Brief History of Computer Architecture

Elissaveta Arnaoudova

People about Computer Architecture:

- Baer: “The design of the integrated system which provides a useful tool to the programmer”
- Hayes: “The study of the structure, behavior and design of computers”
- Hennessy and Patterson: “The interface between the hardware and the lowest level software”
- Foster: “The art of designing a machine that will be a pleasure to work with”

The different usages of the term:

- The design of a computer's CPU architecture, instruction set, addressing modes
- Description of the requirements (especially speeds and interconnection requirements) or design implementation for the various parts of a computer. (Such as memory, motherboard, electronic peripherals, or most commonly the CPU.)
- Architecture is often defined as the set of machine attributes that a programmer should understand in order to successfully program the specific computer

- So, in general, computer architecture refers to attributes of the system visible to a programmer, that have a direct impact on the execution of a program.

Brief history of computer architecture

First Generation (1945-1958)

- **Features**
  - Vacuum tubes
  - Machine code, Assembly language
  - Computers contained a central processor that was **unique** to that machine
  - Different types of supported instructions, few machines could be considered "general purpose"
  - Use of **drum memory** or **magnetic core memory,** programs and data are loaded using paper tape or punch cards
  - 2 Kb memory, 10 KIPS

First Generation (1945-1958)…

- **Two types of models for a computing machine:**
  - **Harvard architecture** - physically separate storage and signal pathways for instructions and data. (The term originated from the Harvard Mark I, relay-based computer, which stored instructions on punched tape and data in relay latches.)
  - **Von Neumann architecture** - a single storage structure to hold both the set of instructions and the data. Such machines are also known as **stored-program computers**.
    - **Von Neumann bottleneck** - the bandwidth, or the data transfer rate, between the CPU and memory is very small in comparison with the amount of memory.

NB: Modern high performance CPU chip designs incorporate aspects of both architectures. On chip cache memory is divided into an instruction cache and a data cache. Harvard architecture is used as the CPU accesses the cache and von Neumann architecture is used for off chip memory access.

First Generation (1945-1958)…

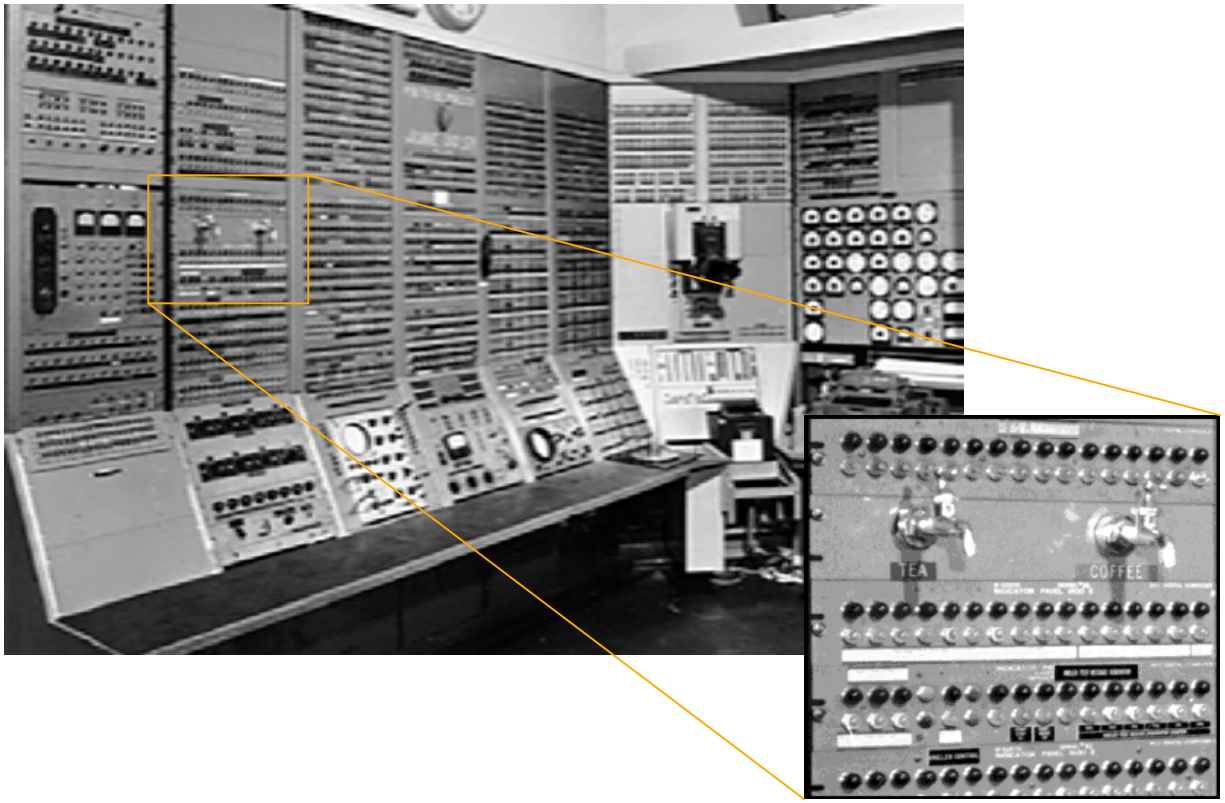
- 1943-46, ENIAC (Electronic Numerical Integrator and Calculator) by J. Mauchly and J. Presper Eckert, first general purpose electronic computer

The size of its numerical word was 10 decimal digits, and it could perform 5000 additions and 357 multiplications per second.

Built to calculate trajectories for ballistic shells during WWII, programmed by setting switches and plugging &
unplugging cables.
It used 18,000 tubes, weighted 30 tones and consumed 160 kilowatts of electrical power.

First Generation (1945-1958)…

- 1949, Whirlwind computer by Jay Forrester (MIT) with 5000 vacuum tubes, main innovation - magnetic core memory



First Generation (1945-1958)…

- 1951 UNIVAC (Universal Automatic Computer) - the first commercial computer, built by Eckert and Mauchly, cost – around $1 million, 46 machines sold
  - UNIVAC had an add time of 120 microseconds, multiply time of 1,800 microseconds and a divide time of 3,600 microseconds, used magnetic tape as input

- 1953, IBM's 701, the first commercially successful general-purpose computer. The 701 had electrostatic storage tube memory, used magnetic tape to store information, and had binary, fixed-point, single address hardware.

- IBM 650 - 1st mass-produced computer (450 machines sold in one year)

Second Generation (1958-1964)

- **Features**
  - **Transistors –** small, low-power, low-cost, more reliable than vacuum tubes,
  - Magnetic core memory
  - Two's complement, floating point arithmetic
  - Reduced the computational time from milliseconds to microseconds
  - High level languages
  - First operating Systems: handled one program at a time
  - 1959 - IBM´s 7000 series mainframes were the company´s first transistorized computers.

Second Generation (1958-1964)…

- **IBM 7090** is the most powerful data processing system at that time. The fully-transistorized system has computing speeds six times faster than those of its vacuum-tube predecessor, the IBM 709. Although the IBM 7090 is a general purpose data processing system, it is designed with special attention to the needs of the design of missiles, jet engines, nuclear reactors and supersonic aircraft.

  Contains more than 50,000 transistors plus extremely fast magnetic core storage. The new system can simultaneously read and write at the rate of 3,000,000 bits per second, when eight data channels are in use. In 2.18 millionths of a second, it can locate and make ready for use any of 32,768 data or instruction numbers (each of 10 digits) in the magnetic core storage. The 7090 can perform any of the following operations in one second: 229,000 additions or subtractions, 39,500 multiplications, or 32,700 divisions.

  - Basic Cycle Time: 2.18 μSecs

Second Generation (1958-1964)…

Third Generation (1964-1974)

- **Features**
  - Introduction of integrated circuits combining thousands of transistors on a single chip
  - Semiconductor memory
  - Timesharing, graphics, structured programming
  - 2 Mb memory, 5 MIPS
  - Use of cache memory
  - IBM's **System 360** - the first **family** of computers making a clear distinction between architecture and implementation

Third Generation (1964-1974)…

- The **IBM System/360** Model 91 was introduced in 1966 as the fastest, most powerful computer then in use. It was specifically designed to handle high-speed data processing for scientific applications such as space exploration, theoretical astronomy, subatomic physics and global weather forecasting. IBM estimated that each day in use, the Model 91 would solve more than 1,000 problems involving about 200 billion calculations.

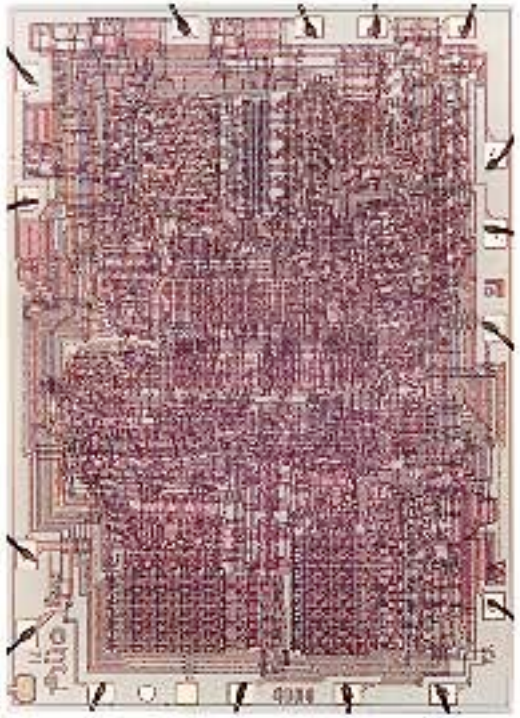
Fourth Generation (1974-present)

- **Features**
  - Introduction of **Very Large-Scale Integration (VLSI)/Ultra Large Scale Integration (ULSI)** - combines millions of transistors
  - Single-chip processor and the single-board computer emerged
  - Smallest in size because of the high component density

  - Creation of the Personal Computer (PC)
  - Wide spread use of data communications
  - Object-Oriented programming: Objects & operations on objects
  - Artificial intelligence: Functions & logic predicates

Fourth Generation (1974-present)…

- **1971** - The **4004** was the world's first universal microprocessor, invented by Federico Faggin, Ted Hoff, and Stan Mazor. With just over 2,300 MOS transistors in an area of only 3 by 4 millimeters had as much power as the ENIAC.
- 4-bit CPU
- 1K data memory and 4K program memory
- clock rate: 740kHz
- Just a few years later, the word size of the 4004 was doubled to form the 8008.



Fourth Generation (1974-present)…

- **1974 – 1977** the first personal computers – introduced on the market as kits (major assembly required).
  - **Scelbi** (SCientific, ELectronic and BIological) and designed by the Scelbi Computer Consulting Company, based on Intel's 8008 microprocessor, with 1K of programmable memory, Scelbi sold for $565 and came, with an additional 15K of memory available for $2760.
  - **Mark-8** (also Intel 8008 based) designed by Jonathan Titus.
  - **Altair** (based on the the new Intel 8080 microprocessor), built by MITS (Micro Instrumentation Telemetry Systems). The computer kit contained an 8080 CPU, a 256 Byte RAM card, and a new Altair Bus design for the price of $400.

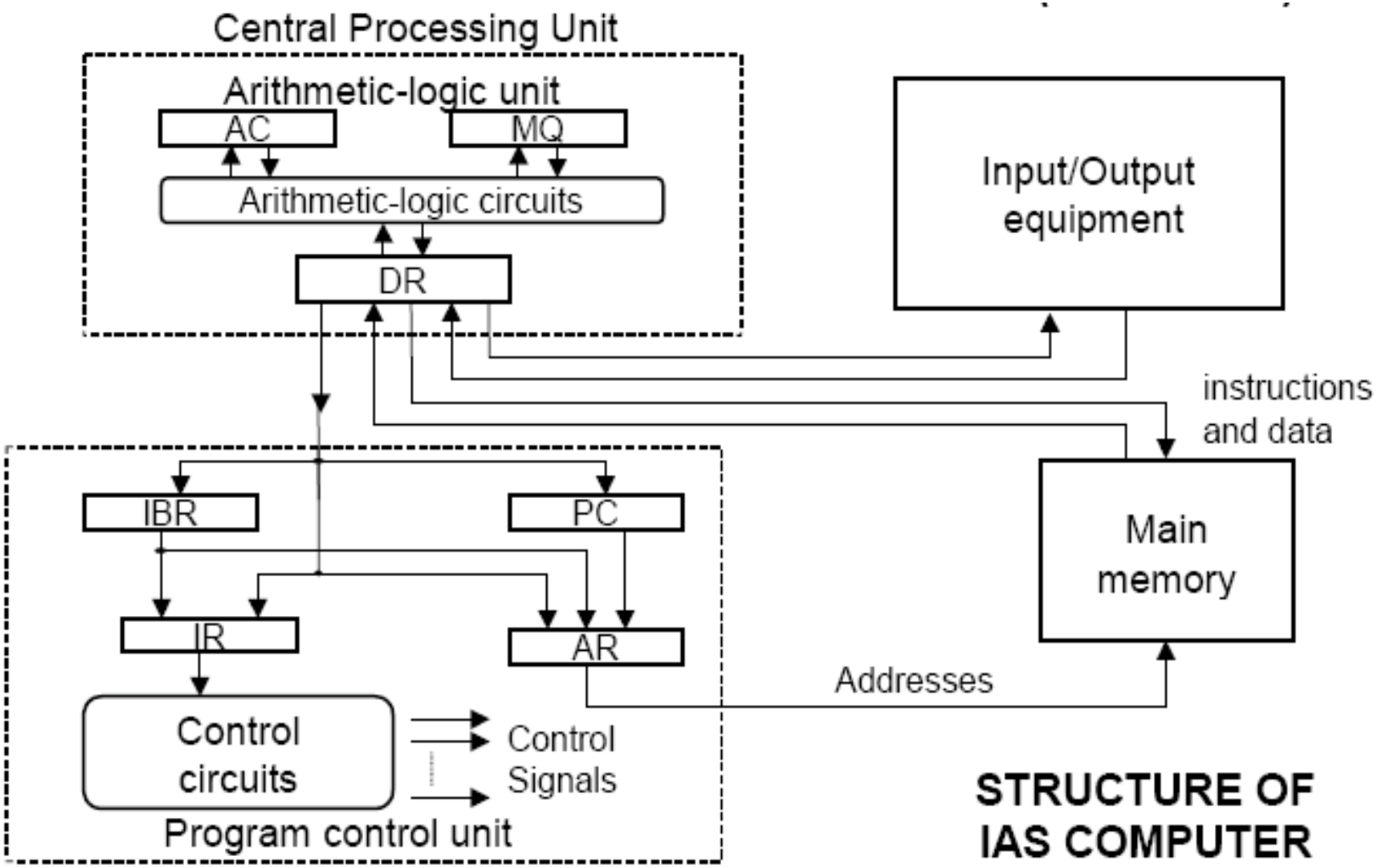
Fourth Generation (1974-present)…

- **1976** - Steve Wozniak and Steve Jobs released the **Apple I** computer and started Apple Computers. The Apple I was the first single circuit board computer. It came with a video interface, 8k of RAM and a keyboard. The system incorporated some economical components, including the 6502 processor (only $25 dollars - designed by Rockwell and produced by MOS Technologies) and dynamic RAM.
- **1977** - **Apple II** computer model was released, also based on the 6502 processor, but it had color graphics (a first for a personal computer), and used an audio cassette drive for storage. Its original configuration came with 4 kb of RAM, but a year later this was increased to 48 kb of RAM and the cassette drive was replaced by a floppy disk drive.
- **1977 - Commodore PET** (Personal Electronic Transactor) was designed by Chuck Peddle, ran also on the 6502 chip, but at half the price of the Apple II. It included 4 kb of RAM, monochrome graphics and an audio cassette drive for data storage.

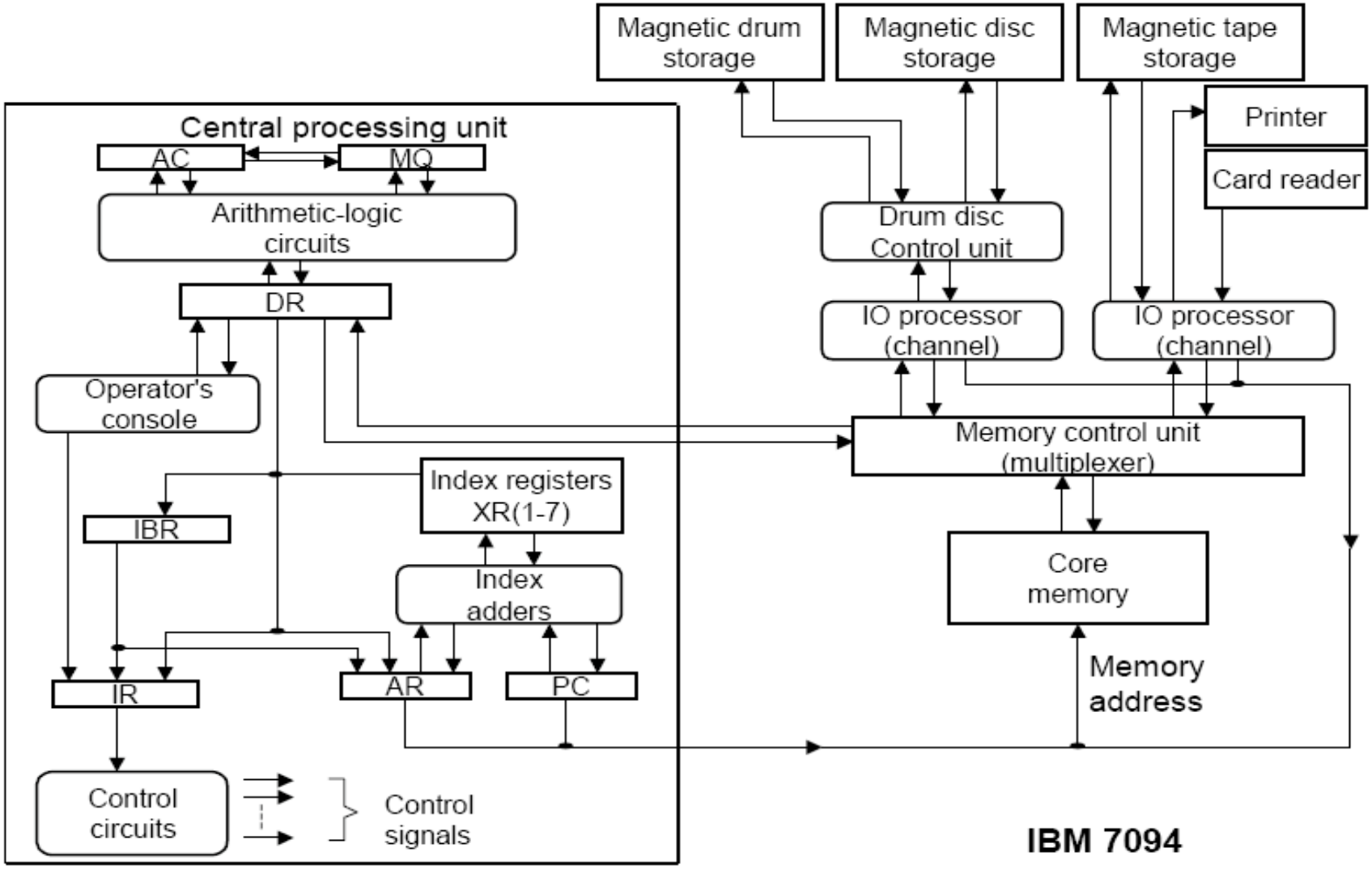
Fourth Generation (1974-present)…

- **1981 -** IBM released their new computer **IBM PC** which ran on a 4.77 MHz Intel 8088 microprocessor and equipped with 16 kilobytes of memory, expandable to 256k. The PC came with one or two 160k floppy disk drives and an optional color monitor.
  - first one built from off the shelf parts (called open architecture) and marketed by outside distributors

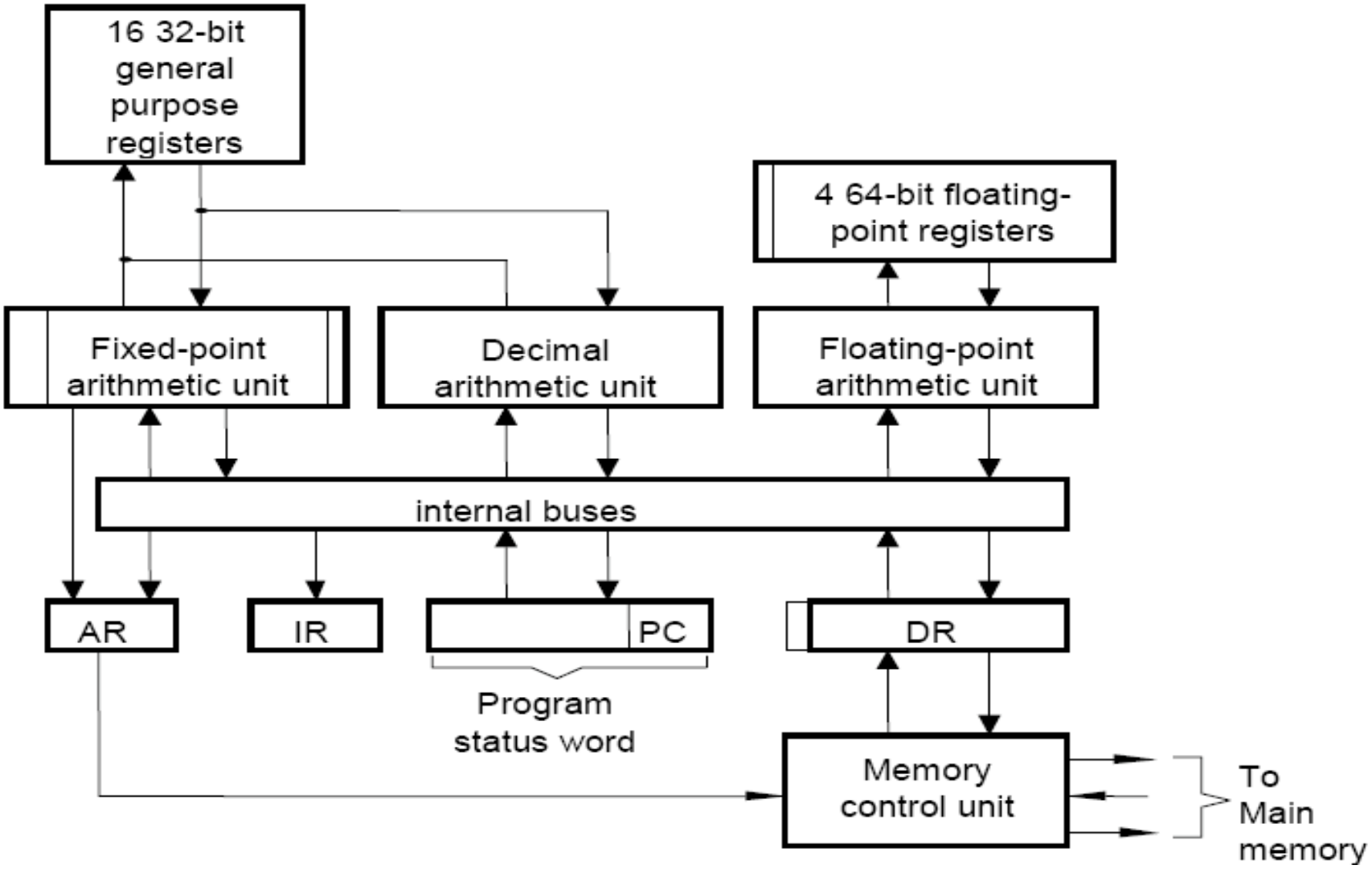
First Generation (1945-1958)



Second Generation (1958-1964)

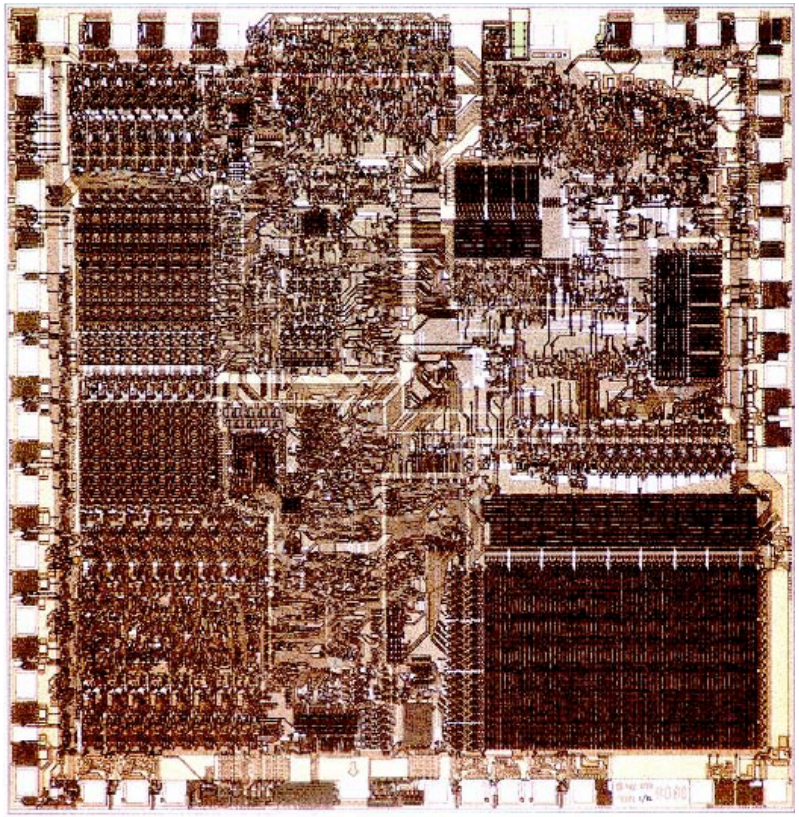


Third Generation (1964-1974)



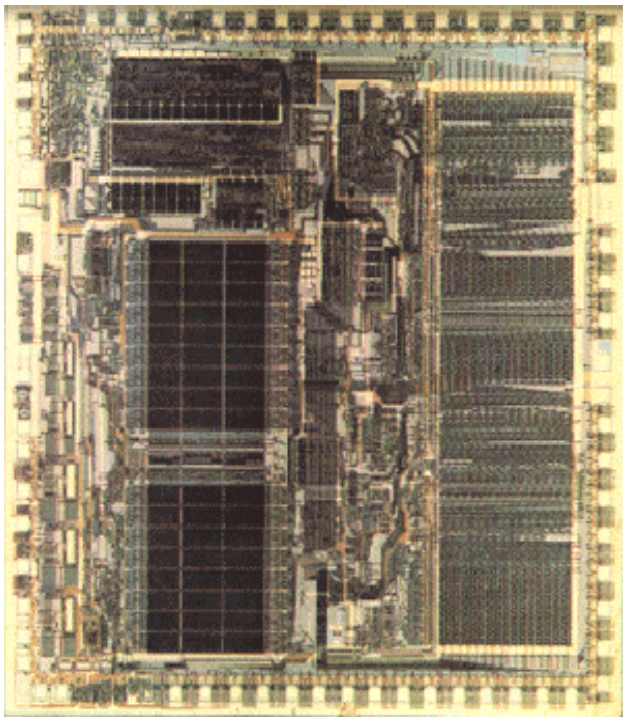
1974-present

- **Intel 8080**
  - **8-bit Data**
  - **16-bit Address**
  - **6 μm NMOS**
  - **6K Transistors**
  - **2 MHz**
  - **1974**



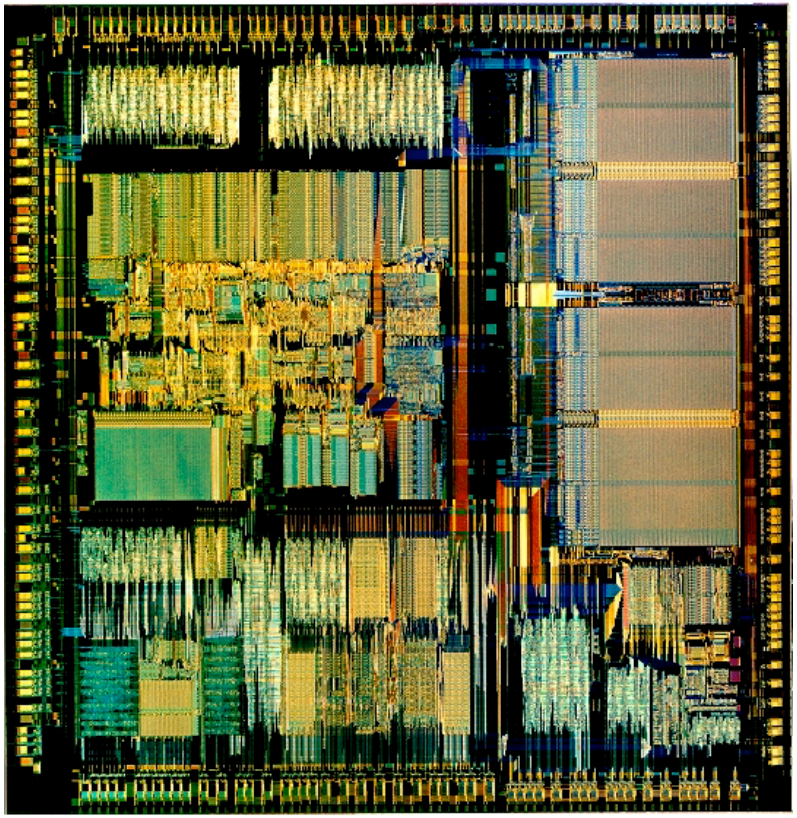
1974-present

- **Motorola 68000**
  - **32 bit architecture internally, but 16 bit data bus**
  - **16 32-bit registers, 8 data and 8 address registers**
  - **2 stage pipeline**
  - **no vertual memory support**
  - **68020 was fully 32 bit externally**
  - **1979**



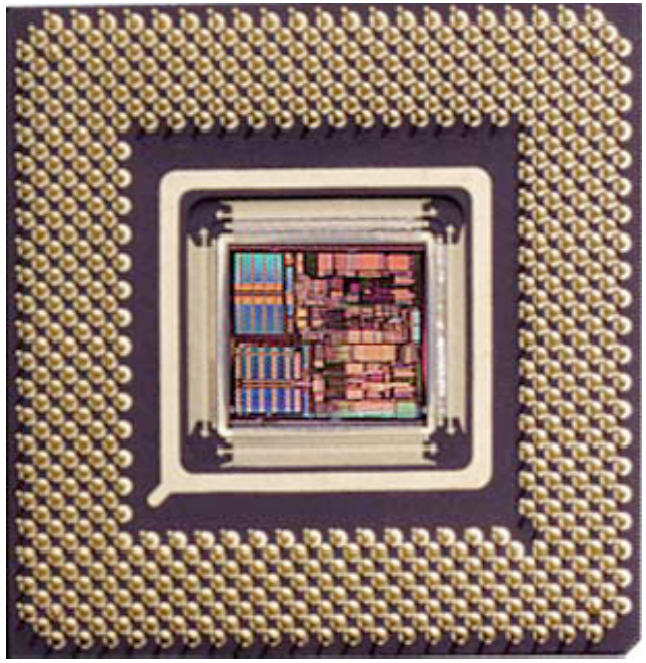
1974-present

- **Intel386 CPU**
  - **32-bit Data**
  - **improved addressing**
  - **security modes** (kernal, system services, application services, applications)
  - **1985**



1974-present

- **Alpha 21264**
  - 64-bit Address/Data
  - Superscalar
  - Out-of-Order Execution
  - 256 TLB entries
  - 128KB Cache
  - Adaptive Branch Prediction
  - 0.35 μm CMOS Process
  - 15.2M Transistors
  - 600 MHz



- http://inventors.about.com/library/blcoindex.htm
- http://www.chick.net/wizards/whirlwind.html
- http://www.columbia.edu/acis/history/36091.html
- http://bwrc.eecs.berkeley.edu/CIC/archive/cpu_history.html
- http://www-03.ibm.com/ibm/history/exhibits/mainframe/mainframe_PP7090.html

Computer Networks

Paul Maynard

What is a network?

- A network can be defined as a system of interconnected communication stations.(1)
- Communication stations are commonly referred to as nodes. They can be:
  - Computers
  - Printers
  - Terminals

What good is a network?

- Resource sharing
  - Printers
  - Storage
- Movement of data
  - Quick movement of information
- Cost cutting

Network Topologies

- Mesh
- Bus
- Ring
- Star
- Hybrid

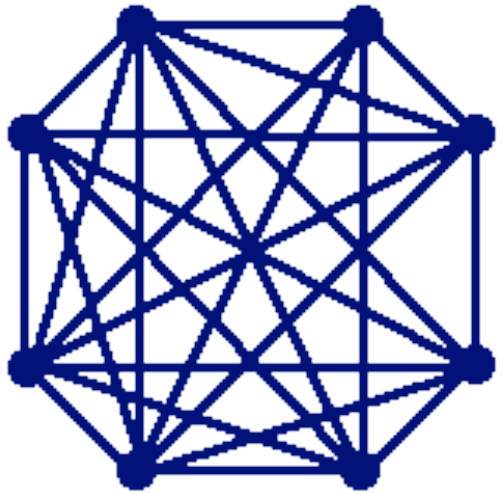
Mesh Topology

Advantages

- Directly connect the nodes.
- Simple to implement and troubleshoot

Disadvantages

- Connections
  - (x * x – 1)/2
    - X is the number of nodes



Mesh Topology

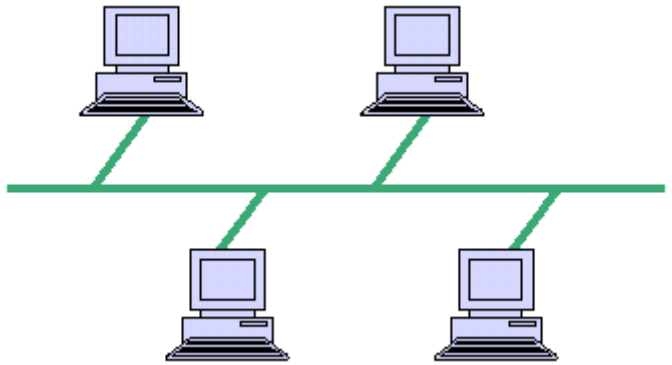
Bus Topology

Advantages

- Scalability
- Shorter cables

Disadvantages

- Trouble shooting



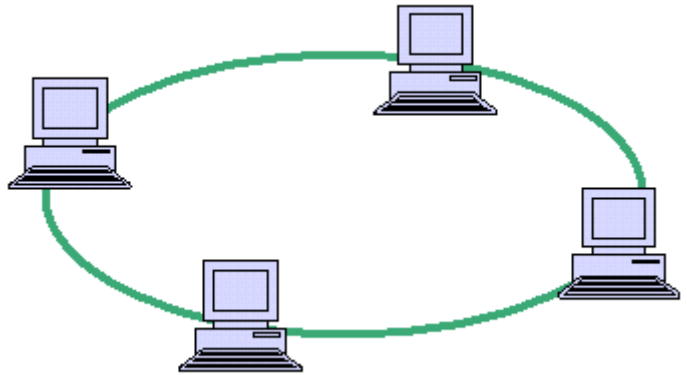
Ring Topology

Advantages

- Distance between nodes
- Higher throughput than bus

Disadvantages

- Failure
- Trouble shooting
- Slow node problem



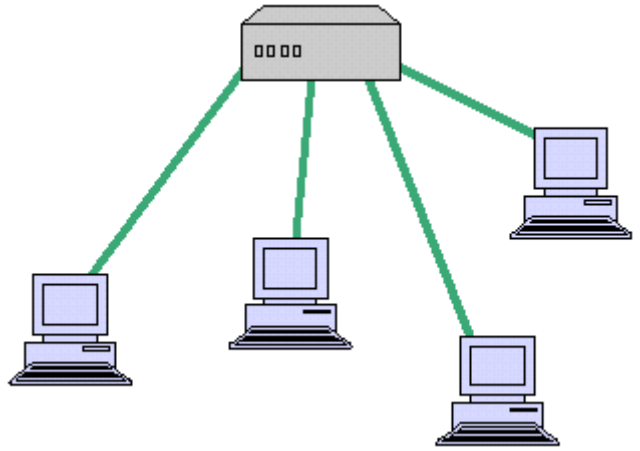
Star Topology

Advantages

- Management
- Scalability

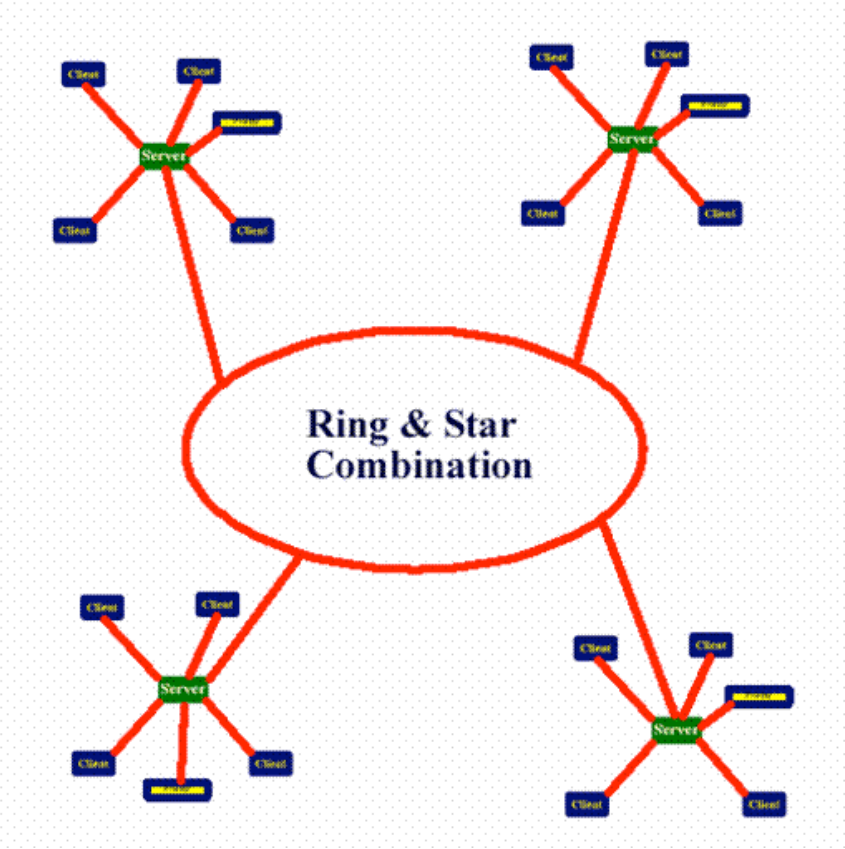
Disadvantages

- Single point of Failure



Hybrid Topology

- Other topologies combined
- Came from want to connect existing networks together



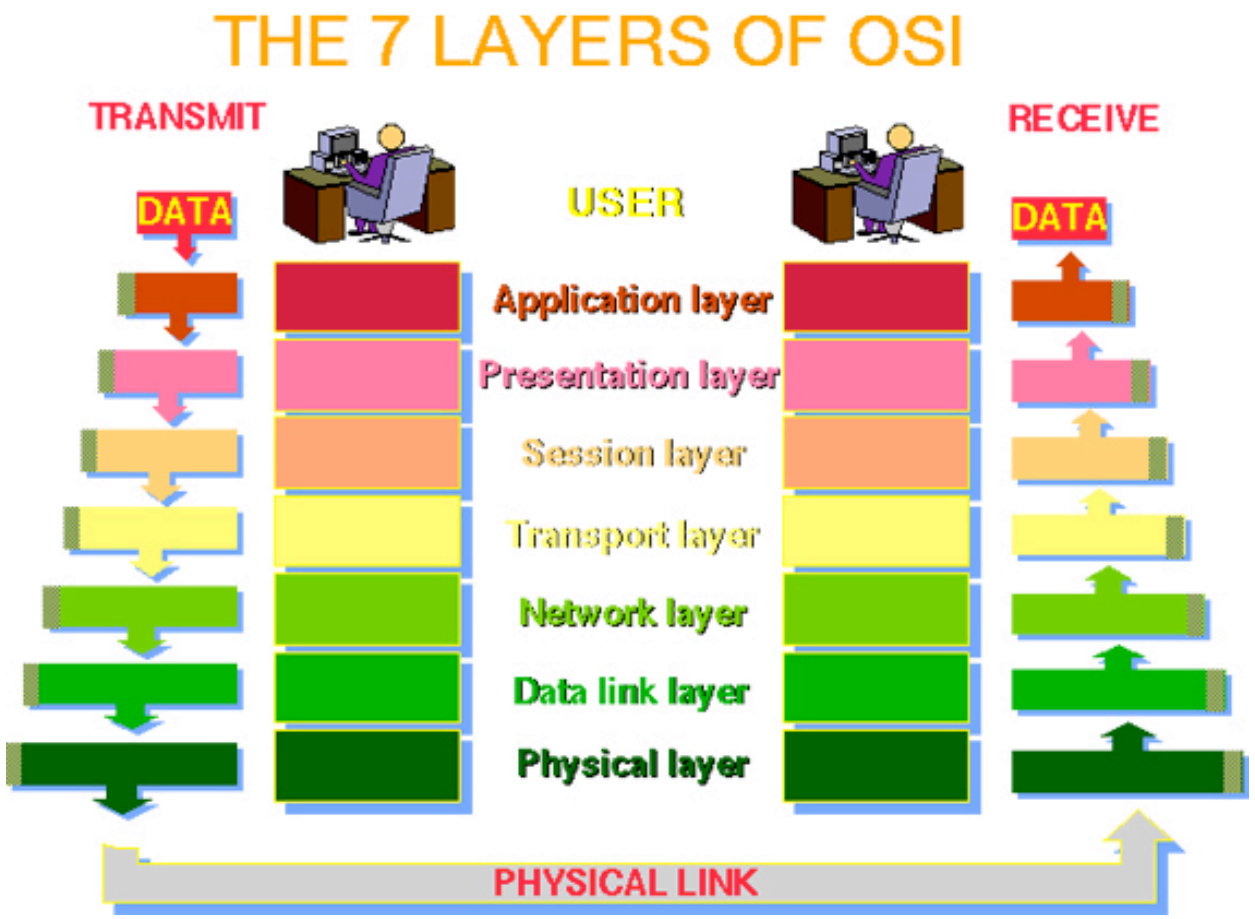
The OSI Model

- What is it?
  - Open systems interconnection
  - General specifications for networks
- Who created it?
  - Mainly ISO
  - IEEE, ANSI, EIA, and others

The 7 OSI Model Layers

- Physical
- Data link
- Network
- Transport
- Session
- Presentation
- Application

The OSI Model



Physical layer

- The conversion of the data into a sending stream
  - Electrical pulses
  - Radio waves
  - Light pulses
- Defines the transmission medium
  - Copper, air, or fiber optics
- Defines how the transmission will be done

Cable Type Comparisons

Cable Type Comparisons							
Type	**Speed**	**Distance**	**Installation**	**Interference**	**Cost**	**# of nodes per segment**	**# of nodes per network**
10BaseT	10 Mbps	100 meters	Easy	Highly susceptible	**Least expensive**	1 computer	
100BaseT	100 Mbps	100 meters	Easy	Highly susceptible	More expensive than 10BaseT		
STP	16 to 155 Mbps	100 meters	Moderately Easy	Somewhat resistant	More expensive than Thinnet or UTP		
10Base2	10 Mbps	185 meters	Medium Difficulty	Somewhat resistant	Inexpensive	30	1024
10Base5	10 Mbps	500 meters	More difficult than Thinnet	More resistant than most cable	More expensive than most cable	100	300
Fiber Optic	100 Mbps to 2 Gbps	2000 meters	Most difficult	Not susceptible to electronic interference	**Most expensive type of cable**		

Data link and Network layers

Data link

- Error control
- Logical synchronization
- Data flow
- Method of data transmission

Network

- Message delivery
- Connection control
- Controls the route
- Does the name translation

Transport and session layer

Transport

- Makes sure the message is received without error
- Divides the message into packets
- Controls the quality of service

Session

- Connection and disconnection
- Binding
- Interface between connections

Presentation and application layer

Presentation

- Encryption
- Type conversions

Application

- Basic application protocol
  - HTTP
  - SMTP
  - FTP
  - IRC

Network classifications

- LAN
- GAN, WAN, and MAN
- CAN

LAN – Local Area Network

- Small usually connects nodes in a single building.
- Most common LAN technologies
  - Ethernet
  - Token Ring
  - FDDI – Fiber Distributed Data Interface

GAN, WAN, MAN, and CAN

- Larger classifications of networks
- Spread over a larger areas

- CAN – Campus area network
- MAN – Metropolitan area network
- WAN – Wide area network
- GAN – Global area network

References

- Data & Network Communications – Miller
- http://compnetworking.about.com/od/networkdesign/l/aa041601a.htm
- http://www.webopedia.com/quick_ref/OSI_Layers.asp
- http://www.geocities.com/SiliconValley/Monitor/3131/ne/osimodel.html
- http://students.juniata.edu/rappdl3/hybrid.gif

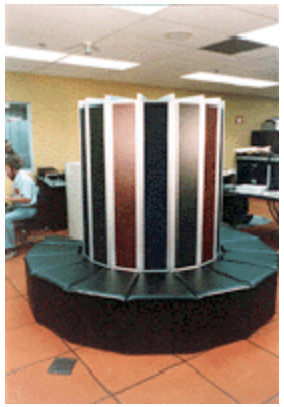
Supercomputer

Nathan Liang

Overview

- What are Supercomputers?
  - Most advance and powerful
- Why were they built?
  - to solve problems that are too complex or too massive for standard computers
- Some supercomputers are single computers consisting of multiple processors; others are clusters of computers that work together

Supercomputers introduced in the 1960s.



Cray-1 was designed by Seymour Cray at Control Data Corporation (CDC).
It could perform over a hundred million arithmetic operations per second.

- CDC's early machines were very fast single processors.
- In the 1970s most supercomputers were dedicated to running a vector processor, and many of the newer players developed their own such processors at lower price points to enter the market.
- In the later 1980s and 1990s, attention turned from vector processors to massive parallel processing systems with thousands of simple CPUs; some being off the shelf units and others being custom designs.
- Today, parallel designs are based on "off the shelf" RISC microprocessors, such as the PowerPC or PA-RISC.

Supercomputer Architecture

- Vector computers use a very fast data “pipeline” to move data from components and memory in the computer to a central processor.
  - a single super-fast processor with all the computer's memory allocated to its operation
- Parallel computers use multiple processors, each with their own memory banks, to 'split up' data intensive tasks.
  - A parallel machine has multiple processors, each with its own memory

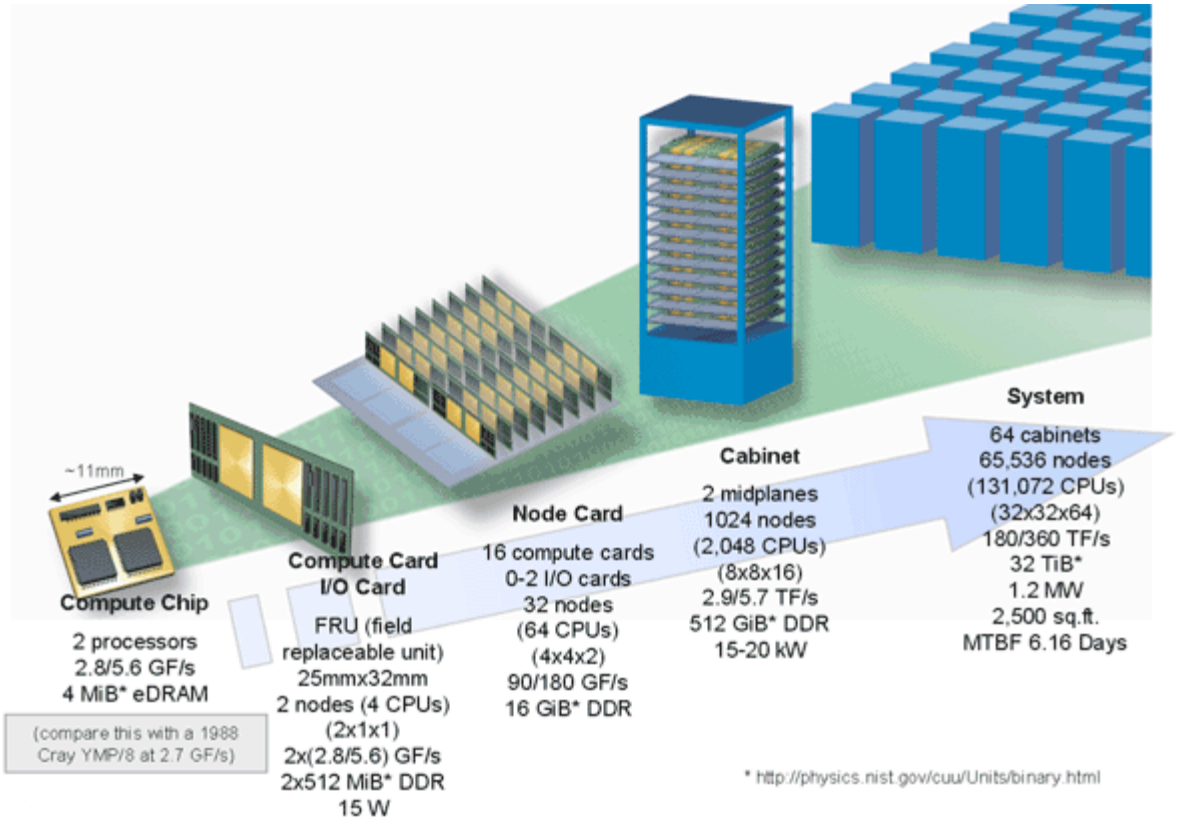
- Vector machines are easier to program, while parallel machines, with data from multiple processors (in some cases greater than 10,000 processors), can be tricky to orchestrate

Some supercomputer designs feature

- **network co-processors**
  - Problem: When sending and receiving data at the rate necessary for high-performance networking, it's common for a single processor to become heavily loaded with communications interrupts that take away too many cycles from primary computing tasks.

- **Supercomputer Interconnects**
  - utilize specialized network interfaces.
  - support high bandwidth and very low latency communication.

Blue Gene


~11mm
Compute Chip
2 processors
2.8/5.6 GF/s
4 MiB* eDRAM
(compare this with a 1988 Cray YMP/8 at 2.7 GF/s)
Compute Card
I/O Card
FRU (field replaceable unit)
25mmx32mm
2 nodes (4 CPUs)
(2x1x1)
2x(2.8/5.6) GF/s
2x512 MiB* DDR
15 W
Node Card
16 compute cards
0-2 I/O cards
32 nodes
(64 CPUs)
(4x4x2)
90/180 GF/s
16 GiB* DDR
Cabinet
2 midplanes
1024 nodes
(2,048 CPUs)
(8x8x16)
2.9/5.7 TF/s
512 GiB* DDR
15-20 kW
System
64 cabinets
65,536 nodes
(131,072 CPUs)
(32x32x64)
180/360 TF/s
32 TiB*
1.2 MW
2,500 sq.ft.
MTBF 6.16 Days
* http://physics.nist.gov/cuu/Units/binary.html

- Interconnects join nodes inside the supercomputer together
  - Compute nodes
  - I/O nodes
  - Service nodes and network nodes
  - Special purpose nodes
- Supercomputer nodes fit together into a network topology
  - Hypercube, mesh, and torus
- Message Passing
  - Supercomputers that utilize message passing require routing to ensure the individual pieces of a message are routed from source to destination through the topology without creating hotspots (bottlenecks).
- Supercomputers utilize various network protocols.
  - Application data communication generally take place at the physical and data link layers.
  - I/O and communications with external networks utilize technologies like HIPPI, FDDI, and ATM as well as Ethernet.
  - Do not utilize wireless

Processing Speed

- Supercomputer computational power is rated in FLOPS (Floating Point Operations Per Second)
- The first commercially available supercomputers reached speeds of 10 to 100 million FLOPS
- the Cray C90 (built in the mid to late 1990s) has a processing speed of *only* 8 gigaflops. It can solve a problem, which takes a personal computer a few hours, in .002 seconds.

- **vector processor**
  - a CPU design that is able to run mathematical operations on a large number of data elements very quickly
  - Cray90 can do 7.8 gigaflops

- **Parallel computing**
  - the simultaneous execution of the *same task* (split up and specially adapted) on multiple processors in order to obtain faster results.
  - Cray SX-6 A design of using parallel computing with multiple Vector processors

- **BeoWulf**
  - Originally developed by Donald Becker at NASA
  - a design for high-performance parallel computing clusters on inexpensive personal computer hardware
  - A *Beowulf cluster* is a group of usually identical PC computers running FreeBSD or another open source Unix operating system, such as Linux or OpenBSD
  - They are networked into a small TCP/IP LAN, and have libraries and programs installed which allow processing to be shared among them.
  - It uses parallel processing libraries include MPI (Message Passing Interface) and PVM (Parallel Virtual Machine) to permit the programmer to divide a task among a group of networked computers, and recollect the results of processing

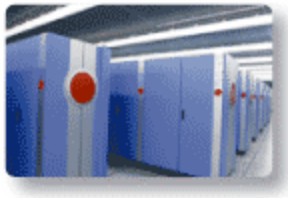
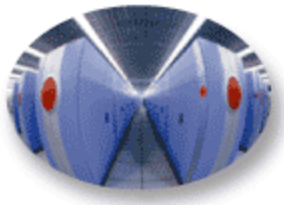
University of Kentucky KLAT2

- University of Kentucky' supercomputer brought down the price to $650 per GigFlops.
- To function as a supercomputer, the PCs within a cluster need to be interconnected by a high-performance network.
- Instead of expensive gigabit network hardware, KLAT2's network uses lots of cheap 100Mb/s Ethernet hardware in a new configuration called a "Flat Neighborhood" network

Some Fastest Supercomputers

- Today's fastest supercomputers include IBM's Blue Gene and NASA’s Columbia, NEC Earth Simulator, and Barcelona’s (Spain) MareNostrum
- These supercomputers are usually designed to carry out specific tasks
  - Example:

    NEC Earth Simulator is located at the Earth Simulator Center in Kanazawa-ku (ward), Yokohama-shi, Japan. The computer is capable of 35.86 trillion (35,600,000,000,000) floating-point calculations per second, or 35.86 TFLOPS. The system was developed for NASDA, JAERI, and JAMSTEC in 1997 for climate simulation.

It consists of 640 nodes with eight vector processors and 16 gigabytes of computer memory at each node, for a total of 5120 processors and 10 terabytes of memory. The system has 700 terabytes of disk storage and 1.6 petabytes of mass storage in tape drives

– Another example:

  IBM's Blue/Gene

  this high-end computer architecture can be scaled to the hundreds of teraflops

- Supercomputers are called upon to perform the most compute-intensive tasks of modern times.
- Modeling of real world complex systems such as fluid dynamics, weather patterns, seismic activity prediction, and nuclear explosion dynamics represent the most modern adaptations of supercomputers.

Reference

- Top 500 (www.top500.org)
- Cray Inc (www.Cray.com)
- BeoWulf (www.beowulf.org)
- Earth Simulator (www.es.jamstec.go.jp)

The Cell Chip

- Cell Chip Specifications
- Inside The Cell Chip
- Cell Vs. AMD64, Intel
- What the Future Holds

Speed and Ability

- 64 Bit Instruction Set
- Clock Speed over 4GHz
- 70 Watts @ 4GHz with 1.1V
- 180 Watts @ 5.6 GHz with 1.4V

Floating Point Operations

- 256 GFlops throughput from its 8 SPE's (SP)
  - Speed for the PS3
  - not fully IEEE754 compliant / rounding modes
- 25 GFlops throughput from its 8 SPE's (DP)
  - Accuracy for computer systems
  - IEEE859 compliant / IBM's Main contribution

Comparing Cell Chip To Competitors

- Each Cell Chip
  - Capable of 25~30 GFlops/s (DP)
- Each CPU on Earth Simulator
  - Capable of 8 GFlops/s (DP)
- Each Itanium2 Processor
  - Capable of 6 GFlops/s (DP)
- Each IBM Power 4 Processor
  - Capable of 5.2 GFlops/s (DP)

I/O Speeds

- 25 GB/s memory bandwidth to Rambus XDR Ram
- 44.8 GB/s outbound and
  32 GB/s inbound bandwidth
  for I/O interface total bandwidth of 76.8 GB/s
- To Balance processing power and data bandwidth each cell chip has Redwood Rambus Asic cell on one side and a Yellowstone high bandwidth DRAM on the other side.

Diagram Names

- (PPE) - The PowerPC Processing Element
- (SPE) - The Synergistic Processing Element
- (EIB) - The internal Element Interconnect Bus
- (MIC) - The shared Memory Interface Controller
- The L2 Cache – Main Memory 512KB
- The FlexIO interface

Inside the Cell Chip

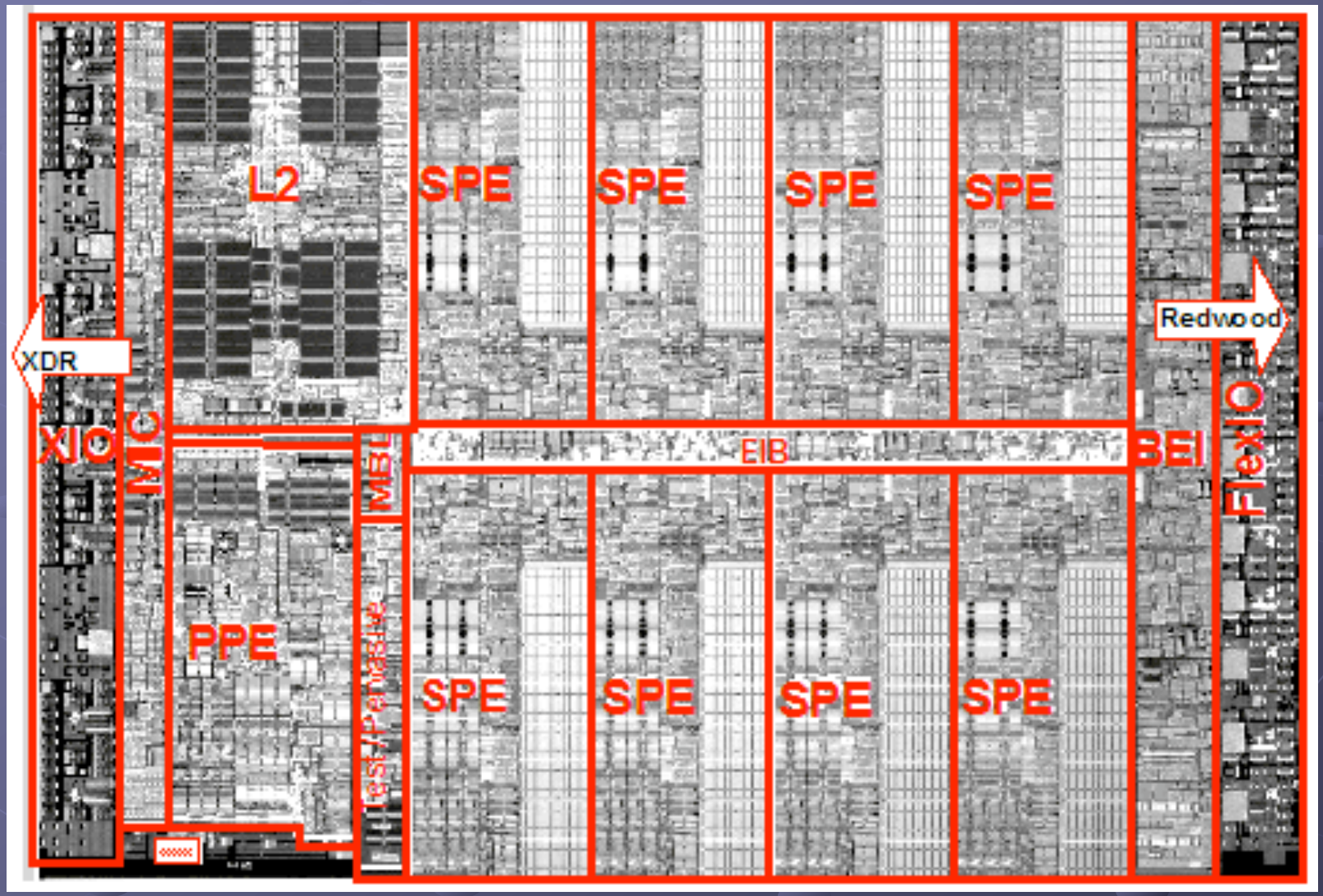
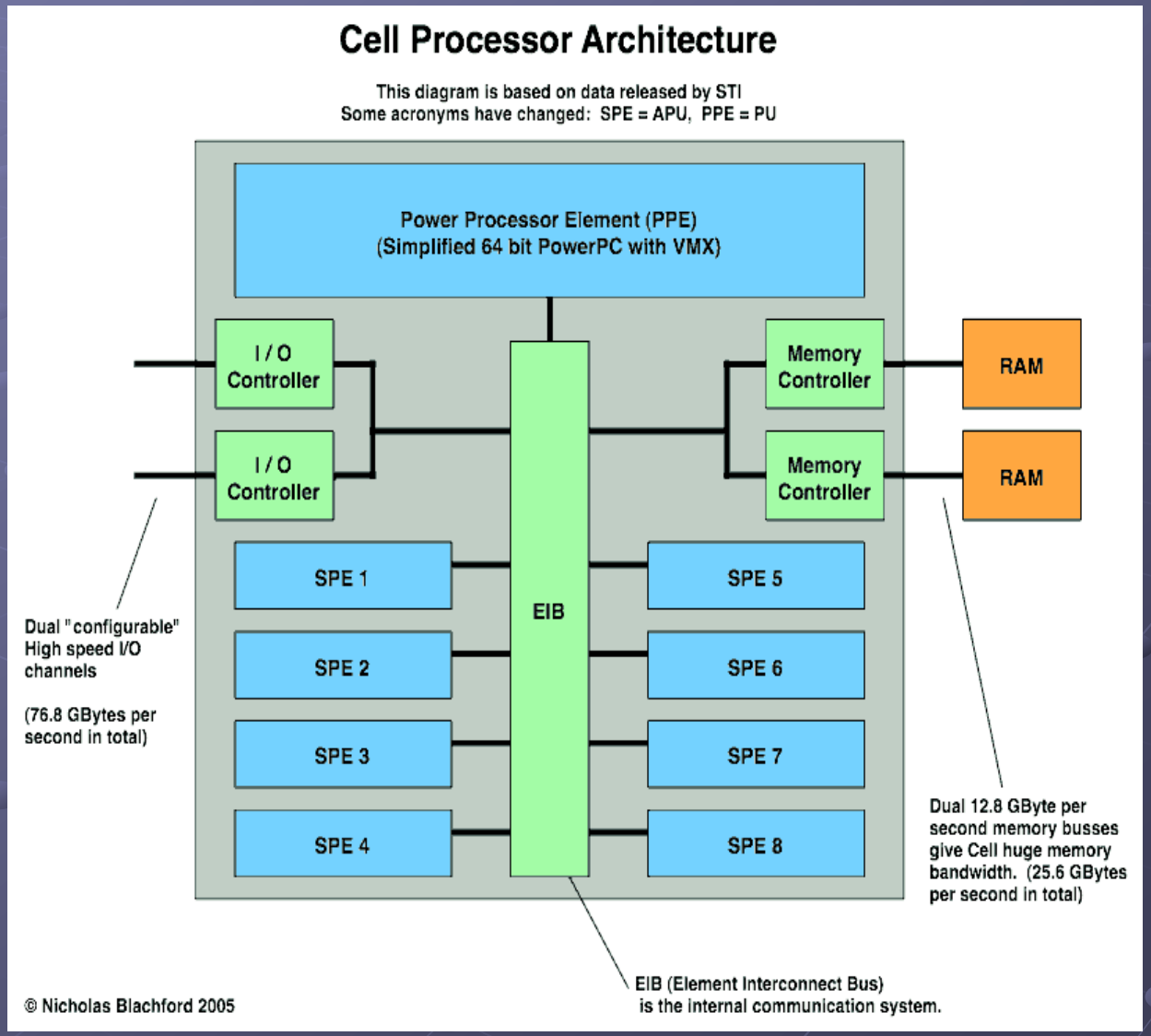


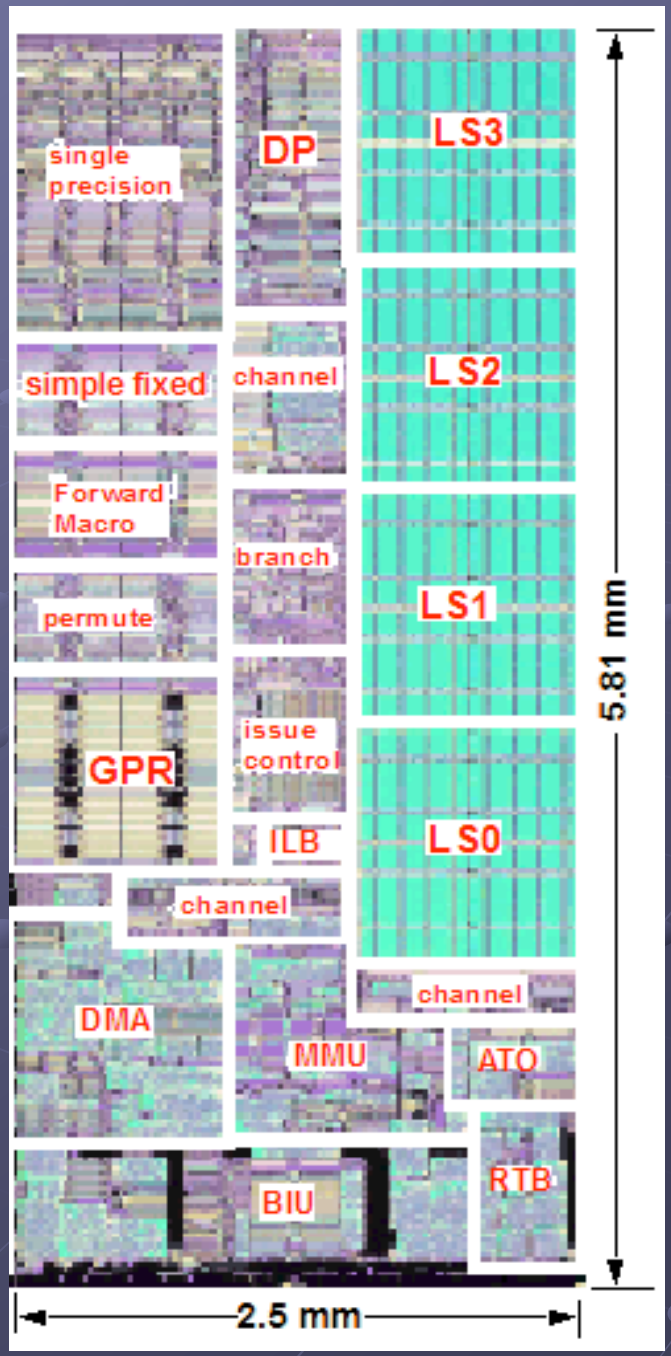
Diagram View

- PPE – 64bit PowerPC? core
  --(most likely)
- PS3 will most likely have 3
- PPE is the Controller of Each SPE



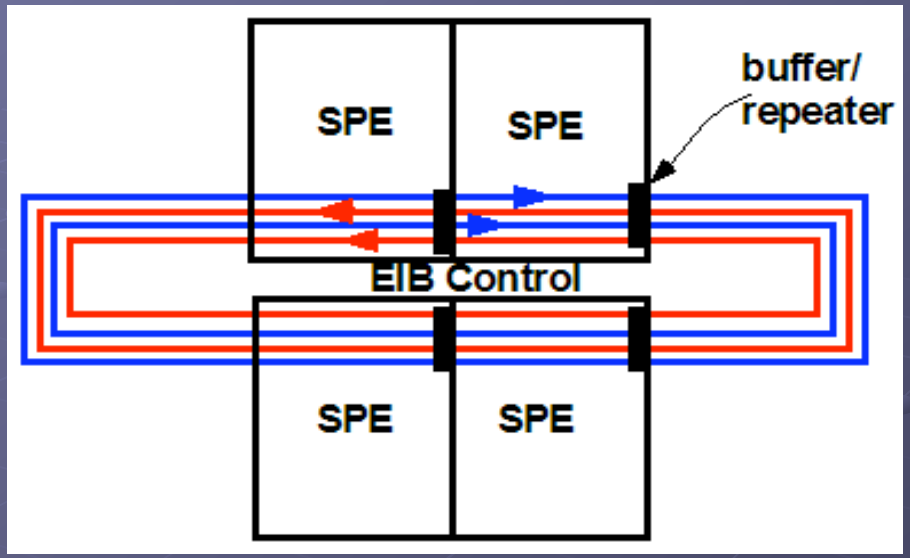
SPE Architecture

- Each SPE Has
  - 128x128 bit registers
  - 4 floating point units
  - 256 KB local memory instead of a cache (LS) Load Store 64KB each
- Local Memory
  - SPE operates on registers which are read from or written to the Load Store (LS)
  - Local Store Memory can access main memory in blocks of 1024 bits
  - SPE's cannot act directly on Main memory



Element Interconnect Bus

- EIB – Ties together all of the processing, memory, and I/O elements in the Cell
- Four concentric rings that is routed through portions of the SPE
- Each ring is 128 bit wide interconnect
- Data going to the (LS) unit can be off-loaded at the BIU outgoing data is placed on the EIB by the BIU



- Data moving from one SPE to another requires a repeater
- EIB is specifically geared toward the scalability of the Cell processor

Cell Vs. Itanium, AMD 64

- AMD 64 Opteron
  - 64 bit
  - 2.4GHz
  - 4 GFlops/s
  - 25.6 GB/s bandwidth I/O
- Itanium 2
  - 64 bit
  - 1.5GHz
  - 6 GFlops/s
  - 6.4 GB/s bandwidth I/O

Cell Chip
- 64 bit
- 4.7GHz
- 25~30 GFlops/s
- 76.8 GB/s bandwidth I/O

A Bad Future For Cell?

- The PC has seen off every competitor that has crossed its path no matter how good the OS or hardware was.
- All major advances have been taken from the competitors and used for themselves to stabilize their market again.
- The major problem that the Cell has is software.
- No matter how powerful the Cell is, it is worthless without the proper software.

Good Future for Cell?

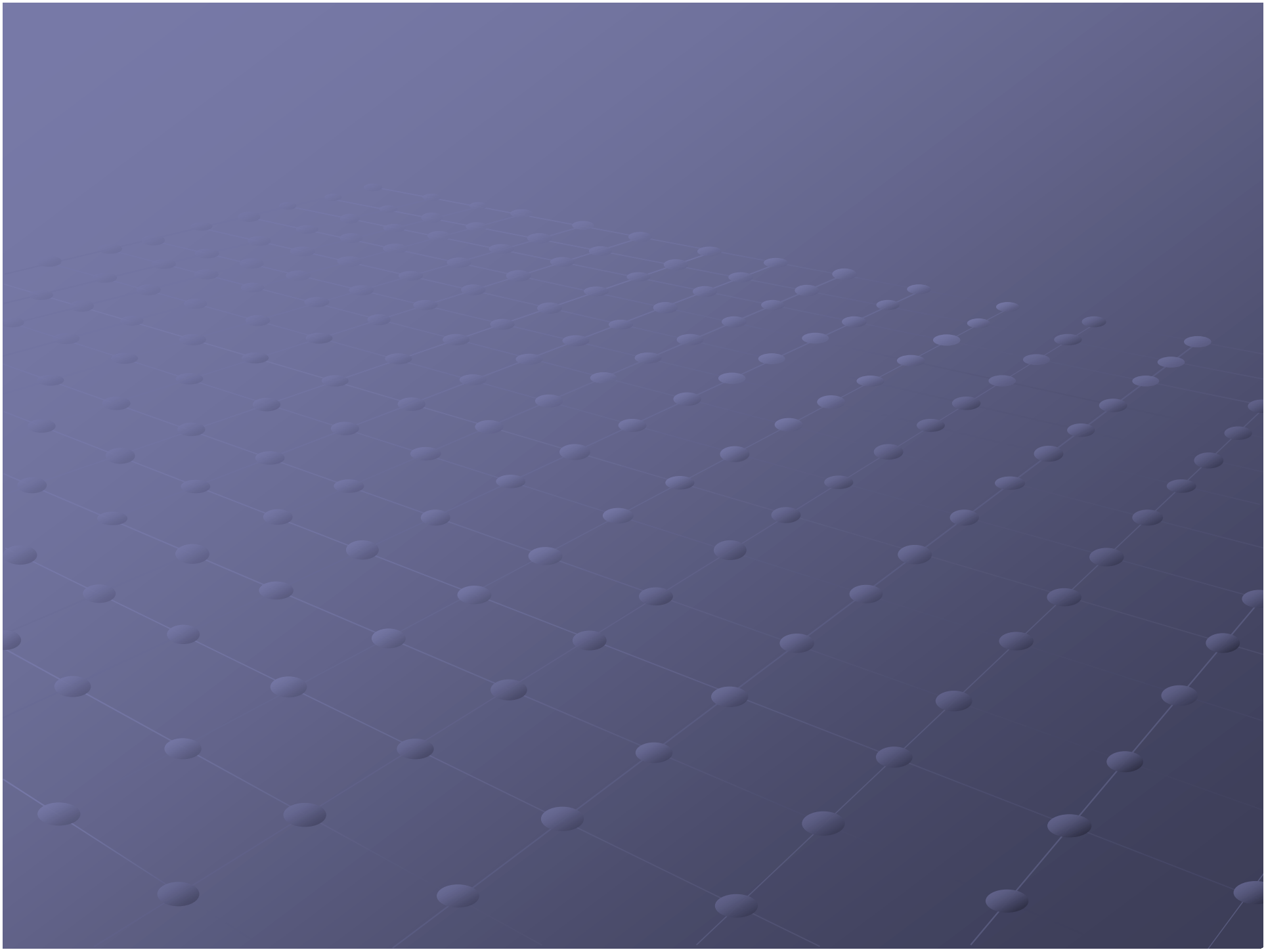
- Linux - could save and make a place for Cell with open source programmers.
- Cost - since Sony, IBM, and Toshiba are all mass producing cost will be cheap
- Power - this chip will run on small amount of power so ideal for smaller devices.
- Possible to Put a cell chip on a PC card and run PS3 games on it. Even low end computers could run high end games easily.

Were Will Cell Start

- Sony – PS3, also using it in HDTV's Along with other visual devices
- Toshiba – HDTV's along with other visual devices
- IBM – Work Stations, Servers, Supercomputers.

Were Will Cell End

- After they produce for these products the list could be endless, with the right market and software the cell could take over the computer processor for the next few years.
- Or it could not move into the computer field at all and overdone by AMD or Intel in a few years
- But hopefully this new processor will not miss the market like the Itanium and many other processors did and will make a huge impact on the future of super computing.



Resources

- Real World Tech

http://www.realworldtech.com

- AMD home site

http://www.AMD.com

- Intel home Site

http://www.intel.com

- Nicholas Blachford

http://www.blachford.info/computer/Cells/Cell0.htm