# PARALLEL PROCESSING

**Definition :** The parallel processing is an efficient form of information processing which emphasizes the exploitation of concurrent events in the computing process. Concurrency implies parallelism, simultaneity and pipelining. Parallel events may occur in multiple resources during the same time interval. Simultaneous events may occur at the same tine instant and pipeline events may occur in overlapped time spans.

## EVOLUTION OF COMPUTER SYSTEMS

Over the past decades the computer industry has experienced four generations of development. The vacuum tubes from 1940 – 1950, diode and transistors from 1950-1970, small and medium scale integration from 1960-1970 and very large scale integrations from 1970 and after etc.. Hence increases in device and speed and reliability and reductions in hardware cost and physical is enhancing the computer performance. But better devices are not the main factor contributing to high performance. Also stored program concept of Von Neumann has been recognized. A modern computer system is really a composite of such items like processors , memories, functional units, interconnection networks, compilers ,operating systems, peripheral devices , etc…

Hence to design a powerful and cost effective computer system and use it for efficient programs to solve a problem , the user must understand the hardware and software system and its structures. They develop a computing algorithms to be implemented on the machine with user oriented language. These discipline help the technical scope of the computer architecture. A good computer architect should master in all these disciplines.

## GENERATIONS OF COMPUTER SYSTEMS

Here we divide computer generations into five types as follows :

**(i) The first generation(1938-1953) :** The introduction of the first electronic analog computer in 1938 and the first electronic digital computer called electronic numerical integrator and computer (ENIAC) in 1946 was the marked the beginning of first generation computer. Electromechanical relays used as switching devices and vacuum tubes are also used. These devices were interconnected by the wires. Hardware components were expensive at this time. Here hardware components forced CPU structure to be a bit-serial. Arithmetic is done on a bit by bit floating point basis.

The binary coded machine language is used in early computers. In 1950 the first stored program computer EDVAC(electronic discrete variable automatic computer) was

developed. It uses the system software to relive user burden. By 1952, IBM had announced its 701 electronic calculator.

**(ii) The second Generation(1952-1946) :** Transistors were invented in 1948. The fist transistorized digital computer called TRADIC was developed in bell laboratories in 1954. Discrete transistors and diodes are used in TRADIC. At this time printed circuits are used and magnetic core memory was introduced. Assembly languages were used until the development of high level languages. Fortran in 1956 and Algol in 1960 was introduced.

In 1959, IBM and Sperry Rand started stretch project. These were first two computers attributable to architectural improvement. The Larc had an independent I/O processor which operated in parallel with one or more processing units. The first IBM scientific and transistorized computer called IBM 1620, was developed in 1959. COBOL was invented I 1959.

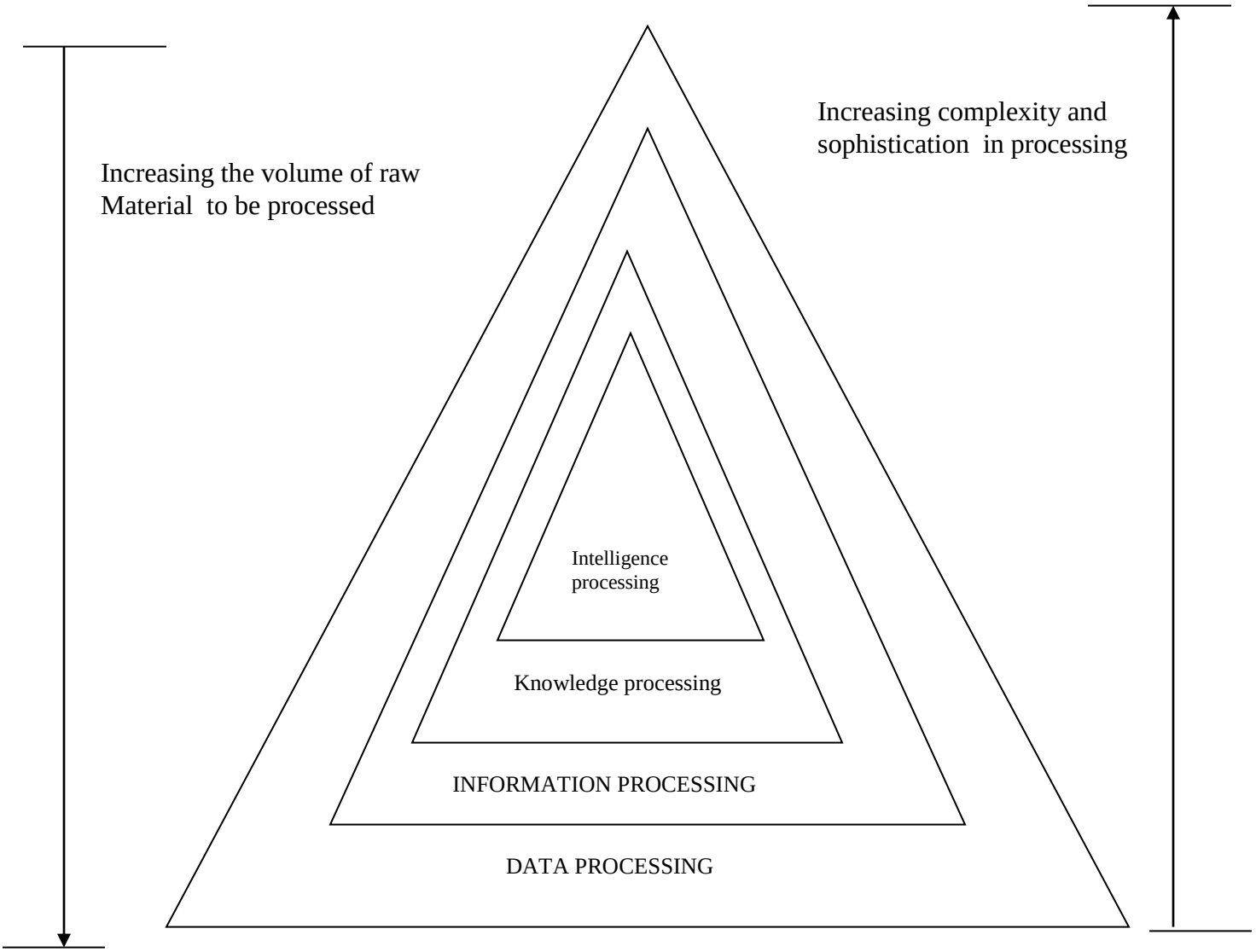
**(iii) Third Generation (1962-1975) :** This generation was marked by the use of small scale integration(SSI) and medium scale integration(MSI) circuits as basic building blocks. Multilayered printed circuits are also used. Core memory was used in CDC-6600. In 1968 many fast computers , like CDC-7600 is to replace the core with solid state memories. High level languages were greatly used with intelligent compilers at this period.

Multiprogramming was developed this time to allow the simultaneous access and execution of program segments. Many high performance computers like IBM 360/91, Illiac-IV, TI-ASC, Star-100 etc./. and several vector processors are developed. Virtual memory was developed at this period.

**(iv) Fourth Generation(1972-present) :The** present generation computers use large scale integration(LSI) circuits for both logic and memory sections. High density packaging is also used. High level languages are extended to handle both scalar and vector data. Most operating systems are time-sharing. They use virtual memory. High speed mainframes and super computes like Univac – 1100/80,Fujistu M 382, IBM 370/168 MP and IBM 3081, Burroughs B-7800 and Cray X-MP were invented. A high degree pipelining and multiprocessing is introduced in the commercial super computers. A massive parallel processor(MPP) was designed in 1982. This MPP contains 16,384 bit-slice microprocessor is under the control of one array controller for satellite image processing.

**(v) The Future :** Computers to be used in 1990's may be next generations. Very large scale integration(VLSI) chips is used along with design. Multiprocessors like 16 processors in S-1 project at Lawrence Livermore National laboratory is used. Cray-2 is using four processors, to be delivered in 1985. More than 1000 mega float point operations per second(Mega flops) are expected in super computers.

**TRENDS TOWARDS PARALLEL PROCESSING**




From an application point of view, the mainstream usage of computers in experiencing a trend of our ascending levels . They are

- Data processing
- Information processing
- Knowledge processing
- Intelligence processing.

Here the relation ship between the data, information , knowledge, and intelligence is as shown above. Here the data space is the largest including the numeric numbers, I

various formats, character symbols, multidimensional measures. Data objects are considered mutually unrelated in the space. Huge amounts of data are being generated daily in all walks of life, science and business.

An information is a collection of data objects that are related by some structure or relation . Therefore information in terms form a subspace of the data space. Knowledge consists of information items plus some semantic meanings. Thus knowledge items for a subspace of the information space. Finally the intelligence is there, it is derived from the collection of knowledge items. It is the innermost of the diagram.

Computer usages is started with the data processing, which is still a major task of computers. With more and more data structures developed many users are shifting to computer roles from pure data processing to information processing. Most of today's computing is still confined within these two processing levels. A high degree parallelism is found in these levels. From these levels they introduced a new level called knowledge processing level.

Today's computer can be made very knowledgeable but are far from being intelligent. Intelligence is very difficult to create, its processing is ever more. So today's computes are very fast and have many reliable memory cells to be qualified for data-information-knowledge processing. But none of the existing computers can considered as a intelligent thinking systems. Also computers are still unable to communicate with human beings in natural forms like speech and written languages, pictures, and images, documents and illustrations. Computers are far from being satisfactory in performing theorem proving , logical inference and creating thinking. Many scientists feel that the degree of parallelism is exploitable at the two high processing levels like intelligence and knowledge than data and information level.

From an operating point of view , the computer systems will fall in four phases in chronological order, they are

1. Batch processing
2. Multiprogramming
3. Time sharing
4. Multiprocessing.

In these four operating modes, the degree of parallelism increases sharply from phase to phase. The general trend is to emphasize parallel processing of information.

The highest level of parallel processing is conducted among multiple jobs or programs is through multiprogramming. , time sharing, and multiprocessing. This level requires the development of parallel processable algorithms. The implementation of parallel algorithms depends on the allocation of limited hardware and software resources to multiple programs being used to solve a large computation system. The next level of parallel processing is conducted among procedures or tasks with in the same program. The third level is to exploit concurrency among multiple instructions. Data dependency analysis is often performed to reveal parallelism among instructions. Vectorization may

be desired among scalar operations using DO loops. Hence to sum up these , the parallel processing can be challenged in four programmable levels , they are :

- Job or program level
- Task or procedure level
- Interinstruction level
- Intrainstruction level.

The highest level is often conducted algorithmically. The lowest intra-instruction level is implemented by the user hardware means. The hardware roles increase from high to low levels. Where as software implementations are increase from low to high levels. The trade-off between hardware and software approaches to solve a problem is always a controversial one. As hardware cost declines and the software cost increases This trend is supported by the increasing demand for a faster real-time , resource sharing and fault-tolerant computing environment.

The above characteristics suggest that parallel processing is a combined field of studies. It requires a broad knowledge , and experience in all aspects of algorithms, languages, software , hardware, performance evaluation, and computing. Hence to achieve parallel processing we require a lot of development with all above characteristics plus cost-effective computer programs.

## PARALLELISM IN UNIPROCESSOR SYSTEMS

Most general purpose uniprocessor system have the same basic structure. The development of parallelism in uniprocessor helps to increase the power and bandwidth of the computer, mechanisms etc.. Here we discuss about uni-processor architecture as follows :

### Basic Uniprocessor Architecture

A typical uniprocessor computer consists of three major components, they are main memory, central processing unit and I/O subsystem. The architectures of two commercially available uniprocessor computers are given below to show the possible interconnection of structures among the three sub systems. This diagram shows the architectural components of the super minicomputer called VAX-11/780. It is manufactured by Digital equipment company. The CPU contains the master controller of the VAX system. There are sixteen 32-bit general purpose registers, one of which serves as the program counter(PC).There is also a special CPU status register containing the information about the current status of the processor and the program which is to be executed. The CPU contains the arithmetic and Logic a unit(ALU) with optional floating pint accelerator and some local cache memory . The CPU is intervened by the operator through the console connected to a floppy disk. The CPU, the main memory and the I/O subsystem are all connected to common bus called synchronous backplane

interconnect(SBI). Through this bus , all I/O devices can communicate with each other with the CPU or with the memory. The peripheral storage or I/O devices can be connected directly to SBI through the unibus and its controller.



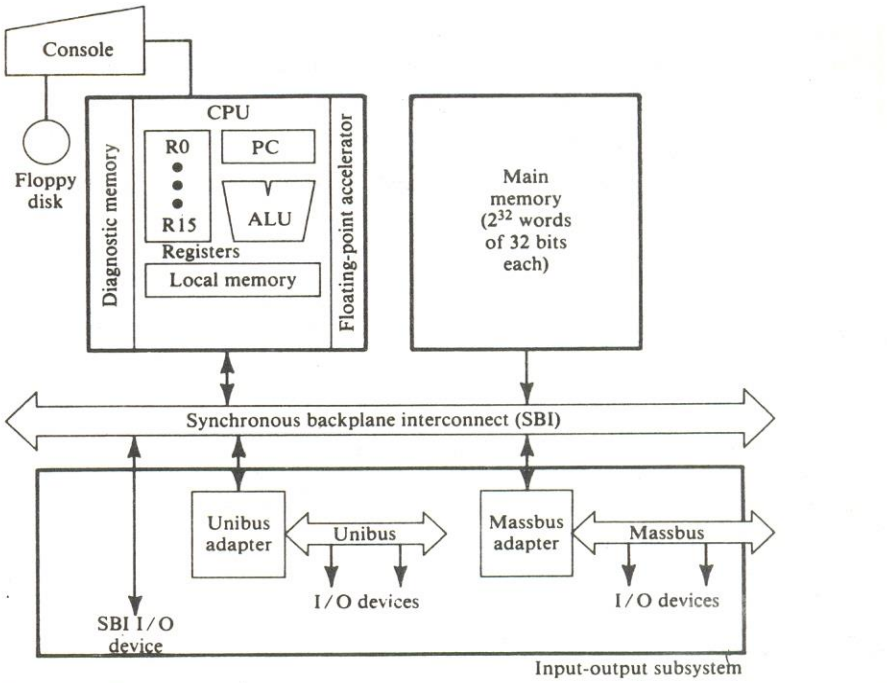



**Figure 1.3 The system architecture of the supermini VAX-11/780 uniprocessor system (Courtesy of Digital Equipment Corporation).**

## Parallel Processing Mechanisms

A number of parallel processing mechanism have been developed in uniprocessor computers. We identify them in the following six categories , they are :

- Multiplicity of functional units
- Parallelism and pipelining within the CPU
- Overlapped CPU and I/O operations
- Use of a hierarchical memory system.
- Balancing of subsystem bandwidths
- Multiprogramming and time sharing

## Multiplicity of functional units :

The early computer had only one arithmetic and logic unit in its CPU. The ALU could easily perform one function at a time, rather a slow process for executing a long

sequence of arithmetic logic instructions. In practice , many of the functions of the ALU can be distributed to multiple and specialized functional units which operate in Parallel. The CDC-6600 has ten functional units built into its CPU as in the diagram.

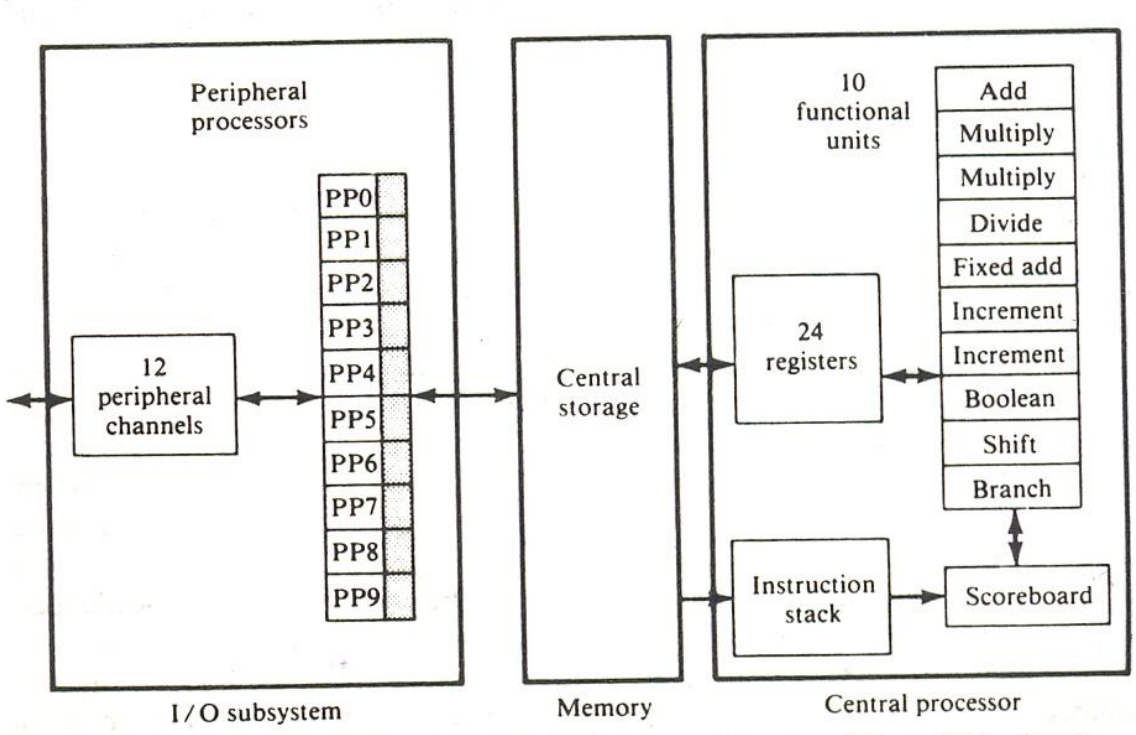



**Figure 1.5 The system architecture of the CDC-6600 computer (Courtesy of Control Data Corp.).**

**These** ten units are independent of each other and may operate simultaneously. A scoreboard is used to keep a tack of the availability of the functional units and registers being demanded. With 10 functional units and 24 registers available, the instruction issue rate can be significantly increased.

Another good example of a multifunction uniprocessor is IBM 360/91 ,which has two parallel execution units (E units ) one for arithmetic and another for floating point arithmetic. Within the floating point E unites there are two functional units, one for floating point add-subtract and another for floating point multiply-divide. The IBM 360/91 is a highly pipelined, multifunction, scientific processor.

**Parallelism and pipelining within the CPU :** Parallel adders, using such techniques called carry-lookahead and carry-save are not built into all ALUs. This is in contrast to the bit-serial adders used in the first generation machines. High speed multiplier recoding and convergence division are techniques for exploring parallelism and the sharing of hardware resources for the functions of multiply and divide.

Various phases of instruction execution are now pipelined that includes instruction fetch , decode, operand fetch and arithmetic logic execution, and store result. To facilitate overlapped instruction executions through the pipe, instruction prefetch and data buffering techniques are developed.

**Overlapped CPU and I/O operations :**

The I/O operations can performed simultaneously with the CPU computations by using a separate I/O controllers, channels or I/O processors. The direct memory access(DMA) channel can be used to provide direct information transfer between the I/O devices and the main memory . The DMA is conducted on a cycle stealing basis, which is apparent to the CPU. Also the I/O multiprocessing like the use of 10 I/O processors in CDC-6600 can speed up data transfer between the CPU and outside world.

**Use of hierarchical memory System**

**The** CPU is about 1000 times faster than memory access. A hierarchical memory system can be used to close up the speed gap. Computer memory hierarchy is as shown in the diagram.

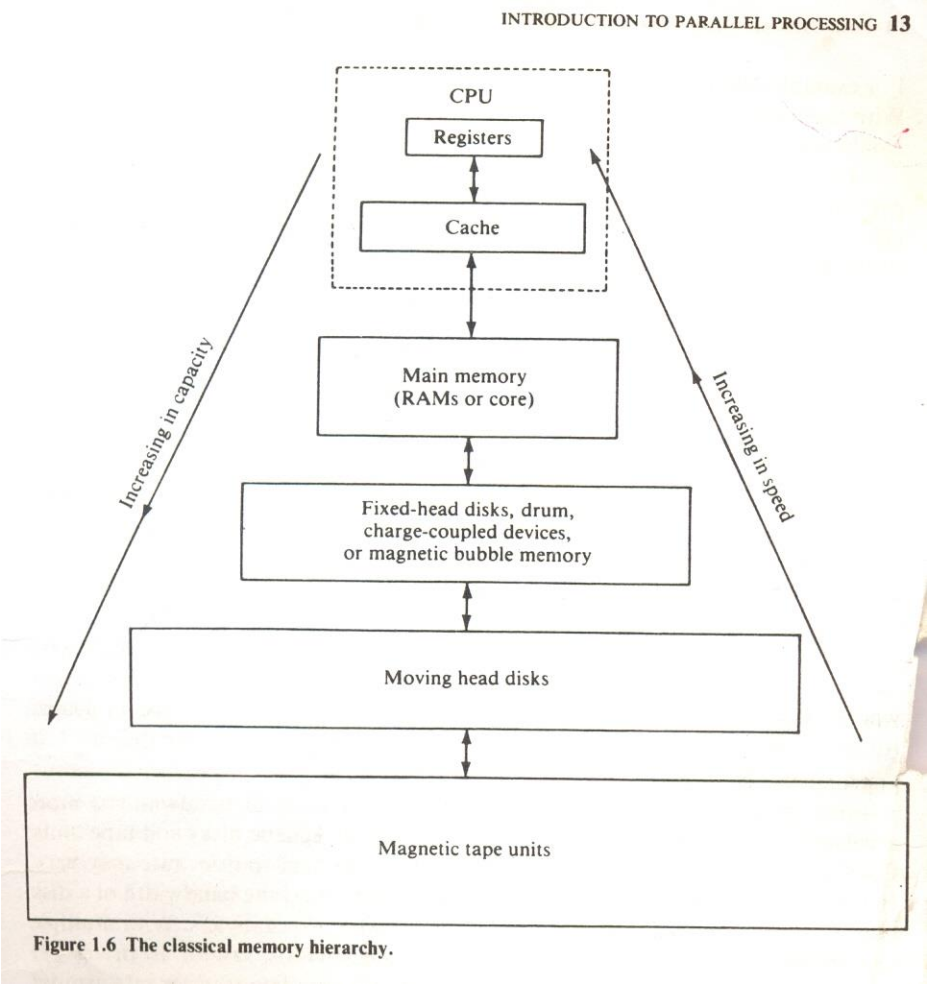


Figure 1.6 The classical memory hierarchy.

The innermost level is the register files directly addressable by the ALU. The cache memory can be used to serve as a buffer between the CPU and main memory. Block access of main memory can be achieved through multiway interleaving across parallel memory modules .

**Balancing Of subsystem Bandwidth**

In general, the CPU is the fastest unit in computer with a processor cycle of $t_p$ of tens of nanoseconds. The main memory has a cycle time $t_m$ of hundreds of nanoseconds and I/O devices are the slowest with an average of access time $t_d$ of few milliseconds. It is observed that

$t_d > t_m > t_p$

For example, the IBM 370/168 has $t_d$ of 5 ms , $t_m$ = 320 ns and $t_p$ = 80 ns. With these speed gaps between the subsystems, we need to match their processing bandwidth in order to avoid a system bottleneck problem.

The bandwidth of the system is defined as number operations performed per unit time. In the case of main memory system the memory bandwidth is measured by the number of memory words that can be accessed per unit time. Let W be the number of words delivered per memory cycle $t_m$ . Then the maximum memory bandwidth Bm is equal to

$B_m = W/ t_m$

The bandwidth of the processor is measured as the maximum CPU computation rate $B_p$ . For example it is 160 megaflops in the Cray- 1 and 12.5 million instructions per second in IBM 370/168.
Also the utilized CPU rate is $B^u_p \leq B_p$.

Hence the utilized rate is measured as $B^u_p = \frac{R_w}{T_p}$

**Bandwidth balancing between CPU and memory**

The speed gap between the CPU and the memory can be closed by using the fast cache-memory between them. The cache should have an access time of $t_c = t_p$ . A block of memory words is moved form the main memory into the cache so that immediate instructions or data can be available most of the time from the cache. The cache serves as a data or instruction buffer .

**Bandwidth balancing between Memory and I/O devices**

Input/output channels with different speeds can be used between the slow I/O devices and the main memory. These I/O channels perform buffering and multiplexing functions to transfer the data from multiple disks into the main memory by stealing cycles from the CPU. Further more intelligent disk controllers or database machines can be used to filter out the irrelevant data just off the tracks of the data. This filtering will easy the I/.O channel saturation problem. The combined buffering, multiplexing, and filtering operations can provide faster, more effective data transfer rate , matching that of the memory.

**Multiprogramming and Time Sharing**

When there is only one CPU in a uniprocessor system, we can still achieve a high degree of resource sharing among many user programs. The multiprogramming and time sharing are the software approaches to achieve the concurrency in a uniprocessor system. Here we use three notations like i , c and o which represents input , compute and output.

**Multiprogramming**

Within the same time interval, there may be multiple processes active in a computer. Competing for memory, I/O and CPU resources. We are aware of the fact that some computer programs are CPU bound and come are I/O bound. We can mix the execution of various types of the programs in the computer to balance bandwidth among the various functional units. The program interleaving is intended to promote better utilization through overlapping of I/O and CPU operations.

When ever a process called P1 is tied up with I//O operations, the system scheduler can switch the CPU to process P2. This allows the simultaneous execution of several programs in the system. When P2 is done, the CPU can be switched to P3. Note that the overlapped I/O and CPU operations and the CPU wait time are greatly reduced. This interleaving of CPU and I/O operations among the several program is called multiprogramming.

**Time Sharing**

Multiprogramming on a uniprocessor is centered around the sharing of the CPU by many programs. Sometimes , a high priority program may occupy the CPU for the long time to allow others to share. This problem can be overcome by the method called timesharing. It extends from the multiprogramming by assigning a fixed or variable time slices to multiple programs. Or equal opportunities are given to all programs competing for the use of the CPU.

The time sharing use of CPU by multiple programs in a uniprocessor computer creates a concept called virtual processors. The Time sharing is particularly effective when

applied to computer system which is connected to many interactive terminals. Each user at a terminal can interact with the computer . Time sharing was first developed for a uniprocessor system. It is also extended to multiprocessor system.

## PARALLEL COMPUTR STRUCTURES

Parallel computers are those systems that uses parallel processing. The basic features of parallel computers are listed below, they are

(i) Pipeline computers
(ii) Array processors
(iii) Multiprocessor systems.

A pipeline computer performs overlapped computations to exploit temporal parallelism. An array processor uses multiple synchronized arithmetic logic units to active spatial parallelism. A multiprocessor system achieves asynchronous parallelism through a set of interactive processors with shared resources.

### Pipeline Computers

The process of executing an instruction in a digital computer involves four steps, they are (i) Instruction Fetch from main memory, (ii)Instruction Decode(ID) to identify which operation is to be performed, (ii)Operand Fetch(OF), if needed in execution , (iv) Execution (EX) of the decoded arithmetic logic operations. In nonpipelined computers these four steps must be completed before the next instruction can be issued. But in a pipelined computer , successive instructions are executed in an overlapped fashion. The following diagram shows the process.

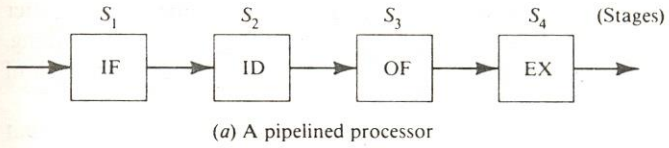
In this diagram the four pipeline stages IF,ID,OF and EX are connected or arranged in a linear cascade. The two space diagrams show the difference between overlapped execution and sequential non overlapped execution.

The instruction cycle consists of multiple pipeline cycles . A pipeline cycle can be set equal to the delay of the slowest stage. The flow of data from state to stage is triggered by a common clock of the pipeline. Also the operation of all the stages is synchronized under a common clock control. Interface latches are used between adjacent stags to hold the intermediate results. For the nonpipelined computer, it takes four pipeline cycles to complete one instruction. Once the pipelines is filled up the output result is produced from the pipeline on each cycle.

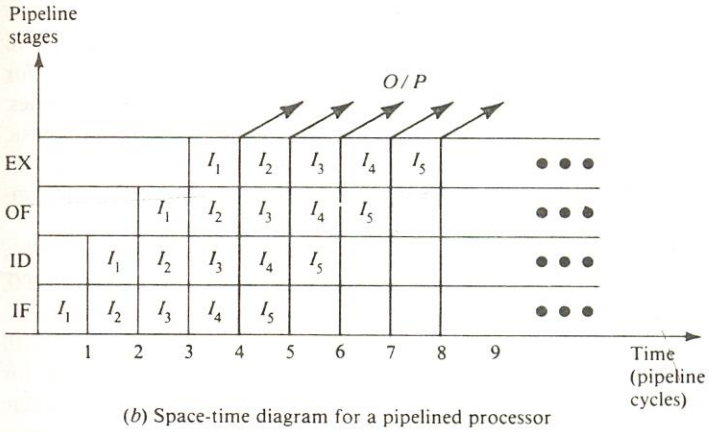
Due to the overlapped instruction and arithmetic execution it is obvious that pipeline computers are better tuned to perform the same operations repeatedly through the pipeline. Whenever there is change of operation we say that from add to multiply, the

arithmetic pipeline must be drained and reconfigured. Which causes extra delays . Hence pipeline computers are more attractive for vector processing.

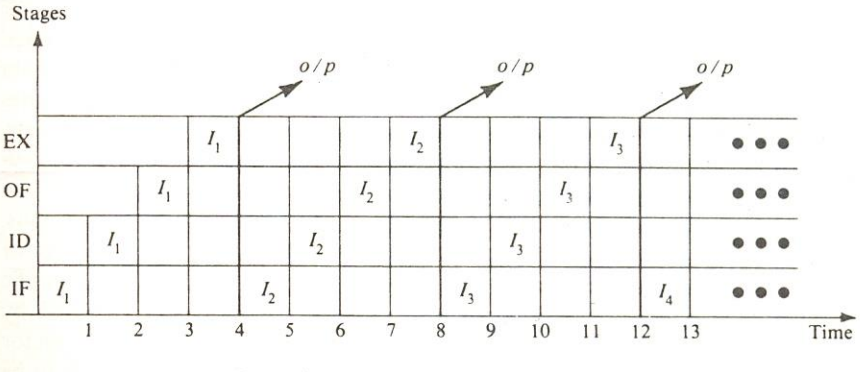





(*a*) A pipelined processor




(*b*) Space-time diagram for a pipelined processor




(*c*) Space-time diagram for a nonpipelined processor

**Figure 1.10 Basic concepts of pipelined processor and overlapped instruction execution.**

## Array Computers



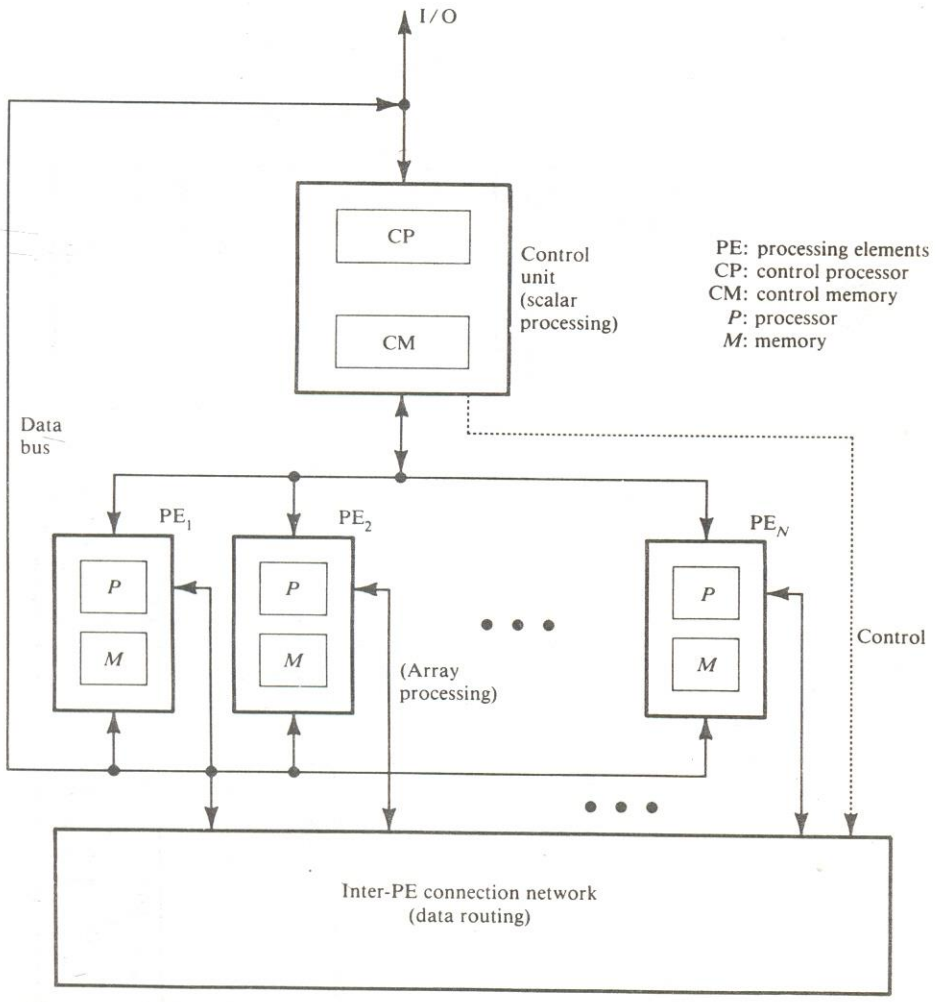



Figure 1.12 Functional structure of an SIMD array processor with concurrent scalar processing in the control unit.

**An array processor** is a synchronous parallel computer with multiple arithmetic logic units, called processing elements(PE). It can operate in parallel in a lock-step fashion. By replication of ALUs, one can achieve the spatial parallelism. The PEs are synchronized to perform the same function at the same time. An appropriate data routing mechanism must be established among the PE's.

A typical array processor is as shown in the diagram. Here scalar and vector type instructions are directly executed in the Control unit. Each PE consists of an ALU with registers and a local memory. The PE's are interconnected by a data-routing network. The interconnection pattern to established for specific computation is under program control . Vector instructions are broadcast to the PE's for distributed execution over

different component operands fetched directly from the local memory. The PE's are passive devices with instruction decoding capabilities.

Also the associative memory which is content addressable will be treated in the context of parallel processing. The array processors designed with associative memory is called as associative processors. The parallel algorithm on array processors will be given for matrix multiplication, merge, sort, and fourier transform.

## MULTIPROCESSOR SYSTEMS

The research and development of multiprocessor system are aimed at improving throughput, reliability, flexibility, and availability. The basic multiprocessor organization contains two or more processors of comparable capabilities. All processors share access to common sets of memory modules, I/O channels, and peripheral devices. Most importantly the entire system must be controlled by a single integrated operating system which provides interaction between processors and their programs. Besides the shared memories and I/O devices, each processor has its own local memory and private devices. The interprocessor communications can be done through the shared memories or through the interrupt network.

Multiprocessor hardware system organization is determined by the interconnection structure to be used between the memories and processors . The three different interconnections are

- Time-shared common bus
- Crossbar switch network
- Multiport switches.

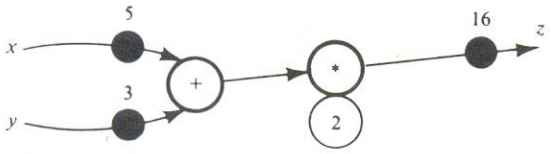
## Data flow and new concepts

**(i) Data flow computers :** The conventional von Neumann machines are called control flow computers. Because instructions are executed sequentially as controlled by a program counter. Sequential program execution is slow. To exploit maximal parallelism in a program the data flow computers are used. The basic concept of data flow computer is to enable the execution of an instruction whenever its required operands are available. Thus no, program counters are needed in data-driven computers. Instruction initiation depends on data availability, independent of the physical location of an instruction in the program. Also instructions in a program is not ordered.

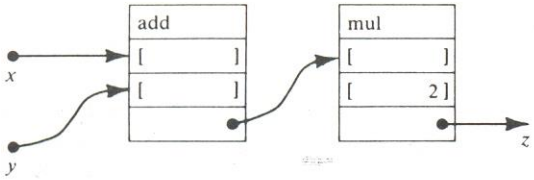
The execution follows the data dependency constraints. There is a maximum concurrency achieved in this computer.

Programs for data-driven computations can be represented by data-flow graph as in the following diagram.

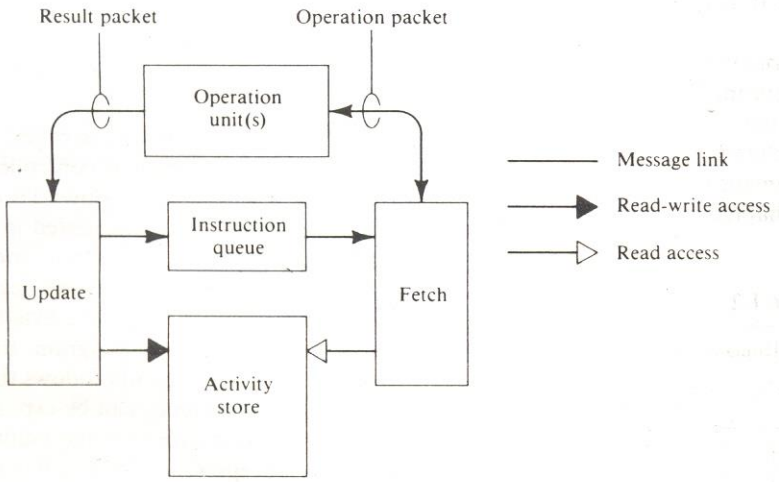





(*a*) Data flow program graph




(*b*) Template implementation




(*c*) Basic data flow mechanism

An example of data flow graph is as shown above. . To calculate the following expression the above graph is used

$Z = (x + y) * 2$

Here each instruction in a data flow graph is implemented as a template, which consists of an operator, operand receivers, and result destination. Operands are marked on the incoming arcs and results are defined as outgoing arcs. The template is also as shown above(b). Also the basic method of the execution of a data flow program is as shown above(c) . Here activity templates are stored in the activity store. Each activity template has a unique address which is entered in the instruction queue when the instruction is ready for execution. The instruction fetch and data access are handled by fetch and update units. The operation unit performs specified operation and generates the result to be delivered to each destination.

**VLSI computing Structures**

The advent of VLSI technology has created a new architectural horizon in implementing parallel algorithms directly into the hardware. The new high resolution lithographic technique is used to fabricate the $10^5$ transistors in an NMOS chip. It has been projected that by the late eighties it will be possible to fabricate VLSI chips which contain more than $10^7$ individual transistors. The use of VLSI technology in designing the high performance multiprocessors and pipelined computing devices is currently under investigation in both industrial and university environments. The multiprocessors are expected to be interconnected. Also pipelining makes it possible to overlap I/O with internal computations. Pipelined multiprocessing is the future of VLSI computing structure. Most proposed VLSI arithmetic devices are for vector and matrix type computations.

## ARCHITECTURAL CLASSIFICATION SCHEMES

There are three architectural classification schemes, they are Flynn's classification based on multiplicity of instruction streams and data streams. Feng's scheme bases on serial versus parallel processing . Handler classification is determined by the degree of parallelism and pipeline.

## MULTIPLICITY OF INSTRUCTION AND DATA STREAMS

In general digital computers may be classified into four categories, according to the multiplicity of instruction and data streams . This scheme for classifying computer organizations was introduced by Michael J.Flynn. The computing process is the execution of instruction on a set of data. The term stream is used here to denote a sequence of items as executed or operated upon a single processor. Instruction or data are defined with respect to the machine. An instruction stream is a sequence of instruction executed by the machine. The data stream is a sequence of data including input, partial and temporary results.

Computer organizations are characterized by the multiplicity of hardware provided to service the instruction and data streams. Here there are Flynn's four machine organizations :

- Single instruction stream – single data stream(SISD).
- Single instruction stream – multiple data stream(SIMD).
- Multiple instruction stream- single data stream(MISD).
- Multiple instruction stream – multiple data stream(MIMD).

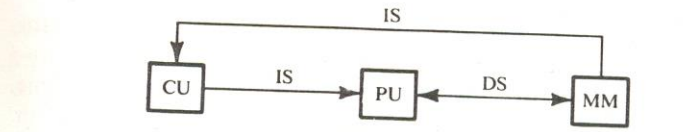
These organizations are illustrated in the following diagram. Here only three types of system components are needed in the illustration Both instructions and data are fetched from the memory module. The instructions are decoded by the control unit, which sends the decoded instruction stream to the processor units for execution. Data streams flow between the processors and the memory is in bidirectional manner. Also multiple memory modules may be used in the shared memory sub system. Here we define above four classes

**SISD COMPUTER ORGANIZATION :** This organization in the above diagram(a) represents most serial computers used today. Instructions are executed sequentially but my be overlapped in their execution Most SISD uniprocessor systems are pipelined. An SISD computer may have more than one functional unit. These functional unit is under the supervision of the one control unit.

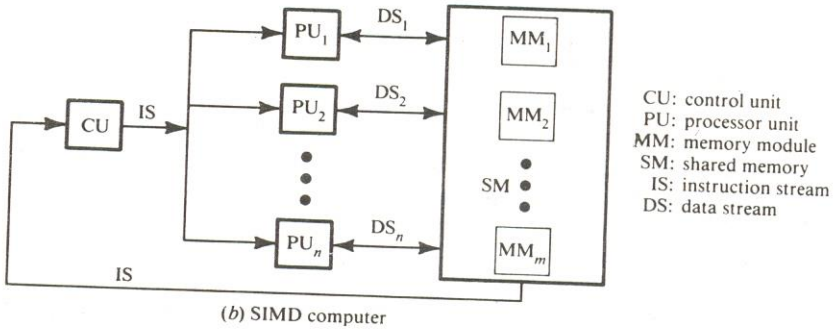
**SIMD COMPUTER ORGANIZATION : This class corresponds** to the array processors (diagram b) Here there are multiple processing elements which is supervised by the same control unit. ALL PE's receive the same instruction broadcast from the control unit but operate on different data sets from data streams. The shared memory subsystem may contain multiple module.

**MISD COMPUTER ORGANIZATION :** This organization is conceptually illustrated in the diagram(c) . Here there are n processor units, each receiving distinct instructions operating over the same data stream. The results of the one processor become the input of the next processor in the macro pipe. This structure receive less attention and has been challenged as impractical in some application.

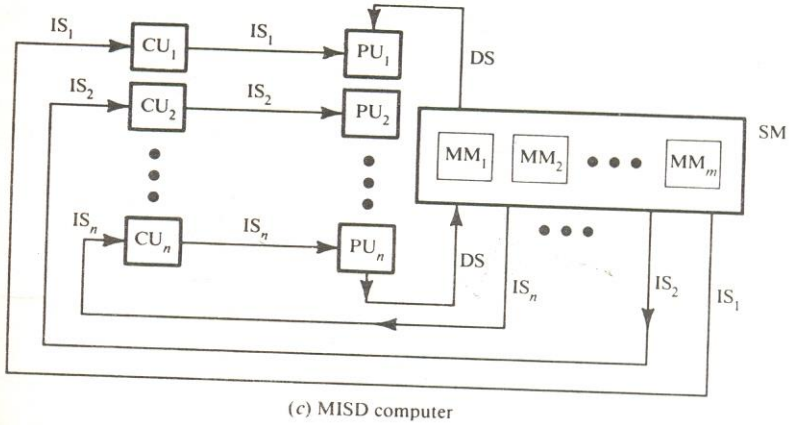
**MIMD COMPUTER ORGANIZATION :** Most multiprocessor systems and multiple computer systems are come under this category. An intrinsic MIMD computer implies interactions among the 'n' processors because all memory stream are derived from the same data space shared by all processors. If the 'n' data streams were derived from disjointed subspaces of the shared memories. Then we call it as multiple SISD(MSISD). An intrinsic MIMD computer is a tightly coupled if the degree of interactions among the processors is high. Otherwise we call it as loosely coupled.



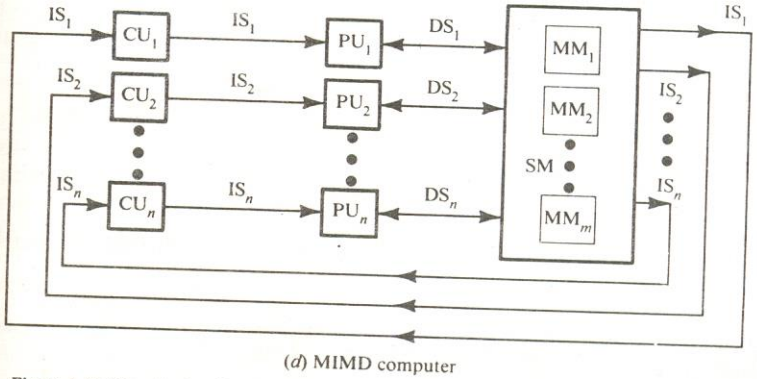

(*a*) SISD computer




(*b*) SIMD computer




(*c*) MISD computer




(*d*) MIMD computer

**Figure 1.16 Flynn's classification of various computer organizations.**

## Serial Versus Parallel Processing

Tse-yun Feng has suggest the use of degree of parallelism to classify various computer architectures. The maximum number of binary digits (bits) that can be processed within a unit time by a computer system is called the maximum parallelism degree P. Let Pi be the number of bits that can be processed within the ith processor cycle. Consider T processor cycles indexed by I = 1,2, …. T , the average parallelism degree Pa is defined as

$$Pa = \frac{\sum Pi}{T}$$

In general $Pi \leq P$.

If the computing power of the processor is fully utilized then we have Pi = P . the utilization rate depends on the application program being executed.

From the diagram(not in our document) which demonstrates the classification of computers by their maximum parallelism degree. The horizontal axis shows the word length n. The vertical axis correspond to the bit-slice length m. Both these lengths measures are in terms of the number of bits contained in a word in a bit slice. A bit slice is a string of bits, one from each of the words at the same vertical bit position.

Here we find maximum parallelism degree P(C) of a given computer system C which is represented by the product of word length 'n' and bit-slice length ,m'

$$P(C) = n.m$$

The pair (n,m) corresponds to the computer space in the co-ordinate system(refer diagram)

From this diagram we see there are four processing methods, they are

- ✓ Word serial bit serial(WSBS)
- ✓ Word parallel and bit serial(WPBS)
- ✓ Word serial and bit parallel(WSBP)
- ✓ Word parallel and bit parallel(WPBP).

WSBS has been called as bit – serial processing because it process one bit at a time(n=m=1). WPBS(n=1 ,m >1) is called as bis(bit slice) processing because m-bit slice is processed at a time. The WSBP(n >1, m=1) is called as word – slice processing and mostly used because one word of n bits is processed at a time. Finally WPBP(n >1,m>1) is a fully parallel processing in which an array of n.m bits is processed at one time, the fastest processing.

**PARALLELISM VERSUS PIPELINING**

Wolfgang handler has proposed a classification scheme for identifying the parallelism degree and pipelining degree built into the hardware structures of a computer system. He considers parallel-pipeline processing at three subsystem levels, they are :

- Processor control Unit(PCU)
- Arithmetic Logic Unit(ALU)
- Bit-level circuit(BLC).

The functions of PCU and ALU should be clear. Each PCU corresponds to one processor or one CPU. The ALU is equivalent to processing element. The BLC corresponds to the combination logic circuit which needed to perform the 1–bit operations in the ALU.

A computer system C can be characterized by triple containing six independent entities they are :

T ( c ) = < K x K' , D x D' , W x W'>

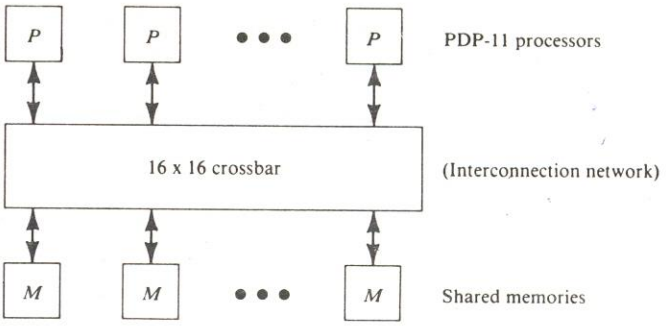
Where K – the number of processors (PCU) with in the computer.
D- the number of ALU's under the control of one PCU.
W- The word length of ALU or PE.
W' – The number of pipeline stages in all ALU's or in a PE.
D' – the number of ALU's that can be pipelined.
K' – the number of PCU's that can be pipelined.

Several real time computer examples are used to clarify the above parameter descriptions. Here we see on example as follows The Texas instrument's advanced scientific computer(TI-ASC) has one controller for controlling four arithmetic pipelines, each has 64-bit word lengths and eight stages. Hence we have

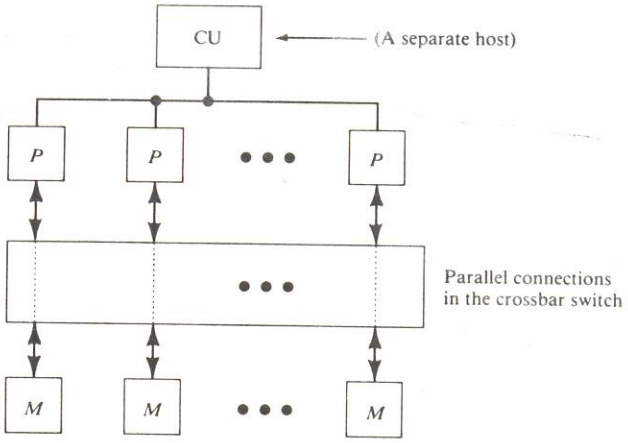
TI(ASC) = < 1 x 1 , 4 x 1 , 64 x 8 > = < 1,4, 512 >

Here we have a sample system which is the C.mmp multiprocessor system developed at Carnegie-Mellon university. The system is used in number of ways. The system consists of 16 PDP-11 minicomputers of a word length of 16 bits. It is operate in MIMD mode. They are also operate in SIMD mode. It provides all the minicomputers that are synchronized by one master controller as in the diagram. These systems are rearranged to operate in a MISD mode also in the diagram. Hence based on these three configurations we have computing system which has three parts having operator + to separate them.

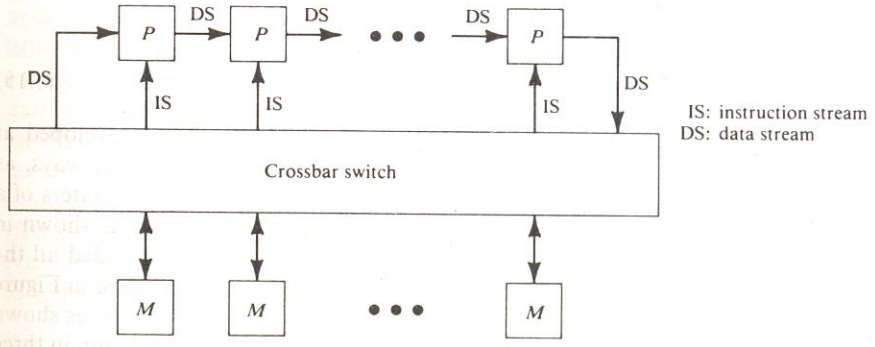
T(C.mmp) = < 16 , 1, 16 > + < 1 X 16 , 1 , 16 > + < 1 , 6 , 16 >




(*a*) $T(16, 1, 16)$ for MIMD mode




(*b*) $T(1, 16, 16)$ for SIMD mode




(*c*) $T(1 \times 16, 1, 16)$ for MISD mode

**Figure 1.18 Operation modes in C.mmp system (all double-arrowed paths are for both IS and DS).**

## PARALLEL PROCESSING APPLICATIONS

Fast and efficient computers are in high demand in many scientific, engineering , energy, resource, medical, military, artificial intelligence and other basic areas. Large-scale computations are often performed in these application areas. Parallel processing computers are needed to meet these demands. To design a cost effective super computer or to better utilize an existing parallel processing system, one must first identify the computational needs of important applications. Large-scale scientific problem solving involves three interactive disciplines , they are theories, experiments, and computations as in the diagram.



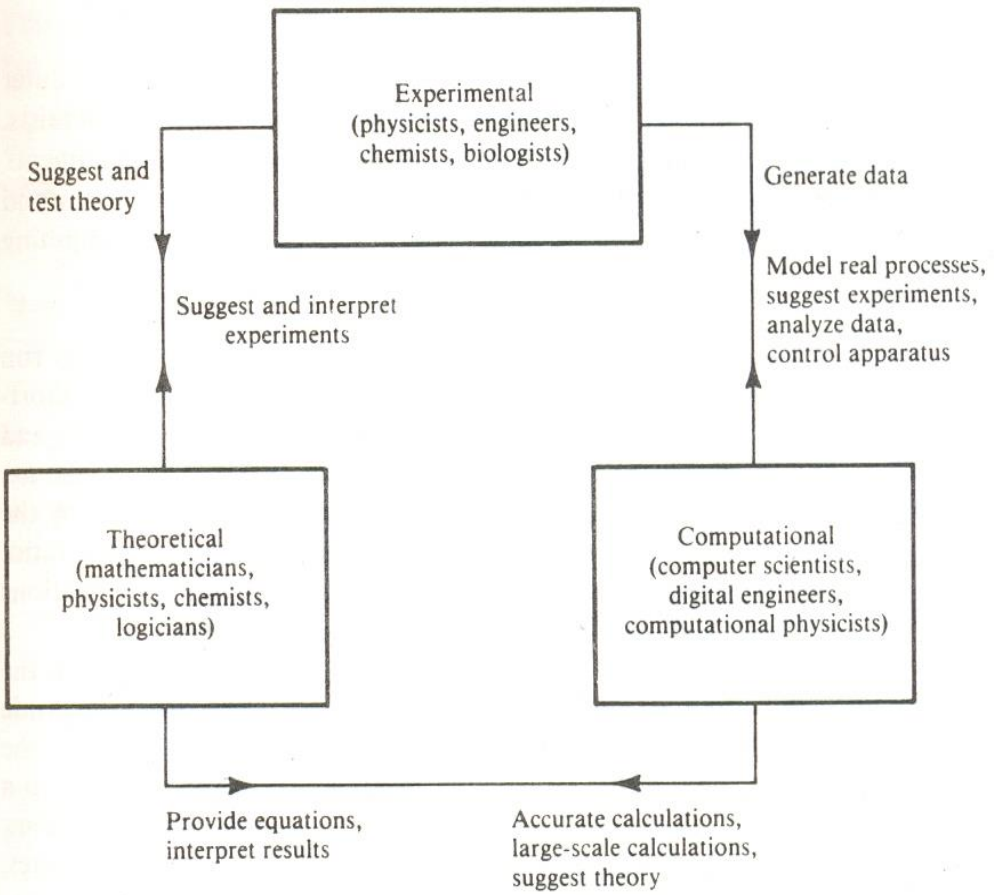



**Figure 1.19 Interaction among experiments, theories, and computations to solve large-scale scientific problems (Courtesy of Rodrique et al., *IEEE Computer*, 1980).**

Here theoretical scientists develop mathematical models that computer engineers solve numerically. The numerical results may often suggest new theories. Experimental

science provides data for computational science. Using computer simulations which has several advantages

1. Computer simulations are far cheaper and faster than physical experiments.
2. Computers can solve a much wider range of problem than specific laboratory equipments.
3. Computational approaches are only limited by computer speed and memory capacity, while physical experiments have many practical constraints.

Theoretical and experimental scientists are users of large program codes provided by the computational scientist. The codes should produce accurate results with minimal user effort. The computer scientists must apply advanced technologies in numerical modeling, weather forecasting, software development. Here we see some applications. as follows :

## PREDICTIVE MODELING AND SIMULATIONS

Multidimensional modeling of the atmosphere, the earth environment , outer space and world economy has become a major concern of world scientists. Predictive modeling is done through extensive computer simulation experiments, which often involve large-scale computations to achieve the accuracy and time. The examples of predictive modeling are :

**(i)** **Numerical weather forecasting :** Weather and climate researchers will never run out of their need for faster computers. Weather modeling is necessary for short range forecasts and for long range hazard predictions like flood, drought and environment pollutions. The weather analyst needs to solve general circulation model equations with the computer. The atmospheric state is represented by the surface pressure , wind field, and temperature, so on. The computations is carried out on a three dimensional grid that partitions the atmosphere vertically into K levels and horizontally into M intervals of longitude and N intervals of latitude . Here a fourth dimension is added as the number P of time steps used in the simulation. Using a grid with 270 miles on a side, a 24-hour forecast would need to perform about 100 billion data operations. Increasing the forecast by halving the grid size on all four dimensions would take the computation at least times longer. The 100 megaflops machine like a Cray-I would take 24 hours to take forecast.

**(ii)** **Oceanography and astrophysics : Since** oceans can store and transfer heat and exchange it with the atmosphere, a good understanding of the oceans would help in the following areas :

- Climate predictive analysis
- Fishery management
- Ocean resource exploration
- Coastal dynamics and tides.

Oceanographic studies use grid size on a smaller scale and a time variability on a larger scale than those used for atmospheric studies. To do a complete simulation of the Pacific ocean with adequate resolution.

The formation of the earth from planetestimals in the solar system can be simulated with a high speed computer. The dynamic range of astrophysics studies may be from billions of years to milliseconds . Interesting problems include the physics of supernova and dynamics of galaxies.

**(iii) Socio economics and government use : Large** computers are in great demand in the areas of econometrics, social engineering , government census, crime control, and the modeling of the world economy for the year 2000. In the United States, the FBI uses large computers for crime control, the IRS uses a large number of fast mainframes for tax collection and auditing. There is no doubt about the use of supercomputers for the national census and general public opinion polls. It was estimated that 57 percent of the large scale computers manufactured in the US have been used the US government ,

**Engineering design and Automation**

Fast supercomputers have been in high demand for solving many engineering design problems, such as the finite element analysis needed for structural designs and wind tunnel experiments for aerodynamic studies. Also industrial development like the use of computers to advance automation, artificial intelligence and remote sensing as follows :

- **Finite element analysis - The design of dams** , bridges, ships , supersonic jets, high buildings, and space vehicles requires the resolution of a large system of algebraic equations or partial different equations. Conventional approaches using pre developed software packages require intolerable turnaround times. Many researchers and engineers have attempted to build more efficient computers to perform finite element analysis or to seek finite different solutions. Computational engineers have developed finite-element analysis code for the dynamic analysis of structures. High-order finite elements are used to describe the spatial behavior. Vectorization procedures can be used to generate the element stiffness and mass matrices.
- **Computational aerodynamics – Large scale** computers , have made significant contributions in providing new technological capabilities and economics in pressing ahead with aircraft and spacecraft lift and turbulence studies. NASA's Ames Research Center is seeking a supplement to Illiac-IV to do three dimension simulations of wind tunnel tests at gigaflop speeds. Two gigaflops supercomputers, known as Numerical Aerodynamics simulation facilities(NASF) have been proposed by Burrough's corporation and by the control data corporation.
- **Artificial Intelligence and automation -** Intelligent I/O interfaces are being demanded for future supercomputers that must directly communicate with human beings in images, speech and natural languages. Here we have some lit of

intelligence which demands parallel processing , they are (i) Image processing, (ii) Pattern recognition, (iii) Computer vision, (iv) Speech understanding, (v) Machine inference, (vi) CAD/CAM/CAI/OA,(vii) Intelligent robotics, (viii)Expert computer systems, (ix) Knowledge engineering.

Special computer architectures have been developed for some above machine intelligence applications. Japan launched a national project to develop the fifth generation computers to be used in 1990's. The Japanese predict this new generation to possess highly intelligent input-output systems . The projected computing power of a system being developed is 100 mega to 1 giga logical inferences per second(LIPS). The time to execute one logical inference equals to executing of 100 to 1000 machine instructions. This machine should be able to execute 10,000 to 1 mega million instructions per second.

- **Remote sensing applications :** Computer analysis of remotely sensed earth-resource data has many potential applications in agriculture, forestry, geology, and water resources. Explosive amounts of pictorial information need to be processed in this area. For example a single frame of LANDSAT imaginary contains 30 billion bytes, it takes 13 such images to cover the state of Alabama. NASA has ordered a massively parallel processing(MPP) for earth resources satellite image processing. The MPP has a computing rate of 6 billion 8-bit integer operations per second.

**Energy Resource Exploration**

Energy affects the process of entire economy on a global basis. Computers play an important role in finding the oil and gas and the management of their recovery. Here we see some applications they are :

- **Seismic Exploration :** Many oil companies are investing in the use of attached array processors or vector supercomputers for seismic data processing, which accounts for about 10 percent of the oil finding costs. Seismic exploration sets of a sonic wave by explosive or by jamming a heavy hydraulic ram into the ground and vibrating it in a computer controlled pattern. A few thousand phones scattered about the spot are used to pick up the echos. The echo data are used to draw two-dimensional cross sections that display the geometrical underground level. A typical field record for the response of the earth to one sonic input has 3000 different time values each at about 48 different locations. This produces about 2 to 5 million floating point numbers per kilometer along a survey line.
- **Reservoir modeling:** Supercomputers are being used to perform three dimensional modeling of oil fields. The reservoir problem is solved by using the finite difference method on the three dimensional representation of the field. Geologic core samples are examined to project forward into time the field's expected performance. Presently at least 1000 flops needs to be processed per data point in the three-dimensional model of an oil field. This means the

superpower computer must be employed to achieve an accurate performance evaluation in a reasonable time period for the large field.

- **Plasma fusion power :** Nuclear fusion researchers are pushing to use a computer 100 times more powerful than any existing one to model the plasma dynamics in the proposed Tokamak fusion power generator. Magnetic fusion research programs are being aided by vector supercomputers at the Lawrence Livermore national laboratory . The potential for magnetic fusion to provide an alternate source of energy has become closer as a result of the cooperative effort of the experimental program with the computational simulation program. Also synthetic nuclear fusion requires the heating of plasma to a temperature of 100 million degrees. This is the costly effort. The high temperature plasma consisting of positive charge ions and negative charged electrons is confined.
- **Nuclear Reactor Safety –** Nuclear reactor design and safety control can both be aided by computer simulation studies. The studies are , (i) On-line analysis for reactor condition, (ii) Automatic control for normal and abnormal operations, (iii) Simulation of operator training, (iv) Quick assessment of potential accident procedures. To implement these operations we use TRAC code which is developed to simulate the non-equilibrium , non-homogeneous flow of high temperature water and steam.

**Medical, Military and Basic Research**

**In the** medical area , fast computers are needed in computer assisted tomography, artificial heart design, liver diagnostics, brain damage estimation and genetic engineering. Also military defense is used by supercomputers for weapon design , simulation etc.. Here se see some applications :

- **Computer Assistant Tomography -** A human body can be modeled by computer assisted tomography(CAT) scanning. The Mayo clinic in Rochester , Minnesota is developing a research CAT scanner for the three dimensional , stop action , cross action viewing of the human heart. Also Courant Institute of Mathematical sciences , research scientists are using the array processor for three dimensional modeling for the blood flow in heart. Cross sectional CAT images used to take 6 to 10 minutes to generate on a computer. Using the dedicated array processor, the processing time can be reduced to 5 to 20 s. The image reconstruction of human anatomy in present CAT scanners is two dimensional.
- **Genetic Engineering –** Biological systems can be simulated on supercomputers . Genetic engineering in advancing rapidly in recent years . There is a growing need for large-scale computations to study molecular biology for the synthesis of complex organic molecules, like proteins . Crystallography is also used . A highly pipelined machine, called Cytocomputer has been developed at the Michigan Environmental Research Institute for biomedical image processing. It can be used to search for genetic mutations.

- **Weapon research and defense -** Military research agencies have used the majority of the existing supercomputers. The first Cray-I was installed at the Los Alamos Scientific Laboratories in 1976. By 1981 , four upgraded Cray – I had been acquired by Los Alamos. The following is the defense – related military applications of supercomputers. They are

  - ✓ Multiwarhead nuclear weapon design (Cray-I)
  - ✓ Simulation of atomic weapon effects by solving hydro dynamics and radiation problems.
  - ✓ Intelligence gathering , like radar signal processing on the associative processor for the antiballistic missile.

- **Basic Research Problems -** Many of the application areas are related to basic scientific research. Here below some several additional areas that demand of the use of super computers. **They are :**
  - Computational chemists solve problems on quantum mechanics , statistical mechanics, polymer chemistry , and crystal growth.
  - Computational physicists analyze particle tracks generated in spark chambers , study fluid dynamics , examine quantum field theory and investigate molecular dynamics.
  - Electronic engineers solve large-scale circuit equations using the multilevel Newton algorithm, and layout the VLSI connections.