CMU-ITC-83-027

PRELIMINARY TECHNICAL PLAN

FOR THE

JOINT VENTURE

BETWEEN

CARNEGIE-MELLON UNIVERSITY

AND

INTERNATIONAL BUSINESS MACHINES CORPORATION

Draft of January 9, 1983

Douglas Van Houweling

This Technical Plan is a preliminary document which attempts to outline the the technical approach Carnegie-Mellon University and International Business Machines Corporation intend to use in their joint venture focused on networked personal computing. The plan has been contributed to by a large number of individuals from CMU and IBM. The author's main contribution has been to record and document those contributions here. Since this joint venture has the objective of advancing the state-of-the-art in computing system design and use, the technical plan will be continuously and extensively revised. The system design is at an early stage, and substantial changes in this plan will be necessary as system development continues. Because of its preliminary nature, this document is understood to be confidential and is not to be transmitted or quoted outside CMU and IBM.

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

This document summarizes the technical plan that has been agreed upon between Carnegie-Mellon University (CMU) and the International Business Machines Corporation (IBM) for a joint venture in the development of an integrated personal computing environment. This development effort is directed at creating a computing system based on personal computers, a communications network, and a central computing facility. These elements of computing equipment are to be integrated into a single computing system throught the devlopment of a comprehensive software system. IBM expects to market the resulting system to the higher education community.

To place the technical plan into perspective, the objectives for the system are first discussed in detail. Achieving these objectives will create a system with much higher performance and broader applicability to higher education than any extant system. The system should make possible fundamental improvements in the capability to provide higher education.

Based on the objectives, the majority of the document summarizes the system's design. The design is explicitly related to the system objectives, and deals with the personal computer workstation, the network, the cluster facilities, the central computing facilities, and the software. The crucial role of the system's software is highlighted.

Since it will be two years before the new system will be available for pilot use, a transition strategy is required. The transition must support not only CMU's needs, but also provide a means of developing instructional computing applications software for the new system prior to its availability. The need for compatibility between the transition system and the new system is discussed, and the tools planned for the transition system are discussed.

Finally, based on the new system's objectives and design as well as the transition plan, the implementation of the new system is discussed. Most of the stress is placed on the implementation of the software, since that is the most challenging aspect of the effort and the portion of the effort to be undertaken jointly by CMU and IBM. Based on a discussion of implementation strategies, highlights of the implementation schedule are outlined.

The picture that emerges is of a very major effort which is not devoid of risk. The overall design and implementation plan deals with this risk in several ways. First, it contemplates heavy use of existing software. Second, it prioritizes the various software development efforts to reduce the probability that large amounts of effort will be expended on less needed facilities when more important facilities need development. Third, it includes a strategy for incremental development which will allow the system to become available in segments across a four-year period. CMU and IBM have examined the risks, and have concluded they are manageable. More important, the payoff that will result from the availability of the integrated personal computing environment is enormous.

2. SYSTEM OBJECTIVES

Carnegie-Mellon University (CMU) has joined with the International Business Machines Corporation in a joint venture to develop the potential of personal computing. The joint venture exploits the capabilities of CMU in integrated personal computing systems software and of IBM in the development of high performance personal computing systems, communications networks, and large scale computing systems. The joint venture's technical objective is the development and trial deployment of an integrated computing system which combines the advantages of personal computers with modern timesharing computing technology. The resulting system will support applications not presently feasible with either personal computing or timesharing technology. The characteristics of the integrated personal computing environment (IPCE) determine its design features. The characteristics include:

1. Reduced resource contention -- The ideal computing system's accessibility and performance is unrelated to the number of simultaneous users and their applications. Timesharing systems are especially subject to poor performance and accessibility because of resource contention, while a single-user personal computer is not.

2. Easy information sharing -- Subject to the author's restrictions, the ideal computing system should provide any user access to the information generated by any other user. The problems of security and system-wide identification of information must be solved while providing each user with physical access to information stored in the system. A high-quality timesharing system provides easy information sharing for a limited community of users, while a standalone single-user personal computer does not.

3. Ease of use -- The computing system must be easily used by non-programmers as well as programmers. In particular, it should be possible to utilize specific applications without comprehensive knowledge of the system's structure and other applications. Ideally, the system should customize itself to each particular user's characteristic pattern of use.

4. Context preservation and switching -- The computing system must be able to maintain several contexts simultaneously, and preserve them across sessions. For instance, it should be possible to interrupt a document composition task in order to retrieve information from the library database and then re-enter the composition task as if there had been no interruption. Furthermore, it should be possible to suspend the current task environment and re-enter it at another time and place.

5. Geographic flexibility -- A user should be able to work at any location equipped with an appropriate workstation without physically moving storage media such as floppy disks or tapes. This use should extend to off-campus locations as well as laboratories, classrooms, dormitories, and offices on campus. In the latter part of the decade, the system should be

capable of providing a large subset of applications anywhere in the world. Initially, the emphasis will be on service to users located within ten to fifty miles of campus.

6. Reliability -- The system must be as dependable as any utility on which we rely for our everyday work. If it is less reliable, it will not be used for important and critical work when other, more reliable, facilities are available.

7. Multi-media capability -- The system must be capable of assisting in the generation, storage, retrieval, and communication of audio and video forms of information as well as the more traditional forms of written information. In particular, the system should be designed to interface with telephone and television systems.

8. Cost-effectiveness -- The computing system must be affordable for college and university use. It should be deployable across a broad continuum of investment per student to support varying institutional strategies for computer use in higher education.

9. Expandability -- System configurations should range from tens of workstations up to thousands, and the system should be capable of growing smoothly from a small to a large system.

10. Flexibility -- The system should be capable of supporting a wide range of applications in instruction, research, and administration. It should support those applications using the most appropriate computing technology. to avoid the necessity of developing new applications software whenever possible.

11. Compatibility -- The system must provide an interface to the existing timesharing systems and enable a smooth transition from the current computing environment to the new one. Such compatibility will reduce the need to develop new applications software. The system must conform to software and communications standards to allow smooth inclusion of new technology as it becomes available in the future.

3. SYSTEM DESIGN

Meeting the system objectives requires a fundamental change in strategy regarding the geographic distribution of computing power. The computing system will have many components -- personal computer workstations, a communications network, cluster facilities, and a central computing facility. All of these components will be supported and combined into the overall system through an integrated software environment. This section provides a description of these components and the strategic rationale for their capabilities and inclusion in the system.

3.1 PERSONAL COMPUTER WORKSTATION

The component of the integrated personal computing environment that has attracted the most attention is the personal computer workstation. The use of personal computers as workstations in the new system is undoubtedly the fundamental strategic change from the current timesharing computing environment. In the timesharing environment, the computing power is provided by a shared computer located centrally. In the integrated personal computing environment, almost all of the computing power will be provided by microprocessors in the personal computer workstations. In combination with the other components of the system, this redistribution of computing power is the key to realizing many of the objectives of the integrated personal computing environment.

- Reduced Resource Contention -- Each user of the integrated personal computing environment will have a powerful computer dedicated to his or her own use. As a result, the performance seen by that user will be unaffected by the computational demands of other system users. Furthermore, since it is CMU's goal to provide at least one such personal computer workstation for each individual using computing, the user will always have access to computing.

- Ease Of Use -- The personal computer workstation will have adequate computing capacity to implement a user interface which depends on artificial intelligence techniques to provide interactive computing adapted to the needs of each particular user. By keeping a record of the user's interaction with the system, the software will provide extensive help by coaching a user who is exploring a new and unfamiliar tool. As the user becomes more experienced with the application, the system will automatically interact in a terser, less time-consuming fashion. This approach incorporates work currently underway in the CMU Computer Science Department, in the Graceful Interaction project and the Gandalf project. Both of these projects focus on developing user interfaces which support the user when help is needed but do not impede the experienced user.

Such an adaptive user interface requires substantial computational power and is therefore difficult to realize in a timesharing computing system. The timesharing system designer's objective is to provide adequate response time to a large number of users. As a result, priority is given to user interfaces which minimize the amount of power required for each interaction. The precise opposite design criteria applies in a system based on personal computing. Between the time a user request is satisfied and the user formulates the next request, the computational power of the personal computer workstation is not in use. These extra computing resources can be applied to providing a high quality user interface.

- Context Preservation And Switching -- The best of the currently available timesharing systems support the feature of context preservation and switching. But the complexity and performance penalty is relatively high. A timesharing system is already required to support a context for each user of the system and to switch among those contexts at a high rate of speed. Preserving multiple contexts for each of those users adds a further level of complexity and system overhead. A personal computer based system can utilize the techniques already well understood and mastered for timesharing to provide the facility to switch among tasks for a single user rather than for multiple users. In other words, the segmentation of the overall computing load among many personal computers allows a relatively straightforward implementation of context preservation and switching.

- Reliability -- Since there will be a large number of personal computer workstations in the overall integrated personal computing environment, a breakdown in any single personal computer will affect only the user of that computer at that moment. This can be dealt with by simply moving to another personal computer or exchanging one personal computer for another while the first is being repaired. A timesharing system idles tens of terminals if the central computer fails and the cost of the central computer makes it economically difficult to provide a backup unit in case of failure.

- Multimedia Capability -- Providing a multimedia environment for the user requires that large quantities of data be transmitted rapidly between the computer and the workstation. For instance, providing high resolution animated video requires that information equivalent to 225 large books be transmitted between the computer and the workstation each minute. Physical separation of the workstation from the computer virtually prohibits this type of display because high bandwidth communication facilities are very costly. The personal computer workstation, however, incorporates the computing in the same unit as the display, making such data transmission economic. Furthermore, since a workstation will be available to each individual, it can also serve as a telephone and television, providing the potential for integrating the telephone and television function into the workstation.

- Cost Effectiveness -- The distribution of computing power into the personal computer workstations reduces the overall cost of the computing system in at least two ways. First, large quantity production of very large scale integrated circuits has made it relatively inexpensive to add a powerful microcomputer to a computer terminal. Since the computer terminal would be required in any case for access to a timesharing system, the incremental cost of deploying computing power in this fashion is relatively small. Furthermore, the fact that personal computer technology is becoming a consumer commodity with very large production volumes is rapidly reducing costs.

- Expandability -- While some of the advantages of the integrated personal computing environment will not be achieved until every user has his or her own personal computer workstation, many are achievable when these workstations are shared by a number of users. As a result, it is possible to deploy the integrated personal computing environment at varying levels of intensity. An institution which places relatively low priority on access to computing might only deploy one such workstation per twenty students. Another institution, such as Carnegie-Mellon, might strive to provide at least one workstation per student. Since the other elements of the system will be designed to accommodate up to 10,000 personal computer workstations and can be sized to efficiently accommodate a system with as few as 50 such workstations, it should be possible for an institution to start with a relatively small investment and increase that investment smoothly across a period of time.

- Flexibility -- The large number of personal computers now being deployed throughout society has led to an explosion of computing software for personal computers. The unique characteristics of personal computers has resulted in certain software packages being available on personal computers first, and only later on timesharing systems. The most striking example, of course, is Visicalc. Visicalc requires a high data rate between the display and the computer to function effectively, and was therefore developed first for personal computers. It still exists for only a few timesharing systems. The use of personal computer workstations should make this cornucopia of software available to the users of the integrated personal environment.

The personal computer workstations now available do not, by and large, have sufficient capabilities to realize all of the above advantages. As a result, CMU and IBM have jointly specified a target for the personal computer workstation. The characteristics of the workstation that have been specified are as follows:

- Modularity -- The personal computer workstation must be designed so that it can be expanded in a modular fashion. This modular approach is required by the cost effectiveness and flexibility goals stated above. The basic workstation supplied to each student and faculty member will probably not contain high resolution color display capability, a local printer, or high speed numeric computation capability, but these capabilities will be required for certain applications. Furthermore, the capability to interface the personal computer workstation to laboratory equipment will be crucial for many research and teaching purposes. It is plainly not feasible to provide all of these functions on every personal computer workstation. But, some of them will be required on some of the workstations. As a result, a modular design for the workstations is a necessity.

- Powerful Processor -- The maintenance of a high quality, rapid response user interface incorporating multiple media requires very substantial computing power. Experimentation in the SPICE project at CMU's Computer Science Department indicates that at least one half million instructions per second of processing power is required and it is desirable to have a processor of two to four times that speed. The initial personal computer workstation will be based on a Motorola 68000 processor with a processing speed of approximately one half million instructions per second. The processing power expected to be available at the time of full deployment is in the neighborhood of 2 million instructions per second.

- Virtual Memory -- Maintaining multiple contexts and switching rapidly between them is most easily and efficiently implemented in a system that has a very large address space. Since such a large address space cannot be economically realized through providing real memory for the entire address space, virtual memory is necessary. The initial personal computer workstations will have a virtual memory of sixteen million characters, and the personal computer workstations expected to be available in 1986 will have a substantially larger virtual memory. The initial implementation of virtual memory will be with the Motorola 68010 microprocessor supported with special purpose memory mapping hardware.

- High Resolution Graphics -- The capability to implement a multimedia user interface requires that the system include high resolution graphics. The objective is a graphic display approximately the size of a typewritten page with resolution of approximately 800 pixels horizontally and 1,000 pixels vertically. This display resolution will be capable of reproducing not only high resolution graphics, but also a printed page incorporating multiple type fonts. Combined with multifont laser printers, the high resolution display will allow a writer to compose each page of a document on the screen and see it exactly as it will be printed. The high resolution display is important not only for its support of graphics and document preparation but also because of its capability to support the context switching user interface. The high resolution display allows simultaneous display of multiple viewing areas, one for each application currently active. These viewing areas can be adjusted in size and overlaid, depending on their priority at any given instant. In other words, the computer screen can be treated very much as the surface of a desk is usually treated, with multiple tasks generating several overlaid stacks of documents and information, arranged to conform to the user's activities.

- Graphics Input -- The user interfaces that have recently been developed by Xerox and are demonstrated most clearly in their Star professional workstation derive much of their power from their use of a graphics pointing device to determine the activities which the computer is to undertake. On systems of this type, rather than typing commands on a keyboard, the user points at a menu of commands displayed on the screen, and the computer executes those commands. The result is a much higher rate of information exchange between the user and the computing system. Furthermore, the graphics input capability will enable the user to draw pictures on the computer screen so that a broad range of applications ranging from the incorporation of diagrams into documents to computer assisted painting become possible.

- Video Capability -- With the advances in video disk technology and high definition television, an increasingly large amount of video information will become vital to the everyday activities of students and faculty. For instance, a history of art course may well maintain a library of artwork on video disk. In the near future, the most feasible way of transmitting these video images may be through standard cable television technology. It is therefore important that the personal computer workstations be able to display a television image. This television image display should be able to coexist with information generated by the personal computer. In other words, it should be possible to annotate and edit video displays.

- Audio Capability -- The personal computer workstation requires audio capability to serve two purposes. First, the user interface should include the capability to generate spoken responses to the user and, as voice recognition technology improves, intercept and act on spoken commands from the user. Second, the system needs to be interfaced to the telephone system. In particular, the personal computer should be capable of answering a telephone call and recording a message. It should also be capable of including an audio message as part of a document to be transmitted from one individual to another. Finally, there are applications in the arts, especially music, which require the capability to synthesize tones and sounds. In particular, there is a project currently underway at CMU which has the objective of using personal computers in a network to synthesize the sound of a full orchestra based on a score input to the computing system.

3.2 NETWORK

Many of the characteristics of the integrated personal computing environment require that the major source of computing power is the personal computer workstations. The use of personal computer workstations alone, however, is not sufficient to realize the characteristics of the integrated personal computing environment. While the personal computer workstations provide significant improvements over a timesharing system, other beneficial aspects of timesharing are lost in a system of personal computers. The data communications network is therefore a key element in realizing the overall objectives of the computing system. The relationship of the computing network to the relevant system characteristics follows:

- Reduced Resource Contention -- Moving most of the computational power from a central shared computing system to microprocessors resident in each of the personal computer workstations eliminates a major source of contention for resources. If, however, the network which interconnects the personal computers, cluster facilities, and central computing system is not adequate to provide a performance margin during peak usage, it could become a new bottleneck for a number of the network elements which are shared by multiple personal computer workstations. The network, therefore, needs to be designed in such a way that its capacity can grow in proportion to the number of workstations. To support future growth, the network must also have capacity more than adequate for current applications.

- Easy Information Sharing — A system which uses personal computer workstations to provide the majority of computing power does not necessarily achieve the objective of information sharing. In fact, most personal computers currently available possess only rudimentary facilities for transmitting information generated on a particular personal computer to any other computer in the system. These facilities are rudimentary in three areas. First, there is little provision in most of the systems for high speed physical communication of information. Second, the information used by any particular personal computer is usually contained on physical storage media resident in that personal computer and is therefore not available to other computers unless explicitly transmitted to some central data storage facility. Third, the conventions used to identify information are designed only to facilitate the identification of information within each personal computer and do not facilitate system wide identification of information. The network plays a key role in addressing these obstacles to easy information sharing. The communications network will interconnect all of the personal computer workstations, clusters, and the central computing facility. It will have adequate speed and capacity to transmit one half million characters per second between each personal computer and the remainder of the system. This rate of data transfer will make it feasible for most of the personal computer workstations to store all of their files in file servers attached to the network rather than in disk storage integral to each workstation. As a result, the data generated by a user at any workstation is stored in the computing system via the network and is, therefore, available for retrieval from any other workstation.

- Ease Of Use — The communications network will provide the physical capability required to make each personal computer workstation an integral part of the overall computing system. Using the communications network, a user at a personal computer workstation will be able to access the library database automatically, without becoming aware that the library database is resident in a computer at the central facility rather than in the personal computer workstation. In other words, the complexity of using other computing resources is greatly reduced by having each personal computer continuously attached to the network.

- Context Preservation And Switching -- As outlined above in the discussion of the personal computer workstation, the capability to preserve contexts and switch between them requires that the personal computer workstations have a virtual memory architecture with a large address space. An obstacle to this approach is the capital and operating expense required to provide a mass storage device to support the virtual memory in each of the personal computer workstations. The network will have adequate capacity to allow the personal computer workstations to page virtual memory over the network to a shared page server.

- Geographic Flexibility — The network will have adequate capacity to transmit the full context of a session from one paging server to another. As a result, a user can work for a period of time at one personal computer workstation, then move to another personal computer workstation located some distance away across the campus and retrieve the exact context that was left behind on the first workstation.

- Reliability — Because the data communications network is the single component of the integrated personal computing environment that has the capability to render the entire system inoperative, it is imperative that it be designed with redundant signal paths and provision for automatic isolation of faults.

- Multimedia Capability — The data communications system must have adequate capacity to transmit digitized video and audio. Furthermore, wherever possible the wiring for data communications should be consolidated with wiring for analog video and telephone transmission, to enable the ultimate merger of these communication media.

- Cost Effectiveness — The high capacity of the data communication system allows a number of key cost effectiveness objectives to be realized. The most important is the ability to specify personal computer workstations without local data storage. Perhaps as important, the data communication system will be designed to a widely accepted standard. This will encourage volume production of the important components, thereby lowering the cost of connecting all of the system components.

- Expandability — The data communications system should be capable of expansion. It therefore needs to be designed as a hierarchical system to which additional capacity can be added as the overall computing system is expanded.

- Flexibility — The data communications network is the key to providing flexibility and support for a wide variety of applications. For instance, the data communications network can make it possible for a personal computer workstation to provide the user interface while a central computing system provides the database service for an application such as library information retrieval. In addition, the data communications system will make it possible for a wide variety of special purpose computing facilities to be interconnected using standard, well defined interfaces.

- Compatibility — Again, the data communications network provides the key to compatibility. In particular, the network will make it possible for any personal computer workstation in the integrated personal computing environment to simulate a terminal to the existing timesharing computing systems. It will furthermore allow existing computing terminals to be connected to various facilities wherever the applications can reasonably be supported with such terminals.

As a result of Carnegie-Mellon's and IBM's investigations of data communications, we have jointly agreed to proceed with a communications network based on the submissions that IBM made to the IEEE 802 Network Standards Committee. In particular, the data communications network will be capable of very high speed — in the range of a million characters per second. It will connect all of the computers in the computing environment and will have standard interfaces so that other computers can be connected with minimum difficulty. It will be capable of supporting file access and virtual memory paging. It will provide support to the existing computing equipment and terminals on the campus and it will be hierarchically organized for easy expansion and reduced contention. Finally, the data communications network will be extended off the campus into the city through the use of the facilities of AT&T and/or Warner Cable. Discussions are now underway with both of these organizations to work out business arrangements for the use of their technology. It appears that technological capabilities exist in both the telephone and the cable television network to support this geographic expansion of the computing network.

3.3 CLUSTER FACILITIES

The cluster facilities serve an important economic function in the overall integrated personal environment. The most important application is to provide public or semi-public shared access to personal computer workstation options and peripherals which are at that time too expensive to include in each individual personal computer workstation. A representative cluster facility in a classroom building, therefore, might include a set of personal computer workstations with high resolution color displays, very fast scientific computation, generation of complex sound waveforms and local data storage. A cluster facility in the laboratory might include machines with some of the same features but also configured to interface to and control laboratory equipment for automating student laboratories. Shared use of these more expensive personal computers will make it possible to justify their substantial capital cost. The modularity of the personal computer workstations will allow for this expansion of these individual systems.

In addition to providing a capability for shared access to high function personal computer workstations, the clusters will also typically include peripheral equipment such as laser printers, input scanners for input of printed text to the system, and the capability to read data from floppy disks, tapes, and other media such as video disks which may appear in the future. For instance, at the end of each corridor in a dormitory it may be wise to deploy a cluster facility with several high function personal computers, a laser printer, and an input scanner. Thus, when users of the network need to produce printed output or input documents to the system, they would be able to route the print to the cluster and then walk down the hall a few moments later and pick it up.

The clusters will also provide for continued face to face group interaction among users of the computing systems. While this will prove necessary for certain types of classroom exercises, it is also a valuable for the university environment at large. Carnegie-Mellon's existing experience with its timesharing computing systems indicates that the public and semi-public terminal rooms are important sources of social interchange and co-teaching. As a result, it is expected that a university-wide system will have clusters of personal computers before individual personal computers become available, and that these clusters will never be replaced by the individual machines. Rather, the clusters will supplement the individual personal computers. This natural migration from a cluster based, shared computing environment to a more individual computing environment is also economically desirable because the early personal computer workstations may be too expensive to dedicate to individual students, and may best be made available through shared cluster facilities.

3.4 CENTRAL COMPUTING FACILITIES

The central computing facility almost certainly will grow over the entire development period of the integrated personal computing environment. The growth will be smaller and much more specialized than it would be if the central facility were to continue to provide the majority of the computing capability. It should be noted that the "central" facility may not be provided by a co-located concentration of large computers. Technology may dictate a centrally-managed facility which consists of a geographically distributed group of hardware installations. Areas which will continue to require substantial central capacity are large scale on-line and archival storage, large-scale computation, and costly specialized devices.

3.4.1 Large scale on-line and archival storage

Current experience with professionals who have unconstrained access to computing facilities indicates that they easily can utilize three million characters of on-line storage in support of their writing, administrative, and research activities. For Carnegie-Mellon University to achieve this standard, twenty billion characters of storage are required. Current plans are to distribute the system's file storage across a number of file servers, each of which would support twenty to one hundred personal computers. The files in these file servers would be periodically backed up through the network to a central file for reliability and transportability. Approximately ten percent of the files generated each month will be archived for possible reference months or even years later, which will necessitate an archival system capable of growing at the rate of approximately twenty billion characters per year.

3.4.2 Large scale computing

Despite the fact that the personal computers with appropriate floating point options will soon be available that can execute several million floating point operations per second, a minority of tasks will require the capability to execute a single instruction stream at an even higher rate. In addition, substantial computing power will be required to manage the large body of on-line and archival storage and distribute it to the file servers and personal computers. The personal machines will also make substantial use of the central facility to acquire, develop, distribute, and store software. Finally, a minority of users will need access to the rich software environment characteristic of a mature computing system because less-used software will not have been converted to execute on the personal computers.

The central facility will need high reliability because it will play a critical role in the functioning of the overall environment. As a result it should contain a fully duplexed set of equipment designed so that loss of some element will cause only a temporary interruption in service. Based on current usage projections, it appears that the facility will need a processing capability of at least twenty million instructions per second if it is to support the volume of use expected at Carnegie-Mellon University. The needs of users with large-scale numerical tasks indicates that a floating point processor which can execute about twenty million floating point operations per second will also be desirable.

3.4.3 Specialized devices

The heavy anticipated use of graphics and text processing facilities requires that facilities be available to produce high quality final output. In addition, facilities will be needed to input data from the printed page, from various types of magnetic media, and probably from the new video disk technology. In particular, the following equipment will be needed:

1. Multi-font and graphics capable laser printer,
2. Phototypesetter,

3. Color graphics capable printer,
4. Color graphics transparency camera,
5. Color, high resolution, large format plotter,
6. Microform printer,
7. Character page reader, and
8. Tape, floppy disk, and video disk drives for media conversion.

As technology advances and costs change it is likely that additional devices will be needed at the central facility and that some of the devices will be low enough in cost to migrate to local clusters or into the personal computers themselves.

3.5 SOFTWARE

All of the components of the integrated personal computing environment discussed above are based on state-of-the-art computing equipment and data transmission facilities. The most innovative requirement for implementation of the integrated personal computing environment, however, is the development of software to tie all of the elements together into an integrated system. While there are currently efforts underway at a number of research institutions to generate software of this type, the first known attempt to reduce this research to practice in a large integrated computing system is the foundation of the Carnegie-Mellon/IBM Joint Venture. The software provides the last and crucial link in realizing the characteristics desired for the integrated personal computing environment.

3.5.1 Software Features

The key features of the software will be:

- System Wide File Access — As described above, the combination of the communications network and the personal computer workstation will result in each user's data being stored in the system rather than in his personal computer. In other words, data in the integrated personal computing environment, while organizationally the property of the individual user and/or the group or groups with which he or she shares data, will be operationally resident in the overall computer system. When combined with system-wide naming, residence of information in the system rather than in the personal computer makes it possible to implement easy information sharing, context preservation and switching, geographic flexibility, and reliability in a cost effective manner.

- System Wide Naming — One of the greatest challenges in implementing a computing system with thousands of users and billions of characters of information is developing an approach to find information stored in the system. This capability is absolutely vital if the desired characteristics are to be achieved. Therefore, the software will include the capability to identify and find data according to a variety of search patterns. In order to implement this difficult task, the file structure in the system will itself be contained in and organized under a powerful database system.

- Artificial Intelligence In The User Interface -- As outlined above, the nature of a system which relies on personal computer workstations for its computational power allows the use of artificial intelligence techniques to enhance the user interface. This will provide the ease of use so vital to broadening the user community in coming years. In particular, it is expected to be helpful in providing high quality computing facilities to users in the Fine Arts and Humanities.

- Automatic Access To Network Resources -- Software will also provide the user with a single system image. In other words, even though the overall integrated personal computing environment will consist of thousands of machines, network facilities, and peripheral devices, it will be made to appear to the user as a single integrated system. For instance, when a faculty member uses his personal computer workstation to assist in assigning grades to his class, he or she will not be aware that the personal computer workstation, communicating through the network, is invoking procedures to retrieve information from the student database on a central computer and subsequently updating that database with information supplied by the personal computer workstation. The same type of integrated system image will be provided for a large number of complex applications including electronic document transmission (mail) between users and access to the resources of the library.

- Resource Sharing -- One of the outgrowths of automatic access to network resources is that a number of resources will be designated as shared resources. The most trivial example is probably a laser printer located in a cluster facility or the central computing facility. A less trivial, more complicated example is the sharing of a large database for administrative purposes. The system will be designed so the personal computer workstations have shared access to such resources.

- Multiple Contexts -- Perhaps one of the most important aspects of the software environment will be its ability to support simultaneous multiple contexts for each user. As a result, a user who is currently editing a document does not need to lose his or her place because of an interruption from a phone call and a resulting inquiry to a database unrelated to the document being edited. The context of the document is maintained while the interrupting task is carried out and is then available for further work. The ability to maintain multiple contexts will be available not only across time on a particular personal computer workstation but also across space since the network will make it possible to preserve contexts and transmit them from one personal computer workstation to another.

3.5.2 Software System Design

Meeting the requirements of the software system for the integrated personal computing environment requires a carefully structured layered design for the system, tool, and application software. Such a layered design has the structure of an inverted pyramid in that each layer depends on and is supported by the previous layer.

In particular, the first layer will consist of routines whose details are dependent on the system's computer hardware. This layer will be the only layer of software which is written in a machine specialized fashion. It is, therefore, the only layer which will need to be rewritten if the overall computing environment is moved to a new type of computing hardware. Routines for memory management, task switching, interrupt handling, device handling, process management, interprocess communication, keyboard input, display management, and file input and output will be included in this layer.

The communications subsystem which will tie all of the computing facilities in the integrated personal computing environment together will be constructed on the layer above the hardware dependent software.

A set of global utilities which use the services of the communications subsystem and the hardware dependent software will comprise the next level. Name and authority servers to find and authorize access to information in the system; file and output servers to retrieve information and output it to printers, plotters, and phototypsetters; and system-wide database and archiving systems to organize and provide for the permanent storage of data will be implemented on this level. The display window manager will also be at this level and will allow the personal computer display to be broken into several areas to serve multiple tasks.

On top of these levels, finally, the user accessible utilities and tools will be built. The utilities and tools will include a document formatting system; a screen editor; several high level language development systems for languages such as Pascal, Fortran, and Logo; a loader and link editor to combine routines into executable packages; and a library of routines to link the high level language facilities to the lower level global utilities in the system.

Based on the availability of these utilities and tools, another set of programs will be implemented which provide on-line documentation, word processing, electronic mail, bulletin board, on-line help, calendar management, statistical analysis and graphics presentation. These software facilities will also provide the foundation for development of a system which provides access to the library database, and an on-line assistance capability for users who need help and guidance. The top-most level of the system will be the user command language which will be capable of automatically customizing itself to the needs of each user based on that user's expertise in various applications.

The layered approach to development of software has several advantages. First, it means that programmers working on particular applications can use the tools already available at a lower level and do not have to "reinvent the wheel". Second, common use of the utilities and command language will provide users with a common user interface as they move from one application to another. The need to continually learn new command languages and new means of invoking and controlling particular applications will be eliminated. Finally, the layered software development approach requires well defined interfaces between the various components of the system, facilitating the replacement or modification of components in response to technological opportunities and user needs. For instance, moving to a different type of personal computer workstation will require revision of only a limited part of the overall system. The software environment will therefore not become obsolete because when more advanced computing system hardware becomes available.

Another aspect of the software system design is the approach taken to providing access for other computing equipment. This requirement is prompted by two characteristics of the university computing environment. First, Carnegie-Mellon and most other universities have a large variety of existing computing equipment which will continue to serve an important segment of the user community's needs over the coming five to ten years and this equipment must be accessible and usable through the integrated computing environment. Second, the needs of researchers on campus will not in all cases be met by the personal computers and central computer facilities that are part of the integrated personal computing environment. As a result, it must be possible to attach equipment to the computing environment that is not part of the integrated system.

These facilities will be provided in two fashions. First, provisions will be made in the software for the transfer of files to and from any computer connected to communications network. At the beginning, there will be large amounts of file transfer traffic to and from the existing Digital Equipment Corporation computing systems at Carnegie-Mellon.

Second, the network will provide for the connection of standard computer terminals to the communications network for access to applications that are not part of the integrated personal computing environment. During the early years of implementing the environment this access will serve as an invaluable tool for maintaining continuity of service. In later years, it will reduce the need for reworking existing application so that one or two specialized users will continue to have access to them.

Third, it will be possible for the personal computer workstations to simulate terminals to the existing computing. This capability, in combination with the file transfer, will make it possible for existing applications to stay in use for as long as it is technically desirable and coexist with the integrated personal environment.

4. TRANSITION PLAN

Since the schedule does not project the availability of the personal computer workstation and associated integrated system software in a form suitable for production use until August of 1985, an important element of the plan is the transition between the current timesharing systems and the integrated personal computing environment. The transition plan is based on early availability of personal computers and a network of more limited capability than that which will be supplied as part of the integrated personal computing environment. The objective for the transition plan is to gain experience with the use of personal computers in the university environment and to provide a vehicle for developing applications which can be moved with very little effort to the integrated personal environment when it becomes available.

The transition system equipment configuration will include personal computers, the existing central computing facility, and cluster local area networks. At least initially the number of applications executable on the personal computers will be relatively small.

4.1 PERSONAL COMPUTERS

The transition personal computers will be based on the Motorola 68000 microprocessor. They will have approximately one half million characters of local semi-conductor memory, a monochrome graphics display, and a local disk storage unit. The software for these personal computers will be a version of UNIX^{tm}. Each personal computer will be directly connected to the existing central computer facility through the same communications facility now used to accommodate computer terminals. Furthermore, it will connected to another personal computer that serves as a common data storage and file server utility.

4.2 NETWORK

As outlined above, the network for the transition computing environment actually will consist of two separate networks. The first will be the existing network already in place through which the personal computers will be attached to the existing central computing facility. That network consists of asynchronous 9,600 baud RS-232 EIA standard communication circuits. These communication circuits are connected to a large Micom port selection and contention unit which makes it possible to connect any of the circuits to any of the eleven computers in Carnegie-Mellon's central computing facility.

The second network will consist of a local connection operating at a much higher data rate between a group of ten to twenty transition personal computers and a file server. This file server will provide access to programs and data which need to be available to all of the personal computers in the cluster. Storage of data particular to the needs of any one user will be on the file of that user's personal computer. Thus, for instance the files required to support the software of the UNIX^{tm} operating system would reside on the cluster file server while the file which contained a document being revised by a particular user would reside on the disk file in the user's personal computer.

4.3 APPLICATIONS

The transition personal computing system has two important objectives. The first is to provide the university with additional computing in a fashion that begins to prepare the university for a personal computer based system. The second is to provide a base for development of new instructional applications which take advantage of the power of personal computing. In order to achieve the first objective several basic applications will be developed by the Computation Center and be available by August 1983. These applications include:

- Distributed editor -- One of the most expensive uses of the current central computers at Carnegie-Mellon is the editing of text. The editor used at Carnegie-Mellon, EMACS, is a very powerful full screen editor. Since it has to interpret every character the user inputs and very often has to update the terminal screen in major ways, it consumes large amounts of computing power and requires fast data communications facilities. Because the editor is extremely interactive, it does not perform well when the central computing facilities are heavily loaded. The plan is to provide a compatible editor in the personal computer workstation which automatically accesses a file on the central computer, does the editing on the personal computer, and then returns the file to the central computer. This strategy should result in better performance of the editor and more efficient use of the central computing facilities.

- Document Production -- Another heavily used facility on the central computers is the SCRIBE document production facility. Compatible systems with slightly less capability are available for personal computers and will be made available in the transition system. As a result, an author will be able to use the editor to produce the text of a document and the document formatter to examine drafts of that document as it is being edited. Only when the document is ready for final production will it need to be returned to the central computing facility to be printed on the high quality laser printers at that facility.

- Programming Languages — The Computer Science Department has developed an advanced programming language environment used to support beginning instruction in programming at Carnegie-Mellon University. That environment now operates on two medium-scale timesharing systems under the UNIX^{tm} operating system. It is expected that that environment will be transferred to the transition personal computers as soon as possible thereby freeing the timesharing systems for other uses and improving the service to students in programming instruction courses. In addition to the Computer Science PASCAL programming environment, programming language facilities for developing programs in FORTRAN and C will be available.

- File Transfer — A program will be available which allows files to be transferred between the transition personal computers and the central computers through the terminal network. It will, therefore, be possible to share information developed by a user at one personal computer with a user at another personal computer and it will be possible to develop input to applications that operate on the central computing facility using the editing capabilities of the personal computers.

- Terminal Simulation — A program will be written for the transition personal computers that allows them to accurately and efficiently simulate a computer terminal. As a result, any application not available on the transition personal computer but available on the central computing facility will be available to a user of the personal computer. The personal computer user will simply enter a command which causes the personal computer to simulate a terminal, and then access the central timesharing computing facility in the normal way. This terminal simulation capability is obviously a key element in the overall transition plan.

5. IMPLEMENTATION

5.1 STRATEGIES

By any standard, implementation of the integrated personal computing environment is a large task. The CMU/IBM plan for implementation places the responsibility for computing hardware and network development with IBM in the Communication Products Division. The software development effort is to be carried out jointly between IBM and CMU at Carnegie-Mellon University. The budgeted level of effort for the activity at Carnegie-Mellon is thirty-three people over a five year period. Of those thirty-three individuals, twenty will be professional software developers, and thirteen will be support and management. In total, then, CMU and IBM expect the software development effort to consume one hundred person years of professional software development effort.

Previous experience with efforts of this type have shown that it is just as difficult to produce a large computing software system with too many people as it is with too few. Nevertheless, the task of developing the software described above at this staff level will be challenging. Several specific strategies will be used to achieve success.

First, the overall development effort can be divided into three parallel tracks. As a result, the development group will be broken into three teams which work cooperatively but in parallel to achieve the overall schedule. One team will be responsible for the hardware dependent software and global utilities. Another team will be responsible for the user accessible utilities and application software. The third team will be responsible for the analysis, design, and installation of the interconnecting network.

Second, the system will be built by adapting, wherever possible, existing software. This software will come from two sources. The first source is the wide variety of UNIX^{tm} software that is used widely throughout the academic community. The second source is the Carnegie-Mellon University Computer Science Scientific Personal Interactive Computing Environment (SPICE) Project which has developed operating systems and global utilities code for a distributed system environment.

Based on these sources of existing software it will generally be possible to begin work with tested software and then move and adapt that software to the new hardware used by the project. Subsequently, the software will be tested in the new environment, released for limited use, and finally made generally available. In other words, we expect to do very little development of operating system software from scratch.

The third strategy is the utilization of an incremental development approach in which each important module of software is developed and tested insofar as possible independently of other modules. The result will be that difficulties in one area have minimal impact on other areas. In addition, this approach provides a continual flow of useful and usable output from the project to support parallel efforts to development educational and research applications.

5.2 SCHEDULE

While the overall schedule for the development and deployment of the integrated personal computing environment describes a continuous activity, it is best understood in terms of the calendar targets by which we expect to achieve the important goals.

- October 1982 -- Execute joint venture agreement with IBM.
- January 1983 -- 30% staffing including Director, Associate Director.
- March 1983
  - System development schedule completed.
  - 50% staffing.
  - Development computing system installed and operating.
  - Initial meeting of consortium members.
  - Social science evaluation work begins.
- September 1983
  - New building completed and occupied.
  - Stand alone personal computer workstation operational.
  - Operating system kernel and file server complete.
  - Editor and display manager usable.
  - PASCAL programming environment usable.
  - Specifications for help facilities, terminal emulation, and communications network final.
  - 100 transition personal computers installed for campus use.
  - Distributed editor and terminal simulator for transition personal computers in production use.
  - Transition personal computers available to consortium for application development.

May 1984

- Multiple workstations on local network operational.
- Print server, Fortran, C operational.
- Preliminary electronic mail available.
- Limited campus access to new system.

- September 1984
    - Multiple local networks on backbone network.
    - Instructional programming environment complete.
    - Command language complete.
    - Connectivity to central systems.
    - Refinement of consortium applications begins.
    - Begin moving instructional applications to new system.
    - 150 additional transition personal computers installed.
    - Operating system refinement begins.
    - Final system documentation effort begins.

- May 1985
    - Document production system complete.
    - Electronic mail complete.
    - Terminal simulation complete.
    - File transfer complete.
    - Help system complete.
    - Campus network installed.
    - New central computing facility installed.
    - Preliminary library system access.
    - Development of administrative applications begins.

- September 1985
    - 1000 advanced personal computer workstations installed.
    - 20 file servers installed.
    - 10 cluster facilities installed.
    - Library system access complete.
    - Test installations off campus.
    - Final system documentation complete.

September 1986

- Incoming students receive advanced personal computer workstations (1200 additional advanced personal computer workstations installed).
- 24 additional file servers installed.
- 20 cluster facilities installed.
- 200 off campus installations.
- Educational applications from transition system and consortium available.
- Pilot installations available to consortium.

- September 1987
    - Integrated personal computing environment available from IBM.

- Total population of advanced personal computer workstations reaches 3400 (more than the number of telephones at CMU).
- 20 file servers installed
- 10 cluster facilities installed.
- Total off campus installations reaches 600.

- October 1987 -- IBM Agreement period ends.