithms developed for the offline analysis chain. The CDF group has the opportunity to choose between emulators (e.g., 3081/E's) or single board microcomputers with the performance enhancements now being developed by the Fermilab Advanced Computer R&D Project.

Once this choice is made, the CDF group should more than likely use these processors as part of their offline data processing in order partially to off load the central processor. Section VI discusses these problems in greater detail.

The tasks recommended to be performed will undoubtedly lead to greater responsibilities and work load placed upon the Computing Department as well as on other groups in the laboratory. These cannot be assumed without a significant increase in staff and some increase in the percentage of the laboratory budget which is devoted to computing with most of this increase earmarked for the Computing Department.

All the issues mentioned in this Introduction are discussed in greater detail in the body of the report and in the Appendices. However, before listing the committee's recommendations, one general observation must be made. Fermilab, understandably, has concentrated its main effort over the past few years in bringing to fruition the most advanced accelerator of the decade and is still deeply immersed in this task. The fact however remains that the laboratory has fallen behind in its overall computing capability although maintaining an adequate level of raw computing capacity. It now should continue the process of

catching up, and this should remain the laboratory's top computing priority for the next few years.

In order to implement the recommendations of this committee, we propose the formation of a standing Computer Coordinating Group appointed by and reporting to the director. The committee should be composed of high-energy experiment users, someone from theory, members of the Computing Department, and engineering, accelerator and administrative users. The members should be people with major computing responsibilities within their groups and should be prepared to meet often and on a regularly scheduled basis. The chairman should be a senior member of the laboratory staff. This committee should be charged with 1) reviewing all activities associated with computing at Fermilab, 2) making recommendations for future acquisitions and growth, 3) assisting the director in keeping the user community well informed. In addition, this committee will define, manage, and review working groups set up to pursue special computing system development task force projects, examples of which are described in Appendix F.

A visiting committee should be appointed by the director annually to review the progress that has been made in achieving Fermilab's computing goals. This committee should be composed mostly of knowledgeable people from outside Fermilab but with some staff and user membership. This kind of review should help Fermilab form a long term view of its computing environment.

Recommendations

1. Significant additional computing capacity and capability beyond the present procurement should be provided by 1986. A working group with representation from the principal computer user community should be formed to begin immediately to develop the technical specifications. High priority should be assigned to providing a large user memory, software portability and a productive computing environment.

2. A networked system of VAX-equivalent super-mini computers should be established with at least one such computer dedicated to each reasonably large experiment for both online and offline analysis. The laboratory staff responsible for mini computers should be augmented in order to handle the additional work of establishing, maintaining and coordinating this system.

3. The laboratory should move decisively to a more fully interactive environment.

4. A plan for networking both inside and outside the laboratory should be developed over the next year.

5. The laboratory resources devoted to computing, including manpower, should be increased over the next two to five years. A reasonable increase would be 50% over the next two years increasing thereafter to a level of about twice the present one.

6. A standing computer coordinating group, with membership of experts from all the principal computer user constituents of the laboratory, should be appointed by and report to the director. This group should meet on a regularly scheduled basis and be charged with continually reviewing all aspects of the laboratory computing environment.

II. FUTURE CENTRAL COMPUTER NEEDS

This section deals with the forseen demands that will be placed on the Central Computer Facility at Fermilab over the period 1984-1987. This period will be crucial to the laboratory with the startup of the Doubler and Collider providing greatly enhanced physics possibilities. In addition there is the prospect of work commencing on a new super-collider, which may entail more detailed design studies to be performed than has previously been the case because of the size and costs involved.

The estimates presented below are based on the replies to a survey of all scheduled experiments (See Appendix C) for the coming running periods and also from a survey of non-experimental computer users, in particular the accelerator designers and theorists. It is important to realize that the results of such surveys are based on users own conceptions of what they will require in the future and, in particular, may be biased by one person's viewpoint. In some cases the estimates received were very vague, but in others the estimates have a more solid foundation since the experiment concerned is an extension of an existing one. The quoted CPU requirements are those on the Fermilab Central Computer Facility. Thus, where an experiment expects to utilize other computer installations, either off-site, or as trigger processors, the CPU power required by these installations is not included in the totals.

Over the next four year period the following amounts of computer time are forseen for each of the major usage areas. (The numbers are in Cyber-175 equivalents.)

	1984	1985	1986	1987 †
Proton Area	3	4	5	5
Neutrino Area	1	1	1	1
Meson Area	1.5	1.5	1.5	1.5
Collider	1	3	10	16
Accelerator	1	1	1	1
Theory	*	*	*	*
Sums	7.5	10.5	18.5	24.5

*Estimated theory requirements are relatively small

† No fixed target program has yet been approved for 1987

Comments and Conclusions

The estimates in the above table are significantly different from a simple extrapolation of present trends for a number of reasons:

1. The fixed target experiments have increased in complexity and enhancements to the beam lines have allowed previously mutually exclusive experiments to coexist.

2. The Collider experiments, CDF and D0, with their high trigger rates and high particle multiplicities, will require large amounts of computer power. For CDF, estimates of the required power required to analyse the data accumulated over a one year period, given inefficiencies both in the data acquisition and reconstruction, arrive at a number of 10 Cyber equivalents. The numbers entered in the summary tables show 8 of these equivalents being operated by the Computer Department, while the remaining 2 are within CDF itself. The exact allocations of these resources and the form which they may take are the subject of active discussions between CDF and the Computer Department.

   The D0 experiment is in a far greater state of flux at the moment. The indications are that the number of electronic channels will eventually rival CDF. Thus the figures for D0 are assumed to be the same as for CDF, but

delayed by one year.

3. The conception of what a modern computer environment can provide is changing rapidly, perhaps more so now than at any previous time. This puts great emphasis on a high degree of interaction with the computer to enhance the productivity of the programmer or physicist. An example is the great use of highly interactive graphics programs by the experiments at CERN. This item is discussed in more detail in other sections of this report.

These figures indicate that, once the accelerator is working well, the presently planned computer capacity (including the present acquisition) will rapidly become saturated by the fixed target program. Significantly greater computer power will need to be installed by 1986 with the commencement of the collider experimental physics program.

It is unlikely that large mainframe computers will be cost effective in meeting all the needs of 1986 and 1987. Instead, the Committee anticipates that by 1987 at least 1/2 of the total can be met in this way, the remainder will require alternative solutions. These may include the development of "farms" of emulators such as the 3081/E or microprocessor based systems such as those proposed by the Advanced Computer R & D Project. VAX style super-minis can also provide some of the required CPU power.

The problem of assessing computational requirements for theorists is bound to be an exercise in guesswork. An experiment, before it comes anywhere near making use of extensive computing, involves the commitment of a large amount of manpower, financial support, precious machine time, etc. Assuming that the planning has been competent, that implies the existence of a more or less known amount of data, of which a reasonable proportion must be analysed in order to achieve any scientific return. In theoretical work the inertia is much less. For those theoretical studies which involve large amounts of computation, the computing itself is the major cost, and so to estimate the amount needed is like estimating what budget is "required" by high energy physics.

Theoretical computing needs fall into the following categories.

1. General purpose computing done quite effectively on a VAX or 4341 and presumably eventually on work stations. This accounts for a large proportion of theoretical computing measured in terms of scientist time.
2. Conventional number crunching, such as numerical integration of quantum electrodynamic expressions, solving integral or differential equations as in scattering theory or orbit dynamics or magnetic field calculations. These problems have geometrically increasing requirements.
3. Monte Carlo calculations, especially for lattice gauge theories. Here the growth in requirements is compounded from the need to increase accuracy (quadratic) and the need to increase dimensionality and lattice size (highly geometric). This is currently the area where the greatest anticipated consumption of resources is expected.
4. Symbolic algebraic manipulations (Macsyma, etc.). These typically have exponentially growing requirements. While these calculations have not been major resource consumers to date, the demand could quite suddenly explode if some particular QCD or other calculation becomes fashionable.

The evidence all points to the conclusion that demand for theoretical computing is potentially open ended, is likely to show strong fluctuations in direction depending on fashion, and is likely to be strongly influenced by what was made available in previous years. For the time being, a prudent suggestion would be to allow it to track some small fraction of the more easily modelled experimental usage.

Other major needs for computing in support of theory may perhaps be handled through the use of supercomputers at outside sites. Special processor projects may also be appropriate; the NNCP project of G. Fox at Caltech, the Gibbs project of K. Wilson at Cornell, and the SU3 project of N. Christ at Columbia being examples of what could result from strong local interest.

III. HIGH ENERGY PHYSICS TASKS AND THE COMPUTING ENVIRONMENT

The principal elements in a modern "computing environment" are discussed, and the role of each element in the overall computing task of a major experiment is outlined below. Particular attention is given to the "interactive environment" which users require for program development and many analysis-related tasks.

A. Computing Environment

The term "computing environment" refers to the ensemble of facilities which physicists have available to them to carry out all the software-related design, simulation, and analysis tasks for their experiments. Successful completion of the task means that the data are completely analyzed and physics results are extracted with minimum delay, within the confines of available money and manpower. In order to set the proper future course for computing at Fermilab, we attempt to specify guidelines for the highest quality computing "environment", in which a balanced, integrated set of hardware and software facilities are made available to the physicists, allowing them to work most efficiently.

The evolution of the computing environment in HEP has been driven by the exponential increase in size and complexity of the largest experiments, which have now reached 2000 cubic meters of detector volume, and 3×10^5 electronic channels. The software base to be developed now runs to several $\times 10^5$ lines of Fortran code, and as many as 10^6 geometric and calibration constants are needed to describe the detector for analysis and simulation. Although the number of physicists on each experiment has increased to over a hundred for CDF and UA1, and to 160 - 260 for the LEP experiments, the development of the software base falls to a relatively small number of people, typically no more than about 20 full-time equivalents.

The quality of the software development environment will therefore determine whether the experiment will be ready on time – ready, that is, to extract physics results early in the run.

The basic model we will use for future computing at Fermilab consists of a strong central computing facility which is coupled via networks to a number of auxiliary computing systems, including super-minicomputers, special processors for compute-bound jobs, and specialized devices for graphics and large-volume program development. The heterogeneous nature of the overall system is an unavoidable consequence of the variety of tasks which a large collider-detector group faces in carrying out the experiment. A representative list of these tasks is given in paragraph C of this section.

There are potential problems with this model, particularly in "integrating" (providing effective interfaces between) the diverse parts into a coherent overall computing "system". The core of the issue is whether the special features of a given new device (such as an emulator, hardwired processor, or new computer operating system) are worth the effort required to implement efficient and user-friendly interfaces to the conventional parts of the computing environment, namely the central mainframe and super-mini clusters.

Nevertheless, it is the conviction of the committee members that a strong central system, augmented by – and networked to – sets of super-minis and special purpose fast processors, is essential for detectors on the scale of modern collider experiments as well as the very large fixed target experiments. This is the best generally accepted way to provide sufficient computing capacity and capability and a wide enough range of facilities to fulfill the physicists' and engineers needs at reasonable cost.

B. Interactive Environment

The term "interactive environment" will be used in general to describe the editing, file handling, job submission, on-line running and system monitoring facilities which an individual physicist using a keyboard and a screen encounters. The "screen" may be a terminal of varying degrees of intelligence, a specialized graphics device with complex hardware and software functions built in, or even a stand alone computer with its own operating system. In a more restricted

sense, a truly interactive system allows a user to run jobs online and to alter their execution through input from his keyboard (or other input device).

The phrase interactive computing apparently means different things to different people, depending on which system(s) the person uses. The following list should help to distinguish what is meant by the evolution of the "interactive computing environment". We define four levels of increasing sophistication:

1. The lowest level of interactive computing gives the user the capability of terminal supported text editing, batch job submission and retrieval. This is the current level of support at CERN under WYLBUR. The file editing system at this level is typically a line editor.

2. The next level is all of the above plus interactive program running. By interactive running it is meant that one can write to the terminal screen using instructions in a job whose execution can be started from the terminal without entering the usual job queues. One can also have the running program accept input from the terminal during its execution. A particularly important use of these functions is interactive job debugging, in which the running job may be interrupted at specific "break points", and the values of variables in the program may be examined or altered.

   This is the current level of support at SLAC on IBM mainframes and on VAX's running VMS. According to the definition, the DESY IBM system under NEWLIB, and Fermilab's CYBER's are at this level. Because of the limited region size and other restrictions, however, DESY and Fermilab effectively function at level 1. Editing facilities at this level typically include a full screen editor in which a page of a file (typically 20-25 lines on a VAX and up to 43 lines on an IBM) may be stored locally in a buffer in the terminal, and changes anywhere on the page may be entered on the screen in any order, until the page is sent to the supporting computer to be updated.

3. The third level of interactive computing is associated with interactive graphics in the sense of being able to support a graphics screen with a

light pen, mouse or tablet. One can input commands from the keyboard of by referring to screen locations and inputting symbols from the mouse (for example). This level of support is currently available on a limited number of systems which function mainly at level 2, often with the aid of special graphics programs which run on a mini-computer. Full support for this level has not been achieved without substantial work on the part of the local computer experts.

4. The last level is reserved to characterize the environment on upper-end personal work stations. It is important to realize that this level, which exists nowhere on large mainframes or super-minis at the moment, may become the standard sometime between 1985 and 1990. As discussed in some detail in Section VII the distinguishing feature of this level is the handling in the work station itself of a large number of full-scale processes which run in parallel. The user may view the output from and may interact with several processes at once through windows on the screen.

An interactive environment does not imply that there is no need for the batch environment as well. In fact most capacity of a mainframe will be consumed in batch with jobs lasting hours at a time. Ideally, the batch environment should be a similar as possible to the interactive environment, sharing the same command syntax, data files, etc. This is needed so that once programs are debugged interactively, they can be sent to batch processing with minimum changes.

In specifying a future computing environment for HEP, one would certainly like to aim for level 4 by 1986. The problems and pitfalls for achieving this in a large-scale coordinated analysis effort are unfortunately many. We should therefore aim for facilities at least at level 3 in any major computing acquisition beyond 1985, and allow for level 4 on a limited number of work stations if users desire it.

Level 2 is the minimum level at which users function efficiently. The user at this level or above is far more productive than he is at level 1 because routine jobs such as editing, file handling, and the execution of utilities are far easier and

quicker. The entire process of program debugging goes much faster, particularly for large jobs which fail only after significant execution time.

Without an interactive environment at level 2 or higher, the user must adjust to the relatively slow process of job setup, batch running, and job retrieval by reducing the number of jobs submitted and the amount of work done. Many tasks which could be done in minutes at the higher levels stretch into hours, and in extreme cases into days.

Fermilab should fully implement a level 2 interactive environment as soon as possible.

C. HEP Computing Tasks

In order to emphasize the need for diverse computing facilities, and to illustrate the broad scope of analysis-related work faced by large experiments like CDF, we enumerate some of the tasks:

1. Test beam analysis for the development of data bases characterizing detector response and calibration.
2. Development of online and offline event filtering algorithms.
3. Development of calibration and monitoring programs to test and check detector performance during running, and to track the variation of critical detector parameters with time.
4. Generation of event samples for a variety of physics processes of interest.
5. Development of a complete detector simulation, including all available survey, detector calibration, field map, and detector response information. This simulation may incorporate special simulations of electromagnetic and hadronic showers, to be validated through test beam data.
6. Development of pattern recognition and reconstruction algorithms for multi-track events, and for energy flow patterns in calorimeters.
7. Trigger and online filter validation and checking by offline programs.
8. Specialized detector design studies (using specialized shower simulations

such as CALOR or EGS for example).

9. Production running of Monte Carlo and detector simulation jobs to establish data bases of simulated events for each physics process of interest. In some cases, this running is continuous, as new physics ideas, analysis ideas, and an improved knowledge of the detector lead to an ongoing demand for new simulated event samples.

10. Production analysis, including the final stages of event filtering, pattern recognition, and reconstruction. Beyond the early states of the experiment this will not be among the most labor-intensive of the physicists' activities, although it will consume the largest part of the computer time required by experiments such as CDF and many fixed-target experiments. The ongoing development of the analysis chain, which is the driving force in the continued demand for new simulated event files, consumes far more effort.

11. Physics analysis. The majority of physicists in each experiment, even if not involved in developing the mainstream of the production programs, will participate in the analysis of one or more physics reactions of interest. As the experiment reaches a stage of stable operation, and as the production program stabilizes, this becomes the main experimental activity.

12. Event scanning and interactive analysis. Recent experience of UA1 and UA2 shows that it is important for large collider detectors to scan the events in final candidate samples using sophisticated interactive graphics facilities, as described in detail below. Fixed target experiments at Fermilab have shown variable degrees of interest in interactive graphics, but the need is well established for large solid-angle colliding beam experiments at $e^+ e^-$ as well as $p\bar{p}$ machines.

It is clear that the above list is necessarily incomplete, and that it is biased in terms of the needs of the largest colliding beam experiments. It should serve, however to illustrate the diversity of analysis activities which will be the central concern of experiments at Fermilab in the period of 1985 - 1990.

D. Quality of the Computing Environment

The quality of the environment depends not only on the overall capacity expressed in CPU cycles but on the computing "capability" which depends on a number of crucial characteristics of the software systems which are essential to increase human productivity. Optimum choices of computing systems for the lab are thus a compromise between the capacity per unit cost, and the operating systems' friendliness, flexibility and "sophistication" which serve to determine its capability. A variety of special devices such as hardwired processors or emulators may also be incorporated for increased computing capacity at low cost, but most likely at some inconvenience to the user relative to general purpose computers.

Once sufficient CPU power is provided, quality of the computing environment should be the prime consideration, as measured by the following criteria:

— Does the environment provide an efficient means for doing all the essential data-taking, reduction, and analysis - related jobs?

— Are the hardware facilities fully utilized without unduly sacrificing convenience or quick availability of computing power to users?

— Are the facilities sufficiently centralized, at least in part, to allow for effective use of limited manpower?

— If the facilities are distributed, as in the case of many super-minis around the site, are the overall hardware and software maintenance tasks reasonable?

— Does the environment (ensemble of computing facilities) function as a flexible, integrated whole? Can the user choose a part of the total system best suited to perform a particular task efficiently (such as emulators, or work stations) and not be subject to excessive "overhead" ? For example, will the user have to prepare specially formatted input or output files, or will he have to transport large amounts of data across slow or unreliable links?

E. Principal Components

The main components of a modern HEP computing environment (ca. 1985 -1990) are:

1. A large mainframe. This need not be a single processor, but must be capable of supporting upwards of 150 users logged on simultaneously, on terminals or work stations. It should present a "single system image" to the user, meaning that he need not concern himself with which processor he is communicating with or on which one the jobs he submits will run. It should however possess certain features namely:

   (a) A memory management system which provides each user with a multi-megabyte memory space and which includes sufficient physical memory to allow several of these programs to execute simultaneously without significantly degrading system performance.

   (b) Facilities for including the machine in a network accessible by computers of other manufacturers, using standard protocols. Links to VAX computers are particularly important.

   (c) Facilities for providing a high-level interactive environment to the users, as specified above.

2. Banks of special purpose processors for "number crunching". These may be single board level micro-computers or emulators which are capable of running any program which runs on a mainframe, or the processors which are specialized for a small set of well-defined algorithms.

3. A distributed set of super – minis such as VAX 11/780's (or their successors which are about 4 times faster). One set of these machines will be used mainly for online, and another larger set mainly for offline applications.

4. A set of computers at experimenters' home institutions, mainly VAX's (or their successors).

5. Personal work stations. These devices can take one of several forms, according to the experimenters' particular needs and desires:

(a) Specialized graphics stations.

(b) Stations which are integrated into the operating system of a major computer manufacturer.

(c) Stations which function as a complete computing system without mainframe support.

These are described in more detail in Section VII.

6. Networks of intercomputer links at several levels:

(a) Local area networks (LAN's) within counting rooms, single buildings, or wherever the network extends over not much more than 1000 meters. These links are for high speed file transfer as well as for work station interconnections. Effective speeds (not raw line speeds) of 1 Mbit/sec or more are desirable.

(b) Wide area networks covering the lab site. Once again, speeds of 1 Mbit/sec are desired; perhaps as high as 10 Mbits/sec peak for urgent short file transfer.

(c) Long range low speed links to the outside world. Dedicated links of up to 56 kbaud may be used over a few hundred miles. Worldwide links of up to 9.6 kbaud may be used through TELENET and TYMNET (X.25 links), for interactive remote terminal sessions as well as for transfering small files. File transfer is also provided, in a more convenient and perhaps more reliable fashion, by BITNET. This rapidly growing network will probably provide extensive access to European as well as U.S. sites by 1985.

(d) Satellite links covering the whole speed range from 9.6 kbaud through 1.5 mbaud. These links are generally very reliable and they are now quite cost competitive, especially for traffic across the U.S. Service is provided through short ground lines to nearby major earth stations, or (more recently) through the local installation of relatively low-bandwidth small radius dishes.

F. Computers at Experimenters' Home Institutions

The computers at physicists' home institutions (mainly VAX's) will be used extensively for program development and for physics analysis. The use of these computers is restricted, however, by their widespread distribution and the fact that no single home-based facility may be comparable in computing power to the central facilities at the lab site. Use of home-based facilities for large scale production would be quite inefficient in terms of time and manpower, unless high speed links and corresponding network hardware and software interfaces were provided. Management of a large-scale distributed production process, with data analyzed at a number of institutions, would also present new and complex problems of coordination.

Some universities involved in experiments at Fermilab (as well as other HEP labs) are, however, setting up emulator facilities at home capable of handling a significant part of the production analysis of even the largest experiments. This could be a cost effective way for them to lessen the computing load on the lab's central facility, and the work load on the support staff.

G. Conclusions

We have presented a model for a future computing environment which incorporates a central facility networked to several auxiliary computing systems, including a VAX network, emulator or special processor facilities, and (possibly) local area networks of Personal Work Stations.

The balance between these different elements is difficult to specify solely on the basis of the merits of each type of system for a given computing activity as perceived by one or more Committee members. It is recommended therefore that work projects be set up, allowing users of all major experiments to evaluate the uses of heretofore unfamiliar systems and devices, before committing the lab's resources. The lab should also seek the opinions and direct participation of experts and interested physicists at other HEP labs.

Since the next major acquisition, beyond the current RFP, is planned for late 1986 the work projects could extend until late 1985 before the final specification

of the new system is formulated. By this time, it is hoped that the new system architectures for large computing facilities are also on the market so that their advantages and limitations can be realistically evaluated.

IV. EXPERIMENTAL AREA COMPUTERS

Many of the fixed target experiments have reached the point where a single PDP11 can barely manage the data acquisition and recording of events. Very little computing time is available for running special monitoring and calibration programs or doing even crude online data analysis to verify that the equipment is working properly. Detailed data analysis, for example full track reconstruciton on a subset of events, is difficult to accomplish with the existing computing power.

To remedy this situation, the Computing Department has undertaken the so-called "Connected Machines Project". Experiments will be given additional computers, VAX 11/780's or in some cases PDP11's, to augment their online computing capability. The intent is for one PDP11 to manage data acquisition and event recording while passing samples of data over a high speed link to a data analysis machine which would be a VAX in the case of a large experiment.

The committee makes the following observations and recommendations concerning the Computing Department's plans for future online computing at Fermilab:

1. For most large experiments, a VAX 11/780 is necessary in order properly to monitor and calibrate the detector systems while running. A partial physics analysis is the best test of data integrity and this type of analysis is the only possible with computing power on the order of a VAX. The committee strongly endorsed the Computing Department's plan to provide this capability to all major experiments.

2. Offline software developed on the VAX could also be useful in online data analysis. Certain useful interactive debugging and analysis tasks (e.g. obtaining calibration constants) could be implemented on the VAX. This would reduce the demand for this kind of service on the central facility and leave that facility free for data crunching. Experiment dedicated VAX's would also help experimenters develop hardware and microproces-

sors to do part of their offline analysis. Experiment 690 is an example of an experiment that intends to use processors connected to a VAX for offline analysis. For these possibilities to be fully realized, all long running programs need to have access to a dedicated VAX for the duration of the experimental program. The Computing Department and the laboratory are urged to acquire enough VAX's to make this possible. The alternative, shuffling computers from experiment to experiment, is costly both in setup time, exposure to damage in transit, and will prevent stable, long-running experimental programs from making full use of their machines.

3. Networking of the onsite VAX's to the central facility software development VAX's is both necessary and logical. Further networks to home institution VAX's would be very desirable and open up further opportunities for more off-line analysis exclusive of the central computer. The committee strongly recommends that Fermilab tie all the experimenter VAX's and the Central Facility together in a network.

4. The committee strongly urges that the necessary personnel be added to the Computing Department staff fully to support networking the Experimental Area VAX's to the other machines onsite and to a wide area network.

5. The future plans should pay more attention to the use of the collection of VAX's during lengthy shutdowns. The fixed target program will be off for long periods of time - for example, during collider runs. Thus, there will be different levels of activity on the various VAX's. It is reasonable to assume that some of these will be somewhat underutilized. A study should be made to see how this computing resource can be made more generally available.

6. The VAX's stationed at experiments should be supplied with a set of peripherals that allow full interactive use of the VAX/VMS system to develop software, analyze data and monitor the experiment. These peripherals should include sizeable disk storage (500 mbytes), reasonable memory size (2-4 mbytes) and at least one dedicated 6250 bpi tape drive. The goal is

to provide a system in which jobs can be run and debugged interactively and with full screen editing capability. (Level 2 interactive computing, as defined in Section III.) Since graphic displays with interactive event scanning capability is a primary function for online and any interactive offline system, these experiment VAX's should be provided with good graphics hardware and software.

7. Online software should be developed for these experiment VAX's that takes full advantage of the VMS system and multi-tasking capability. The current VAX-Multi software is a start but an ambitious program of online software development should be vigorously pursued. Close ties with the CDF efforts should also be made.

Fermilab is moving in a good direction, one that will provide adequate computing power to monitor and calibrate large experiments. Its approach is flexible in the sense that it is prepared to consider several connected computers (some of which may be suppled by Users) sharing the problem of taking data, monitoring equipment and analyzing, some data on-line. It seems, however, that it is not acquiring enough VAX's to support the experimental program. It may be that 25% more funds in this area will be the difference between a merely adequate on-line computing environment and a much better one.

V. NETWORKING

There are various kinds of Networking and each satisfies different needs. As with the words 'interactive computing', 'computer networking' means different things to different people. Let us try to understand networking better by saying that are two different kinds:

1. Terminal to computer (local computer or remote computer).
   - to use the computer as primary computer resource.
   - to make enquiries of and occasional use of that computer.
2. Computer to Computer (local computer or remote computer).
   - to send mail/files/messages without logon.
   - to receive the above asynchronously

Now let us explore which existing facilities are available and at a reasonable cost.

Terminal to Fermilab Computer :

- Interlab 9600 baud terminal to computer digital switch, i.e. MICOM. This has been installed at Fermilab for quite some time and at other laboratories like SLAC and CERN.
- Dial-in to digital switch over standard telephone lines. Standard telephone lines will support 1200 baud. This speed is tolerable for use of terminals at home using remote logon. Problem is the expense of telephone call from long distances. Thus this service is good for local calls, home to lab, and good enough for occasional calls at long distance. Fermilab currently supports this service.
- Public packet switching networks to digital switch, i.e. Tymnet or Telenet. These public networks require rental of equipment on host computer but offer much lower cost to long distance user. Use of these networks is the

only way to make transatlantic computer calls at a reasonable cost. Fermilab does not support this service apparently because of lack of interest by remote users who could benefit. SLAC, CERN, Rutherford, Saclay, and even the Mark J VAXs at DESY and CalTech all support this type of connection. Fermilab should take another look at supporting this service since it costs only about $1000/month.

— Leased lines. At modest cost, $10,000/year within U.S., one can lease a 9600 baud line from a remote institution to Fermilab. When coupled with Statistical Multiplexors at each end, several terminals all appear to be runing at 9600 Baud (as long as the users at each terminal don't ask for output exactly simulaneously). This service is being supported by Fermilab currently, as well as by SLAC. This is a very cost effective service for users remote from the laboratory and who need to use a laboratory computer extensively.

Terminal at Fermilab to remote computer:

— Dial out via digital switch. Fermi supports this currently, but only at 300 Baud because the telephone lines used are FTS, which apparently can't support 1200 baud. Obviously, the laboratory could not support unrestricted calls on commercial telephone lines, but some arrangement should be made to allow 1200 Baud dial out service via commercial grade lines.

— Dial out via digital switch to Tymnet/Telenet lines. This would allow Fermilab terminals to connect to remote computers as a user of Tymnet/Telenet. Thus connections to SLAC, and Europe can be made. SLAC, DESY, CERN, and other users have this possibility now, Fermilab doesn't.

The use of these services allows terminal to computer connection without much effort. Which a particular user chooses, depends on the quality of service he is willing to pay for and the frequency of expected use. They do serve the purpose of immediate access to/from Fermilab with other unexpected sites and with sites only occasionaly used. Their cost is sufficiently low so that one need

not hesitate about implementing them.

It is also possible to do mail and file transfer using a combination of the above services. There are problems with each, however. One needs to have an account and know how to use the remote computer. One can initiate mail to a remote user at will, but one cannot receive mail unless one logs on again, or the responding party logs on to your computer. The problem of everybody having accounts on everybody else's computer and knowing how to use it is a big one, to say the least. File transfer can be accomplished by using a computer to logon to a remote computer with special software. Although this is workable, especially between identical computer types, it is far from completely satisfactory especially for long files (several thousand lines) and at slow speeds (1200 Baud).

Computer to Computer networks :

There are in existence many computer networks using dedicated or dial-up leased lines. These networks allow file transfer with supported software. Mail transfer is a subset of file transfer. The important point is that one need not logon to a remote computer to use the network, although some of the networks also allow remote logon. Examples in the U.S. of such networks are:

— DECNet. DECNet allows mail/file transfer and remote logon between DEC computers; VAX, PDP-11, DEC-10, etc. It will operate over various line speeds and physical media. One example of a HEP DECNet is the SLAC/LBL DECNet which includes all 18 VAXs at SLAC, another 10 at Berkeley, and many other home VAXs of users of SLAC from Southern California to the Midwest (even the CDF VAX is a dial-in node to this Network). The only problems with DECNet are that mail can't be initiated unless the all the required links are operational (making dial-in nodes difficult to use), and the fact that it runs only on DEC computers. Fermilab should link all VAXs and all PDP-11's in the experimental areas together via DECNet, and have at least one of them be a host for linking leased lines from outside the laboratory. A leased line to the Agonne VAX

on SLAC/LBL DECNet, for example, would make the Fermilab DECNet a permanent member of the SLAC/LBL network.

BITNET. BITNET allows mail/file transfer between various computers. It is based on an IBM RSCS protocol and is almost no effort to operate in an IBM VM/CMS environment. RSCS emulating software exists for other computers, however, including VAXs running UNIX or VMS, and CDC CYBER 175. Remote logon is possible between consenting machines running VM/CMS, although this is not encouraged. BITNET currently consists of 99 computers at 23 Universities, of which many have high energy physics groups. BITNET may also be soon extended to Europe and unless the various PTT's block its use in Europe, will be extended to CERN, DESY, and Rutherford. Fermilab should join BITNET, and has already taken steps to do so. At least several of the key VAXs at Fermilab should also be on BITNET, and Fermilab users should be encouraged to link their VAXs to BITNET by making a link to the nearest BITNET node which might be as close as the University computer center.

- Other networks: Other established networks have more restrictive use or more limited community. ARPANET was originaly form to link DoD contracters or university DoD research centers. The military part of ARPANET seems to be in the process of splitting off to another network now, leaving ARPANET for more general use. Argonne and Brookhave are both ARPANET nodes, but they also may become BITNET nodes soon. CSNET is a network of computer science departments which uses ARPANET, TeleNet and dialup for transportation media. It is restricted to computer science research. The largest network is probably USENET which links over 1600 sites via dial-up lines. Unfortunately, all are UNIX sites, an operating system rarely used in HEP. It is recommended that Fermilab not concern itself with these other networks.

VI. SPECIAL DATA PROCESSORS

Introduction

Nothing is more flexible or as easy to use than a general purpose computer. However, it has always been clear that one can process data with higher speeds and with much lower hardware costs with a processor designed and built for the purpose of running one algorithm. In the past, because raw data volume was sufficiently small and special purpose processors were sufficiently inflexible and hard to use, most data analysis was done on general purpose computers. Special purpose processors were mostly used as a low level trigger in relatively simple experiments.

In spite of the ever increasing sophistication of lower level triggers, today's and foreseen future detectors will generate a much greater volume of data. This makes it extremely difficult to justify continuing to process all the data on standard mainframes due to the very large investment that would be required for their procurement. The data volume would in fact exceed known technologies for permanent storage in some experiments, such as CDF, without some processing on-line to the experiment.

Fortunately, data processors designed by the HEP community have reached a level of flexibility, by being programmable, such that they can start to play at least part of the role such that had previsously been the domain of real computers. This is evident with the development of processors such as the M7 machine at Fermilab, the ESOP at CERN, and the fast-rising use of emulators such as the 168/E at SLAC, CERN, DESY, and elsewhere. Many of the largest existing experiments depend on these processors for much of their main-stream data filtering and/or production analysis, and many of the largest experiments now under construction will depend on new processors such as the 3081/E.

There are a number of reasons that these processors are more cost effective for data processing than large computers. These reasons also lead to the fact

that they are generally more difficult to use than computers. The usefulness of a processor project will depend on how well these difficulties are overcome in the system integration. The most important reasons are outlined in the next paragraphs, where emphasis will be placed on the off-line processing. The special case of on-line processing will be treated later.

Terminology

To make this discussion clearer, we need to first define what is meant by special purpose processors, etc. A special purpose processor (SPP) is one with specially designed hardware to do or assist in doing some predefined algorithm. Its programmability varies greatly depending on its design. For example, the Columbia (Knapp-Sippach) hardware processor for E-690 reconstructs events at a speed equivalent to 100 Cybers. Programmability, however, is only achieved by moving wires in the processor or changing the contents of lookup tables in the processor's memory. The M7 processor (Droege et al) or the ESOP processor (CERN) do not require changes in wiring, but programming is done in the fairly difficult microcode of specialized instructions.

Another category of processors is emulators. They are specially designed processors that are capable, with some limitations, of running almost any program that was developed on the computer being emulated. For example, the 168/E emulator at SLAC runs FORTRAN programs developed on an IBM mainframe at speeds up to 1/2 of an IBM 370/168 for a hardware cost of only $ 5,000. A strong point for emulators is that they usually are fast in carrying out floating point instructions.

A new category of processors may best be described as Single Board Computers (SBC) which use microprocessor chips, such as the 68000 with its associated memory, on a single or few circuit boards. Programmability depends on the compiler that runs on it or that runs on another computer (cross- compiling). The speed of SBCs will be at most 1/2 a VAX 11/780 or about 40 times less than the installed CYBER capacity at Fermilab based on manufactorer's announcements for products to be delivered in 1984.

The Fermilab Advanced Computer Project's current plan is to use SBCs as a base and to argument their capability with hardware subroutines. If these subroutines go beyond simple math functions, the Committee would classify the result as a special purpose processors for reasons that will become clear below.

A single emulator or single board computer alone is not sufficient to make a large impact on the already installed capacity at Fermilab. Both of these types of processors, however, can be used in a multiple processor system. For example, 6 3081/E's in a multiple processor system will have the same CPU capacity as the 3 CYBER 175's.

Applicability Criteria

Simple criteria which can be used to judge how applicable a proposed processor or processor system will be for a particular environment are discussed below.

Special purpose processors, single board computers, and emulators generally do not directly support the large scale I/O peripherals that computers do. Thus a general purpose computer, either large or small, is needed to support these I/O peripherals for the processor, examples of which are 6250 BPI tape drives, Gigabyte disks, printers, etc. It also means that since data will need to be transferred to and from the processor system by a computer, the amount of computer time one saves by using the processor system must be large compared to the amount of time to do the I/O. Otherwise it would be far more efficient to use the computer in the first place. It also means that the user needs a certain amount of code preparation for one part to run on the I/O computer and the other part to run on the processor system. Thus the role of special purpose processors, single board computers, or emulators is restricted to the CPU intensive part of the work load, which would take up a significant portion of the overall computing time available if run on a general purpose computer.

Special purpose processors, single board computers, and emulators generally do not have the hardware support needed to run multiple jobs concurrently. Therefore, once a long production job is in the system, they cannot be used to run short test jobs. Also, while a processor is waiting for I/O, it can not be used

on another job. One could not run a computer center this way, and this restricts their use for program development. Emulators have an advantage over special processors in that program development may be done on a "host" mainframe, and the program may be down-loaded into the emulator for production runs. Program development for a single board computer can also be done on another computer if the compilers or cross compilers exist and computers are sufficiently compatible.

In general special purpose processors are not fully programmable, in the sense that a computer can mix jobs from different physics groups without hardware or software changes to the computer. A general purpose computer can even take a different version of the program with each run within a single job. This is certainly a problem with special purpose processors. They may achieve higher speeds at low cost in some cases but it may take from weeks to many months to change the program running on the system, and even a year or more to design, build, debug, and make production units of specialized hardware algorithms.

Single Board computers and emulators do not have this problem, since their hardware and software are not job specific, except in that the user code is generally split between the code on processors and the code on the computer used to drive the I/O, for reasons mentioned above. Emulators such as the 3081/E should be capable of running the vast majority of HEP analysis programs now on IBM hosts without major rewrites. One should keep in mind that the experience of all existing colliding beam detectors is that the program is in state of continuous development over a period of years before, during and after the initial data taking. The programs may not reach completely stable state until the time the detector is turned off and dismantled.

General purpose computers come with a considerable amount of support for hardware and software from the manufacturer and laboratory staff. It is very rare that the user need take measures to verify that the hardware or system software is working correctly. With special purpose processors or emulators, such verification measures must be taken. With single board computers more is available from one vendor but one needs to verify that the compiler is correct. At

least in the case of emulators, the verification may be done by simply running the same job on a sample of events on the computer emulated to see if the results are bit for bit identical. When the verification is not successful, then the hardware and system software support that is available in a processor system becomes extremely important in rapidly locating and fixing the problem.

The designers of special purpose processors or emulators may have imposed other restrictions or limitations for the sake of speed, cost, or simplicity. For example, the 168/E emulator has a data memory limitation of 1/2 Megabyte, which forces the user to do overlays if he wants to run the complete event reconstruction code of a large detector. Special purpose processors such as the M7 and the matrix multiplier at Columbia, do arithmetic only in 16 bit integer.

Future processor projects are, however, removing these restrictions as their impact is being fully realized and the cost of components (memory in particular) comes down. For example, the 3081/E emulator will initally have 3.5 Megabytes of memory (14 later) and several special purpose processors are planning the implementation of floating point.

In the case of the largest new experiments like CDF, however, emulators, single board computers, and special processors will provide the only way in which the bulk of the "production analysis" can be carried out at reasonable cost. This is because the large number of electronic channels, the geometrical complexity of the detector, and the particle multiplicity per event make the final stages of event filtering and reconstruction extremely time consuming. These special devices are also essential for extensive Monte Carlo event generation and detector response simulation, as a large pool of simulated events is needed for studies of detector acceptance and reconstruction algorithm performance.

Note that the need for substantial computing power may occur years before the start of data taking, particularly for experiments requiring complete simulation studies of electromagnetic and hadronic showers. This need is naturally fulfilled by emulators or single board computers, as sufficient mainframe CPU

power is not likely to be available at this stage of the experiment's development and special purpose processors are not likely to be running yet.

The use of these processors in the on-line environment has most of the same problems as the off-line case. However, in the on-line case, many of the points are even more demanding. The data rate on-line is in real time, i.e. one must make a decision on an event, on average, before the next event arrives. For large detectors, such as CDF, the data rate is so great that it is impractical to put it all on tape. This high data rate may force one to use highly specialized, and extremely fast processors to reduce the data rate to a manageable level.

During a data taking period, the changes in the programs to make them perform satisfactorily must be made almost immediately. Once the data taking period has started (and one knows it will end on schedule) one has very little time to make necessary changes. The verification of the hardware/software is crucial, since there is no going back on rejected events that were not put on tape.

However, for the on-line environment the lack of a sophisticated I/O structure for large scale peripherals and the lack of multiple user operating system is an advantage. They can both be very efficiently interfaced to the data acquisition hardware, which is very important in maintaining the high data rates.

Conclusion

The best use of special purpose processors is in a situation where the processor and experimental apparatus can be designed in conjunction. For a dedicated application where the analysis problem is fairly stable, the disadvantage of limited programmablity is not very important and the increased processing speed makes the special purpose processor option very attractive. The use of these processors also opens up new physics possibilities for observing rare processes in large backgrounds. The Columbia processor for E-690 mentioned above is one example.

Conversely, in any situation where all the algorithms cannot be defined beforehand, or where large and diverse programs are needed (as in the offline analysis of a collider detector) special purpose processors are of strictly limited ap-

plicability.

Emulators and single board computers may be used in any production running situation which is CPU intensive. Only the split between the part of the program running on the host for purposes of input and output and the part running on the emulators needs to be fairly stable.

VII. PERSONAL WORK STATIONS

One of the central issues in specifying the interactive environment is the use of "Personal Work Stations". Although this term is often used to describe any "screen-plus-keyboard" setup which possesses significant computing power, we will reserve the term "work station" for single user devices which are at least as powerful as the new wave of 32-bit microprocessor-based personal computers. Our "work stations" are therefore at least one step above IBM PC's, for example.

Personal Work Stations of many varieties, levels of sophistication, and spanning a large cost range are currently available:

(a) Specialized graphics stations with sophisticated software and hardware support for 3-dimensional transformations, a large number of colors and very high resolution (at least 1000 x 1000 pixels refreshed at least 30 times per second). A Super-mini "host" is required to drive the station.

(b) Stations whose operating system is entirely compatible with the mainframe to which they are connected. These stations may have little graphics capability compared to the stations described above. The emphasis may be to offload the central CPU from routine tasks associated with program development and the running of small jobs that can be handled by a 32-bit microcomputer. Graphics facilities may be supplemented in this case by a second special-purpose low-cost graphics "station" (which would most likely not satisfy the "Work Station" specification defined in Section III).

(c) Stations which function as a complete computing system without mainframe support. Each station is a complete general-purpose computer in the VAX 11/750 or even 11/780 class in terms of CPU power. The distinguishing characteristic is that each station provides the user with multi-processing capability, and sophisticated high speed graph-

ics approaching the specialized graphics stations. Upper-end work stations (e.g. APOLLO) also provide CPU power and systems facilities rivaling a VAX 11/780, plus interactive facilities for program development which are beyond anything available on terminals attached to a mainframe. Stand-alone stations are nevertheless designed for a single (hard working) user, so large peripherals cannot be attached to every station in a cost effective way. Stations are therefore linked into a Local Area Network where large disks, tape drives, etc. are interfaced to the LAN through special file-server processors using standard bus systems (such as Multibus).

The role of personal work stations has been discussed extensively among a number of committee members. There is no doubt that upper-end Personal Work Stations show substantial promise as valuable tools for the prime program-authors and software system developers. The interactive computing and graphics facilities cannot be matched on terminals attached to mainframes. Of particular interest are the most powerful stations of the "stand-alone" type such as APOLLO's. The APOLLO nodes are connected in a high speed (12 Mbit/sec) network, to which one may attach file server computers which provide an interface to large disks and tapes. Each node can run up to 16 processes simultaneously under a virtual memory operating system. Computing power per node is now approaching a VAX 11/780 with a bit-slice arithmetic plus cache memory "performance enhancement board". The interactive system is supported by a "window manager" which allows the user to view several processes at once, and by friendly editors and symbolic debuggers at least equal in quality to VAX/VMS. Local processing power, plus separate memory and bit-slice support for the display, provide for very rapid response. The multiple windows plus rapid response lead to more efficient working methods for software-development professionals. The degree of advantage for more casual users such as the average physicist involved in physics analysis is more open to question.

Graphics facilities of APOLLO's include a large (to 19") high-resolution (to 1024 x 1024) bit-mapped screen, with up to 8 color planes (256 colors) in the

DN600. Vectors may be drawn at 10^6 pixels per second, and "bit blocks" may be transferred to the screen at 160 Mbits/sec. The high speed allows three dimensional transformations to be performed in real time (though not smoothly for complex objects). Complex two dimensional drawings should be producible much faster than on Tektronix storage scopes, and without the need for mainframe computer support. Further improvements now under development include a "geometry engine" that will provide hardware support for three dimensional transformations, allowing full speed animation. The graphics facilities and local processing power of APOLLO stations make them very attractive for interactive event processing, which includes both event scanning and "online" analysis. Just how much sophistication in the interactive graphics is useful or "necessary" is still an open question. Complex three-dimensional static color event and detector displays have recently been shown to be extremely useful by UA1. The utility of animation, often stressed by work-station vendors, for HEP analysis is not immediately apparent. Past history has shown, however, that creative users often find such uses, and soon after demonstrate that they are "essential".

Software packages already exist for APOLLO's for a broad range of CAD/CAM and CAE (Computer Aided Engineering) applications. Many of these track the evolution of sytems dynamically over time, and present the results in terms of animated pictures. A number of similar applications in the field of accelerator design are easy to imagine, such as particle tracking and the turn-to-turn evolution of the beam envelope in phase space.

In evaluating upper-end stand alone stations such as APOLLO's, it must be pointed out that the graphics resolution and display speed are not quite up to the level offered on top-of-the-line specialized graphics stations (e.g. Megatek) at present, although independence from mainframe or super-mini support is a distinct advantage. For smaller scale applications, cheaper graphics stations (less than $5k vs. $50k or more for a color APOLLO), may fulfill the users' needs.

In spite of all the attractive features of APOLLO's and other stand-alone stations, several members of the committee expressed considerable reservations on their use. The key issue is the need for a uniform working environment. Over-

heads associated with working with two operating systems, transferring programs and data files have already been mentioned.

There are really two very different, almost complementary issues involved. Both are concerned with reducing time wasted in migrating from one device to another. The first is the migration of the user; users need a familiar language and outwardly familiar tools. Exactly how they are implemented and whether they are efficient are not particularly important as long as the majority of what the user types at the keyboard is familiar. Even a certain number of variant dialects is tolerable. This has strong implications for what is meant by "user friendly". Usually this means a system which a novice user can quickly and easily learn to use. This is important in practice because most of us who may use a given system only occasionally act as novices each time we use it. If the user environment was really uniform across all machines which we use ease of learning would be much less important than convenience of use for the regular user.

There is no reason why there should be a unique correspondence between a particular vendor machine or particular operating system and a particular user interface. In fact with modern command language processors emulation of simple commands from a given environment is fairly trivial, and more serious emulations are feasible with manageable effort. Furthermore, vendors seem to be waking up more and more to the fact that emulating each other is likely to gain sales.

The other compatibility problem is migration of the codes, that is, getting codes written on one machine to run and give the correct answers on another. In this case rigorous portability is essential, and is usually well worth coping with different user interfaces. To some extent the problem is that of developing strictly portable FORTRAN or other language, but much more is involved: particularly the question of libraries, linkages, buffer sizes, etc. all of which 'trivial' differences can cause havoc with large codes.

Since this compatibility is much harder to implement between different families of machines, it is probably to be considered the dominant requirement.

The conclusion is that general purpose code development and debugging will

probably have to wait for mainframe vendors to supply their own work stations. There is of course hope for some degree of integration of APOLLO's or similar devices into the major operating systems from APOLLO or small start-up companies. It is likely, however, that basic VAX/VMS and IBM VM/CMS work stations will appear before existing upper-level stations move significantly toward integration into a DEC or IBM operating system.

In order to keep in contact with work station developments, and to evaluate their advantages for heavy interactive graphics and major code development projects, it is recommended that a work station project be established and maintained. APOLLO's, SUN's (under UNIX), specialized graphics stations, and new DEC and IBM stations should be evaluated by a member of the computing staff in conjunction with interested physicists.

If one forgoes stand alone work stations, even for interactive graphics, the cost of special purpose stations of sufficient resolution and quality (color, hardware support for 3D, etc.) should be carefully evaluated. Work stations integrated into the major manufacturers' operating systems are not expected to be sufficiently sophisticated for HEP applications for large collider experiments for some time to come.

It should also be mentioned that the whole field of personal work stations is evolving extremely rapidly. The CPU power of an upper-level work station is expected to reach 2 IBM 370/168 equivalents by 1990. As the processing power of a single station increases, it will become natural for a variety of jobs, even jobs of "medium" length according to current standards, to be run on the stations. The bit-mapped raster graphics screen which is now largely confined to work-stations will soon spread to the field of personal computers: within approximately 2 years the average personal computer (replacing present-day IBM PC's for example) is expected to incorporate a 1000 x 800 bit-mapped screen as standard equipment.

Appendix A

Committee Membership

J. Ballam [SLAC] Chairman	G. Lynch [LBL]
J. Butler [FNAL]	H. Newman [California Institute of Technology]
K. King [Cornell University]	R. Peierls [BNL]
P. Kunz [SLAC]	D. Quarrie [FNAL]
D. Linglin [CERN - Annecy]	M. Shaevitz [Columbia University]

Record of Meetings :

May 31, 1983 FNAL
July 14 and 15, 1983 FNAL
Sept. 12 and 13, 1983 FNAL

Presentations to Committee

1. CDF Computing Plans	D.Quarrie
2. UA1 Computing Experience	D. Linglin
3. LEP - 3 Computing Plans	H. Newman
4. TPC - PEP4 Computing Experience	G. Lynch
5. FNAL Computing Department Review and New Computer Acquisition	A. Brenner
6. FNAL User Experience	S. Pordes
7. FNAL Advanced Computer Project	T. Nash

Documents Submitted

1. Report of the Evaluation Sub-Group of the Ad Hoc Committee.
2. Minutes of the CDF Computing Needs Committee.
3. Fermilab's Advanced Computer R & D Program.
4. Theorists Computing Requirements.

5. Fermilab CYBER 175 Computing Facility.
6. Computing Requirements for Engineering and Accelerator Problems.
7. Report of the LEP Computing Planning Group.
8. The LEP-3 Computing Task.

Appendix B

Fermilab Computing Strategy for the Mid – 80's

The Current State of Affairs

The Fermilab Central Computing Facility currently consists of a complex of three Cyber 175 computers which were installed at the end of 1978. They are currently quite saturated, although normal priority production jobs typically do turn around overnight. Interactive (terminal - CRT) jobs are responsive, but are limited in both memory allocation and in CP time for each user session. The latter is hardly a constraint, but the former is becoming a more and more serious limitation to the user community.

About 90% of the CP time is utilized by the experimental physics community setting up for and analyzing data. The remaining 10% is mostly utilized for accelerator design, structural analysis, magnet design and other engineering functions. A small fraction currently is utilized by the theoretical physics community. A substantial number of non-physics users also utilize the facility and constitute a third or more of the interactive population. Their use of CP time, however, is negligible. These include most of the secretaries at the laboratory and a small number of people doing computer aided design (CAD) work.

All users on the site with direct hard-wired connections to the computing facility, as well as those off-site accessing the facilities via common carrier lines, have access to the computers under management control of the Computing Department as well as a few other computers. This is accomplished with a very modern MICOM port selector "cross-bar" switch which gives good flexibility to the user community. Bandwidths up to 9600 baud are possible. Thus, from any given terminal, access is possible to any of the connected computers, or with a dial-out capability over common carrier lines to any of a limited number of allowed facilities outside the Laboratory.

Currently, there is no direct connection to ARPANET or to any of the value

added networks. We are currently negotiating to become a participant in the BITNET network. ARPANET is accessible when required through a connection to Argonne. The primary reason for the lack of connections here is that every time an analysis has been made as to the apparent need and the cost for the service, we have always come out on the negative side. We have tried very hard to organize our external users around one of the other value added network. However, every time we try that and come up with a slightly more costly arrangement to them than we currently have (with a promise that it will go down), we have not been able to get enough takers to pursue this approach. This is an effort that we are continuing and imagine we will commit during the next fiscal year. We have recently committed and started installing a broad-band multiple T1 (1.54 MHz) system connecting the Hi-rise and the general BO/Industrial Area complex. There will also be installed in that complex a satellite Micom port selector. That will give us broadband communications for a number of different lines between the BO/Industrial complex and the Hi-rise. Furthermore, that connection will be extended during the next year further out to the other three experimental areas. Some of these connections will be used for machine-to-machine communications; in the first instance between the CDF data acquisition and off-line VAX's.

Projections for the immediate future in terms of raw computing power required are as difficult as ever to make, but for the near-term early Tevatron period, the techniques we have used heretofore should continue to work. Thus, we project a need for a factor of two more computing than we have right now before the end of calendar '85. By that time, Tev I will start operating and the projections further into the future are very much more difficult to make. Without doubt the growth slope will start increasing at that point. Thus, by the end of 1986 we will be in a very serious overload condition and will require major additional equipment again.

Our system is nicely backed up for archival purposes with the Automatic Tape Library (ATL). This system is fully operational and is an enormous aid in terms of storing and retrieving files with a minimum of hassle for the operations group and with minimum delay for the user. The ATL in this environment is very effective. It is able to contain at current rates, one year's worth of active disk files.

For interactive terminals, we have uniform across the Laboratory a large number of dumb ADM-3A terminals. A substantial number of these have graphics board extensions. The terminals have been very effective, inexpensive and stable. New competition has come into play in the last year or so with somewhat more powerful terminals. We have slowly moved in that direction as well. We have also been searching for an appropriate terminal on which to allow for output previewing. To date, we have not found an appropriate terminal.

With respect to the question of upgrading our facilities to take advantage of modern interactive graphics techniques, over the years we have developed our own internal global device independent graphics system (DIGS) similar to what most of the laboratories have done, each in its own individual style. For interactive graphics, we mostly use 4010 protocols and we support a number of different terminals using 4010 graphics mode, including Retrographics Corporation enhancements to the dumb ADM-3A terminals that we use. These have found widespread and satisfactory use, at a very good price, across the whole Laboratory. The device independent graphics system also allows us to handle a variety of on-line graphics devices including a number of Calcomp drum plotters and two high quality Benson Varian medium speed electrostatic plotters. The latter, in particular, have been very successful, and a direct connection through a rasterizing device to a Cyber channel has been very effective in supplying high quality graphics output with very good turnaround.

Recently, after a search for more than two years, we have chosen a low cost, reasonably high resolution, high quality color graphic system. A number of these Envision terminals are on order and before the end of the summer will start being allocated to the user community. As part of this process, we are setting up a do-it-yourself transparency and slide making facility for the convenience of the users.

In other dimensions, we have made available simple data base management systems, likely to be convenient for the management of scientific as well as administrative data. We also are in the process of introducing new scientific high-quality printing terminals so that papers and letters can be output directly with an extended character set including Greek and mathematical symbols. Associated with

this, we are also having developed a new dumb terminal which will present both Greek and mathematical symbols directly as typed.

The Near Term Plan

The plan conceptually at the time of the FY79 acquisition was to add "a fourth mainframe" during the FY83/84 time period to enhance the currently-installed system. All the necessary connections were provided at the time of that acquisition to make this possible. Currently, the "fourth mainframe" in the sense described does not exist as a marketplace item. Thus, the enhancement originally planned to extend the life of the existing system in a way which would be responsive to the increased demand of our user community is not possible. Thus, we have broadly opened a competitive procurement to obtain enhancement at the same level, i.e., to at least twice that which we currently have installed to carry us through this early Tevatron period. Since a shortage of computing cycles is viewed as the greatest short term (and indeed longer term) need, the RFP procurement most strongly values this aspect of the system. Bids were received from Amdahl, CDC, Cray and IBM.

CDC's bid for a dual CPU Cyber 875 system with 750K words of memory has been selected. The delivery of a first CPU is scheduled for December, 1983. The remaining equipment and software is scheduled for delivery over the next 24 months. The total installed computing power in the Central Computing Facility at that time will be the equivalent of seven Cyber 175 CPU's. In addition to the CPU's and associated channels and controllers that are required, additional peripheral devices will come with the system. The installed disk capacity will double in size from about 15 gigabytes to a total of 30 gigabytes. Also, the high-speed printing capacity will be doubled. As part of this acquisition, during the next year a major upgrade of the operating system from NOS 1 to NOS 2 will also occur. With this change will come a number of added features giving rise to a more powerful operating environment for the user community.

The Cyber 875 will be connected to the existing and other local systems via a Network Systems' Hyperchannel connection. We also have committed and are

awaiting delivery of two quite large VAX 780 computers which will be installed in a cluster network themselves as well as being separately connected to the Hyperchannel. This cluster will serve as an alternate front end to the system currently installed as well as whatever new system results from the current acquisition. It is our intention to continue this expansion of options for alternate interactive mechanisms using the existing old architecture machines and the computers about to be acquired as the computing workhorse engines, behind these new interactive systems.

The programmed time for all this to happen and start being a useful tool to the user community is early in 1984. Somewhat longer range is an intention to introduce Ethernet-type connections throughout the Hi-rise for general interconnectability via a gateway to the central local area network. The details of this connection are only conceptualized and are not yet detailed enough to proceed. It is only recently that Ethernet has stabilized and become a real product, supported by industry with real hardware and software.

With respect to graphics enhancements, we are watching very closely the status of the U.S. Standards Committee relative to both GKS and Core (the U.S. nominal standard). We fully expect GKS or some minor variant of it to be adopted as the U.S. standard, and as soon as that happens we will put major effort behind changing our local, unique DIGS standard to that one.

With respect to the current explosion of powerful desktop workstations, we are continuously tracking the availability in the marketplace in this arena. Clearly it is exploding with new ideas and new options at a very rapid rate. Currently, our perception is that there is no obviously correct system to which Fermilab should commit. Furthermore, we fully expect to have our hands full making work all the new items we have coming in with the current new acquisition. It is our intention to continue to track the availability and performance of workstations. Probably next year when we have stabilized our new acquisition and integrated it into a total system and made it useful to the user community, we will turn our attention once again to the workstation question. Hopefully new and stabilized workstation equipment will then be available, and we will evaluate seriously the choice of workstation(s) appropriate to the Fermilab needs. As indicated earlier, we will continue to expand

and increase the level of integration of a number of networking systems that we use. We will connect to some x.25 value added network, become a participant in the Bitnet network, install an Ethernet system, enlarge the Hyperchannel network and connect BO/Industrial Area to the Central Laboratory via a T1 broad band connection. The latter will then be expanded to other experimental area locations and also to the cross gallery.

We are also interested in expanding our support for engineering computing during this period. The Director has recently polled the Fermilab community on this question and many people are now trying to formulate their needs in this area.

Longer – Range Expectations

The major problem with any medium term solution now is that none of the normal architectural enhancements, neither IBM-like architectures nor any of the other extant architectures, is likely to bring with it at a cost that is even imaginably tractable, the necessary computing which we will require, once the Tevatron starts functioning and producing data at the rates expected. Even assuming that we do everything we can to preprocess in real time as much of the data as possible, we will still need a factor over our current computing level exceeding 10 and possibly approaching 100 to handle the expected data rate as we approach the end of the decade. We will start needing more than the factor of two that we are currently installing as early as the end of 1985. No announced product is likely to be able to provide at reasonable cost the level of computing that we project as our need. On the other hand, there are a number of relevant activities currently in progress, with lots of interest on the part of government, laboratories, and universities starting over two years ago. The message is being heard, the potential Japanese commercial encroachment also is having useful effect in inducing the various companies in this marketplace to come up with new products, some of which may be appropriate for the high energy physics community.

In our long-range ADP plan, we have set aside an amount of $12M in FY83 dollars to accommodate to bringing in major new architectural equipment in FY87.

To do that will mean moving out much, if not all, of the currently installed computing equipment. The question of overlap and availability of equipment and the physical space questions are very serious and complicated issues not yet addressed. Furthermore, the $ 12M budget item should be and is reevaluated every year as we learn more as to what might be available and what it might cost. Numbers approaching up to $ 20M might be necessary to meet our needs by that time, if that turns out to be the right quantum cost to buy the appropriate machine. Of course, in the best of all possible worlds one has a buy-in at $ 12M with in situ expansion as we require it over the years with "just more money." Also, the timing of such a major new acquisition can be adjusted, e.g., moved forward, as we better learn what really are our needs.

As indicated above, there is no collection of machines today which carries with it in a natural way the capability of computing at levels which we require. All indications are that the vector machines, even the dual headed vector machines, i.e., the Cray-XMP, will not solve our problems. Neither do any of the IBM or IBM look-alikes, nor do any of the other available machines that function today. On the other hand, all of the important manufacturers, e.g., Amdahl, CDC, Cray, Denelcor, IBM and Trilogy as well as all three of the major Japanese companies, i.e., Fijitsu, Hitachi and NEC, are committed to big new machines in the relevant time frame. In many cases, however these are vector machines and what performance we would achieve with our job mix is totally unknown to us at this point. Nevertheless, we have been making known, and our benchmarks have demonstrated, just this issue. I believe we're being heard and I believe industry understands that vector machines give rise to the necessary power for only an important subset of the jobs for which they are trying to design their large machines.

In the meantime, we are trying to encourage all of the companies, especially the new ones like Denelcor, whose architecture appears to be better suited to our needs than any other machine, to pursue their efforts with new technology within that architectural construct. We would like to start using a HEP 1 on an experimental basis, either situated at Fermilab, at Denelcor's plant in Colorado or at Los Alamos. It would be very valuable experience if we can get the funds to learn how to use such

a machine and estimate what its capability might be for the future. We could also perform the same studies with other machines as they become available to us. If we are successful in such an approach, for example, by the time the HEP 2 becomes available (currently claimed to be 1985), we might very well be positioned for an appropriate major acquisition.

It also must be understood, however, that if we take the Denelcor route, the chances of there being a sophisticated operating system with all sorts of productivity enhancing tools on it, is very small. The only way in which such a machine will end up being both a useful computing engine and a useful productivity tool for the user is by appropriately networking it with other machines. We are in the process of learning how to network various machines as is the rest of the world just learning. We are sensitive to the availability of all the tools and we believe that we will put together an appropriate mixture of productivity tools connected to computing engines to handle the Fermilab computing needs as they grow.

Appendix C.

Survey Results

The following forms were sent to all spokesmen (or their representatives) for all approved experiments on TEV I and II.

Neutrino Area, Meson Area, Collider

Correspondent:
No. of Physicists:
Average Event Size (kb):
Events/sec:
Days/Year Running:
Running Periods:
Fraction of Data Reconstructed:
CPU Time/Event (Cyber):
Fraction of workload at FNAL
Calibration and Monte-Carlo:
Total forseen useage of FNAL:
General Comments

Proton Area

Experiment:
Respondent:
Data characteristics
- Average event length
- Event rate:
- Running/year:
- Total events:

Analysis states:

Fraction of data reconstructed:

Sociology:

Rapid turnaround:

Monte Carlo analysis:

Calibration:

Database needs:

Local network needs:

General comments:

Detector and data characterisitcs:

Tracking system:

Calorimetry:

Special detectors:

Average Multiplicity:

The data were assumed to be collected over the running periods as follows:

Period I	Oct 83 - May 84
Period II	Oct 84 - May 85
CDF Test	May 85
Period III	Nov 85 - Feb 86
CDF Production	Spring 86
Period IV	Etc.

The results of the responses to the questionnaire are:

Neutrino Area

E649

No. of Physicists:	16
Detector Characteristics:	5000 PWC 500 ADC Flash Chamber
Average Event Size (kb):	100-200
Events/sec:	?
Days/Year Running:	100
Running Periods:	II-III
Fraction of Data Reconstructed:	50-100 %
CPU Time/Event (Cyber):	0.5-1. Osec
Fraction of workload at FNAL:	100%
Calibration and Monte-Carlo:	large
Total forseen useage of FNAL:	0.15 Cyber/year

General:

Semi-visual Detector. Need good interactive graphics on main computer for development of Algorithms and event scans.

E653

No. of Physicists:	15
Detector Characteristics:	2000 DC 4000 Liq.Arg. 3000 ADC (Cal). 10000 ADC
Average Event Size (kb):	20
Events/sec:	5-15
Days/Year Running:	100
Running Periods:	I-II
Fraction of Data Reconstructed:	100 %
CPU Time/Event (Cyber):	1.1 sec
Fraction of workload at FNAL:	20%
Calibration and Monte-Carlo:	small
Total forseen useage of FNAL:	0.25 Cyber/year

General:

Cyber mostly used for reconstruction. Institution computers used for DST analysis and calibration.

E652

No. of Physicists:	20
Detector Characteristics:	8000 DC 800 ADC
Average Event Size (kb):	25
Events/sec:	0.5-3.0
Days/Year Running:	100
Running Periods:	II-III
Fraction of Data Reconstructed:	50-90%
CPU Time/Event (Cyber):	10 sec
Fraction of workload at FNAL:	50%
Calibration and Monte-Carlo:	large
Total forseen useage of FNAL:	0.5 Cyber/year

General:

Comment on networking between exp. VAX, Nevis VAX and central Comp. would be very useful

E640

No. of Physicists:	9
Detector Characteristics:	17000 PWC 2000 DC 150 ADC
Average Event Size (kb):	1-2
Events/sec:	5-7
Days/Year Running:	100
Running Periods:	III+
Fraction of Data Reconstructed:	80 %
CPU Time/Event (Cyber):	1-2 sec
Fraction of workload at FNAL:	50%
Calibration and Monte-Carlo:	large
Total forseen useage of FNAL:	0.3 Cyber/year

General:

Previously doing 50 % on Cyber at LBL but that is going away. Strongly recommend a VAX solution (distributed processing)

E646

No. of Physicists:	15
Detector Characteristics:	15ft Bub. Ch.
Average Event Size (kb):	N/A
Events/sec:	N/A
Days/Year Running:	100
Running Periods:	III
Fraction of Data Reconstructed:	N/A
CPU Time/Event (Cyber):	0
Fraction of workload at FNAL:	0
Calibration and Monte-Carlo:	0
Total forseen useage of FNAL:	0

General:

Scanning and analysis to be done at Nevis (VAX 780)

E636

No. of Physicists:	?
Detector Characteristics	BC Holog.
Average Event Size (kb):	N/A
Events/sec:	N/A
Days/Year Running:	100
Running Periods:	III
Fraction of Data Reconstructed:	
CPU Time/Event (Cyber):	?
Fraction of workload at FNAL:	0
Calibration and Monte-Carlo:	small
Total forseen useage of FNAL:	0.05 Cyber/year

General:

All analysis done outside FNAL on VAXs.
Cyber only for quick turnaround (300 hrs/year)
Helpful if link from exp. Comp to Cyber.

E690

No. of Physicists:	15
Detector Characteristics:	?
Average Event Size (kb):	?
Events/sec:	10
Days/Year Running:	100
Running Periods:	I-II-III
Fraction of Data Reconstructed:	?
CPU Time/Event (Cyber):	?
Fraction of workload at FNAL:	100%
Calibration and Monte-Carlo:	VAX
Total forseen useage of FNAL:	0.15 Cyber/year

General:

Trigger processor reduces trigger rate.
Most offline done on dedicated VAX.

E665

No. of Physicists:	20-40
Detector Characteristics:	8000 PWC 7600 Proptubes ADC (Ring C) DC Streamer
Average Event Size (kb):	20
Events/sec:	$<<1$
Days/Year Running:	100
Running Periods:	III
Fraction of Data Reconstructed:	70%
CPU Time/Event (Cyber):	1.0sec
Fraction of workload at FNAL:	100%
Calibration and Monte-Carlo:	small
Total forseen useage of FNAL:	0.15 Cyber/year

General:

All computing to be done on Cybers plus
1-2 exp. VAXs. Very unsure of both no.
of events and time/event

E632

No. of Physicists:	?
Detector Characteristics:	?
Average Event Size (kb):	?
Events/sec:	?
Days/Year Running:	?
Running Periods:	?
Fraction of Data Reconstructed:	?
CPU Time/Event (Cyber):	?
Fraction of workload at FNAL:	?
Calibration and Monte-Carlo:	?
Total forseen useage of FNAL:	?

General:

Not interviewed

Meson Area

E557/E672

No. of Physicists:	45
Detector Characteristics:	20k PWC 1500 ADC 700 ADC
Average Event Size (kb):	4
Events/sec:	10
Days/Year Running:	100
Running Periods:	I-II-III
Fraction of Data Reconstructed:	60-100 %
CPU Time/Event (Cyber):	4.5 sec
Fraction of workload at FNAL:	30% (rest on VAXs)
Calibration and Monte-Carlo:	large
Total forseen useage of FNAL:	1300hrs/year (0.2 cyber)

General:

Generally happy with present system. Comment on slow turnaround of tape jobs. Would like extra software help from Comp. Dept.
Most of their reconstruction done on VAXs offsite.
Currently feel that they are people limited rather than computer limited.

E706

No. of Physicists:	?
Detector Characteristics:	12000 ADC 10000 PWC 10000 microstrip 10000 ADC
Average Event Size (kb):	?
Events/sec:	?
Days/Year Running:	?
Running Periods:	II-III
Fraction of Data Reconstructed:	?
CPU Time/Event (Cyber):	$\sim$ 0.5 sec
Fraction of workload at FNAL:	?
Calibration and Monte-Carlo:	?
Total forseen useage of FNAL:	$\sim$ 0.25 Cyber/year (??)

General:

Generally good experience with present system, but would love access to a VAX. Need more software support from Comp. Dept.

E605

No. of Physicists:	?
Detector Characteristics:	6000 DC 4000 ADC 300 Proptubes 150 ADC
Average Event Size (kb):	<1
Events/sec:	400
Days/Year Running:	150
Running Periods:	I-II-III
Fraction of Data Reconstructed:	100%
CPU Time/Event (Cyber):	0.1
Fraction of workload at FNAL:	50%
Calibration and Monte-Carlo:	large
Total forseen useage of FNAL:	0.5 Cyber/year

General:

DST analysis performed at home institutions. Moans about poor graphics facilities, poor interactive facilities, program development environment.

E731

No. of Physicists:	?
Detector Characteristics:	Drift Chambers 804 Pbglass array
Average Event Size (kb):	0.8
Events/sec:	100
Days/Year Running:	100 hrs
Running Periods:	II-III
Fraction of Data Reconstructed:	50-75%
CPU Time/Event (Cyber):	0.25
Fraction of workload at FNAL:	100%
Calibration and Monte-Carlo:	large
Total forseen useage of FNAL:	0.25 Cyber/year

General:

Run from Chicago workstation. Would like faster machine with large memory for Monte-Carlo studies, faster turn-around on short jobs needing little CPU. Generally very happy with system.

E704

No. of Physicists:	?
Detector Characteristics:	?
Average Event Size (kb):	?
Events/sec:	?
Days/Year Running:	?
Running Periods:	I-II
Fraction of Data Reconstructed:	?
CPU Time/Event (Cyber):	?
Fraction of workload at FNAL:	?
Calibration and Monte-Carlo:	large
Total forseen useage of FNAL:	0.25 Cyber/year (guess)

General:

Unable to find anyone able to talk about their computer needs. Guess at computer needs

Collider

CDF

No. of Physicists:	150
Detector Characteristics:	100000 channels (mainly ADC and TDC)
Average Event Size (kb):	50-150
Events/sec:	1-5
Days/Year Running:	100
Running Periods:	II-III+
Fraction of Data Reconstructed:	10%
CPU Time/Event (Cyber):	40 secs
Fraction of workload at FNAL:	>80%
Calibration and Monte-Carlo:	large
Total forseen useage of FNAL:	1984 1.0 Cyber 1985 2.0 Cyber 1986 8.0 Cyber

General:

A document is being prepared (for presentation to the committee) on possible scenarios

DO

No. of Physicists:	?
Detector Characteristics:	?
Average Event Size (kb):	?
Events/sec:	?
Days/Year Running:	?
Running Periods:	III+
Fraction of Data Reconstructed:	?
CPU Time/Event (Cyber):	?
Fraction of workload at FNAL:	?
Calibration and Monte-Carlo:	large
Total forseen useage of FNAL:	1985 1.0 Cyber 1986 2.0 Cyber 1987 8.0 Cyber

General:

The DO proposals are in such a stage of flux that it doesn't seem sensible to give hard numbers. The best estimate of Computer requirements seems to be just to use the CDF numbers. Note that

DO should be somewhat later coming on-stream than that CDF (mid-end 1986).

Proton Area

The summary is as follows:-

Expt	Run	#events	time/ev	fraction	CPU hrs	M/C	Total Cyber
E715	I	6×10^7	0.2	0.6	2k	Small	0.2
E621	-II	6×10^8	0.05	1.0	20k	>20k	1.5
E516	II-III	9×10^7	1.0	1.0	25k	Small	3.0
E400	I	1×10^8	0.25	1.0	10-20k	Small	1.1-2.2
E705	II-III	1×10^8	2.0	0.6	30k	15k	5.0
E687	II-III	1.5×10^7	0.25	1.0	10k	Small	1.1

Note that these requirements are spread over a 2-3 year period.

Appendix D

UA – 1 From a Computer and Software Point of View

D. Linglin

1) Introduction

The UA-1 experiment is installed in the long straight section 5 (LSS5) of the CERN SPS collider. Preparation started early in 1978 (committee approval was in June), the detector was installed in 1981 (although not complete) and by then was able to record its first data.

Three data taking periods occurred so far:

- Oct-Dec 81	sum(L)	= 20 μb-1 (on tape)
- Oct-Dec 82		= 18 nb-1
- Apr-Jun 83		$\simeq$ 120 nb-1

To date, the next $\bar{p}p$ period is not scheduled until the fall of 1984.

As it is well known, UA1 is presently one of the largest particle physics experiments worldwide, with more than 150 physicists from 18 different institutes in UK, France, Germany, Italy, Austria, Finland, the Netherlands, Canada and USA (plus CERN of course).

The detector looks like many of the present all-purpose detectors installed at colliders. It is formed of a set of complementary detectors surrounding completely the interaction region from a fraction of a degree upwards, except for a few cracks here and there. A particle emerging from a collision will possibly traverse a large central detector ("Image Chambers"), EM Calorimeters (64 "Gondolas", and 2 "Bouchons" divided into 32 petals each), Hadron Calorimeters (so-called C's and I's, because of their shape, and located behind the gondolas and bouchons respectively), more shielding iron (to be instrumented before next run) and large muon drift chambers

(typically 4×6 meters, each chamber consists of 2 sets of 4 planes, 50cm apart, with 15×4cm section drift tubes). Forward detectors complement this central coverage. The detector weighs almost 3000 tons and, if one excludes these forward parts, forms a box of about 12-15 meters in any direction.

In contrast to e^+e^- detectors, the UA-1 magnet (7 KG in 80 m^3) is a dipole, to provide good forward momentum analysis.

Installed on a large chariot to shuttle back and forth between the "garage" and the "Xptal area" position, the detector is followed by a Mobile Electronic Chariot ("MEC"), 3-story high, in which is located the electronics that cannot be installed too far away, in the UA-1 control room, a hundred meters of cable length away.

2) Triggering system

With an inelastic cross-section of about 50 millibarns and luminosities in the range 1 to $2x10^{29}$ for the last period, one has to deal with 5000 to 10000 interactions per second. There are 3 bunches of protons and 3 bunches of antiprotons, which yield bunch crossings every 7.5 μsec. At present peak luminosity, there is still only one interaction for every 15 crossings and double events are then negligible.

As it will be seen later, it is not possible for the time being to record more than 3 or 5 events per second. Various triggers have then to reduce the rates by a factor of a few thousands, else dead time will do it for you.

Very schematically, there are 3 levels of triggering (plus level zero), each one receiving what has been accepted by the previous level, as usual.

— Level 0 or Pretrigger. It is based on scintillator hodoscopes and picks up the most of the inelastic cross-section amongst all beam-beam crossings.

  The triggering system is activated by a pulse produced by the p and $\bar{p}$ bunches themselves. Every 7.5 μsec, this pulse starts a Pretrigger: Time of flight measurements use ORs of scintillator hodoscopes covering angles from 1 to 18 degree with respect to the beam axis. Carefully turned timing criteria

ensure that the observed outgoing particles originate within the detector, rejecting nearly all of the non-collision background.

— Level 1 consists of two different types of processors:

(a) 2 fast trigger processors (the "UK box") are fed with information from calorimeters and use weighted sums of energies to decide if the data is to be read out before the next colliding bunches arrive. These processors are programmed to select events exceeding various thresholds, as listed below.

(b) In parallel, a special processor (Muon trigger) selects high momentum tracks from the muon chambers pointing to the vertex, within tolerances, and which penetrate the many interaction lengths before the μ chambers. In this processor, hit patterns of wires in one projection are compared to patterns stored in a digital memory for 3-out-of-4 hits forming straight lines.

— Level 2 or FAMP is for muons only. The FAst Muon Processor is based on 7 Motorola 68000 (one per detector module + one coordinator), it performs a more sophisticated programmable selection using drift times and combining the two projections. Although this level was ready, some missing parts made it idle over the last run. The event Read-Out starts only after level 2.

— Level 3, in the counting room, is made of 4 emulators 168E (Actually, there are 5 emulators, but one is reserved for event displays). Each one has about one Mbyte of memory and runs Fortran programs, previously tested on IBM machines, at a speed roughly equivalent to one half of an IBM-168. At this level are used calibration constants and pedestal substraction, hence results are almost identical to the full off-line programs and yield sharp Et cuts without trigger acceptance problems around the threshold values.

The main triggers in use are:

- "Electron"	Et threshold on 2 adjacent cells of central EM calorimeters
- "Jet"	Et threshold on 8 gondolas and 2 C's behind or one quadrant (8 petals and I's behind)
- "Temperature"	Threshold over ΣEt (all barrel calorimeters)
- "Diffractive"	$\geq$ E1 on one forward side, $\leq$ E2 on the other (asymmetry of deposited energy in forward calorimeters).
- "Electron-Jet"	Coincidence of 2 above "Electron" and "Jet" triggers with lower thresholds
- "Single Muon"	
- "Double Muon"	
- "Cosmic"	Cosmic muons in the event Pretrigger window, triggered by special up-down counters
- "Minimum bias"	Only the Pretrigger
- "Elastic"	

All these triggers (except the last two) are run in parallel in the trigger processors. Typically, the single muon trigger accounts for 50% of all triggers (until level 2 is operational).

The last two triggers are used in high beta runs (low luminosity). In these crude calculations:

$$\mathrm{Et} = \mathrm{E}_{meas}\ \mathrm{Sin}(\theta)$$

3) Data acquisition and rates

Information coming from the calorimeters, muon chambers and forward detectors is either compressed by the digital electronics of these detectors or processed through simple algorithm processors, like the LRS 2280 ADC. These detectors, if one excludes forward Image Chambers, generate of the order of 20 kbytes per event in about 10 to 20 msec.

The information originating from the Central Detector is more complex:

For each of the 6000 wires or so, the analog information at both ends is first digitized every 32 nsec by 2 6-bit flash ADCs that measure a non-linear function of dE/dx and the position along the wire through current division. Drift time in each 32 nsec window is measured by a 3-bit TDC with a 4 nsec accuracy, one additional bit giving the t_o. All this information is stored in parallel over the 3.6 μsec of the maximum drift time in the chambers (16 cm maximum drift space) for a total of about 2.5 Mbytes per event. This information is grouped in sets of 60 wires and processed in a Read-Out processor ("ROP"). Each one of the 110 ROPs is made of 2 μP (one Signetics 8x300, 8 bits and 250 nsec cycling time, and one Motorola 68000) and reduces the data from typically 2.5 Mbytes to 50-100 kbytes (zero skipping, reduction of track signals to mean value and width) in about 20 to 40 msec. After the MEC and its 200 CAMAC crates, the subsequent data collection uses the CERN ROMULUS/REMUS standard, based on branch drivers and crates controllers, linking all crates in a tree-like structure. The data streams are channeled through five single branch drivers (3 for the central detector, one for the forward chambers and one for the remaining information), which transmit in parallel the entire event to the Main Control Room (MCR), 100 meters of cable length away.

Computer-controlled router units in the data path can switch the data between two separate vertical highways, permitting additional data acqusition (eg. 168Es) and monitoring subsystems to operate in parallel. Several Super CAVIARs μ-computers (see next Section) thus can run monitoring tasks and display programs for the various detector parts. Most of these displays are located in the UA-1 MCR, others are positioned close to the fast electronics in the MEC.

The transfer time is about 2.5 μsec per 16-bit word on each branch, hence about 40 msec are needed to get the whole event (time fixed by the largest branch), permitting a theoretical maximum of 25 Hz to be transmitted to the emulators or to the acquisition computer.

The third triggering level, composed of the four 168Es, is installed at this level. For the last run, the average time between events in these processors was about 55 msec. To get a reasonable dead time, it is then hard to go above 3 or 5 Hz (with about 1300 events maximum per 6250 bpi tape, a tape is filled every 10 minutes or so with a 2 Hz rate). Each emulator (2 VAX equivalent each) runs a Fortran program, a stripped-down version of the off-line with a reduced HYDRA system, and depending on the thresholds, provides an additional factor 10 or so reduction. The 168E algorithm was a partial calorimeter reconstruction to get an accurate determination of the Et value in the case of the "electron" trigger. In short, it was mainly for the $W \rightarrow e\upsilon$ and $Z^o \rightarrow e^+e^-$ search, but rare triggers, as the double muon, were added directly. However, as a by-product, one selected as well high Et jets (eg. with 10 or 15 GeV localized transverse EM energy) and therefore picked up most of the Et $>$ 30 GeV jets. About 50 to 200 ms (168E-time) were needed to make the decision for the events flagged as "electron" by the first level trigger. For a simple transverse track finding in the Central Detector (unused so far), one would need about 20 ms.

As it was the first run with the emulators, every event was written to tape (Normal "ZO" tapes), while emulator-selected events were also written on a special "ZF" tape. For the last run have been recorded almost 3000 ZO tapes of all kinds of triggers, and slightly over 300 ZF tapes.

Average event size depends of course on the trigger. For example:

- 115 Kbytes for Electron or Jet trigger.
- 160 Kbytes for ΣEt temperature trigger.
- 85 Kbytes for muon trigger, slightly less for Minimum Bias.

Accordingly, the maximum rate the system can handle with reasonable dead time, depends on the trigger mixture and on other parameters like the percentage of events that go on special tape. For the above triggers the rates were:

- 3.5 Hz for Electron or Jet trigger.
- 2.5 Hz for ΣEt temperature trigger.
- 4.6 Hz for Muon trigger, etc...

Before the next $\bar{p}$ period starts, a full double buffering should be operational, dead time will be reduced.

4) On-line computers

After this brief description of the data acquisition system, here is a list of the various computers and microprocessors of the experiment:

(a) The general support of UA-1 is based on 2 Norsk Data ND 100/500, with 2 Mbytes of central memory each. Each system is in fact dual, with the 16-bit ND100 (multi-user, multi-task) front-end of the 32-bit ND500 (multi-user). Overall capacity is about 3-4 VAXs (As a guideline, one can assume hereafter the following rough computing speed ratios: one Cyber 175 = two IBM 168 = 4 NORD 500 = 4 168E emulators = 8 VAXs).

(b) 3 minicomputers Hewlett Packard HP21MX are in charge of 3 specific tasks, to relieve the load on the Nord systems:

— Data Acquisition of the luminosity detectors.

— "Histogramming machine" to follow the behaviour of several thousands of parameters (eg. PM signals).

— Control of the 2 video-tapes units (unused at the moment).

(c) 168E emulators, as described above. With the ND500s, they provide althogether the computing capacity of 12-15 VAXs.

(d) (Super) CAVIARs. More than 10 units of this system are used in the experiment, for control and development. For instance, there is one to

control the PM high voltage system, two to control the central detector, one to control the emulators, etc... The CAVIAR is a DACQ-oriented microprocessor, developed at CERN. Based on the Motorola 6800 and a floating point processor, it has 256 Kbyte RAM, 85 K EPROM, a clock and a fast CAMAC interface, plus an ADLC ("Advanced Data Link Controller"). It is programmable in Bambi, a language similar to BASIC and is available with a large I/O and histogramming library.

(e) Various types of microprocessors:

— The ROPs, ie. the 2 x 110 μP used for CD data reduction.

— The 7 M68000 of the second level muon trigger.

— The GPMCs ("General Purpose Microprocessor Controllers"), and The MUMMs, based on the M6800, and used for various detector controls.

— Other μPs, M6800 or equivalent, are also used here and there, like in "intelligent" terminals the MVIT type. There is also a μP to drive a voice synthetiser, etc...

(f) Graphics: three kilometers away from the experiment, but linked to the Nords through a special fast CERNET link, a 3-D interactive graphics system Megatek, controlled by a VAX 11/780, complements the list above with the possibility of on-line complex analysis and scanning (Fig. 1).

Such Megatek systems, driven by IBM 4341, PRIME or VAXs, now exist in half a dozen institutions within the collaboration (CERN, Birmingham and Rutherford in UK, Saclay and Annecy in France, UC Riverside in the US).

(g) NETWORKS: Of course, the hundreds of systems listed above are linked together in a LAN, mainly based on the CAVIAR ADLC (with protocol based on HDLC). Very schematically, The NORDs are at the top of the structure (with their own internal network NORnet), The CAVIARs are at the intermediate level, and the ROPs, GPMCs, ... are at the lower

level. 168Es are at the "same level" as the CAVIARs; they are connected to the Nords via CAMAC on the one hand, and have access to the 5 Remus branches on the other hand via a CERN special interface.

At the top of the structure,NORDs and VAX are of course connected to the CERNET network, and via the terminal system INDEX, one can access these machines (and from there, μPs) from anywhere at CERN and from most of the outside labs.

Although it starts to be now common place in HEP experiments, every operation in UA-1 can be controlled by tasks activated from terminals. There is in principle no switch to play with, only terminals.

5) Off-line Saga

When UA-1 software activities started, in 1978, there were only half-a-dozen people involved, then it increased with time, to reach more than 30 in 1981. With the analysis of the data, there are now more than 100 people active on the CERN computer centre (and more than 160 registered today).

At the beginning, the IBM 168 and 3032, with Wylbur, were adequate for all the development work (memory size, Wylbur friendliness, etc..). Later, the replacement of these machines by the IBM 3081 (1981) then the Siemens 7880 (1982) was made without any change to the user, except larger CP capacities.

On Fig. 2 is displayed the monthly CP consumption of UA-1 on the CERN IBMs. It must be stressed that the low level of the load until 1982 was partly due to a tight schedule. No extensive simulation was performed until the first data showed up. In addition, there was no noticeable "spike" over that period. However, a small Monte-Carlo production was run on several computers of the Collaboration, to test the chain of reconstruction programs with simulated events. This was done in the first half of 1981, a few months before recording the first real data.

With the tight schedule to build UA-1 (3 years), the priority given by CERN to the whole $\bar{p}$ project and the limited network connections with outside institutions, it was

decided from the early beginning that most of the development work would be done at CERN. One could have developed the software outside more than what was done, without much problem. But, as most of the physicists were physically at CERN anyway for hardware construction, beam tests, installation, on-line developments, etc... the result was that only a small fraction of the code (a few "processors") was written outside. In this case no particular problem emerged.

The basic decisions and rules, agreed also at the very beginning, were to use:

— FORTRAN IV. Fortran 77 was not available on CERN IBMs until 1982, but was available however on many other machines like the VAX, NORDs, UNIVAC, used by the collaboration. Non standard statements were forbidden, except a few well defined cases, like format delimiters, to ensure compatibility between the various computers accessible to the collaboration (IBMs, Cyber, Univac, VAX, NORD, 168E). A few further restrictions were imposed in view of the emulators.

— PATCHY for file handling, and code transfer (CETA format) from computer to computer. Actually, only a small subset of PATCHY was used, as modern editors make part of it rather out-of-date.

— HYDRA for memory manager, I/O packages (computer independent file formats), debugging, free format titles, recording facilities etc. ... Here again, all the features offered by Hydra were not used. Also, the author in charge was asked to develop for UA-1 a new processor package (JQ81).

— HBOOK as standard histogramming package.

— SCRIPT as text processing package (documentation, publications). Later, secretarial work was transferred to dedicated word processors.

As much as possible, Wylbur command files ("Exec files") were prepared to provide friendly basic facilities for file handling, networking, UA-1 documentation, job submission (mainly production), public mail and all kind of distributed information. This was very important, to spread a minimal information without meeting all the time. In addition, it allowed a coordinator to spy partially on the software activities of others and provide some more cohesion and coordination to the system.

In short, UA-1 has made quite a good use of what was available. However, with presently over 160 registered users in the group (of which 100 submit jobs every week), 500 jobs submitted on the average per day (with peaks over 1000), it becomes hard to manage and sometimes brings the CERN Computer Centre to its limits. In particular, more disk and MSS space than presently allocated (500 Mbytes and 4 Gbytes, respectively) would be welcome. This should be done in the near future.

UA-1 PUBLIC CODE:

There are presently about 200,000 lines of public code, to hold mainly Fortran programs and detector basic constants. They are divided into more than 50 PATCHY PAM files (6 or 8 at the beginning, that were split several times), each one existing at any time under one or several different cycles. The subdivision is made according to several criteria:

— the BCD version of a PAM file must not be larger than the maximum allowed by the editors (typically 15,000 lines for CERN Wylbur).

— Not too many people can work on developing a given PAM, typically not more than 5.

— Each PAM must only contain code or information related to a dedicated task, eg. CD reconstruction, calorimeter simulation, muon chamber description titles, etc... Each PAM is under the responsibility of one person (updates). A few conventions are used to ease the work of everybody. Compiled (binary) libraries are made for every cycle of most of the PAM files. Regular clean-ups are performed on old useless cycles.

MAIN UA-1 PROGRAMS:

Basically, one can divide the production programs into few categories:

— SIMULATION. It consists of a chain of 3 to 5 different programs, the full simulation of one Monte-Carlo event takes about 10-20 sec. CP (IBM-168 units).

- PREPROCESSING. A single program to read raw data tapes, apply calibration constants, convert to HYDRA format (Banks) and possibly filter the events through partial reconstruction. 1 to 1.5 sec./event.

— RECONSTRUCTION. The so-called BINGO program performs partial or full reconstruction of the event, after preprocessing. One needs typically 15 sec. CP / event for a high Pt trigger (half of it for Minimum Bias events). This value is mainly due to Central Detector reconstruction, where time is roughly shared half and half between track finding and track fitting. The high number of points per track and the constant field are the two key reasons for limiting the CP time. Remember, an average high pt event yields 60-70 charged tracks in the CD, of which about 2/3 are connected to the primary vertex.

— DST and ANALYSIS programs allow data reduction and basic analysis facilities.

— GRAPHICS. These programs come in addition to the above and are mainly VAX/Megatek oriented (scanning, debugging, ...).

The memory size needed to run any of these programs is typically in the range 1 to 2 Mbytes. About 4 Mbytes are needed for interactive graphics. The memory manager of HYDRA has helped to keep the above numbers to rather low values. These values scale more or less with the size of the events.

TAPES:

The magnetic tapes in use at any level are recorded at 6250bpi density (about 160 Mbytes/tape).

Typical numbers of events per tape are the following:

- Raw Data	1,300	(any 120 Kbytes/event)
- After Preprocessing	1,300	
- After Reconstruction	450	(350 Kbytes/event, including the 120 of preprocessed data)
- DST's	5,000	
- Mini-DST's	30,000	

OFF-LINE VERSUS LAST RUN

During the last $\bar{p}$ period (end of April - beginning of July, 1983), the standard daily scenario was more or less the following: Every late afternoon, SPS engineers were delivering a dense shot of $\bar{p}$'s, after a few adjustments with pilots. This shot lasted overnight and dumped sometimes in the morning, depending on beam lifetime (and hence decreasing luminosity). Typically, luminosity was above 10^{29} at the beginning and was down by a factor 2 to 3 by the time beams were dumped.

Over the 15-18 hours of the shot, the related off-line activities, carried out by a round-the-clock set of about 15 people ("Xpress-line" team), were mainly the following:

1. 168E checks: emulator results were checked against the same program / algorithm run on the NORDs or/and the IBMs. A few minutes CP per job, run on the first hundreds events of every 10 normal tapes.

2. Calibration of calorimeters: programs were run on the NORDs, using special calibration tapes written at the very beginning of the shot, with magnet ON, in the few minutes where central detector was getting ready.

   Calibration file, valid for the shot, was transferred via CERNET on the MSS and installed in the "UA-1 constant base". Only the last version and the 168E version were kept on the NORD (168E constants were not updated for every shot, only if changes were above some thresholds).

3. Calibration of Central Detector: various jobs run on the first 15 normal tapes of the shot delivered, after 4 hours CP (IBM 168 eq.), a calibration file, valid for the shot, and installed as well as a separate element of the constant base.

Altogether, the various calibration files generated for a shot had a total size of a few hundreds Kbytes, that is, of the order of 50 Mbytes for the whole 1983 running period. The size of this constant base is not a major problem.

Only when the above calibration procedure was performed could one proceed with the "Xpress-line" production. Only special ZF tapes were processed at this level (3 to 10 tapes per shot).

4. Preprocessing: calibration constants are applied. This takes slightly over on second per event, hence 1 to 3 hours CP.

5. Reconstruction: This takes about 15 sec. per event. After an initial period where all special events were reconstructed, a higher Et threshold was applied on input to process only a third of the 168E selected events. Then the cuts were directly applied at 168E level, as one gained confidence and luminosity. All this was in fact done to keep CP consumption at affordable levels, typically a few (3 to 5) thousands events, or about 20 hours maximum per shot and $\simeq$ 3% of all recorded triggers.

6. Megatek selection: Higher cuts and CD track matching to select electron and double muon candidates were performed on reconstructed tapes. Cuts were set up such as to yield not too many events for the round-the-clock team of scanning physicists.

OTHER PRODUCTIONS:

Besides "Xpress-line" activities and analysis (non-production activities over the run were at the level of about 20 hours CP / day), other productions were going on with a slightly lower priority:

— Muon filtering on normal tapes, and reconstruction of a subset of the filtered events, to look for $W \rightarrow \mu\upsilon$ and muons inside jets (one way to look for Top).

— Preprocessing of the first few nb^{-1}, to be later reconstructed outside.

— Partial reprocessing of some 1982 data (several different productions).

All these productions were scheduled, to not overload the centre and avoid competition with Xpress-line. At some stage, the number of tape drives was becoming a

problem (despite a dedicated disk UA1 disk of 300 Mbytes for production).

OVERALL CONSUMPTION:

Altogether, summing up the above numbers, 15-30 hours per shot were used for Xpress-line, 0-20 hours/day for other productions, and a background of 15-20 hours/day for other activities, hence a total of 30 hours/day on the average, with peaks at 60 hours, over the past 2 months. The UA-2 experiment was concurrently using comparable amounts of CP time, namely one third of the IBMs, on the average.

As a reminder, CERN IBMs consist of an IBM 3081-K bi-processor and a Siemens 7880 ("IBM compatible"). Each of these three processors is more or less equivalent to 2.3 IBM-168 units (ie. about 7 units in total). In addition to the load originating from Wylbur and CERNET (file transfer activities : MSS, laser printer ...), the system can yield up to 100 hours equivalent per day for batch processing. This corresponds roughly to 420 hours per month per IBM-168 unit in batch mode, or 5000 hours per year.

In addition to its IBMs, the CERN computer centre is running a CDC 7600 with two Cyber front-ends. This system will be replaced in a few months by a dual Cyber 875. This adds presently an extra 3 to 4 units, to become soon about 6 units.

FORECAST UNTIL 1984 AND OUTSIDE COMPUTING

A reasonable estimate yields $\simeq$ 1/4 to 1/3 of the data to be processed for this run. This corresponds very roughly to 4-6000 hours (IBM-168 units), plus the analysis. On the average, data processing represents between one third and one half of the total computing requirement. One then needs between 10000 and 15000 hours to analyse the past run. This time being required over a year or so, this corresponds to 2 or 3 IBM-168 units equivalent, shared amongst several computer centres.

If needed, a factor two increase should be possible, given the capacity of present computer centres available in the Collaboration. But a factor five would be much more difficult, without using dedicated processors (Emulators ..).

Because of the difficulties to export the constant base and large bulks of tapes,

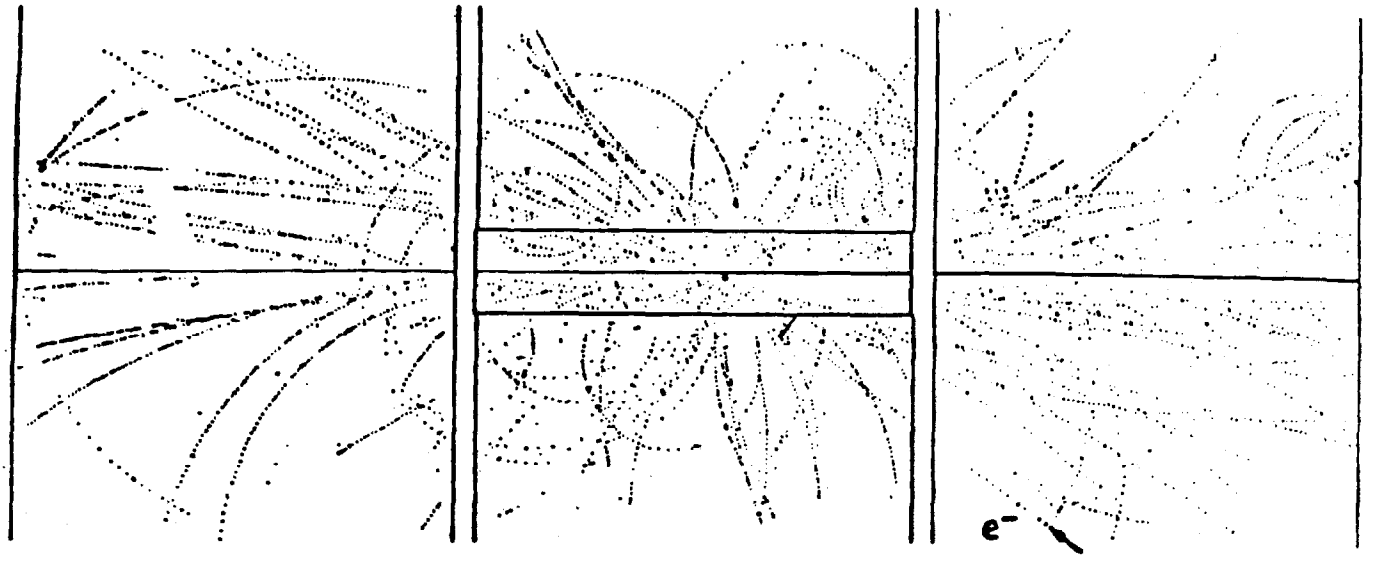
preprocessing (+ filter) of the 3000 normal tapes will be done at CERN, while most of the reconstruction will be performed outside, mainly at Rutherford and Saclay (IBM 3081). A CRAY-1 will also be used at the Ecole Polytechnique, just outside Paris (200 hours allocated), where one will try to vectorize partially the critical loops of the code. There is also the possibility that some time will be made available on the CDC's CYBER 205 near Paris. In addition Saclay is now prepared to run part of the production on two 168E emulators, driven by PRIME (programs downloaded from their IBM 3081). Each emulator has 256 K words of data memory and 128 K words of program memory (1.5 Mbyte total).

Until 1982, almost all the computing has been done at CERN. Only this year one started to export part of the increasing load outside. For the first eight months of 1983, while CERN provided 5-6000 hours, 500 hours were spent in Saclay (out of 1200 hours allocated this year) and more than 2000 hours in Rutherford (IBM-168 units as usual). The ratio of outside / CERN computing should still increase in the months ahead.

6) Conclusion

This short paper is an attempt to summarize the key features of UA-1 that can be of interest to any similar experiment, like CDF, with respect to computer related needs. Although the computer world is moving very rapidly and situations are not the same, one should emphasize a few points not to be forgotten: they are called NETWORKS, GRAPHICS, centralized Mass storage, large memory (including VM) and CP capacities, ... and last but not least, a good set of well organized, motivated people.

EVENT 2958. 1279.



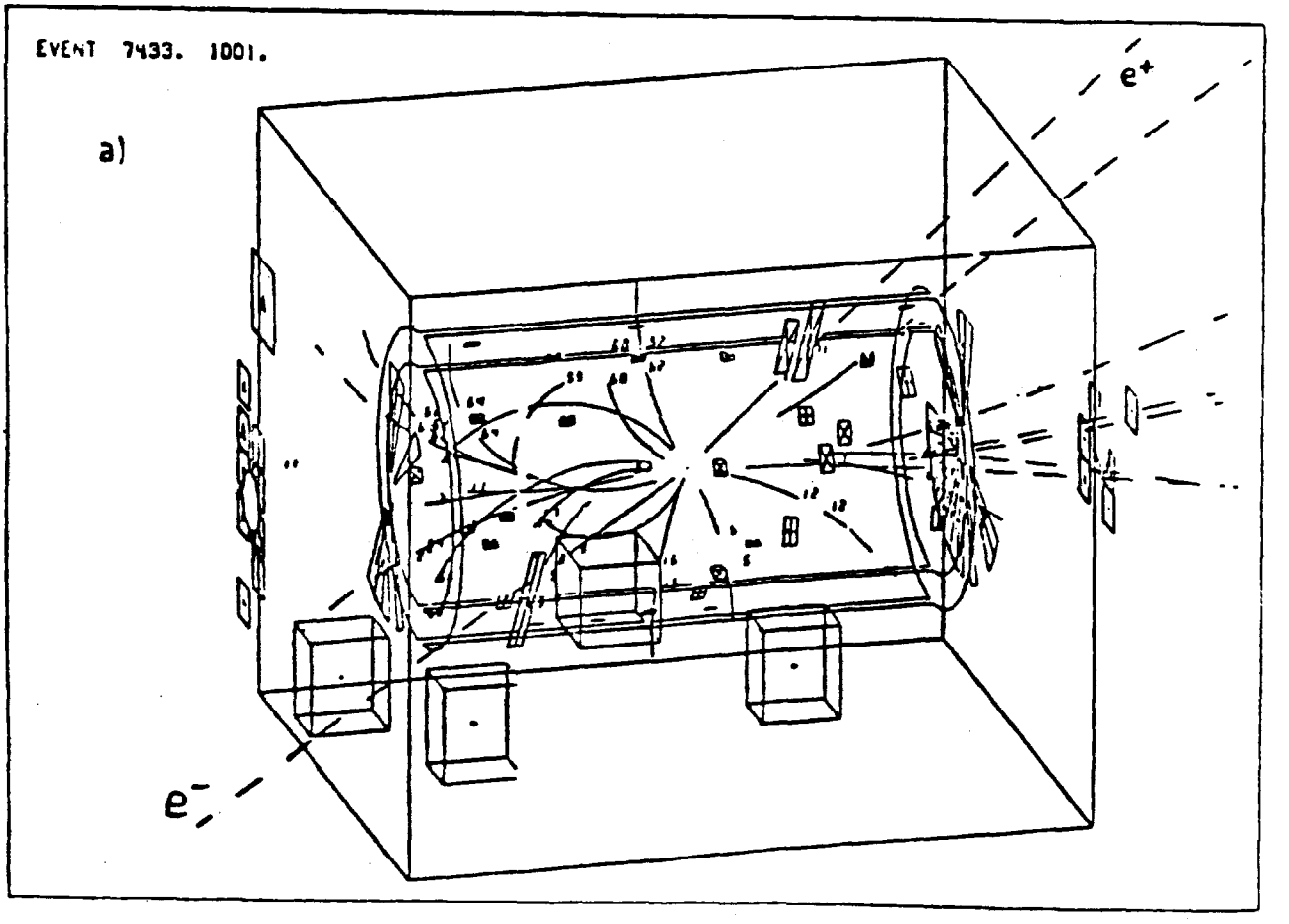
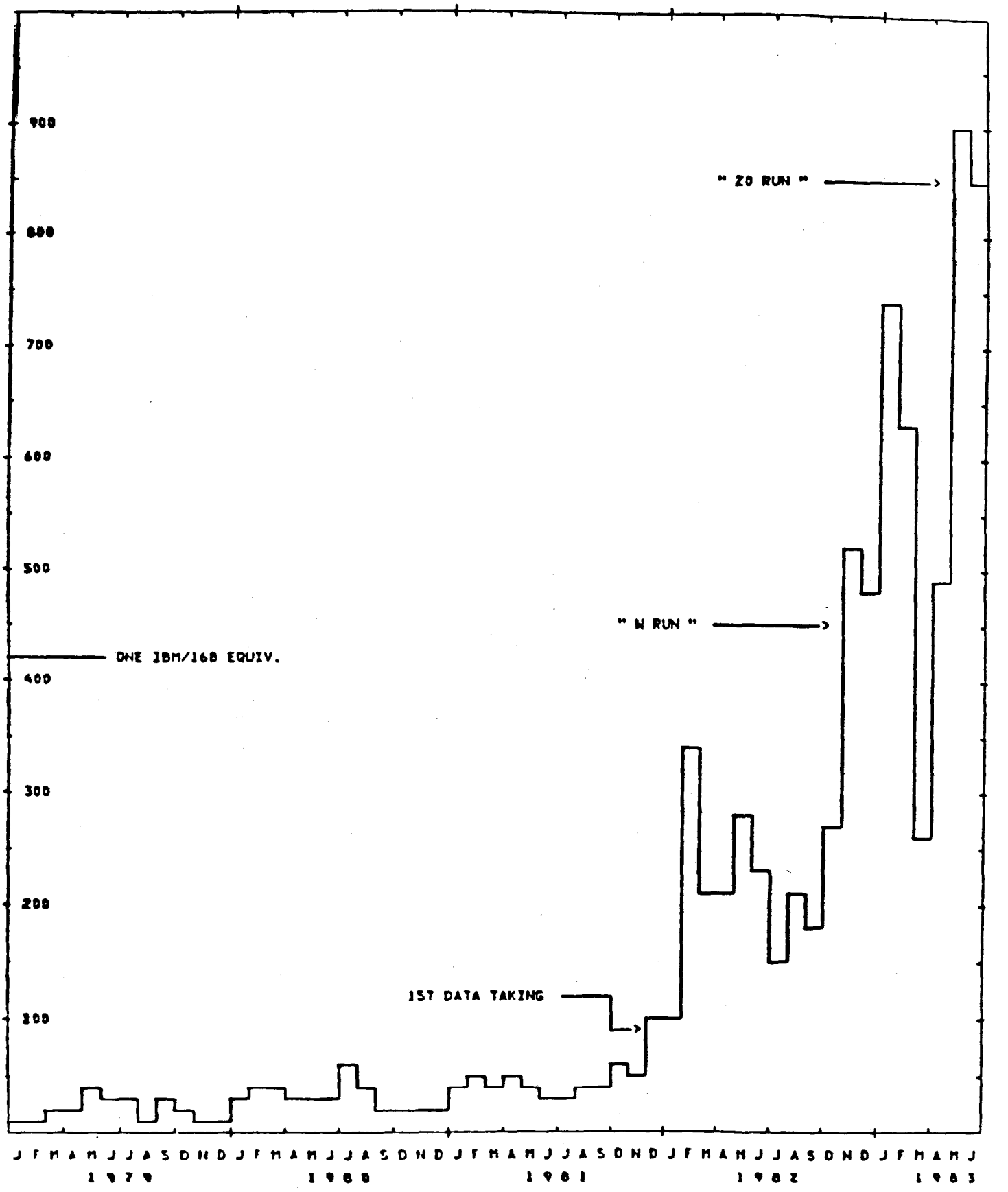


Fig. 1

CERN COMPUTER CENTRE (IBM) : UA-1 USAGE
(CP HOURS PER MONTH - IBM 168 EQUIV.)



1 CRLIP 175 = 2 IBM 168 = 8 VAX (ABOUT)
RE/FS AT CDM / DAY IN THE PAST 2 MONTHS .

FIG. 2

Appendix E
CDF Computer Requirements

1.0 Introduction

This document summarises the results from several discussion sessions by the members of Collider Detector Facility, and in particular a two day workshop held at Fermilab on 4th-5th August 1983, that were held in order to attempt to quantify the Offline Computer requirements. In order better to understand the scale of the problem an event reconstruction scenario was created as the basis for further discussion and study. In addition the relative amounts of CPU power necessary to accommodate the demands made by Monte Carlo studies, program development and physics analysis from DSTs) has been estimated.

2.0 Time Scales

The schedule for CDF is for a preliminary physics run to occur in May 1985, followed by the commencement of regular running periods in January 1986. Not all of the detector will be installed for the May 1985 run, and the main focus will be on understanding the apparatus (both the Collider and CDF) rather than obtaining interesting physics results. However, this short run will provide an excellent opportunity to check out both the online and offline systems and hopefully will also enable some physics studies to be performed. In leading up to this 1985 run, CDF will be mainly performing program development and Monte Carlo studies, while thereafter, data reduction will play a large part in the computing useage. However, program development and Monte Carlo studies will continue to form a significant part of the overall computer requirements. The following discussion concerns the phase after the startup of the regular data taking periods in 1986, rather than the startup phase. It is important to realise that the amount of computer power necessary will not rise instantaneously after May 1985. This is due to the logistics of performing the data reduction and combating the inevitable problems that will occur initially. We foresee that the required amount of computer power will commence to rise soon, reaching its steady-state value sometime in 1986.

3.0 Raw Data

The amount of raw data generated for one event by CDF is expected to be approx. 100-150 KB. This assumes some preprocessing of the tracking information and the suppression of zero data from the calorimetry. Approximately 70% of the event record is expected to contain information from the main tracking chambers (Central, Vertex and Silicon), the remainder being information from the Calorimetry, Muon Tracking Chambers and Trigger. These numbers also include some estimate of noise. This 100-150 KB will be a very rough average, with individual events varying widely in size about this mean. In particular events with full information from even a small region of the Tracking System (e.g. Flash ADC Bucket information) will have sizes greatly exceeding this mean.

On the basis of the 100-150 KB average event size, a conventional 6250 bpi magnetic tape will contain of the order of 1000-1500 raw events (depending on the physical record size) and the logistics of mounting and demounting these tapes will probably dictate that a tape can be filled with raw data at a rate not exceeding 10-20 mins per tape. This implies an event rate that is limited to 1-3 Hz by logistics rather than hardware. The use of other storage media (such as laser discs) will not greatly affect this conclusion (other than allowing many more events to be stored on a physical device) because of the limitations imposed by the desire of reconstructing the events accumulated in one calendar year in the same year. One option that is currently being explored is to have an "express line", operating in parallel to the normal data acquisition (in a similar fashion to UA1), where events with (potentially) great physics interest are identified by the data acquisition system and written onto special tapes at typically 0.1-0.3 Hz. As in UA1, a great deal of the physics interest will initially focus on these events, but events from the normal tapes will be necessary for calibrations and monitoring as well as more detailed physics studies.

As events pass down the reconstruction chain, the event size will increase (for that sample of events retained for their physics, calibration or monitoring interest) through the addition of extra information describing the identification of individual particles or jets. CDF expect to retain the original raw data on the physics DSTs

and estimate the resulting event size will roughly double or treble to approx. 300 KB.

In the following discussions on CPU requirements and Magnetic Tape handling the following assumptions are made on Data Acquisition.

1. Mean Event Rate 1.0 Hz
2. CDF collects data for 4 months per year.
3. CDF is 70% efficient at data collection during these 4 months.

These figures imply that CDF will generate approx. 10000 Raw data tapes per calendar year.

4.0 Offline Data Reduction

A possible scenario for performing the data reduction is described below. All CPU times quoted are for a VAX-11/780 equivalent machine.

1. Raw Data rate 1 Hz. At an average luminosity of $10^{30}/\text{sec}^{-1}/\text{cm}^2$ this corresponds to a cross section of a few ub, while "interesting" physics channels probably have cross sections of at most 10 nb.

2. PASS 1. This comprises various event filtering algorithms which emphasize the use of the calorimetry information and perhaps finding event vertice(s) using the Vertex TPCs. Final Calorimetry analysis (cluster finding) will probably occur here. The expected average CPU time per event is estimated to be 30 secs. It is not yet clear how this number will be affected by the analysis performed by Level 3 of the data acquisition triggering scheme. Current thought is that any algorithms an Level 3 will be limited to a few seconds of VAX-11/780 CPU time per event (see later section) and will therefore not greatly affect the above conclusion. It is also not yet clear what philosophy should be taken - whether interesting events should be tagged or obviously uninteresting ones thrown away. This scenario proposes that approx. 10% of the input events will be retained for further analysis, together with a further 5% being a sample of garbage or uninteresting physics events (e.g. mimimun

bias) retained for calibration and monitoring purposes.

3. PASS 2. This is a full track reconstruction involving a full analysis of the data from the Central, Vertex and Silicon Tracking Chambers. This is expected to take 100-150 CPU seconds per event. It is not yet clear whether all events reaching this stage of the analysis will require full track reconstruction (perhaps only a limited solid angle or Pt range need be considered). All events entering PASS 2 are transmitted to the output tape.

4. PASS 3. This stage will correlate data from the various components (tracking and calorimetry information). The results will be used to sort the events into various classes of output DST's, depending on the expected physics content of each event. This Pass may take several tens of CPU seconds per event, bring the total of PASS2 and PASS3 up to approx. 200 secs per event. Note however, that this CPU power is only needed for 15% of the original events.

The sum of PASS1, PASS2 and PASS3 is then 60 CPU secs per raw event. CDF assumes that we will not be fully efficient at reconstructing the data, expecting false starts and program improvements to require re-analysis of the events. We have conservatively assumed this to double our CPU requirements.

With the assumptions on the trigger rate and running efficiency given above, and with the aim of reconstructing, on a yearly basis, all the data accumulated in real time (i.e. that the data for one calendar year of running should be fully reconstructed within that calendar year), a final figure of 25 VAX-11/780 equivalents is arrived at as being necessary for the data reduction alone.

5.0 Non Data Reduction CPU Requirements

Data Reduction, or event reconstruction, down to the Data Summary Tape (DST) level is only a part of the whole analysis of the data from the experiment. Other major factors are the physics analysis of the fully reconstructed events on the DSTs, Monte Carlo studies, Calibrations and Program Development. The expectation is that the following relative amounts of CPU power will be necessary:-

Data Reduction	DST Analysis	M/C, Calibration, Prog. Dev.
50%	25%	25%

There is not a full concensus on these figures. Particular areas of concern are high statistics Monte Carlo background studies, required to properly understand the signal to background for a variety of exotic physics channels, and the considerable care necessary for the correct calibration of the Tracking Chambers.

However, these figures indicate that our total CPU equirements will be approximately double those for the data reduction alone, bringing the total to approx. 50 VAX-11/780 equivalents.

6.0 Level 3

As yet, there is no real concensus on the particular hardware or software algorithms that will eventually comprise Level 3. However, there is general agreement that CDF soon needs to set specific goals (program size, speed etc.) in order to guide potential developers. At present the following groups are concerned with potential Level 3 hardware:-

1. Fermilab Advanced Computer Group. This group is working towards the creation of a system of parallel microprocessor based (M68000, 80286, etc.) CPUs, each with a large amount of memory (a few MB), connected to FASTBUS for the Level 3 case. Specialised, extremely fast co-processors may also be incorporated into the basic system.
2. 3081/E. SLAC and CERN are both working an a follow up to the 168/E, to be available next year. This solution requires an IBM (or look-alike) for program preparation.
3. DoE has approved a Phase I grant to investigate the incorporation of LSI-VAXs into a FASTBUS environment.

Potential algorithms to be implemented on Level 3 are restricted by the requirement that Level 3 can cope with a 100 Hz input rate, with a reasonably small number of Level 3 "engines" operating in parallel. Possible uses for Level 3 include:-

1. Flag events with multiple interactions.

2. Select "gold plated" events (cut on missing Et, etc.).
3. Calculate primary z vertex of the event. This information will then be used to correct any Et calculations.
4. Reduce the number of false muon triggers. The raw muon trigger rate is expected to be quite large.
5. Implement more sophisticated clustering schemes using the output from the hardware cluster finder.
6. Perhaps make a crude attempt at trackfinding.

However, it is clear that, given the restricted CPU processing time available, that Level 3 will not significantly decrease the processing time required for PASS1, but rather will affect the percentage of events accepted by the PASS1 algorithms.

7.0 The Use of Level 3 Hardware for Offline Data Reduction

One obvious goal is that the hardware developed for Level 3 should also be applicable for production data reduction. The CPU power required for Level 3 is estimated to be approx. 20 VAX-11/780 equivalents, so that a similar configuration dedicated to data reduction should be able, given the appropriate support environment, to handle a significant fraction of this production. Note however, that such a "farm" is only suitable for performing production using stable programs, it is not suitable for performing program development. High statistics Monte Carlo studies may also be efficiently performed in such an environment.

One very important point that must be fully considered is that of verification of the results from such a "farm". Unless the results are bit-for-bit comparable (e.g. by using emulators), very rigid verification tests will be necessary in order to validate results. Another area of concern is access to the calibration and bookkeeping databases. This implies both good networking and support software for such a farm.

8.0 Logistics

The previous sections have highlighted CDF's needs for significant computing power. However, raw computer power is not sufficient in itself. It must be made

available in a usable fashion, and sufficient manpower must be made available to implement the algorithms, submit jobs and chase faults etc. It is unlikely that the 50 VAX equivalents will be made available to CDF in the form of one single mainframe, but rather as an collection of CPUs, comprising CPUs suitable for program development and DST analysis and a "farm" of Level 3 style engines suitable for production data reduction. This collection must be easily accessible and should have access to centralised databases for programs, calibration constants and bookkeeping information. In addition, the collaborating institutions will be able to participate in both program development and DST analysis, although it is not yet clear at what level this activity will occur.

One problem in such an approach is the potentially very diverse machine architectures that must be catered for. This not only affects the program code, such that it must be transportable, but, perhaps more seriously, affects the executive procedures for job submission and access to the calibration and bookkeeping databases. The number of different architectures should be kept as small as possible. The majority of the collaborators have VAX-based computer centers and this, together with the use of a VAX as the online computer, has meant that the program development effort so far has taken place on VAX machines. In order to use the distributed computer power available, we must ensure that the main computer complex is readily accessible by VAX machines.

If CDF is to be successful in processing 10000 raw data tapes per year, considerable resources will need to be dedicated to logistics. Problems associated with mounting tapes, checking the correctness of the results, keeping up to date with the bookkeeping and chasing problems, will probably mean that the ideal goal of keeping up with the incoming data rate will not be achieved in practice, and will require that considerable thought and effort is devoted to achieving close to the ideal. Many of the aspects of handling this amount of data are dealt with best by a computing center having a dedicated staff of professionals available at all times. This has prompted CDF to make the following proposals:-

1. That a high speed link be setup between CDF and the Computing Center such that the raw data be written onto magnetic tapes already resident within

the Computing Center. This has the following advantages:-

(a) No, or minimal, magnetic tape storage is necessary at CDF. Movement of magnetic tapes, and the attendant bookkeeping problem, are considerably reduced.

(b) Magnetic tapes are mounted by Computing Center personnel rather than the people on shift, thus releasing the latter to concentrate on ensuring that the experiment is functioning correctly.

(c) The availability of alternate storage media (such as laser discs) will require an upgrade to the Computing Center rather than CDF, thus allowing such resources to be made available to other experiments.

(d) The possibility of performing the automatic submission and execution of PASS1 as the data arrives may reduce the need for raw data tapes and increase the turnaround of this important first phase of data reduction, providing better feedback of failures in the functioning of the experiment.

Note that the availability of such a link implies also the availability of a considerable amount of disc buffering capability at both ends of the link in order to cope with statistical fluctuations in the event rate and short term "outages" in the link. Longer term disruption to the link service must be catered for by the availability of magnetic tape drives at CDF.

2. That most of the CPU power be managed and operated by the Computing Department rather than CDF. Thus "farms", while being developed by CDF in conjunction with the groups identified in the section on Level 3, should eventually be operated on a day-to-day basis by the Computing Department. This will ensure that such resources become available to other computer users and receive proper preventative maintenance and management.

9.0 Conclusions

The following conclusions can be made:-

1. CDF will require approx. 50 VAX-11/780 equivalents of computer power.

2. Much of the CPU power required should be operated and maintained by the

Computing Department.

3. Logistical problems may well dominate how efficiently such computer power may be utilised.
4. A high speed link between CDF and the Computer Center has many advantages.
5. Good Networking in order to achive efficient access to a centralised database is essential.

As mentioned prev ly, the possibility of an "express line", using a dual stage analysis within Level 3 (either in the software or another level of hardware) to not only reduce the 100 Hz input rate to 1-3 Hz, but to provide a separate stream of potentially interesting events at typically 0.1-0.3 Hz, is being investigated. This might affect some of the CPU power conclusions in the previous sections, but the general feeling within CDF is that much effort will need to be expended on extracting difficult physics from the data, given the successes of UA1 and UA2. It is not clear whether extracting this difficult physics will be possible using such an "express line" approach, but rather all the raw data may need to be processed.

Appendix F

Suggested Topics for Work Projects

One of the major contributions which the proposed Computer Coordinating Group can make to the general health and welfare of Fermilab computing would be to oversee the long-range development of new computing systems and together with the Computing Department to provide practical and creative means of initiating a new system's wide-spread use.

The committee felt that the establishment of Work Projects, each dedicated to a concentrated study of new systems coming on the market, would serve to keep the laboratory aware, in some detail, of what should be taken advantage of for HEP Computing. This would be an excellent way to move this information into all phases of the laboratory's work.

Although decisions as to recommend which projects should be undertaken, and in what order of priority, would be the responsibility of the CCG, the issues raised in this Report suggest that the following topics receive serious consideration:

A. Personal Work Stations, at several levels of sophistication, as specified elsewhere in this report.

B. Interactive Operating Systems such as VM/CMS.

C. Emulators and microprocessor-based systems for offline as well as online application (In cooperation with the FCAG).

D. Dedicated CAD/CAM systems, which overlaps but does not coincide with A. above.

E. Computer networking hardware and software systems, for local area, wide area, and long range networks.

F. Major Software Issues:

1. Use of Data Structures in HEP programs, to describe logical relations between elements in analysis programs, in experimental setups, to maintain inventories and records, etc.

2. Use of Data Base Management Systems.

3. Software Management Systems.

4. Use of languages other than FORTRAN, such as PASCAL, C, or ADA.

Given the rapidity of changes and introduction of new ideas in computing, it would be a very good idea to attempt to carry on two or three projects simultaneously.

Some of these projects may require some reasonable financial commitment while others, such as tracking and actively studying new mainframe architecture may only require personal effort. Other projects may surface once the CCG begins to function.

Appendix G.

Mainframe Developments

There is a clear need for a central computing facility consisting of one or a cluster of mainframes operating with a "single system image", a large pool of peripherals, and a substantial operations, systems, and maintenance staff. The advantages of centralization in terms of cost- and labor-savings are evident in a variety of areas related to mass-production analysis. From the point of view of a user involved in the production analysis, program developments, or physics analysis, working on terminals attached to the lab's central facility also offers a number of strong advantages.

In specifying the choice of mainframe, compatibility of code with other machines owned by collaboration members has become essential. This means that the mainframe should have 32-bit (or 64-bit) word length, which is the de-facto standard among all smaller computers from super-minis down to microcomputers. A virtual operating system is also essential, as indicated above, to allow several of the multi-megabyte jobs typical of the larger present-day experiments to run concurrently, with 100-200 physicists simultaneously logged on via terminals.

The evaluation of the mainframe environment beyond 1986 is complicated by the fact that major changes in the architecture of large computing systems are expected within the next two to three years. The current direction of major vendors is toward hierarchical networks supporting personal work stations, file servers, storage servers and process servers. Each of these "server" systems will perform special functions in parallel, presumably leading to greater flexibility and/or efficiency.

The move towards heirarchical systems in the computing industry will no doubt be evolutionary, and will extend into the 1990's. We cannot expect that the whole computing environment can be implement as a fully-integrated extension of the mainframe operating system of any major manufacturer by 1986, which is the critical time-frame for Fermilab. However, a complete systems under VM/CMS, VAX/VMS, or even UNIX devlope many of the problems of integration now en-

countered in HEP may be avoided.

Table I lists a number of current or announced large high performance machines with some indication of the parameters describing them. (Roughly 10 mips $\sim$ 2-6 mflops in most cases)

Two traditional units of comparison are "mips" ,or millions of machine instructions per second, and "megaflops", or millions of floating point operations per second. Not only is there no direct conversion between these, but the actual rating of a given machine may depend strongly on the nature of the calculations being carried out. Similarly, memory size, per se, does not mean much. An important characteristic is user memory address space, that is, the number of memory locations to which a user program may refer. As discussed above with virtual memory techniques, that is generally far larger than the physical memory, and the performance depends critically on how it is implemented through a combination of cache memory and software.

While the speed and cost-effectiveness of "conventional" mainframes will continue to increase, the main advances in speed are expected to result from increased parallelism of computations. Currently there are a number of computers which provide synchronous parallelism through pipeline or other vector techniques, and which are especially suited to operations on vectors or arrays. These SIMD (single instruction, multiple data stream) devices include both supercomputers such as the Cray and Cyber 205, and some of forthcoming Japanese models, and also a number of more limited, but much cheaper attached processors, which augment the capability of a conventional host. Examples of these include the FPS 164, Star Technology's ST-100, and CDC's Advanced Flexible Processor. It is not at all clear that the performance gains of this kind of architecture offer much to High Energy Physics needs, since the conventional data analysis codes are not easily vectorized. However, it cannot be excluded that new ways of looking at the analysis may alter the picture in the future.

Lastly, there are asynchronous parallel processors, or multiple instruction multiple data stream (MIMD) devices, in which multiple processors or multiple independent instruction streams in one processor cooperate on a single calculation. A considerable amount of R&D activity has gone on in this area for a number of years, but it seems that the Denelcor HEP is the only commercial example. In principle much more powerful than the vector machines, there are formaidable problems in reliable and efficient use of these architectures. Over the latter part of the time period under discussion in this report,however, they may have developed into an effective tool.

Other features than speed may be critical to the effectiveness of future mainframes. Since memory costs can be expected to continue to decrease, memory is not likely to be a critical problem. A very serious question is code compatibility with previous generations of machines and with other networked ones. The trend is strongly towards highly modular bus oriented architectures with much of the detailed design implemented through microcode. With the ability to have alternate microcodes resident in one machine, and with the increasing recognition by vendors of the importance of maintaining code compatibility across and even between product lines, it is to be expected that over the next generation this will become much less of an issue. For the same reason, the sharp boundary between mainframe and supermini, or between workstation and host computer will continue to blur.

Table I Characteristics of Upper – End Computers

Vendor	Model	Options	Maximum Memory	Peak Perf.	Price Range	Comments
AMDAHL	5860		32mb	13mips	$3M	
	5880		32mb	23mips	$6M	
CDC	CYBER875	1 proc	8mb	19mips	$4M	
	CYBER875	2 proc	8mb	32mips	$6M	
	Cyber205	2 pipes	32mb	100-200mfls	$6-8M	
		4 pipes	32mb	200-400mfls	$15M	
CRAY	1S		32mb	160mfls	$8-13M	
	1M		32mb		$4-6M	4Q1983
	2	4-proc	256mb	1-2gfls	$10-20M	1985 (est)
	XMP	2-proc	256mb	400-800mfls	$9-11M	
DENELCOR	HEP1	1-16proc	1gb	10-160mips	$1-16M	
	HEP2	(a)	1gb	250mips	$7M	1986 (est)
		(b)	1gb	500mips		
		(c)	1gb	750mips		
		(d)	1gb	1000mips		
ELXSI	ELXSI64	1 proc		4mips	$0.5M	4Q1983
		10 proc		40mips	$15.0M	
ETA(COC)	CYBER2xx	1 proc		1gfls	$1M	1986
		8 proc		10gfls	$7M	
FUJITSU	VP-200			500mfls	$5M	Japan only
HITACHI	S820			630mfls	$12M	Japan only
IBM	3081 K	2-proc	32mb	19mips	$4.5M	
	3084Q	4-proc	64mb	27mips	$9M	
NAS	AS9080		32mb	20mips	$6-7M	
		IAP		100mfls		IAP not supp. in US
NEC	SX-1			570mfls		1985 Japan
	SX-2			1.3gfls		1985 Japan
SPERRY	1100/91	1-proc	64mb	6mips	$4M	
	1100/94	4-proc	64mb	27mips	$11M	
TRILOGY			256mb	30mips	$5-6M	1985
		IVP		100mfls		

This table was compiled by Alec Grimson of Cornell from a variety of sources. Peak performance data has largely been provided by vendors and should not be accepted as attainable with real problems.

Appendix H

Computing Requirements for Administrative, Engineering, and Accelerator Physics Users

Although the vast majority of CPU cycles at Fermilab are devoted to analysis of data from high energy physics experiments, a significant fraction of the users are involved in other tasks. These tasks may be divided into three broad categories:

1. Administrative : These include word-processing and report generation, electronic mail and filing, database management systems, project management tools such as scheduling programs, and business, accounting and inventory applications.

2. Engineering : These include applications such as stress analysis, magnetic field calculations, heat transfer problems, electronic circuit analysis, computer aided design of both mechanical and electrical systems, printed circuit design and layout, etc.

3. Accelerator Physics: These involve lattice design and associated magnet design, simulation of accelerator and storage ring characteristics, studies of various highly non-linear problems such as beam-beam interactions, beamline transport design, radiation shielding calculations and flux calculations.

All of these users are served at some reasonable level by the existing central computing facility.

Information on the future computing requirements of this rather heterogeneous group of users was obtained from two sources: letters requested by the Director of various department heads at Fermilab and interviews conducted by members of this committee with some of the larger users of the central computing facility.

In assessing the needs of these groups, it is important to note that they

are composed of people who are in general less knowledgeable about computers (and perhaps less interested in them) than the average high energy physics user. Therefore the computer environment is crucially important. There are many examples of powerful computing tools which are not extensively utilized because the pain level for learning them or using them is too high.

The requirements for future computing for administrative, engineering, and accelerator physics applications are:

1. CPU Cycles: The projected need of these groups is one CYBER equivalent. This is to be compared with approximately 0.25 CYBER equivalents currently used for these purposes. Most of the projected increase will be in the area of accelerator design and physics. Fermilab obviously will be strongly involved in the design of the new large accelerator (SSC). An interactive program of lattice calculations and magnet design will be necessary in order to produce a technically correct solution which satisfies budgeting constraints. In addition, extensive studies of the performance of Fermilab's anti-proton source and proton-anti-proton storage rings will need to be undertaken. Several experts have estimated the computing requirements for these tasks and have converged on the estimate given above. Cycles used for engineering computations will increase significantly as more and better programs become available but will remain a small fraction of those devoted to accelerator physics.

2. Memory Requirements: Computers must be able to run jobs requiring up to a few megawords of memory. One of the major problems the engineering and accelerator physics users encounter with the existing central computing facility is the limitation on memory size. Many of the calculations involve the definition of data on two of three dimensional meshes which require large amounts of memory. Other calculations simply use very large, complex codes. The finite element analysis program, ANSYS, is an example of a commercially developed code which had to be converted to an overlay structure in order to fit into the CYBER memory space. Fermilab must have facilities to allow it to use commercially devel-

oped codes, and codes developed at other centers for accelerator physics, without requiring extensive modifications. Fermilab must keep up with the other high Energy Physics Labs and industry in available memory size if it wants to benefit fully from programs developed elsewhere.

3. Machine Architecture: Since the total computing requirements of these users is small compared to the High Energy Physics data analysis load, it is possible to imagine buying computers specially suited to accelerator and/or engineering computations. None of the experts know which of the alternative architecture is most appropriate. Research into this area might be very worthwhile.

4. Text and Graphics Display Equipment: The text and graphics display requirements of these groups are similar to those of physicists doing high energy physics data analysis described in the section "The Computing Environment". In particular, engineering calculations frequently require fairly sophisticated two and three dimensional graphics both for model building and problem definition and for presentation of results.

5. Intermachine Communications: It may turn out that several different types of machine will be required to satisfy all the needs of this heterogeneous group of users. One can even imagine wanting specialized graphics facilities, machines supporting business applications, special computers and hardware for word processing, and computers chosen for their suitability to specific kinds of engineering problems. It is clear that these machines all need to be able to talk to each other in some standard fashion in order for Fermilab to derive maximum use from them.

In conclusion, these users need 1 CYBER equivalent of computing power, with large addressable memory, and excellent text and graphics display facilities. It is possible that some of these needs will be satisfied by several small machines, or even personal work stations, and Fermilab should try to make it very easy to communicate from one machine to any other.