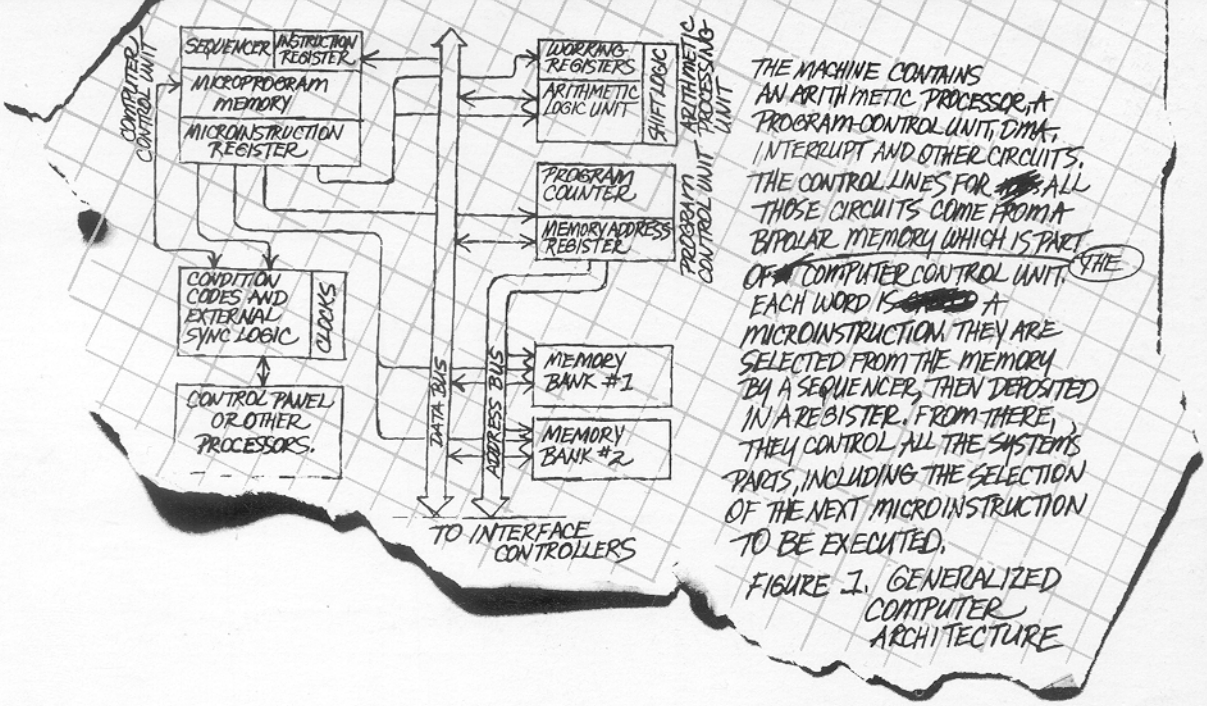


Build A Microcomputer

Chapter 1
Computer Architecture

Advanced Micro Devices

AM-PUB073-1

PREFACE

In this introductory Chapter we intend to:

1). develop a common terminology for future chapters.
2). introduce several stored-program-computer design topics.
3). define some of the computer architect's problems (which will be solved in the subsequent chapters).

In order to achieve these goals, we will start with computer basics. It should be stressed that approaches and solutions can be chosen which are different from the ones described in this and the subsequent chapters. However, the general ideas described will be appropriate to gain familiarity with the micro-programmable bit-slice devices in order to use them in any design configuration.

BACK TO THE BASICS. . .

A STORED-PROGRAM-COMPUTER is defined as a machine capable of manipulating data according to predefined rules (instructions), where the program (collection of instructions) and data are stored in its memory (Fig. 1). Without some means of communication with the external world, the program and the data cannot be loaded into the memory nor can the results be read out. Therefore, an input/output device is required as shown in Fig. 2.

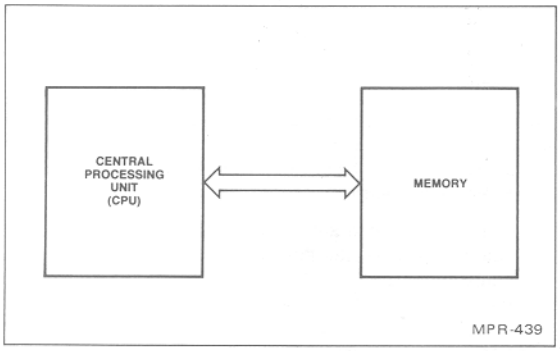


Figure 1. Basic Definition of a Stored-Program-Computer.

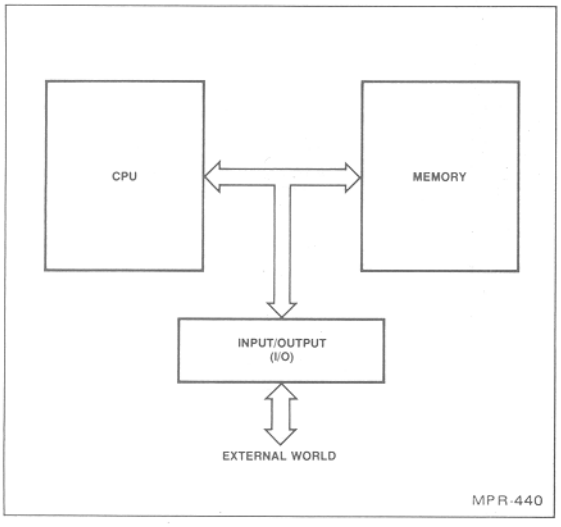


Figure 2. I/O Added to the Basic Stored-Program Computer.

The memory is usually organized in words, each containing N bits of information. A unique address is allocated for each word which defines its position relative to other words. The Central Processor Unit (CPU) usually reads or writes one word at a time by addressing the memory and then when the memory is ready, reading the contents of the word or writing new contents into that word. To perform this operation, two registers are usually used: The Memory Address Register (MAR), which contains the address and the Memory Data Register (MDR) which contains the data (Fig. 3).

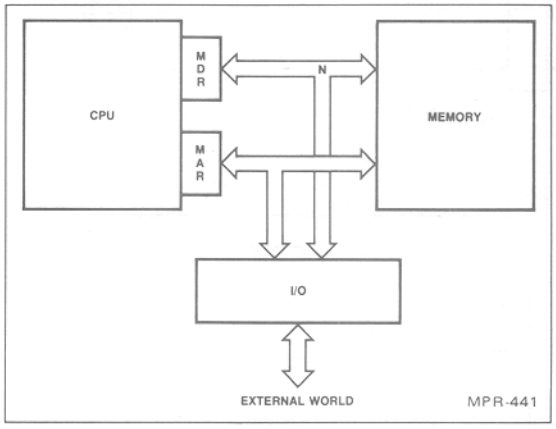


Figure 3. MAR and MDR Depicted for a Stored-Program Computer.

Since accessing a memory (reading from it or writing into it) is usually a relatively slow procedure, it is advantageous to have a few memory locations inside the CPU which can be read from or written into very fast. These locations are usually called Accumulators or Working Registers. Having these fast access registers inside the CPU (Fig. 4) enables many operations to be carried out without referring to the memory (through the MAR and the MDR) and therefore these operations are executed faster.

The unit which actually performs the data manipulation is called the Arithmetic & Logic Unit (ALU). It has two inputs for operands and one output for the result. It usually operates on all the bits of a word in parallel. The ALU can perform all or part of the following operations:

Arithmetic	Logical
Add	OR
Complement	AND
Subtract	XOR
Increment	NAND
Decrement	NOR
	XNOR
	Complement

In some architectures, one of the operands must always be in a special register (accumulator) and the result of the ALU operation is always transferred to this register. In a more general CPU, any two of the internal registers can contain the operands and the result of the ALU operation can be transferred to any one of them.

Another very useful feature of a CPU is the ability to shift the contents of a register or the output of the ALU one or more bits in either direction as shown in Fig. 5.

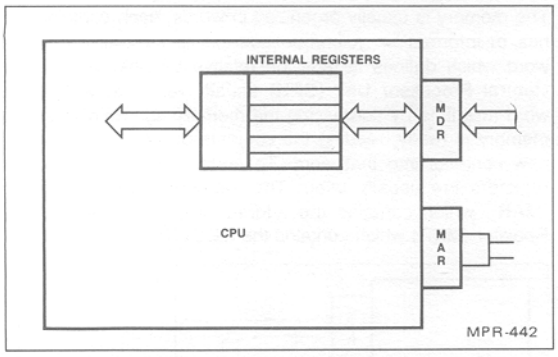


Figure 4. CPU with Internal High Speed Registers.

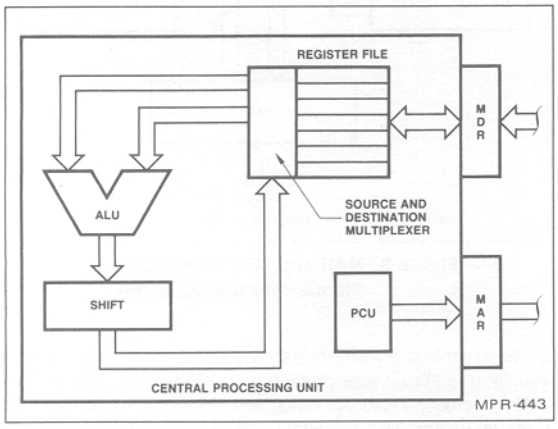


Figure 5. ALU and Shifter Added to the CPU Design.

We now have the elements to do any data manipulation required but we still need a unit which can properly set the MAR in order to find the next instruction of the program in the memory and to find its associated data. This unit is called the Program Control Unit (PCU) and its role is to load the MAR with the correct address in order to find the next instruction or data item or to point to a memory location where a data word should be written.

Often, the program steps (instructions, data) are written in the memory in consecutive locations, starting at address zero or at any other predefined address. The PCU can simply be incremented after each memory access thereby pointing to the address of the next instruction or data item. This counter-type PCU has very little flexibility. Sometimes we wish to change the "normal" flow of the instructions, particularly if we want to enable our computer to "make decisions" according to conditions prevailing at the current execution point. For example, we may want to execute one of two different sequences of instructions depending upon the result of the last operation performed. This is accomplished by loading the MAR with a new value (the address of the next instruction to be executed) rather than incrementing it. This operation is called a BRANCH or JUMP and can be unconditional (which allows execution of a non-contiguous string of instructions) or conditional (depending, for example, on whether the last operation's result was zero or not, was negative or positive, true or false, etc.).

Even more flexibility can be achieved by using a stack (a group of temporary internal or external memory locations) to store vital data. A stack pointer is used to address the memory location currently at the top of the stack. Indirect and relative addressing and other sophisticated addressing modes (all of which can be handled by the PCU) will be discussed later. Meanwhile, Fig. 5 shows the PCU as a part of the CPU.

Executing an instruction in our computer now requires the following steps:

a). The PCU loads the address of the next instruction to the MAR and signals to the memory that a Read is requested. Incidentally, the PCU may be as simple as a Program Counter equal to the address width. The memory loads the MDR with the contents of the location addressed.
b). The CPU decodes the instruction: i.e., (assuming operands are in internal registers) selects the proper registers to feed the ALU, selects the proper function to be performed by the ALU, sets up the shifter to displace the result, if required, and selects the register in which the result should be stored.
c). The ALU performs the function desired.
d). The result is loaded into the destination register.
e). The result is also examined to determine whether a BRANCH is to be performed.
f). The PCU calculates the address of the next instruction, (usually called a "FETCH").

This procedure becomes more complicated if the operands are not stored in the internal registers or if the result is not to be stored in one of them. Let's take an example instruction using relative addressing:

> "Take the first operand from the location specified by the sum of the word after this instruction (immediate) and the contents of register R1; take the second operand from the location specified by the sum of the second word after this instruction and the contents of R2; add the two operands and place the result in the location specified by the sum of the third word after this instruction and the contents of register R3. Then execute the instruction located at the address, which is the sum of the fourth word after this instruction and the contents of register R4 if there is a carry resulting from the addition. Otherwise continue sequentially".

The steps required to execute this instruction are as follows:

a). The PCU loads the address of the next instruction to the MAR, signalling to the memory that a Read is requested. The memory loads the MDR with the contents of the location addressed.
b). The CPU decodes the instruction, i.e., initiates the following steps.
c). The PCU is incremented and the next word is read from the memory.
d). Register R1 and the MDR are selected as source registers, MAR is the destination register.
e). The ALU performs "ADD" and the result is placed in the MAR.
f). The first operand is fetched from the memory and placed, for example, in R5.
g). The PCU is incremented and the next word is read from the memory.
h). Register R2 and the MDR are selected again as source registers and MAR as the destination.

i). The ALU performs "ADD" and the result is placed in MAR.
j). The second operand is fetched from the memory and is placed, for example, in R6.
k). The PCU is incremented, the next word is read from the memory.
l). Register R3 and the MDR are selected as source registers, the MAR as destination.
m). The ALU performs "ADD" and the result is placed in the MAR, which now points to the location where the sum of the operands should be stored.
n). Registers R5 and R6 are selected as sources (they contain the operands), MDR is now the destination.
o.) The ALU performs "ADD" and the result is placed in MDR.
p). A memory write cycle takes place and the contents of the MDR is stored at the desired address.
q). The carry is examined to determine the next step to be performed. Assume there is no carry.
r). The PCU is incremented twice (in order to skip the fifth word of the present instruction). It now points to the address of the next instruction.

As can be seen, 18 steps were used to perform a single addition using this complex relative addressing scheme. Obviously, our CPU needs some kind of "coordinator" which can:

1). Decode an instruction fetched from the memory.
2). Initiate the proper cycle of steps to be performed.
3). Set up the various controls for each step.
4). Execute the steps in an orderly sequence.
5). Make decisions according to the state of various signals (conditions).

We will call this coordinator the Computer Control Unit (CCU) and it is depicted in Fig. 6. Our CPU is now complete (more or less) and we will go into more detail later.

THE MEMORY

Let's now discuss the memory. The information stored in the memory is organized in words, where each word consists of N bits. N may be as small as 8 for very simple processors or as large as 64 in more powerful machines. The most common memory width for minicomputers is 16 bits. The number N is called the width of the memory and the number of bits in the MDR is obviously also N; equal to the width of the memory.

The depth of a memory is the number of words it contains. With a MAR having k bits, 2^k consecutive memory locations can be addressed. The addresses start from zero and range through 2^k-1.

The read access time of a memory directly accessible by the CPU is the time needed from stable address at the memory until the data is properly stored in the MDR. This access time depends on the type of memory used and can be as low as a few tens of nanoseconds and as large as several microseconds. Using high speed memory improves the performance of the computer as less time is wasted waiting for the memory to respond. In general, faster memories are costly, take more PC board area and use more power which results in more heat. A 32 bit wide, 2K (2048) word memory with 50 nanosecond access time may need 10 amps from the +5V power supply and may require a board area of 10" x 6". Yet this is a very small memory space.

It is usually not justified to have very large high-speed memories. Not all the programs and associated data need to reside in this memory at once. We may have the current program (or only a part of it) in the memory while other programs or data files can reside elsewhere and be brought into memory during the appropriate part of the program when needed.

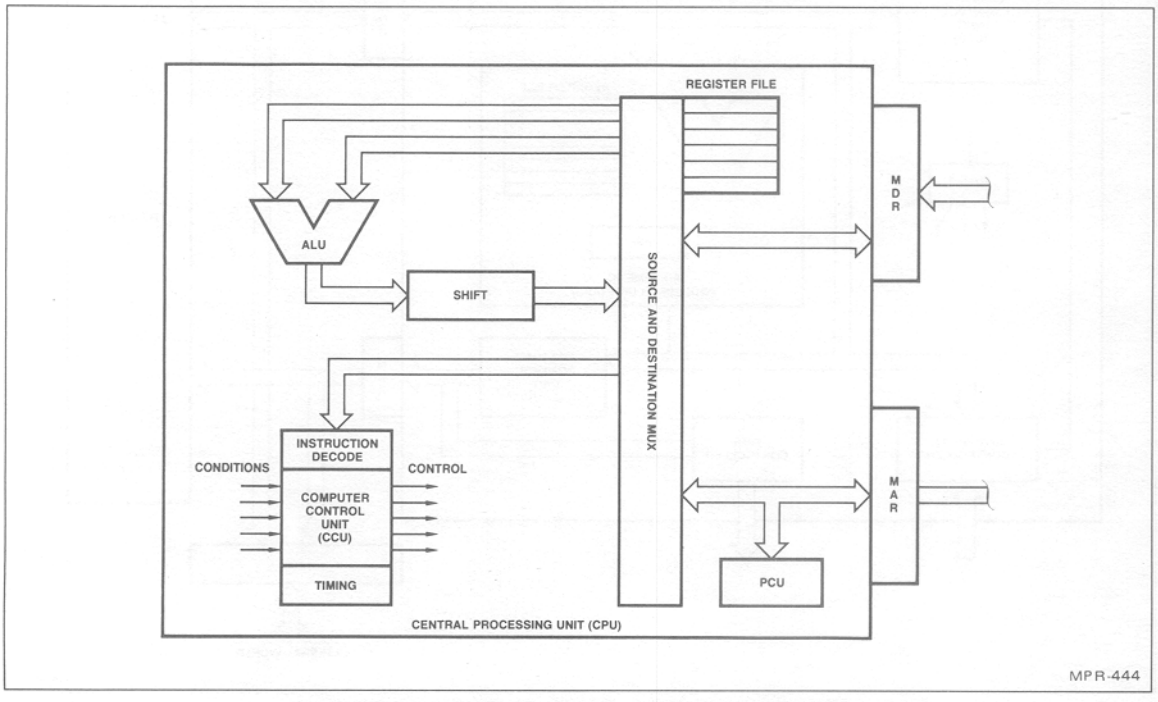


Figure 6. A Computer Control Unit (CCU) Included in a CPU.

This "elsewhere" may be a magnetic tape, cassette, disk, diskette, etc. and we will call it Bulk Memory. The distinctive characteristics of Bulk Memory are:

1). very large capacity
2). non-volatile (retains the information when not in use)
3). not randomly accessible
4). long access time
5). inexpensive (per bit)

Usually, Bulk Memory devices are serially accessible, i.e., the access time for the first word is large, but then consecutive words can be accessed relatively fast.

In a later chapter the most efficient process of communication between the main and the bulk memory, called the Direct Memory Access (DMA), will be discussed in detail.

THE EXTERNAL WORLD

In any useful machine, some means of communicating with the external word is needed. It may be a keyboard, a CRT, a card reader, a paper tape punch or, in a process controller, reading sensors or positioning actuators. The common denominator of almost all of the input/output devices is that they are much slower than the CPU and therefore a timing problem arises; the CPU must know when the I/O device is ready for data transfer. Usually, a signal is sent by the device to the CPU in order to draw its attention. The CPU now can do one of two things:

1). Test this signal periodically and when it is present, jump to a program which handles the data transfer. This type of operation is called "Polling". This technique has two major drawbacks: First, appreciable computer time is spent performing these periodic tests where most of them will fail (no "Ready" signal present). Second, the recognition by the computer CPU of the appearance of a signal is delayed until the CPU arrives at this device in its polling sequence.

   Imagine what will happen if there are a large number of I/O devices. Long latency times (delays) will occur if many I/O devices are busy simultaneously.

2). Include some hardware in the CPU which can sense the presence of a "Ready" signal and interrupt the normal flow of the instructions and force the computer to "Jump" to the I/O service program whenever there is a request. It can even send the CPU to different programs according to the I/O device whose "Ready" flag was detected and even establish priority among the different devices if more than one device would like to have the CPU's attention at the same time. Moreover, under program control, this circuitry can ignore some or all of the signals if the computer CPU must not be interrupted at that time. Obviously by paying the price of very little hardware, we gain enormously in computer performance. We will call this hardware the "Interrupt Controller" and will discuss it thoroughly later.

Our computer is now depicted in Fig. 7. We have included the ALU, the internal register file and the shift circuit in one block, which we call the "Arithmetic Processor Unit."

In the following pages and in the subsequent chapters, we will deal in more detail with each area of the machine.

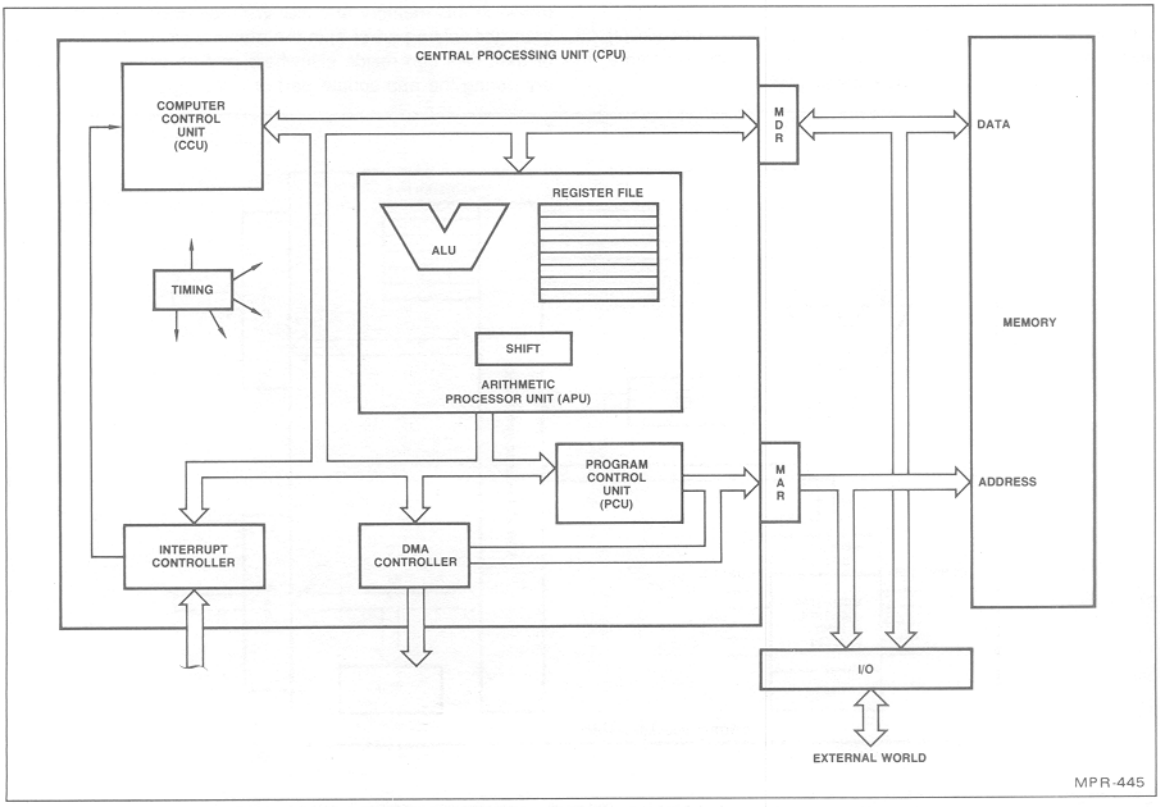


Figure 7. The Stored-Program-Computer with DMA and Interrupt Control Added.

A WORD ABOUT THE INSTRUCTION SET

The internal architecture of the CPU depends to some extent on the instruction set the computer is to execute. If the instruction set is large, some of the instructions usually are more complicated and the computer is more powerful, faster and more efficient. On the other hand, the internal circuitry is also more complicated. Some examples of these tradeoffs are as follows.

ALU Processing Capability:

Although with three basic functions (add, complement, and OR/AND) all the arithmetic and logic operations can be performed, most processors are built to perform subtract, NAND, XOR, etc. This is perhaps the most outstanding example of how performance and speed can be gained with little penalty on the complexity of the machine. With the added features an XOR operation can be performed in one instruction instead of 5.

Data Movement:

Let us assume 4 different computers whose data movement capabilities are described below:

Machine A). A word can be read from the memory and loaded into Register A only. The contents of Register A can be written into the memory, or can be moved into any other register. The contents of any register can be copied into Register A.

Machine B). The contents of any register can be copied into any other register or it can be written into the memory. A word read from the memory can be loaded into any register.

Machine C). As B above but with the added capability to read from one location in memory, to write that word into another location in memory.

Machine D). As C above and also the memory-to-memory operation can be performed on consecutive addresses repetitively. The number of word transfers (or upper and lower address limits) are specified by the instruction.

Machine A has very limited data movement capability. In order to perform an operation on two operands residing in the memory, we have to:

1). Bring the first operand from the memory into Register A.
2). Copy it into another register.
3). Bring the second operand into Register A.
4). Perform the operation required (result in A).
5). Store the contents of Register A into the memory.

If consecutive operations are required with several partial results, the drawbacks of machine A become more annoying, especially if the number of internal registers is small.

Moving a data block from one location in the memory to another location can be performed by one instruction in computer D, but requires the transfer of each word first to an internal register then to the new memory location in machines A, B (two instructions for each word transferred).

Obviously the decoding, multiplexing and sequencing of the computers grow in complexity as we proceed from machine A to machine D. We trade the complexity of hardware versus the software (programming), speed and performance.

Addressing:

The operands for an operation can be found in several ways:

- The operand is an explicit part of the instruction (Immediate)
- The address of the operand is an explicit part of the instruction. (Direct)
- The address of the operand is in an internal register; the register itself is specified by the instruction. (RR)
- The address of the operand is the sum of the contents of an internal register (specified by the instruction) and a number (called the displacement) which is an explicit part of the instruction. (RX)
- The contents of an internal register are added to a number found in an address specified by the instruction. The sum is the address of the operand. (Indirect)
- The contents of an internal register are added to a number which is an explicit part of the instruction. The sum points to the location where the address of the operand is written. (Indirect)
- The contents of an internal register are added to a number which can be found at the location explicitly specified by the instruction. The sum thus formed points to a location where the address of the operand is written.
- Etc.

Many other schemes can be formed by combining the above operations or by chaining them. In every case an "Effective Address" must be found by calculations and/or memory references. Again, we can gain performance by using more sophisticated addressing schemes but we will pay for it by adding complexity to our machine, especially in its control portion.

TIMING, SEQUENCING, CONTROLLING

In the previous paragraphs we have shown that we can gain performance in our computer by having a more complicated instruction set but more complex hardware is required, usually in the CCU. We have also shown an example for an "Add" operation which required 18 precisely controlled steps. Even if we assume that some of them can be performed simultaneously, we will need a multiphase clock to control these steps – something like that shown in Fig. 8. We can now load an instruction register at the beginning of an instruction with the first word of the instruction (the OP CODE) as is shown in Fig. 9. Using the outputs of the Instruction Register (IR_0 to IR_{n-1}), the different phases of the clock and the various condition inputs to the CCU, we can now try to write the logical equations which should satisfy all of the steps of all the instructions of our instruction set. Then use Karnough maps or other techniques to reduce these equations and finally realize them using AND, OR, INVERT gates and Flip Flops. Simple, isn't it? Imagine the complexity of a sophisticated computer and the debugging process it needs!

The question posed immediately is: Isn't there a more organized and more easily understandable way to do that? Or, perhaps, can we have some processor do the job for us? Can't we have some kind of "micro-machine" which can take care of all the timing, sequencing and controlling jobs of our computer – a computer inside the computer? With the advent of the Am2900 family – new Bipolar LSI devices – the answer is: Yes, we can!

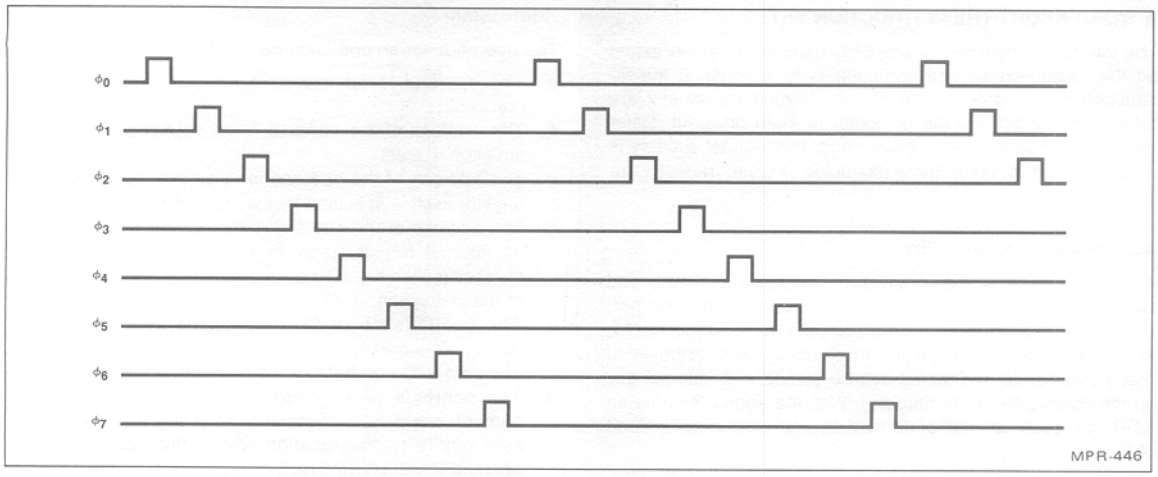


Figure 8. An 8-Phase Clock.

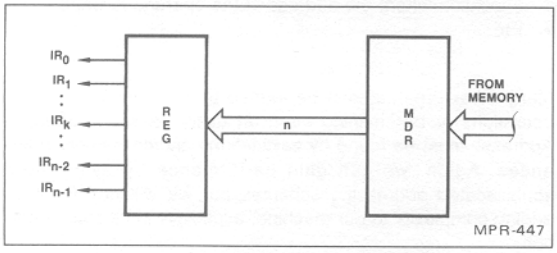


Figure 9. The Instruction Register Bits.

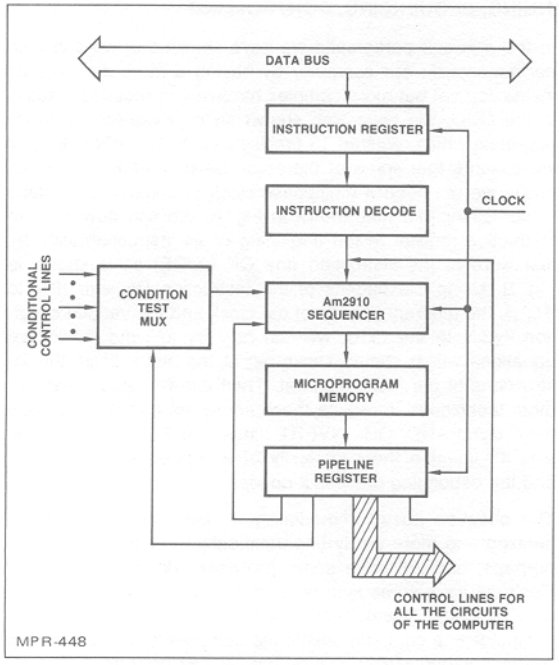


Figure 10. The Micromachine.

THE MICRO-MACHINE

What we need is essentially a machine which can execute a number of well defined sequences. But, remember that this is exactly the purpose of a stored program computer. The only difference between our micro-machine and a general purpose computer is that in the general purpose computer the program to be executed is changed from task to task, while in our micro-machine it is fixed. This allows the use of PROM for its memory instead of the RAM needed in the general purpose (GP) computer. Our Computer Control Unit (CCU) using this micro-machine may now look like Figure 10.

Basically, a microprogrammed machine is one in which a coherent sequence of microinstructions is used to execute various commands required by the machine. If the machine is a computer, each sequence of microinstructions can be made to execute a machine instruction. All of the little elemental tasks performed by the machine in executing the machine instruction are called microinstructions. The storage area for these microinstructions is usually called the microprogram memory.

A microinstruction usually has two primary parts. These are: (1) the definition and control of all elemental micro-operations to be carried out and (2) the definition and control of the address of the next microinstruction to be executed.

The definition of the various micro-operations to be carried out usually includes such things as ALU source operand selection, ALU function, ALU destination, carry control, shift control, interrupt control, data-in and data-out control, and so forth. The definition of the next microinstruction function usually includes identifying the source selection of the next microinstruction address and, in some cases, supplying the actual value of that microinstruction address.

Microprogrammed machines are usually distinguished from non-microprogrammed machines in the following manner. Older, non-microprogrammed machines implemented the control function by using combinations of gates and flip-flops connected in a somewhat random fashion in order to generate the required timing and control signals for the machine. Microprogrammed machines, on the other hand, are normally

considered highly ordered and more organized with regard to the control function field. In its simplest definition, a microprogram control unit consists of the microprogram memory and the structure required to determine the address of the next microinstruction.

The OP-CODE (type of instruction to be executed by the computer) is loaded into the Instruction Register and the Instruction Decoder decodes it. Actually, it generates the microaddress where the first step of the execution sequence for that instruction resides in the microprogram memory. The Am2910 sequencer then generates the microaddress of the next microinstruction. The microprogram data supplies the control signals we need to control all the parts of the computer (and there are a lot of them), including the sequencer itself. When all the steps of a machine instruction are executed, the microprogram will cause the reading (fetch) of the next machine instruction from the computer main memory. Typically, the Computer Control Unit is used to fetch instructions and decode them using a PROM for mapping the op code to the initial address of the sequence of microinstructions used to execute this particular instruction. It will also fetch all of the operands needed by the machine instruction and deliver them to the ALU for processing. An example of the flow of a typical Computer Control Unit is shown in Figure 11.

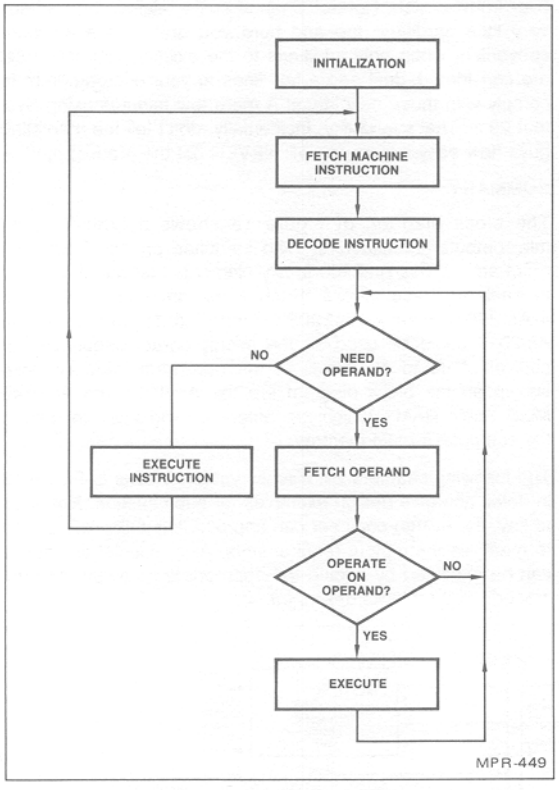
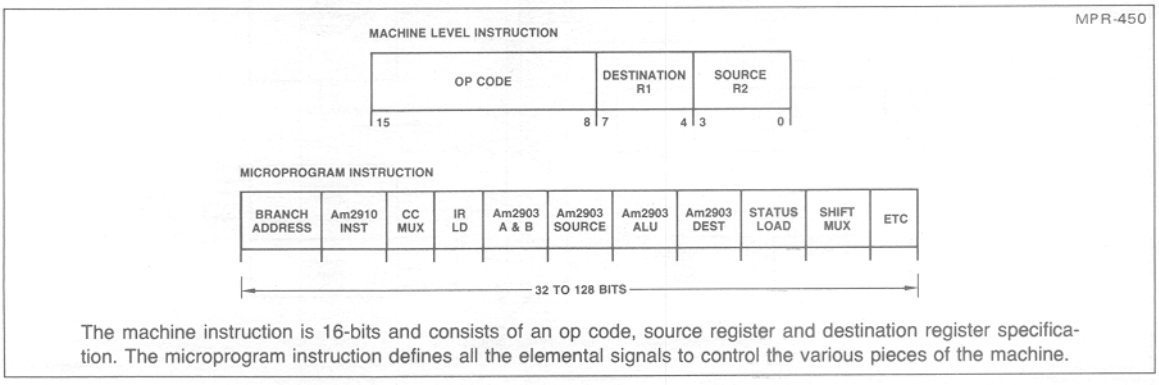


Figure 11. Computer Control Function Flow Diagram.

Assume the OP-CODE of the machine instruction that we fetch is 8 bits wide. This allows us to execute a minimum of 256 different instructions. Assume also that an average of 6 steps are needed to execute these instructions. Even if separate microprogram memory locations are used, a depth of this microprogram memory is only 1-1/2K (K = 1024). But in that case, the sequencer can almost be replaced by a simple counter. Usually we would like to share some micro-routines among different instructions. With very little effort, we can shrink the depth of the microprogram memory of Figure 10 to less than 1/2K. Of course the sequencer will be a little more sophisticated; it will perform conditional Branch and microsubroutine CALL's; but we still don't need the complicated addressing schemes for microprogram control as were described earlier as a part of the machine instruction set.

On the other hand, the width of our microprogram memory may be large – maybe 60 to 100 bits. This will depend on the number of control lines needed in our computer. This is of no great disadvantage since the price of PROM devices is dropping constantly. In a future chapter we will discuss techniques to reduce the depth and width of the microprogram memory to save cost.

It is important to understand the distinction between machine level instructions and microprogram instructions. Figure 12 shows a typical machine instruction for a 16 bit minicomputer that has an 8-bit opcode to identify one of 256 instructions; a 4-bit source register specification to identify one of 16 source registers and a 4-bit destination register specification to identify one of 16 destination registers. The microprogram instruction of Figure 12 may contain from 32 to 128 bits in a typical design; or even more bits in a very fast, highly parallel microcoded machine. This microinstruction word usually will contain fields for the ALU source operand, ALU function, ALU destination, status load enable, shift multiplexer control, bus



The machine instruction is 16-bits and consists of an op code, source register and destination register specification. The microprogram instruction defines all the elemental signals to control the various pieces of the machine.

Figure 12.

cycle control, etc. These fields are used to control the various devices within the machine so that its execution is as desired on each clock cycle. This is more straightforward than using combinatorial logic and yields a more organized design.

Let us now compare the depth-over-width (d/w) ratio of the computer's main memory to that of our microprogram memory.

In the Am9080A type microprocessor, the data field is 8 bits and the address field is 16 bits, allowing direct addressing of 64K locations. The ratio d/w is 8K. In some minicomputers, the data width is 16-32 bits and the addressing capability is 64-128K. The d/w ratio is about the same. In larger computers with 32-64 bit data width, we find 256-512K deep memories or even deeper ones. The d/w ratio again is 8K at least.

On the other hand, the d/w ratio in microprogram memories is seldom greater than a few tens. Even if we assume that it is 2K deep and only 64 bits wide, we arrive at a d/w ratio of only 32; usually it will be around 10. It is much easier to control a machine with a d/w ratio of 10 to 20 than to control one with d/w = 8K.

ONE MORE WORD

We have suggested a replacement of the "random logic" realization of the CCU by a micro-machine. We call this a "Microprogrammed Architecture". Perhaps the biggest advantage of this type of architecture is the ease of structuring the control sequence. We allocate a bit or a group of bits in the microprogram memory to control a certain function (e.g.: ALU source register selection, ALU function, ALU destination selection, condition selection, next address calculation selection, MDR destination selection, MAR source selection, etc., etc.) and for each microstep we write the appropriate state for these bits (LOW-HIGH) into this memory field. Later we will see that automated and sophisticated tools are available to perform this microprogram writing. One such tool is AMDASM™ as available on System 29. But, this is not the only advantage of the microprogrammed architecture.

As nobody is perfect, some "bugs" may inadvertently slip into the design. In a random logic architecture, we will have to redesign and usually rebuild the whole computer. On the other hand, in a microprogrammed machine it is usually sufficient to change a couple of bits in the microprogram to rectify the problem. This is even easier if a RAM instead of a PROM is used during the development and debugging phases. Of course, we must be able to load this memory with the microprogram by some external means. Again, a powerful tool is available: AMD's System/29™.

Finally, let's face the reality: The marketing guys usually change their requirements (i.e., the instruction set) when you are 80% through your logic design. Now you have to start over from scratch. Not so! Change some microcode, perhaps very little hardware too and here you are! It is even more convenient when only additions to the existing instruction set are considered. Just add a few lines to your microprogram to comply with those new ideas! A mere few minutes using System 29 – That's flexibility! Incidentally, don't tell the marketing guys how easy it is or you will NEVER get the product out!!

SUMMARY

The block diagram of Figure 13 shows a typical 16-bit minicomputer architecture. Also identified on this block diagram are various Am2900 family elements that might be used in each of these blocks. Such a design might use either 4-Am2901A's or 4-Am2903's for the data path ALU. An Am2910 could be used as the microprogram sequencer for control of up to 4K words of microprogram memory. Also shown on the block diagram are the Am9130 and Am9140 MOS Static RAM's which are potential candidates for use in the computer's main memory.

The following chapters will discuss various blocks of Figure 13 in detail and give design examples for each section. Needless to say, the design engineer can appropriately tailor any design to meet his throughput requirements. Also, special algorithms can be executed by adding the appropriate hardware and microcode to the blocks described.

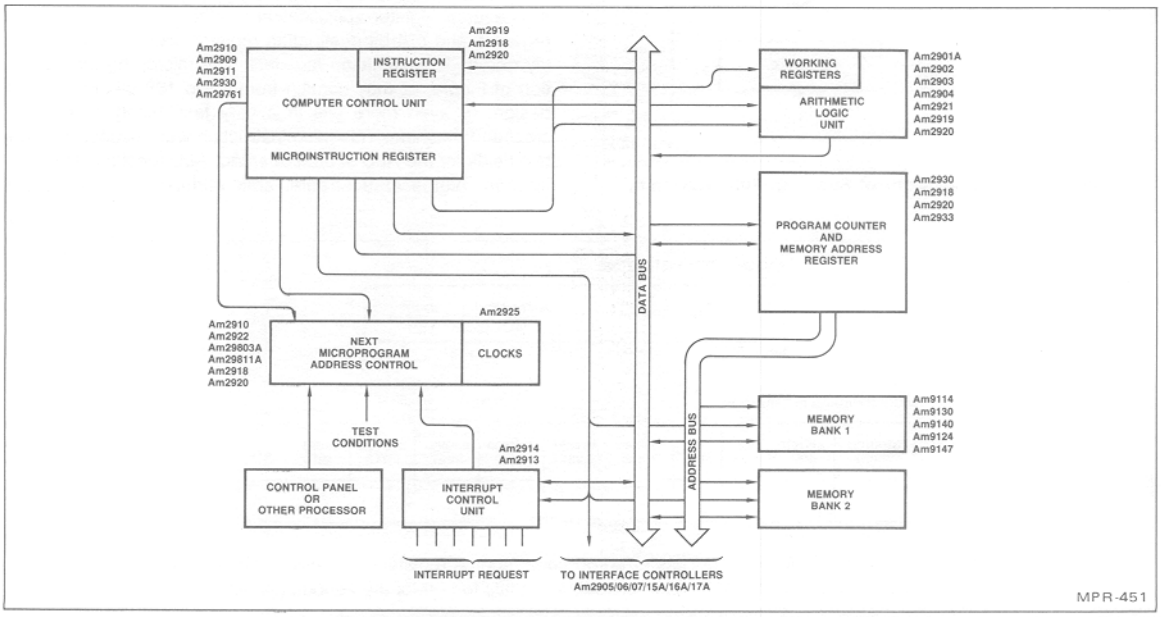


Figure 13. A Generalized Computer Architecture.

ADVANCED
MICRO
DEVICES, INC.
901 Thompson Place
Sunnyvale
California 94086
(408) 732-2400
TWX: 910-339-9280
TELEX: 34-6306
TOLL FREE
(800) 538-8450