Influence of Technology and Software on Instruction Sets: Up to the dawn of IBM 360

*Joel Emer*
Computer Science and Artificial Intelligence Laboratory
M.I.T.

February 11, 2015

http://www.csg.csail.mit.edu/6.823
And then there was IBM 701

IBM 701 -- 30 machines were sold in 1953-54

IBM 650 -- a cheaper, drum based machine, more than 120 were sold in 1954 and there were orders for 750 more!

Users stopped building their own machines.

Why was IBM late getting into computers?

*IBM was making too much money!*

Even without computers, IBM revenues were doubling every 4 to 5 years in 40’s and 50’s.
Computers in mid 50’s

• Hardware was expensive

• Stores were small (1000 words)
  ⇒ No resident system-software!

• Memory access time was 10 to 50 times slower than the processor cycle
  ⇒ Instruction execution time was totally dominated by the memory reference time.

• The ability to design complex control circuits to execute an instruction was the central design concern as opposed to the speed of decoding or an ALU operation

• Programmer’s view of the machine was inseparable from the actual hardware implementation
Accumulator-based computing

- **Single Accumulator**
  - Calculator design carried over to computers

*Why?*

*Registers expensive*
# The Earliest Instruction Sets

*Burks, Goldstein & von Neumann ~1946*

<table>
<thead>
<tr>
<th>Instruction</th>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOAD</td>
<td>x</td>
<td>$AC \leftarrow M[x]$</td>
</tr>
<tr>
<td>STORE</td>
<td>x</td>
<td>$M[x] \leftarrow (AC)$</td>
</tr>
<tr>
<td>ADD</td>
<td>x</td>
<td>$AC \leftarrow (AC) + M[x]$</td>
</tr>
<tr>
<td>SUB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MUL</td>
<td>x</td>
<td>Involved a quotient register</td>
</tr>
<tr>
<td>DIV</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>SHIFT LEFT</td>
<td></td>
<td>$AC \leftarrow 2 \times (AC)$</td>
</tr>
<tr>
<td>SHIFT RIGHT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JUMP</td>
<td>x</td>
<td>$PC \leftarrow x$</td>
</tr>
<tr>
<td>JGE</td>
<td>x</td>
<td>if $(AC) \geq 0$ then $PC \leftarrow x$</td>
</tr>
<tr>
<td>LOAD ADR</td>
<td>x</td>
<td>$AC \leftarrow$ Extract address field($M[x]$)</td>
</tr>
<tr>
<td>STORE ADR</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

*Typically less than 2 dozen instructions!*
Programming: Single Accumulator Machine

C_i \leftarrow A_i + B_i, \quad 1 \leq i \leq n

Problem?

How to modify the addresses A, B and C?
## Self-Modifying Code

<table>
<thead>
<tr>
<th>LOOP</th>
<th>LOAD</th>
<th>N</th>
<th>JGE</th>
<th>DONE</th>
<th>ADD</th>
<th>ONE</th>
<th>STORE</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>LOAD</td>
<td>A</td>
<td>ADD</td>
<td>B</td>
<td>STORE</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>ADD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F3</td>
<td>STORE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LOAD ADR</td>
<td>F1</td>
<td>ADD</td>
<td>ONE</td>
<td>STORE ADR</td>
<td>F1</td>
<td>ADD</td>
<td>F2</td>
</tr>
<tr>
<td></td>
<td>LOAD ADR</td>
<td>F2</td>
<td>ADD</td>
<td>ONE</td>
<td>STORE ADR</td>
<td>F2</td>
<td>ADD</td>
<td>F3</td>
</tr>
<tr>
<td></td>
<td>LOAD ADR</td>
<td>F3</td>
<td>ADD</td>
<td>ONE</td>
<td>STORE ADR</td>
<td>F3</td>
<td>JUMP</td>
<td>LOOP</td>
</tr>
</tbody>
</table>

\[ C_i \leftarrow A_i + B_i, \quad 1 \leq i \leq n \]

Each iteration involves:

- **Total book-keeping**: 1714
- **Instruction fetches**: 108
- **Operand fetches**: 54

Most of the executed instructions are for bookkeeping!

February 11, 2015

Sanchez & Emer
Processor-Memory Bottleneck: Early Solutions

• Indexing capability
  – to reduce book keeping instructions

• Fast local storage in the processor
  – 8-16 registers as opposed to one accumulator
  – to save on loads/stores

• Complex instructions
  – to reduce instruction fetches

• Compact instructions
  – implicit address bits for operands
  – to reduce instruction fetch cost
Modify existing instructions

- **LOAD** \( x, IX \)  
  \[ AC \leftarrow M[x + (IX)] \]

- **ADD** \( x, IX \)  
  \[ AC \leftarrow (AC) + M[x + (IX)] \]

- ...

Add new instructions to manipulate index registers

- **JZi** \( x, IX \)  
  if \((IX)=0\) then  
  \[ PC \leftarrow x \]
  else  
  \[ IX \leftarrow (IX) + 1 \]

- **LOADi** \( x, IX \)  
  \[ IX \leftarrow M[x] \quad (\text{truncated to fit IX}) \]

- ...

Index registers have accumulator-like characteristics

February 11, 2015
Using Index Registers

\[ C_i \leftarrow A_i + B_i, \ 1 \leq i \leq n \]

- **LOADi** N, IX
- **LOAD** LASTA, IX
- **ADD** LASTB, IX
- **STORE** LASTC, IX
- **JUMP** LOOP
- **DONE** HALT

- Program does not modify itself
- Efficiency has improved dramatically (ops / iter) with index regs
- Costs?
  - Complex control
  - Index register computations (ALU-like circuitry)
  - 1 to 2 bits longer Instructions

\[ A_i \leftarrow A_i + B_i, \ 1 \leq i \leq n \]

N starts with -n

\[ \text{instruction fetch: } 5(2) \text{ vs. } 17(14) \]
\[ \text{operand fetch: } 2 \text{ vs. } 10(8) \]
\[ \text{store: } 1 \text{ vs. } 5(4) \]
Operations on Index Registers

To increment index register by k
\[
\begin{align*}
\text{AC} &\leftarrow (\text{IX}) \\
\text{AC} &\leftarrow (\text{AC}) + k \\
\text{IX} &\leftarrow (\text{AC})
\end{align*}
\]
new instruction

also the AC must be saved and restored

It may be better to increment IX directly
\[
\text{INCI} \quad k, \text{IX} \quad \text{IX} \leftarrow (\text{IX}) + k
\]

More instructions to manipulate index register
\[
\begin{align*}
\text{STOREI} &\quad x, \text{IX} \quad M[x] \leftarrow (\text{IX}) \quad \text{(extended to fit a word)}
\end{align*}
\]
...

\emph{IX begins to look like an accumulator}
\Rightarrow several index registers
\Rightarrow several accumulators
\Rightarrow \text{General Purpose Registers}

February 11, 2015
Support for Subroutine Calls

A special subroutine jump instruction

A: JSR F

M[F] ← A + 1 and jump to F + 1
Indirect Addressing and Subroutine Calls

**Indirect addressing**

LOAD (x) means AC ← M[M[x]]

Indirect addressing almost eliminates the need to write self-modifying code (location F still needs to be modified)

**Events:**

- **S1**: LOAD (F)
- **S2**: STORE (F)
- **S3**: JUMP (F)

**Subroutine Fetch**

- A
- F
- A+1
- A+2
- A+3

**Subroutine Store**

- arg
- result

**Problems? ⇒ recursive procedure calls**
Recursive Procedure Calls and Reentrant Codes

*Indirect Addressing through a register*

\[ \text{LOAD } R_1, (R_2) \]

Load register \( R_1 \) with the contents of the word whose address is contained in register \( R_2 \)
Evolution of Addressing Modes

1. Single accumulator, absolute address
   \[
   \text{LOAD } x
   \]

2. Single accumulator, index registers
   \[
   \text{LOAD } x, IX
   \]

3. Indirection
   \[
   \text{LOAD } (x)
   \]

4. Multiple accumulators, index registers, indirection
   \[
   \text{LOAD } R, IX, x
   \]
   or \[
   \text{LOAD } R, IX, (x)
   \]
   the meaning?
   \[
   R \leftarrow M[M[x] + (IX)]
   \]
   or \[
   R \leftarrow M[M[x + (IX)]]
   \]

5. Indirect through registers
   \[
   \text{LOAD } R_I, (R_J)
   \]

6. The works
   \[
   \text{LOAD } R_I, R_J, (R_K)
   \]
   \[
   R_J = \text{index}, R_K = \text{base addr}
   \]
Variety of Instruction Formats

- **Three address formats:** One destination and up to two operand sources per instruction

  \[(\text{Reg op Reg}) \text{ to Reg} \quad R_I \leftarrow (R_J) \text{ op (R}_K)\]

  \[(\text{Reg op Mem}) \text{ to Reg} \quad R_I \leftarrow (R_J) \text{ op M}[x]\]

  - x can be specified directly or via a register
  - effective address calculation for x could include indexing, indirection, ...

- **Two address formats:** the destination is same as one of the operand sources

  \[(\text{Reg op Reg}) \text{ to Reg} \quad R_I \leftarrow (R_I) \text{ op (R}_J)\]

  \[(\text{Reg op Mem}) \text{ to Reg} \quad R_I \leftarrow (R_I) \text{ op M}[x]\]
More Instruction Formats

- **One address formats**: Accumulator machines
  - Accumulator is always other implicit operand

- **Zero address formats**: operands on a stack
  - Stack can be in registers or in memory
    - usually top of stack cached in registers

```
load      M[sp] ← M[M[sp]]
```

Many different formats are possible!
Data Formats and Memory Addresses

Data formats:

- Bytes, Half words, words and double words

Some issues

- **Byte addressing**
  
  Big Endian vs. Little Endian

<table>
<thead>
<tr>
<th>Big Endian</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

- **Word alignment**

  Suppose the memory is organized in 32-bit words.
  Can a word address begin only at 0, 4, 8, .... ?

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>
Some Tradeoffs

- Should all addressing modes be provided for every operand?
  \(\Rightarrow\) regular vs. irregular instruction formats

- Separate instructions to manipulate Accumulators, Index registers, Base registers
  \(\Rightarrow\) large number of instructions

- Instructions contained implicit memory references -- several contained more than one
  \(\Rightarrow\) very complex control

Great variety of instruction sets
The first definition of the Instruction Set Abstraction: IBM 360
The IBM 650 (1953-4)

Magnetic Drum (1,000 or 2,000 10-digit decimal words)

Active instruction (including next program counter)

20-digit accumulator

Digit-serial ALU

[From 650 Manual, © IBM]
Programmer’s view of a machine: IBM 650

A drum machine with 44 instructions

Instruction: 60 1234 1009
- “Load the contents of location 1234 into the distribution; put it also into the upper accumulator; set lower accumulator to zero; and then go to location 1009 for the next instruction.”

- Programmer’s view of the machine was inseparable from the actual hardware implementation
- Good programmers optimized the placement of instructions on the drum to reduce latency!
Compatibility Problem at IBM

By early 60’s, IBM had 4 incompatible lines of computers!

701 → 7094
650 → 7074
702 → 7080
1401 → 7010

Each system had its own

- Instruction set
- I/O system and Secondary Storage: magnetic tapes, drums and disks
- assemblers, compilers, libraries,...
- market niche
  business, scientific, real time, ...

⇒ IBM 360
IBM 360 : Design Premises
Amdahl, Blaauw and Brooks, 1964

The design must lend itself to *growth and successor machines*

- General method for connecting I/O devices
- Total performance - answers per month rather than bits per microsecond $\Rightarrow$ *programming aids*
- Machine must be capable of *supervising itself* without manual intervention
- Built-in *hardware fault checking* and locating aids to reduce down time
- Simple to assemble systems with redundant I/O devices, memories etc. for *fault tolerance*
- Some problems required floating point words larger than 36 bits
Processor State and Data Types

The information held in the processor at the end of an instruction to provide the processing context for the next instruction.

- Program Counter, Accumulator, ...

- The information held in the processor will be interpreted as having data types manipulated by the instructions.

- If the processing of an instruction can be interrupted then the hardware must save and restore the state in a transparent manner.

*Programmer’s machine model is a contract between the hardware and software*
Some things an ISA must specify:

- A way to reference registers and memory
- The computational operations available
- How to control the sequence of instructions

ISA must satisfy the needs of the software:
- assembler, compiler, OS, VM
IBM 360: A General-Purpose Register (GPR) Machine

• Processor State
  – 16 General-Purpose 32-bit Registers
    • may be used as index and base register
    • Register 0 has some special properties
  – 4 Floating Point 64-bit Registers
  – A Program Status Word (PSW)
    • PC, Condition codes, Control flags

• Data Formats
  – 8-bit bytes, 16-bit half-words, 32-bit words, 64-bit double-words
  – 24-bit addresses

• A 32-bit machine with 24-bit addresses
  – No instruction contains a 24-bit address!
IBM 360: Some Addressing Modes

The most common formats for arithmetic & logic instructions, as well as Load and Store instructions

R \leftarrow (R) \text{ op } M[(X) + (B) + D]

a 24-bit address is formed by adding the 12-bit displacement (D) to a base register (B) and an Index register (X), if desired
IBM 360: Character String Operations

SS format: store to store instructions

\[ M[(B1) + D1] \leftarrow M[(B1) + D1] \text{ op } M[(B2) + D2] \]
iterate “length” times

Most operations on decimal and character strings use this format

MVC \hspace{1em} \text{move characters}
MP \hspace{1em} \text{multiply two packed decimal strings}
CLC \hspace{1em} \text{compare two character strings}
...

Multiple memory operations per instruction complicates exception & interrupt handling
IBM 360: Branches & Condition Codes

- Arithmetic and logic instructions set *condition codes*
  - equal to zero
  - greater than zero
  - overflow
  - carry...

- I/O instructions also set condition codes
  - channel busy

- Conditional branch instructions are based on testing condition code registers (CC’s)
  - RX and RR formats
    - BC_ branch conditionally
    - BAL_ branch and link, i.e., R15 ← (PC)+1
    
    *for subroutine calls*
    
    ⇒ CC’s must be part of the PSW
IBM 360: Precise Interrupts

- IBM 360 ISA (Instruction Set Architecture) preserves sequential execution model

- Programmers view of machine was that each instruction either completed or signaled a fault before the next instruction began execution

- Exception/interrupt behavior identical across family of implementations
## IBM 360: Initial Implementations (1964)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Model 30</th>
<th>...</th>
<th>Model 70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory Capacity</td>
<td>8K - 64 KB</td>
<td></td>
<td>256K - 512 KB</td>
</tr>
<tr>
<td>Memory Cycle</td>
<td>2.0µs</td>
<td>...</td>
<td>1.0µs</td>
</tr>
<tr>
<td>Datapath</td>
<td>8-bit</td>
<td></td>
<td>64-bit</td>
</tr>
<tr>
<td>Circuit Delay</td>
<td>30 nsec/level</td>
<td></td>
<td>5 nsec/level</td>
</tr>
<tr>
<td>Registers in Main Store</td>
<td></td>
<td></td>
<td>in Transistor</td>
</tr>
<tr>
<td>Control Store</td>
<td>Read only 1µsec</td>
<td></td>
<td>Dedicated circuits</td>
</tr>
</tbody>
</table>

- Six implementations (Models, 30, 40, 50, 60, 62, 70)
- 50X performance difference cross models
- *ISA completely hid the underlying technological differences between various models.*

With minor modifications, IBM 360 ISA is still in use
IBM 360: Forty-Six years later... 
zEnterprise196 Microprocessor

- 1.4 billion transistors, Quad core design
- Up to 96 cores (80 visible to OS) in one multichip module
- 5.2 GHz, IBM 45nm SOI CMOS technology
- 64-bit virtual addressing
  - original 360 was 24-bit; 370 was a 31-bit extension
- Superscalar, out-of-order
  - Up to 72 instructions in flight
- Variable length instruction pipeline: 15-17 stages
- Each core has 2 integer units, 2 load-store units and 2 floating point units
- 8K-entry Branch Target Buffer
  - Very large buffer to support commercial workload
- Four Levels of caches:
  - 64KB L1 I-cache, 128KB L1 D-cache
  - 1.5MB L2 cache per core
  - 24MB shared on-chip L3 cache
  - 192MB shared off-chip L4 cache

[ September 2010 ]
Instruction Set Architecture (ISA) versus Implementation

- ISA is the hardware/software interface
  - Defines set of programmer visible state
  - Defines data types
  - Defines instruction semantics (operations, sequencing)
  - Defines instruction format (bit encoding)
  - Examples: *MIPS*, *Alpha*, *x86*, *IBM 360*, *VAX*, *ARM*, *JVM*

- Many possible implementations of one ISA
  - 360 implementations: model 30 (c. 1964), zEnterprise196 (c. 2010)
  - x86 implementations: 8086 (c. 1978), 80186, 286, 386, 486, *Pentium, Pentium Pro, Pentium-4, Core i7, AMD Athlon, AMD Opteron, Transmeta Crusoe, SoftPC*
  - MIPS implementations: *R2000, R4000, R10000, ...*
  - JVM: *HotSpot, PicoJava, ARM Jazelle, ...*
Next lecture:
Implementing an ISA