Back to more realistic numbers: 16 MB = $2^{24}$ bytes = $2^{12}$ rows x $2^{12}$ columns. Instead of one 24 to 16,777,216 decoder, we build two 12 to $2^{12} = 4,096$ decoders.

The memory unit also contains a device called the fetch/store controller which determines whether a fetch or store operation will be performed. (See Fig. 5.7 p. 182)

Cache Memory

In addition to the main memory device, modern processors have an additional store of very fast memory called cache.

![Cache Memory Diagram]

Cache memory is made with expensive fabrication techniques and can be accessed at rates up to 10 times faster than RAM.
When an item is needed from memory, the processor first looks in cache. If the item is there, it is accessed at the fastest rate. If not, it is fetched from RAM at the slower rate. The item is then placed in cache, overwriting an older item.

This design is based on the principle of locality, which says that if the processor uses an item, it will probably use it again soon.

Ex.
Suppose the item needed is in cache 70% of the time. (This is called the cache hit rate.) Suppose cache access time is 10 ns, while RAM takes 50 ns to access. The average access time is then

\[(0.70) \times 10 + (0.30) \times (50 + 10) = 25 \text{ nsec.}\]

A 50% improvement over 50 nsec for RAM.

Read 5.2.2 I/O \& Mass Storage p. 184-189
THE ARITHMETIC/LOGIC UNIT (ALU)

The ALU is the subsystem which performs mathematical and logical operations such as addition, subtraction, comparison, etc.

In addition to the circuitry which performs these operations, the ALU contains a number of registers. These are special storage cells which hold the operands of upcoming operations, as well as intermediate results. A typical ALU will have 16-32 registers. These are similar to cache memory in that they allow very fast access.
To perform an operation, the operands are first moved from RAM to the ALU registers. One register is connected to the bus labeled left, the other to the bus labeled right. (A bus is simply an electrical data path which "steps" at a number of locations.) The ALU is then enabled to perform the desired operation. The result is passed to the output bus, and is either placed in another register, or leaves the ALU.

How does the ALU know which operation to perform? One approach is to have the ALU perform all operations, then use a multiplexer to select the output.

Ex. Suppose our ALU does just four operations: a + b, a - b, a < b, a = b.
NOTE THAT IN THESE DIAGRAMS, A SINGLE LINE CAN REPRESENT 8 OR MORE SEPARATE ELECTRICAL PATHS. THIS IS IN KEEPING WITH THE LARGER SCALE, ABSTRACT VIEW WE ARE TAKING. WE WILL GLOSS OVER MANY SUCH DETAILS.

THE CONTROL UNIT

RECALL A FUNDAMENTAL CHARACTERISTIC OF THE VON NEUMAN ARCHITECTURE IS THE STORED PROGRAM: A SEQUENCE OF INSTRUCTIONS STORED AS BINARY VALUES IN MEMORY.

IT IS THE TASK OF THE CONTROL UNIT TO:

1.) FETCH THE NEXT INSTRUCTION
2.) DECODE THE INSTRUCTION
3.) EXECUTE THE INSTRUCTION

UNTIL A SPECIAL INSTRUCTION (E.G. HALT) IS ENCOUNTERED. STEPS (1)-(3) CAN BE THOUGHT OF AS THE BODY OF A REPEAT LOOP.

STORED PROGRAM INSTRUCTIONS ARE ENCODED IN A REPRESENTATION CALLED MACHINE LANGUAGE.
A typical format for a machine language instruction would consist of an operation code followed by one or more address fields.

The operation code (op-code) is a unique unsigned integer code assigned to each operation which the processor can perform.

The address fields contain the memory addresses of the operands. Each address field must be the same width as the MAR.

Ex. Instruction Register (IR)

<table>
<thead>
<tr>
<th>Opcode</th>
<th>Addr1</th>
<th>Addr2</th>
<th>Addr3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bits: 8 24 24 24</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When an instruction is fetched from memory it is placed in a special location called the instruction register (IR).

In this example each instruction occupies 10 bytes of memory, which is the width of the IR. There are at most \(2^8 = 256\) operations, and the memory unit contains at most \(2^{24}\) bytes = 16 MB.
The set of all operations which can be performed by a processor is called its instruction set.

Complex instruction set computers (CISC) recognize 200-400 operations, whereas instruction set computers (RISC) recognize about 30-50 operations.

The trend in recent years has been towards RISC processors which run faster than CISC designs. The trade off is to have longer, more complex machine language programs. This has become more acceptable as memory cost has gone down.

EX

Suppose our machine language instructions consist of an 8 bit op code followed by two 8 bit address fields. (i.e. at most 256 op codes, and 256 bytes in memory.)

Suppose we wish to move the contents of cell 27 to cell 53, and that the op-code for this operation is 15.
15 = 1111, 27 = 11011, 53 = 110101.

The test string for this instruction is then:

<table>
<thead>
<tr>
<th>OP Code</th>
<th>ADDR1</th>
<th>ADDR2</th>
</tr>
</thead>
<tbody>
<tr>
<td>00001111</td>
<td>00011011</td>
<td>00110101</td>
</tr>
</tbody>
</table>

In the sequel we will not express such instructions in binary as above, instead we will use shorthand notation like:

MOVE 27, 53

We must keep in mind though that this shorthand stands for the full binary instruction.

We will use capitals X, Y, Z for cell addresses, and lower case a, b, c for values stored in cells. CON(x) means the contents of the cell with address X.

Assume our ALU contains just one register: R.

Machine language instructions can be grouped into four classes: