Chapter 1

Fundamentals of Computer Design

浙江大学
ZheJiang University

Topics in Chapter 1

1.1 Introduction
1.2 Classes of computers
1.3 Defining computer architecture and What's the task of computer design?
1.4 Trends in Technology
1.5 Trends in power in Integrated circuits
1.6 Trends in Cost
1.7 Dependability
1.8 Measuring, Reporting and summerizing Perf.
1.9 Quantitative Principles of computer Design
1.10 Putting it altogether

Lec2

What’s CA?

What’s a CA designer’s task ?

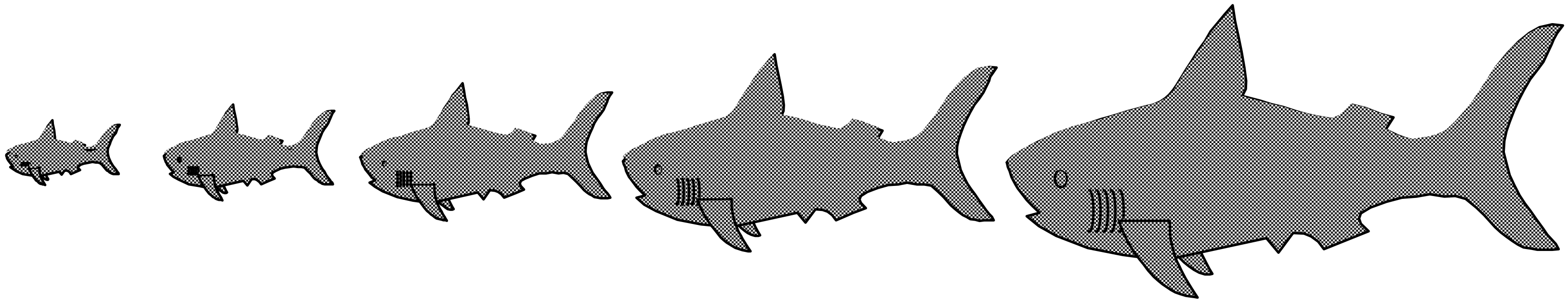
What's Computer Architecture

- **Technology**
  - A key factor in the long run of CPU performance improvement.

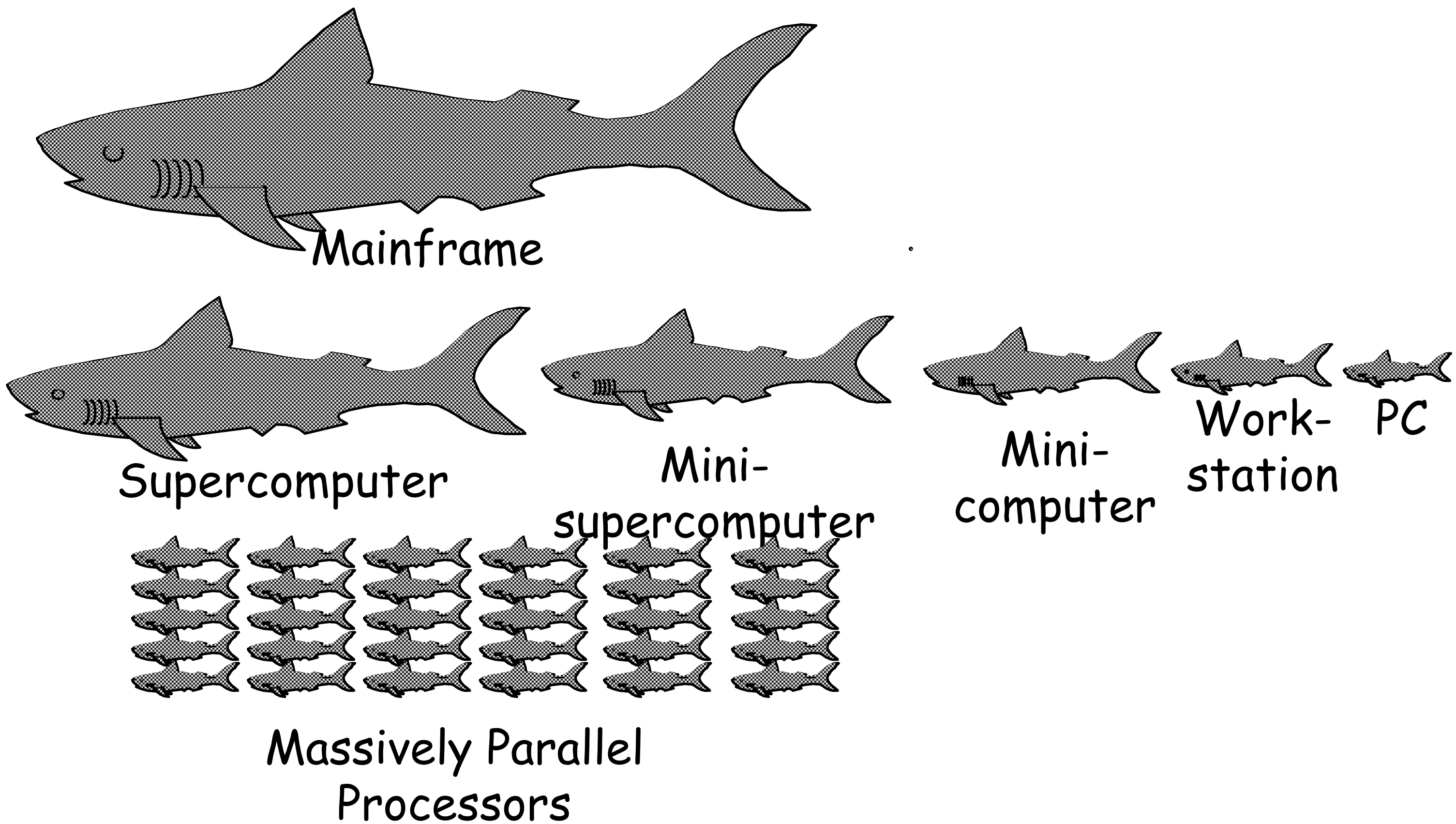
History of the Computer

- Original:

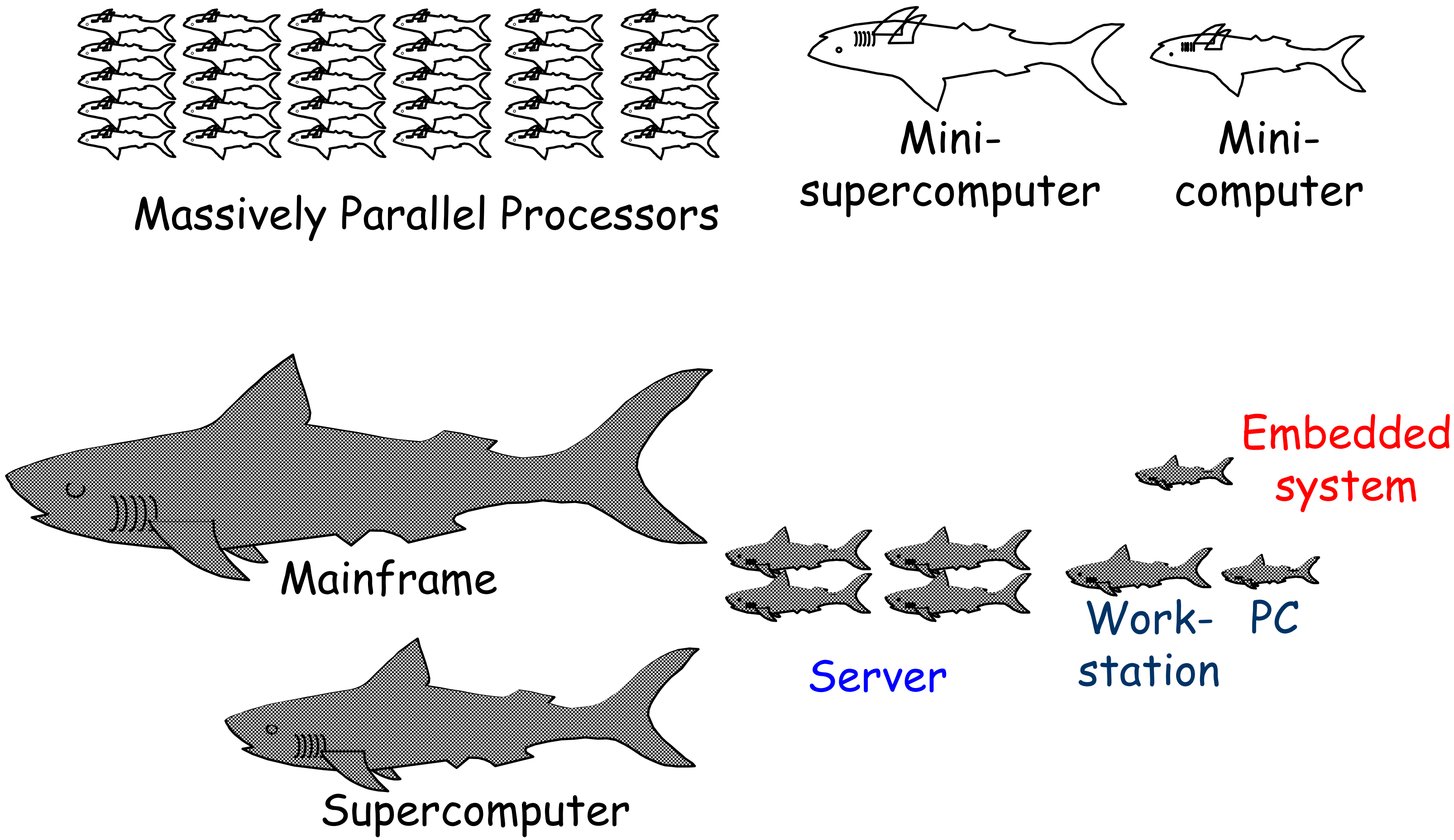
Big Fishes Eating Little Fishes



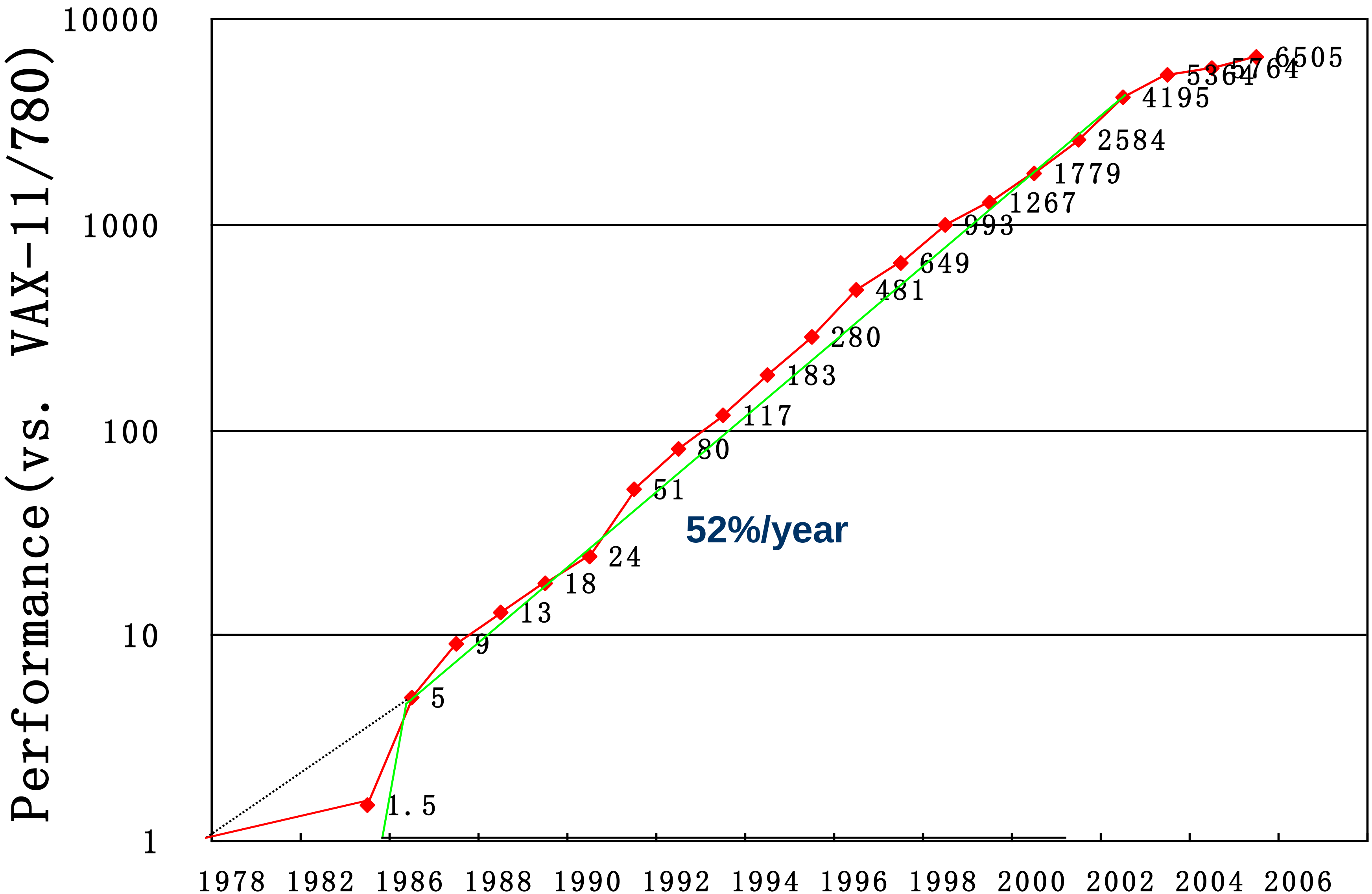
1988 Computer Food Chain



1998 Computer Food Chain



Incredible CPU performance improvement



What the figure tell us ?

- 25%: Technological improvements more steady than progress in computer architecture.

- 52%: After RISC emergence, computer design emphasized both architectural innovation and efficient use of technology improvements.
  - Computer Architecture plays an important role in performance improvement
  - **Pipeline, dynamic scheduling, ooo, branch prediction, speculation, superscalar, VLIW, prediction instructions, ……**

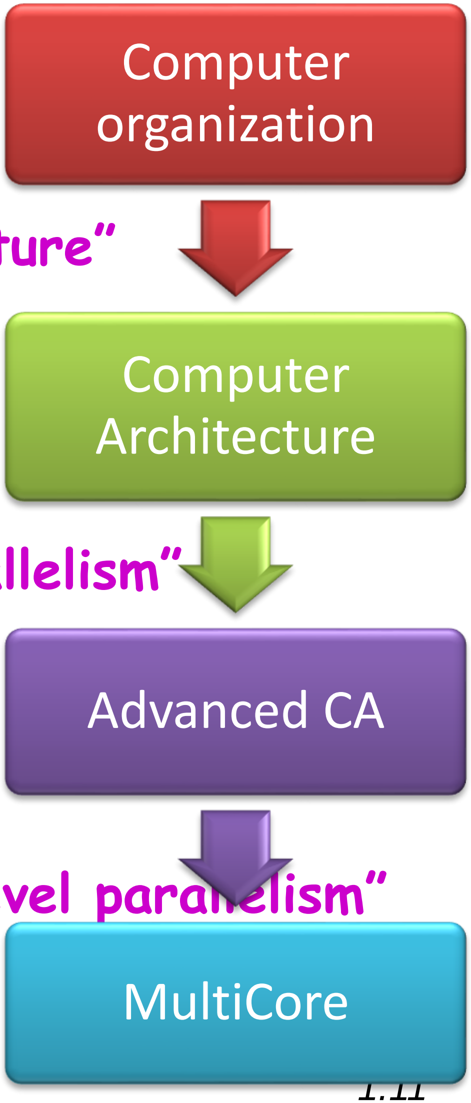
Why Such Change in 60 years?

- Technology Advances
  - CMOS VLSI dominates older technologies (TTL, ECL) in cost *AND* performance
- Computer architecture advances improves low-end
  - RISC, superscalar, OOO, Speculation, VLIW, RAID, …

- Price: Lower costs due to …
  - Simpler development: smaller systems, fewer components
  - Higher volumes: same dev. cost 10,000 vs. 10,000,000 units
  - Lower margins by class of computer, due to fewer services
- Function and Usage
  - scientific computing →non-digital processing → embedded system
  - Rise of networking/local interconnection technology

4 Decades of microprocessor

- The Decade of the 1970's "Microprocessors"
  - Programmable Controller (microprogramming)
  - Single-Chip Microprocessors
  - Personal Computers (PC)
- The Decade of the 1980's "Quantitative Architecture"
  - Instruction Pipelining
  - Fast Cache Memories
  - Compiler Considerations
  - Workstations
- The Decade of the 1990's "Instruction-Level Parallelism"
  - Superscalar Processors
  - Speculative Microarchitectures
  - Aggressive Code Scheduling
  - Low-Cost Desktop Supercomputing
- The Decade of the 2000's "Thread-level/Data-level parallelism"



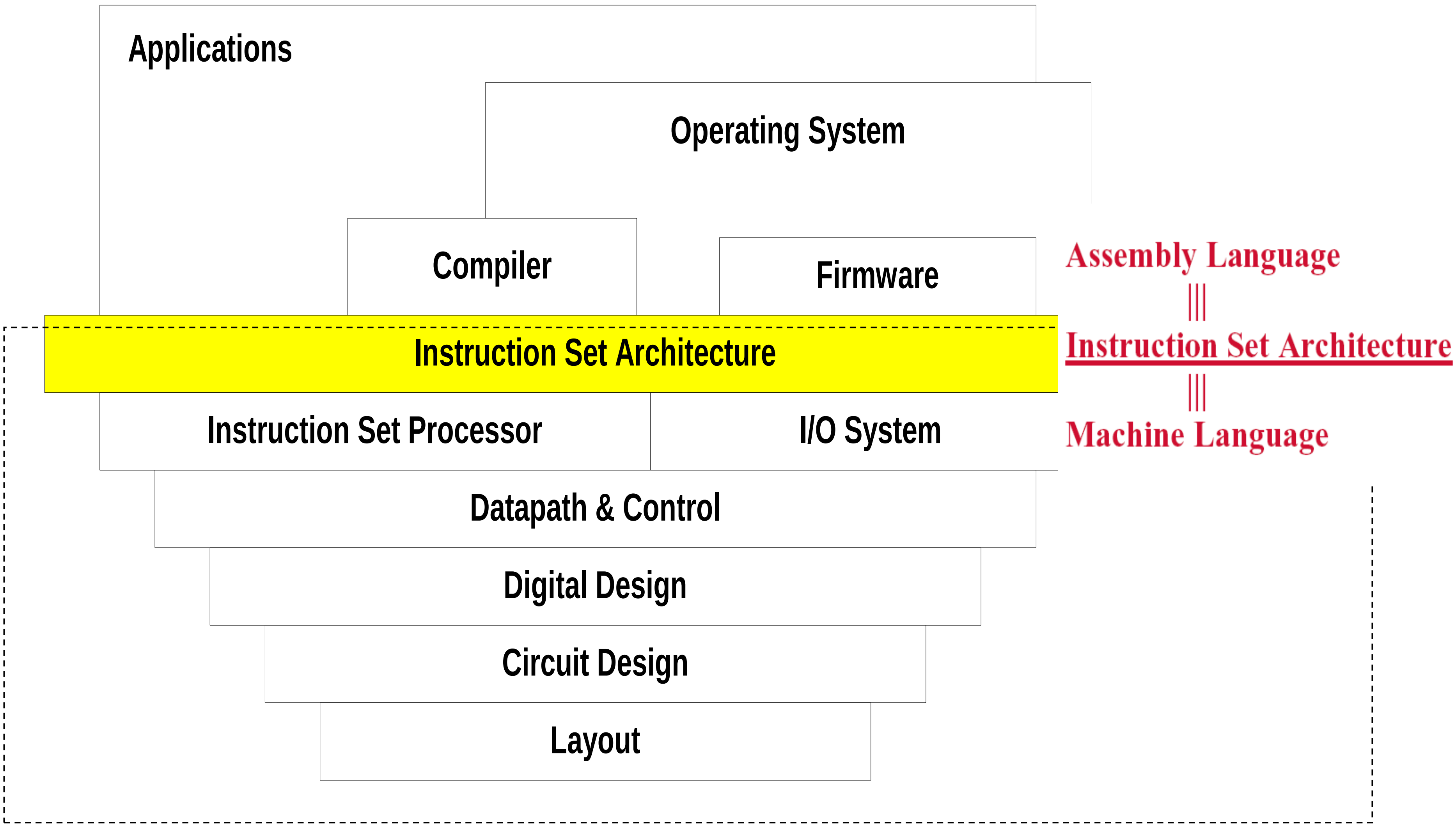
What's Computer Architecture

- **Concept Evolution**

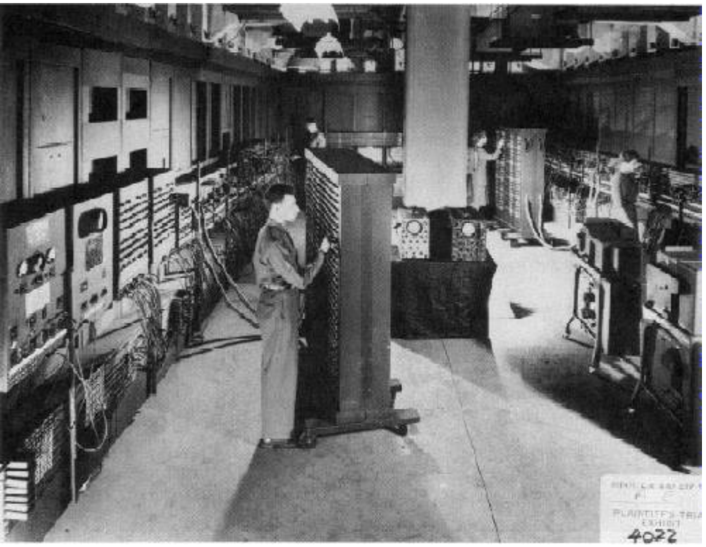
The attributes of a [computing] system as seen by the programmer, i.e., the conceptual structure and functional behavior, as distinct from the organization of the data flows and controls the logic design, and the physical implementation.

Amdahl, Blaaw, and Brooks, 1964

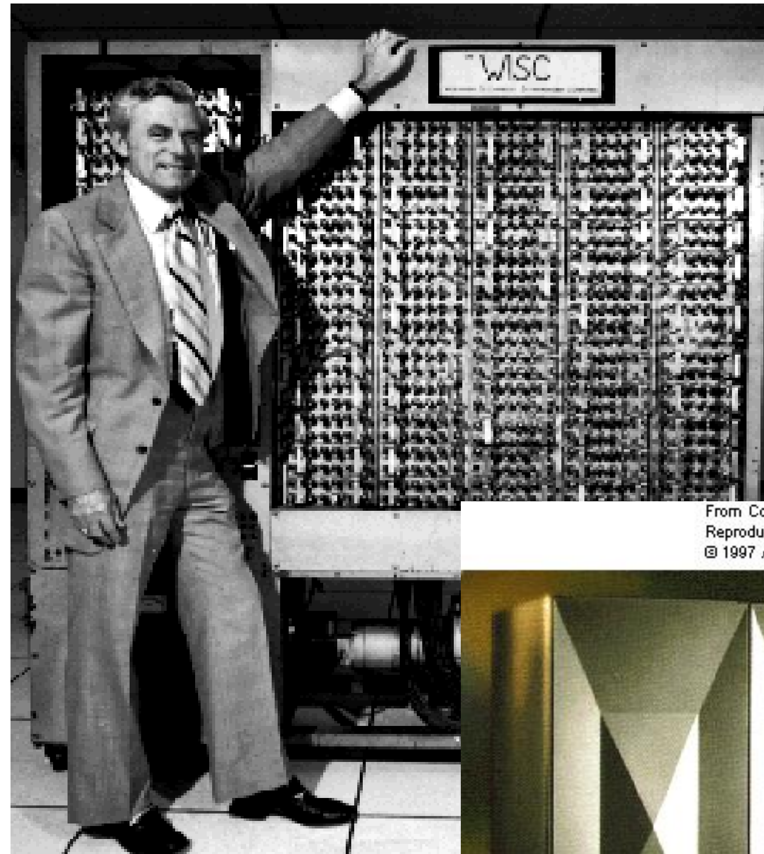
Programmer's perceptual



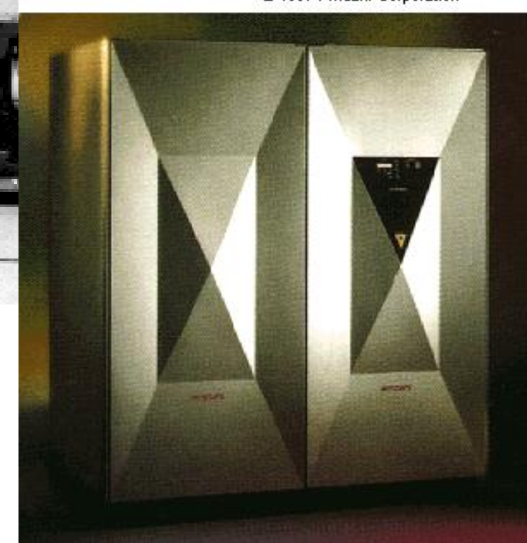
Very different appearance



From Computer Desktop Encyclopedia
Reproduced with permission.
© 1999 Dr. Gene M. Amdahl



From Computer Desktop Encyclopedia
Reproduced with permission.
© 1997 Amdahl Corporation



Very Different ISA

- PDP-11
- IBM 360
- VAX
- CRAY
- …

New Concepts ?

Tradeoff

- *Computer Architecture* is the science and art of selecting and interconnecting hardware components to create computers that meet functional, performance, cost and power goals.

- It is a **blueprint** and functional description of requirements and design implementations for the various parts of a computer, focusing largely on the way by which the central processing unit (CPU) performs internally and accesses addresses in memory.

Computer architecture

- Computer architecture comprises at least three main subcategories:[1]
- ***Instruction set architecture***,
- ***Microarchitecture***, also known as ***Computer organization*** is a lower level, more concrete and detailed, description of the system that involves how the constituent parts of the system are interconnected and how they interoperate in order to implement the ISA.[
- ***System Design*** which includes all of the other hardware components within a computing system such as:
  - Logic Implementation
  - Circuit Implementation
  - Physical Implementation

Seven dimensions of ISA

- Class of ISA
- Memory addressing
- Addressing modes
- Types and sizes of operands
- Operations
- Control flow instructions
- Encoding an ISA

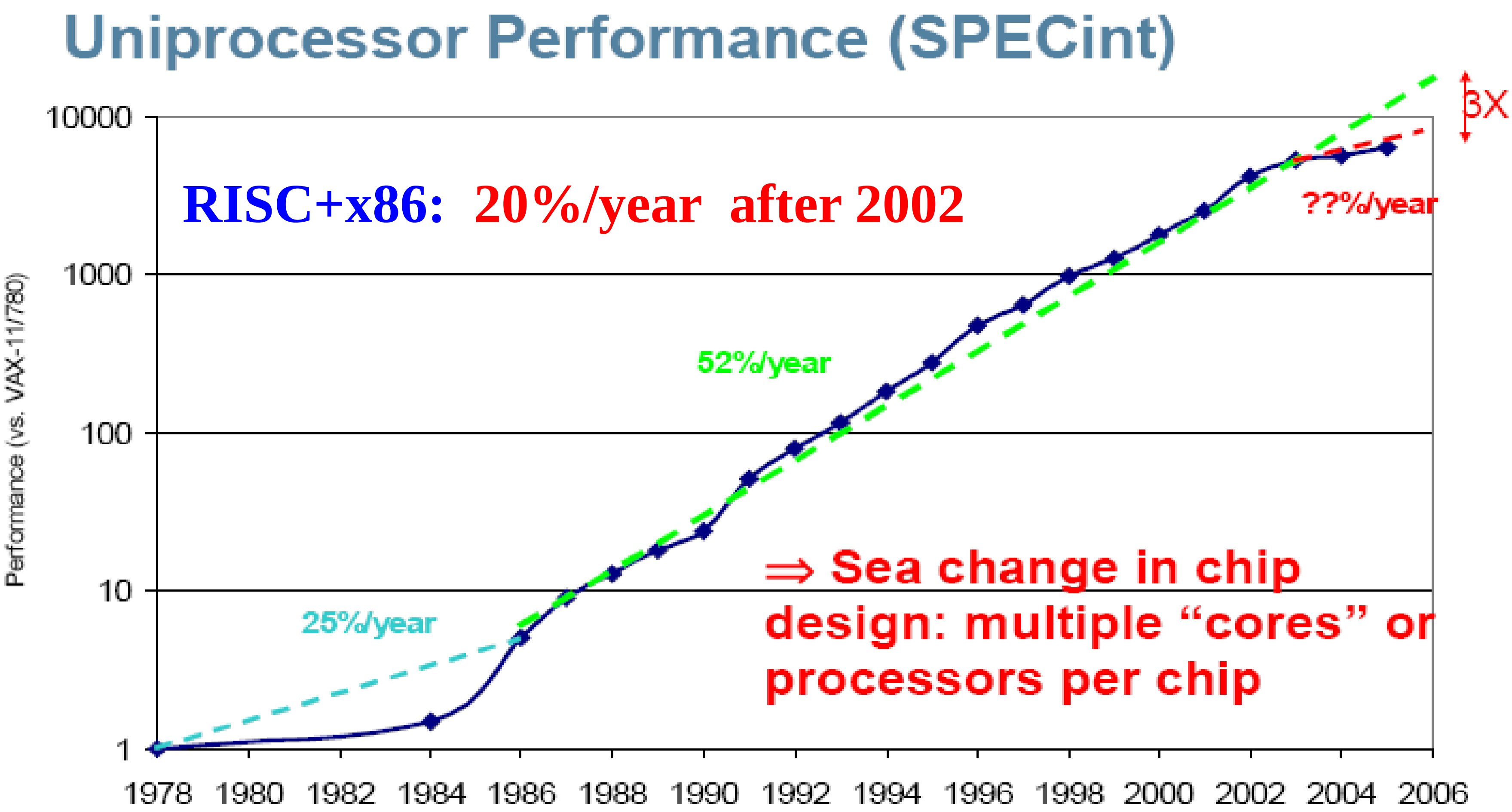
Computer Applications

- **Architects need to understand applications' behavior**
  - **We say we design general purpose processors, but they really focus on specific sets of applications**
  - **Architecture can be tuned to applications**
- **Types of applications today**
  - **Scientific**
    - Weather prediction, crash analysis, earthquake analysis, medical imaging, imaging of the earth (searching for oil)
  - **Business**
    - database, data mining, video
  - **General purpose**
    - Microsoft Word, Excel
  - **Real-time**
    - automated control systems,
  - Others: Games, Mobile

The Task of Computer Design-2

- Determine the important attributes of a new machine to maximize performance while staying with constrains, such as cost, power, availability, etc.
  - instruction set architecture design
  - functional organization
    - High level aspects of computer design, i.e. memory system, bus architecture and internal CPU design.
  - logic design (hardware)
  - implementation (hardware)

What are the CA challenges ?

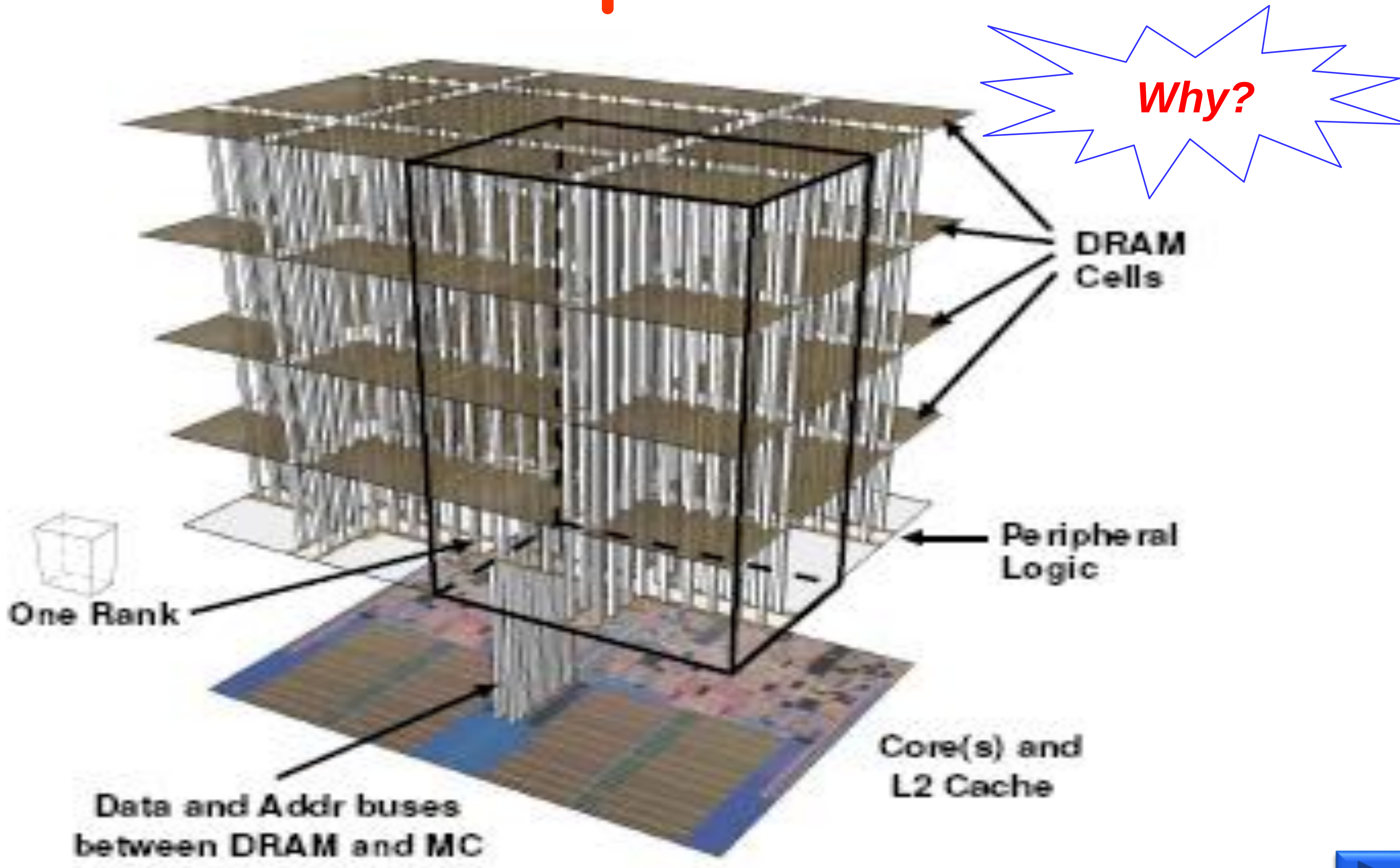


VAX : 25%/year 1978 to 1986
RISC+x86: 52%/year 1986 to 2002

Challenges of "three walls"

- ILP Wall
  - – diminishing returns on finding more ILP HW (Explicit thread and data parallelism must be exploited)
- Memory Wall
  - – growing disparity of speed between CPU and memory outside the CPU chip. Memory latency would become an overwhelming bottleneck in computer performance.
- Power Wall
  - – the trend of consuming double the power with each doubling of operating frequency

Trends of Computer Architecture



Computational RAM / PIM

- Processor in memory (PIM)
  - Processing in memory (PIM, sometimes called *processor in memory*) is the integration of a processor with RAM (random access memory) on a single chip. The result is sometimes known as a *PIM chip*.

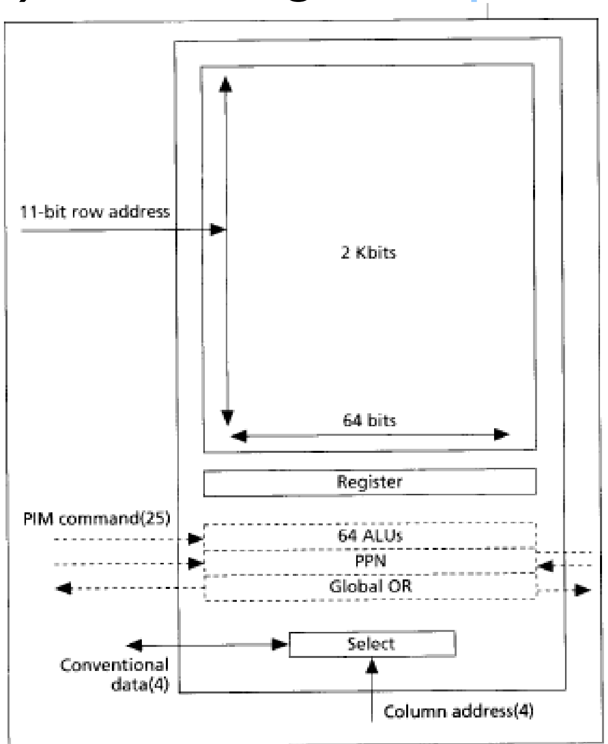


Figure 2. A processor-in-memory chip.

- 1995. 4 IEEE computer

  Processing in memory: the Terasys massively Parallel PIM Array.

Where to explore the parallelism ?

- Implicity, compiler and hardware

  → Explicity, programmer

So, YOU, programmers have to know parallelism in hardware, and to explore parallelism when design Algorithm and programming !

application

Algorithm

Language

Compiler

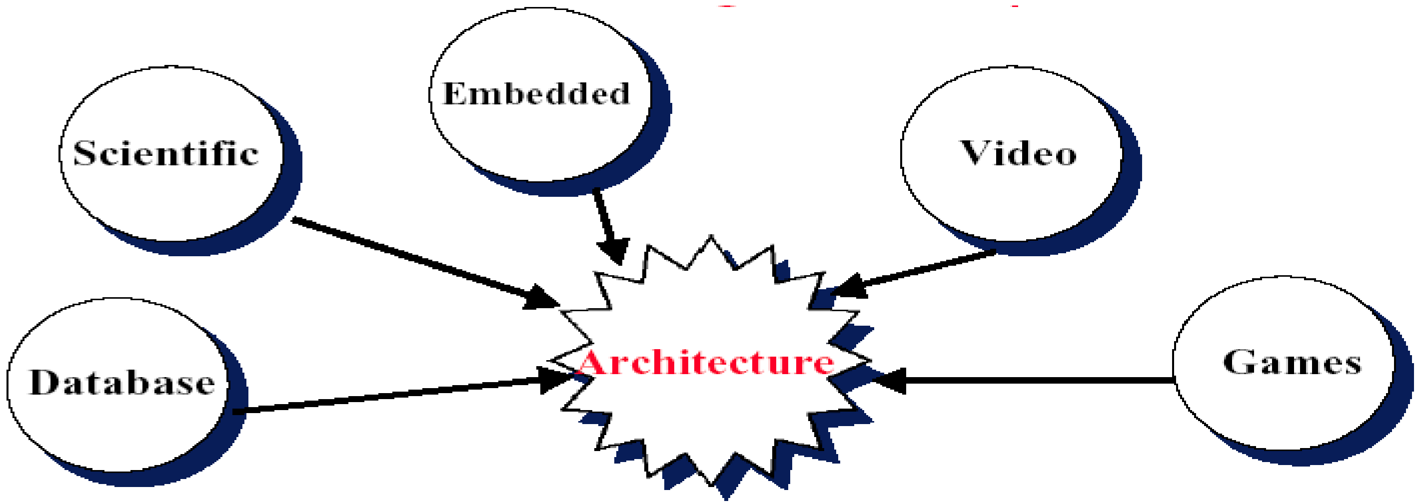
Hardware

Hardware → Hardware and compiler → compiler and programmer

ILP, Loop-level Parallism → TLP, DLP

Architectures are Tuned to Applications

- HP's 1.5 MB cache for transaction processing
- Alpha very fast FP for scientific
- StrongARM for embedded
- Intel MMX for image and video
- Sony EE for graphics rendering
- Applications drive the design of the processor



Something more about CA

ZheJiang University

What are the fastest computers?

  – http://top500.org/ Nov. 2012

- Top500 at Nov.2012

Fastest Supercomputer in the world

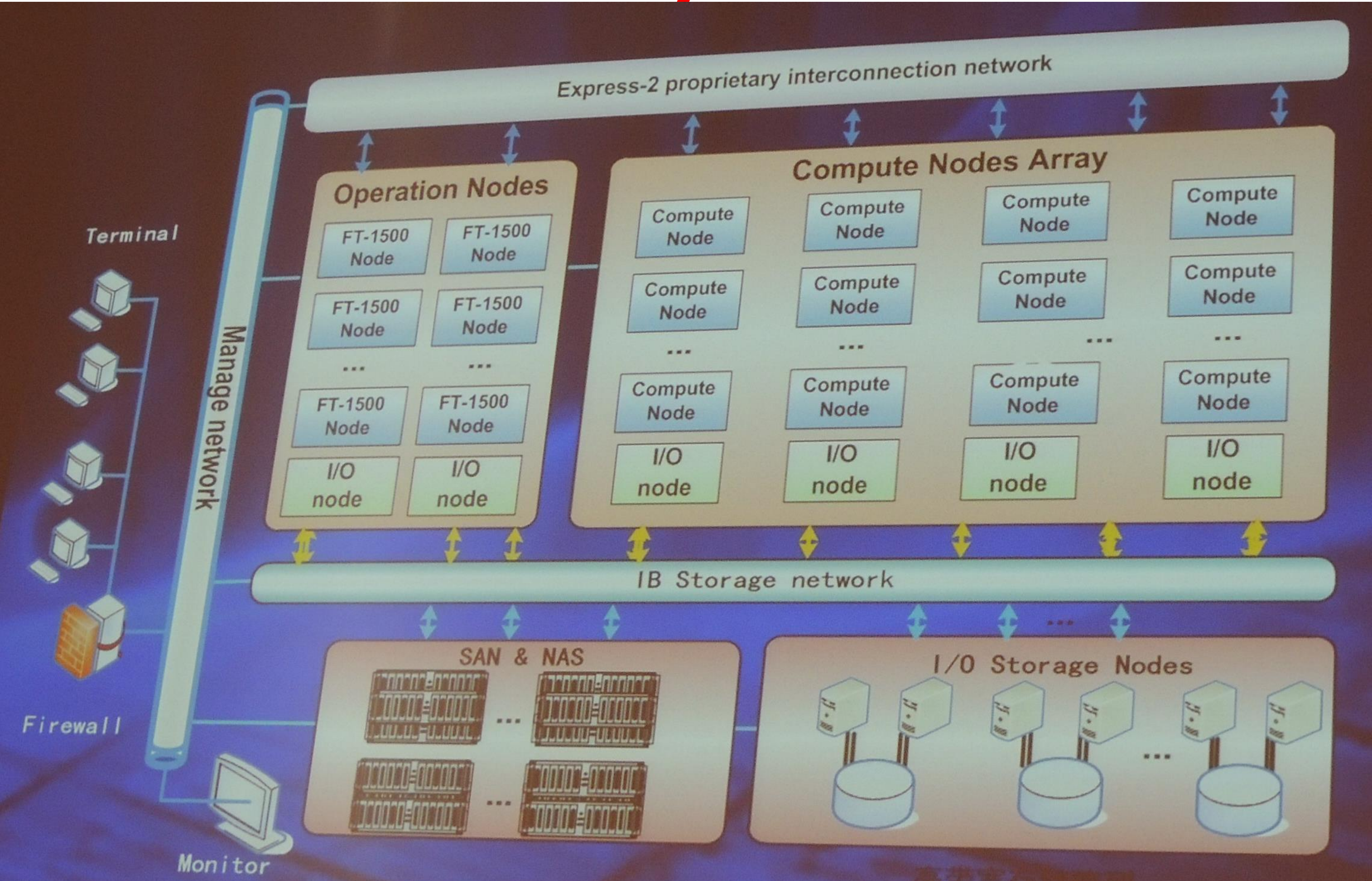
from http://top500.org/ June. 2013

Rank	Site	System	Cores	Rmax (TFlop/s)	Tpeak (TFlop/s)	Power (kW)
1	National University of Defense Technology China	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P NUDT	3,120,000	33,862.7	54,902.4	17,808
2	DOE/SC/Oak Ridge National Laboratory USA	**Titan** - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.	560,640	17,590.0	27,112.5	8,209
3	DOE/NNSA/LLNL USA	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM	1,572,864	17,173.2	20,132.7	7,890
4	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu	705,024	10,510.0	11,280.4	12,660
5	DOE/SC/Argonne National Laboratory USA	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	786,432	8,586.6	10,066.3	3,945
6	Texas Advanced Computing Center/Univ. of Texas USA	**Stampede** - PowerEdge C8220, Xeon E5-2680 8C 2.700GHz, Infiniband FDR, Intel Xeon Phi SE10P Dell	462,462	5,168.1	8,520.1	4,510
7	Forschungszentrum Juelich (FZJ) Germany	**JUQUEEN** - BlueGene/Q, Power BQC 16C 1.600GHz, Custom Interconnect IBM	458,752	5,008.9	5,872.0	2,301
8	DOE/NNSA/LLNL USA	**Vulcan** - BlueGene/Q, Power BQC 16C 1.600GHz, Custom Interconnect IBM	393,216	4,293.3	5,033.2	1,972
9	Leibniz Rechenzentrum Germany	**SuperMUC** - iDataPlex DX360M4, Xeon E5-2680 8C 2.70GHz, Infiniband FDR IBM	147,456	2,897.0	3,185.1	3,423
10	National Supercomputing Center in Tianjin China	**Tianhe-1A** - NUDT YH MPP, Xeon X5670 6C 2.93 GHz, NVIDIA 2050 NUDT	186,368	2,566.0	4,701.0	4,040

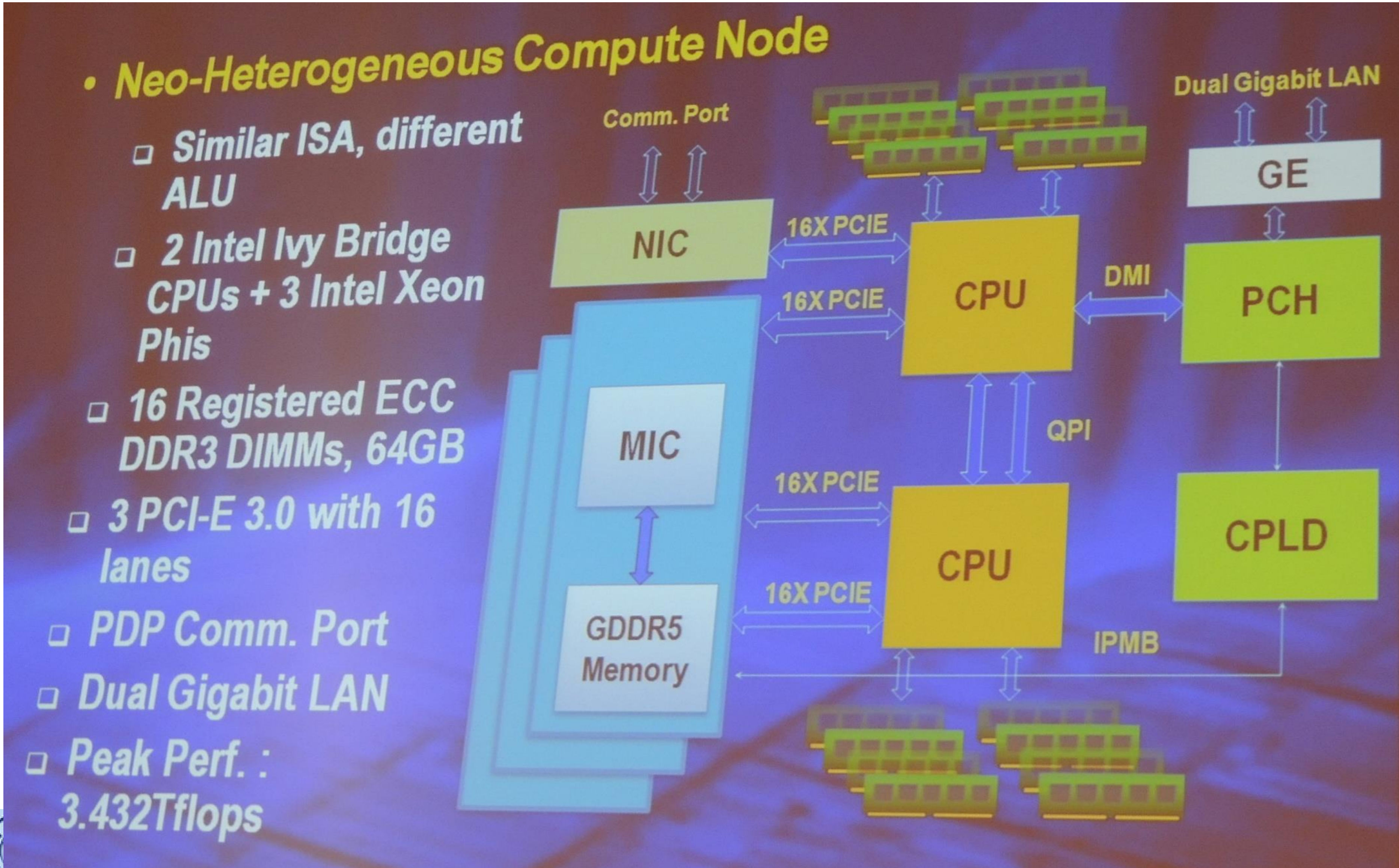
Overview of TH-2

Items	Configuration
Processors	32000 Intel Xeon CPUs + 48000 Xeon Phis + 4096 FT CPUs Peak performance is 54.9 Pflops, sustained performance 33.9PFlops
Interconnect	Proprietary high speed interconnection network TH Express-2
Memory	1.4PB in total
Storage	Global shared parallel storage system, 12.4 PB
Cabinets	125+13+24+8 = 170 Compute/Communication/Storage/Service Cabinets
Power	17.8MW
Cooling	Closed Air Cooling System

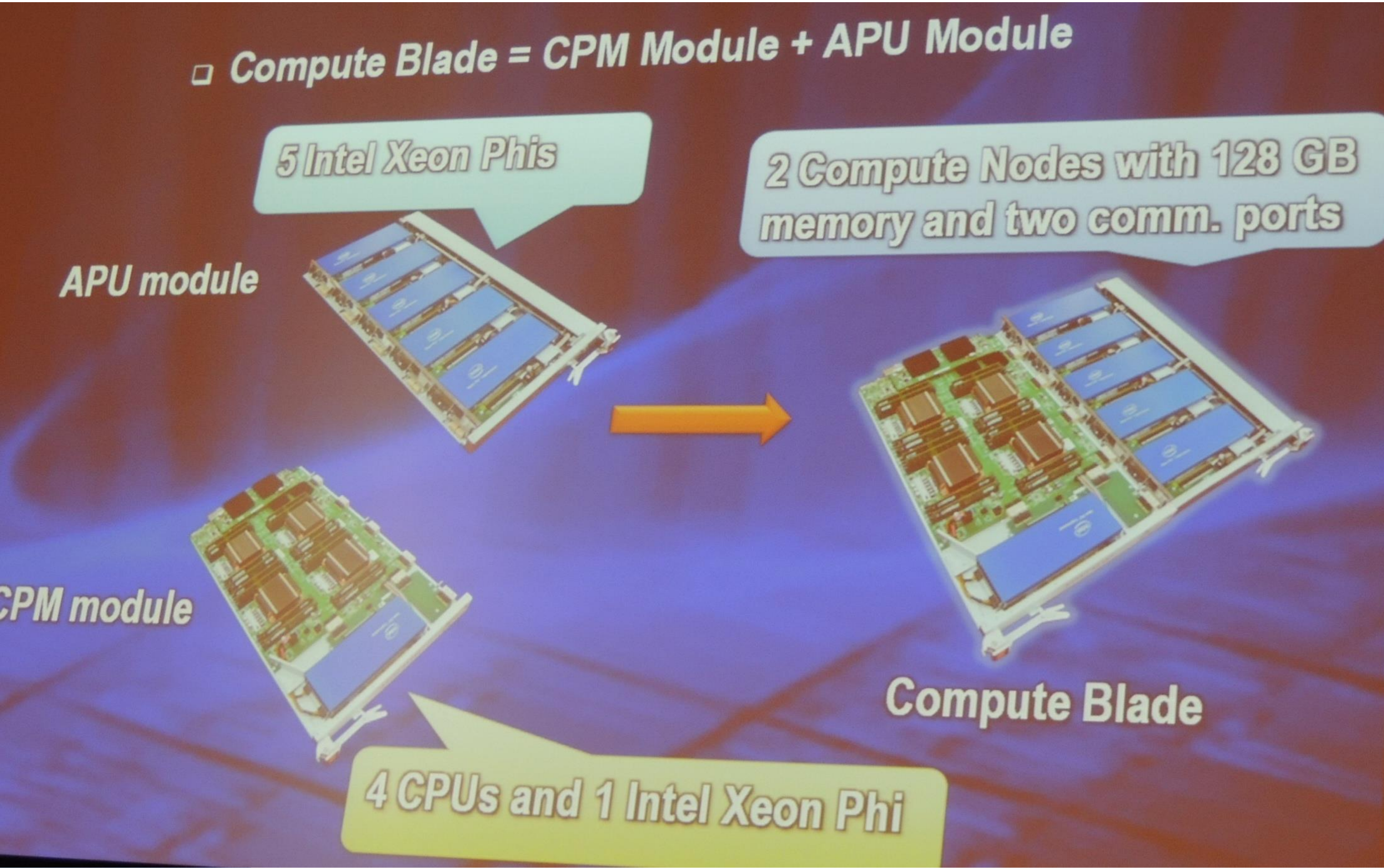
Hardware subsystem of TH-2



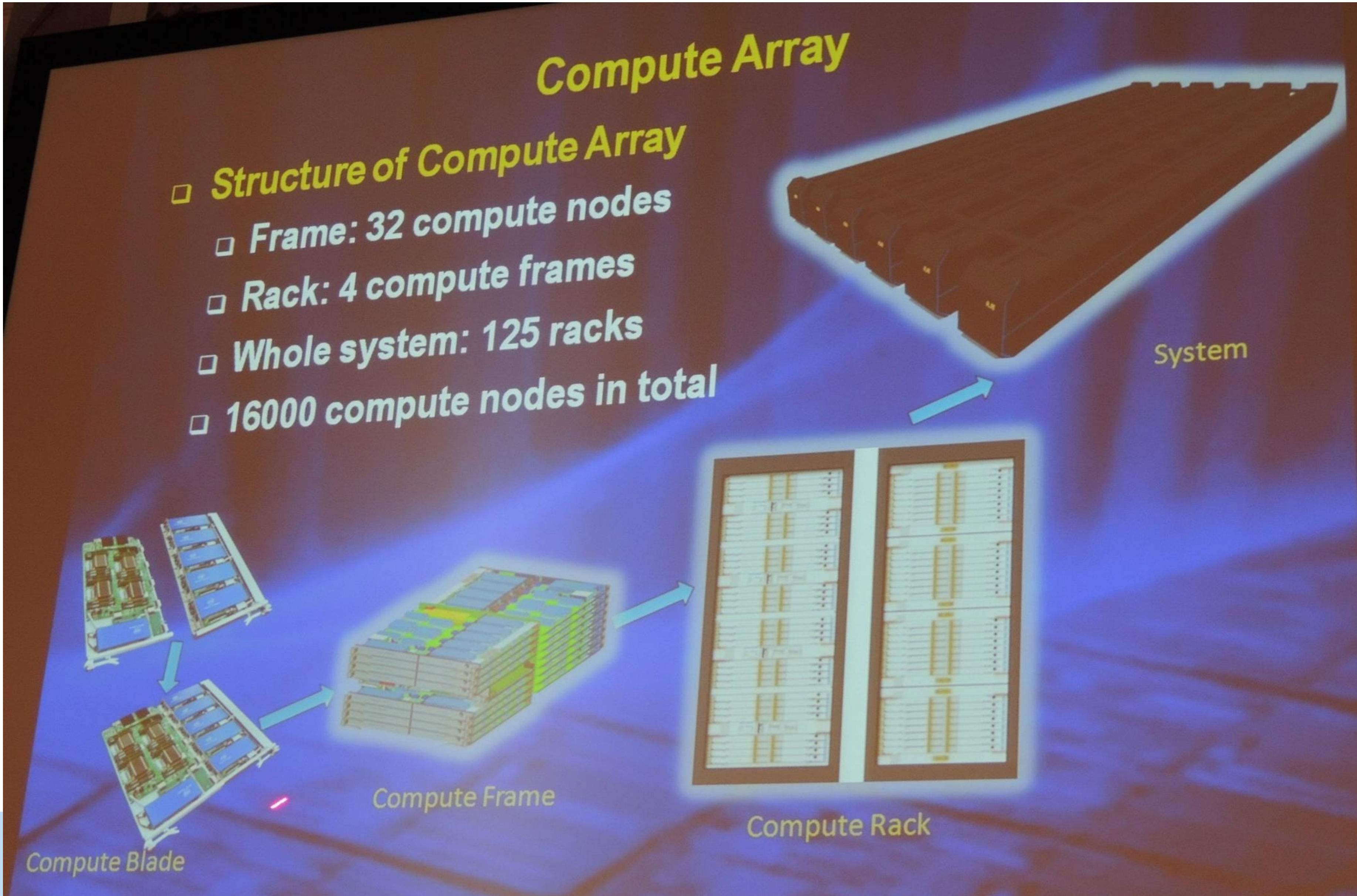
Comuter Node of TH-2



Comuter Node of TH-2



Computer Array of TH-2



What are the fastest computers?

– http://top500.org/ June. 2011

1 K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect

2 **Tianhe-1A - MPP, X5670 2.93Ghz 6C, NVIDIA GPU, FT-1000 8C**

3 Jaguar - Cray XT5-HE Opteron 6-core 2.6 GHz

4 Nebulae - Dawning TC3600 Blade, Intel X5650, NVidia Tesla C2050 GPU

5 TSUBAME 2.0 - HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows

6 Cielo - Cray XE6 8-core 2.4 GHz

7 Pleiades - SGI Altix ICE 8200EX/8400EX, Xeon HT QC 3.0/Xeon 5570/5670 2.93 Ghz, Infiniband

8 Hopper - Cray XE6 12-core 2.1 GHz

9 Tera-100 - Bull bullx super-node S6010/S6030

10 Roadrunner - BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Voltaire Infiniband

Fastest computer in China

- 2011 Tianhe-1A - MPP
- 2008 Dawning 5000A
  - 30720 node * AMD Opteron 1.9Ghz QC
  - Memory: 122.88TB, Infiniband, 180.6 TeraFLOPS
  - OS: Windows HPC 2008
  - Rank 10 in top 500 in Nov. 2008
- 2004 Dawning 4000A
  - 11 TeraFLOPS
  - rank 10 in top 500 in June, 2004
- 2003 ShenTeng6800
  - 5.324 TeraFLOPS
- 2002 ShenTeng1800
  - 2.04 TeraFLOPS
- 2000 YinHe IV
  - 1024个CPU
  - 1 TeraFLOPS

What are the Big Bananas ?

- **Eckert-Mauchly Award**
  - **http://www.computer.org/portal/web/awards/eckert**
  - **Administered jointly by ACM and IEEE Computer Society. The award of $5000 is given for contributions to computer and digital systems architecture where the field of computer architecture is considered at present to encompass the combined hardware-software design and analysis of computing and digital systems.**

Eckert-Mauchly Award Recipients

2011 Gurindar (Guri) S. Sohi
2010 William J. Dally
2009 Joel S. Emer
2008 **Patterson, David**
2007 **Valero, Mateo**
2006 Pomerene, James H
2005 Colwell, Robert P.
2004 **Brooks, Frederick P.**
2003 Fisher, Joseph A. (Josh)
2002 Rau, B. Ramakrishna (Bob)
2001 **Hennessy, John**
2000 Davidson, Edward
1999 Smith, James E.
1998 Watanabe, T.
1997 **Tomasulo, Robert**
1996 **Patt, Yale**

1995 Crawford, John
1994 Thornton, James E.
1993 Kuck, David J
1992 **Flynn, Michael J.**
1991 Smith, Burton J.
1990 Batcher, Kenneth E.
1989 **Cray, Seymour**
1988 Siewiorek, Daniel P.
1987 **Amdahl, Gene M.**
1986 Cragon, Harvey G
1985 **Cocke, John**
1984 Dennis, Jack B.
1983 Kilburn, Tom
1982 **Bell, C. Gordon**
1981 Clark, Wesley A.
1980 Wilkes, Maurice V.
1979 Barton, Robert S.

Big Men in Architecture(1)

- **2007Mateo Valero**

  http://personals.ac.upc.edu/mateo/

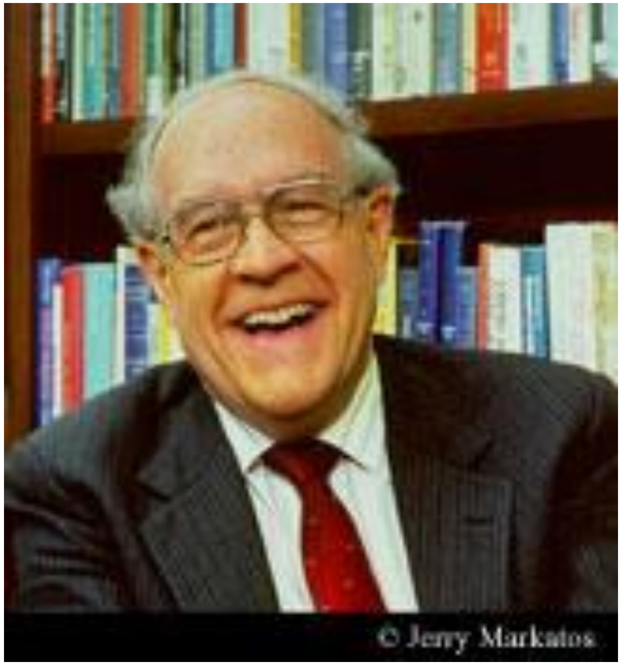
For important contributions to instruction level parallelism and superscalar processor design.

Big Men in Architecture(2)

- 2001 **Hennessy, John**

  For being the founder and chief architect of the MIPS Computer Systems and contributing to the development of the landmark MIPS R2000 microprocessor.

Big Men in Architecture(3)



Frederick P. Brooks

http://www.cs.unc.edu/~brooks/

2004 Eckert-Mauchly Award

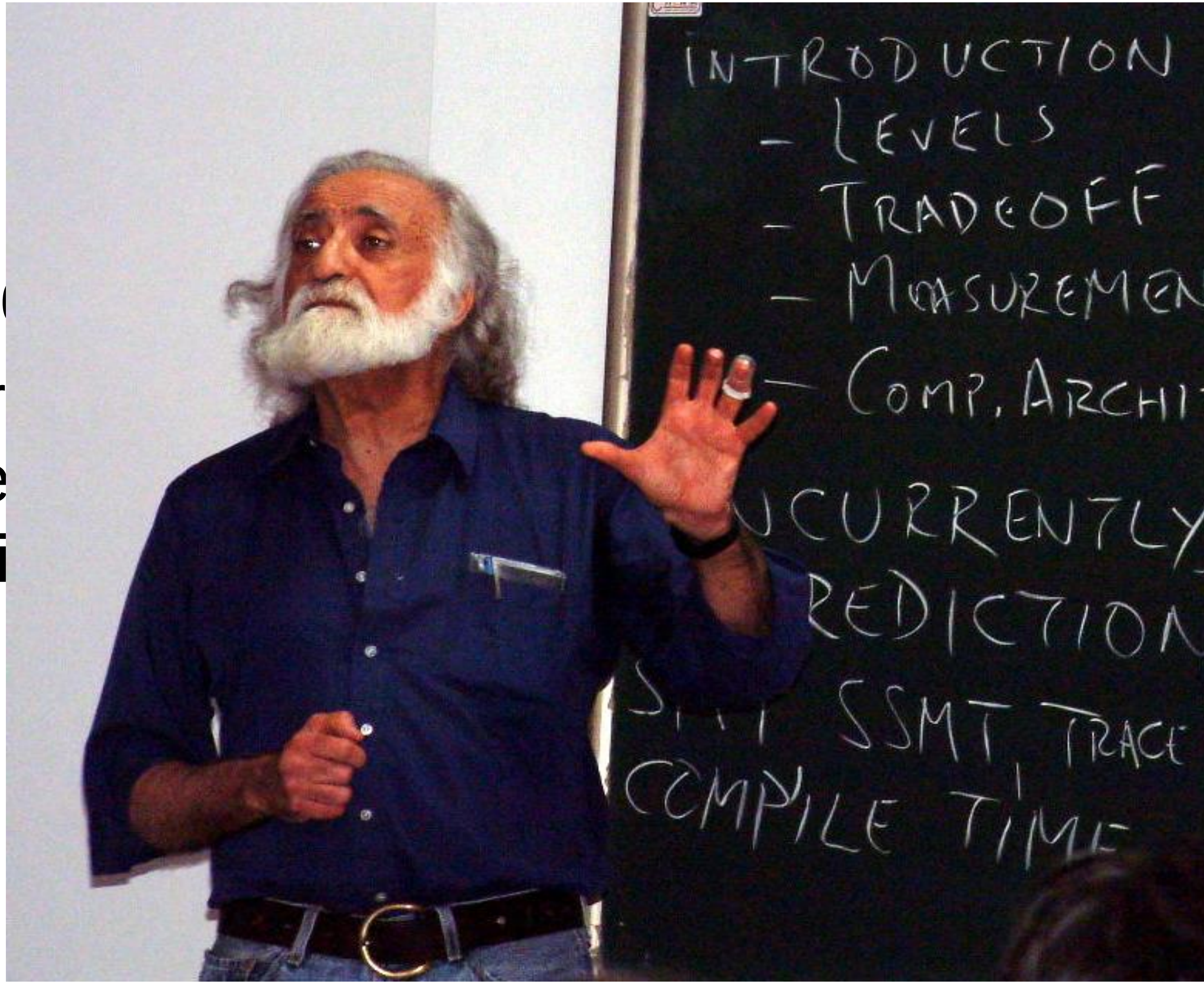
"For the definition of computer architecture and contributions to the concept of computer families and to the principles of instruction set design; for seminal contributions in instruction sequencing, including interrupt systems and execute instructions; and for contributions to the IBM 360 instruction set architecture."

- **1999 ACM Turing Award**

  landmark contributions to computer architecture, operating systems, and software engineering."

Big

- 199
  For
  leve
  desi



Advertising Time ：

Yale Patt will come to Hangzhou

July 29—August 17， 2015

Teaching the course:

introduction to Computer System

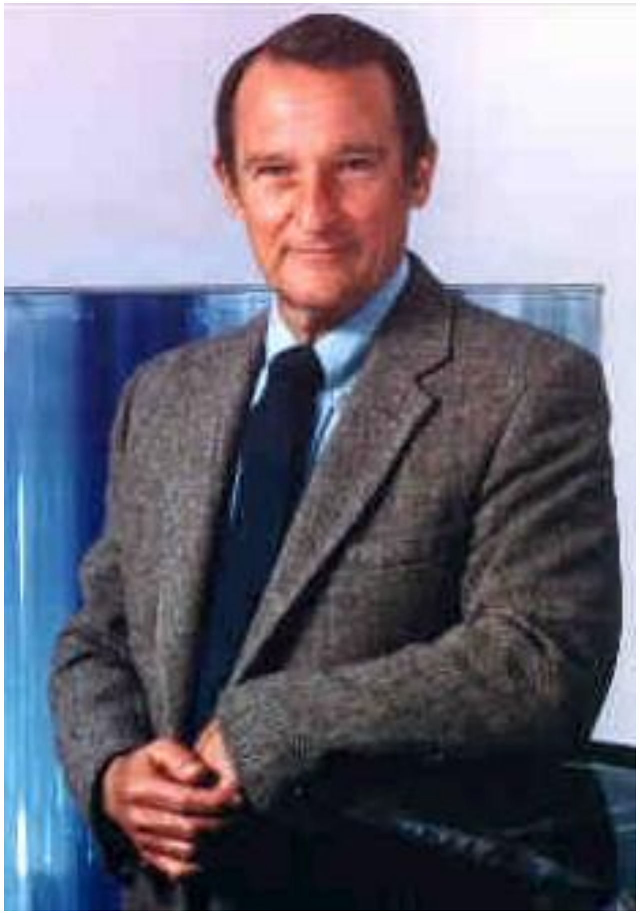
Big Men in Architecture(6)

- 1992 **Michael J. Flynn**

  http://www.cpe.calpoly.edu/IAB/flynn.html
- For his important and seminal contributions to processor organization and classification, computer arithmetic and performance evaluation.

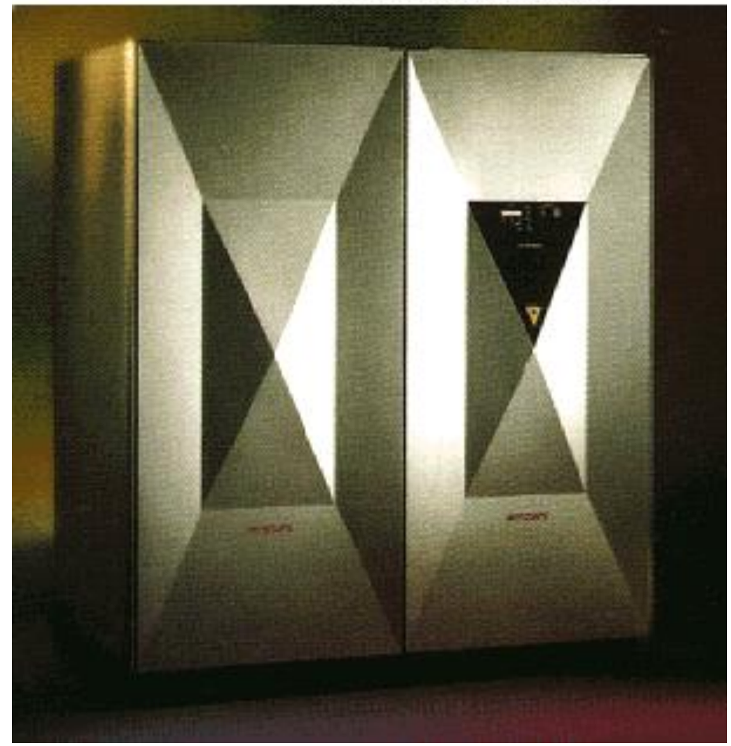
Big Men in Architecture(7)

- 1989 [Cray, Seymour]
- For a career of achievements that have advanced supercomputing design.



Big Men in Architecture(8)

In 1975, Dr. Amdahl stands beside the Wisconsin Integrally Synchronized Computer (WISC), which he designed in 1950. It was built in 1952. (Image courtesy of Dr. Gene M. Amdahl.)



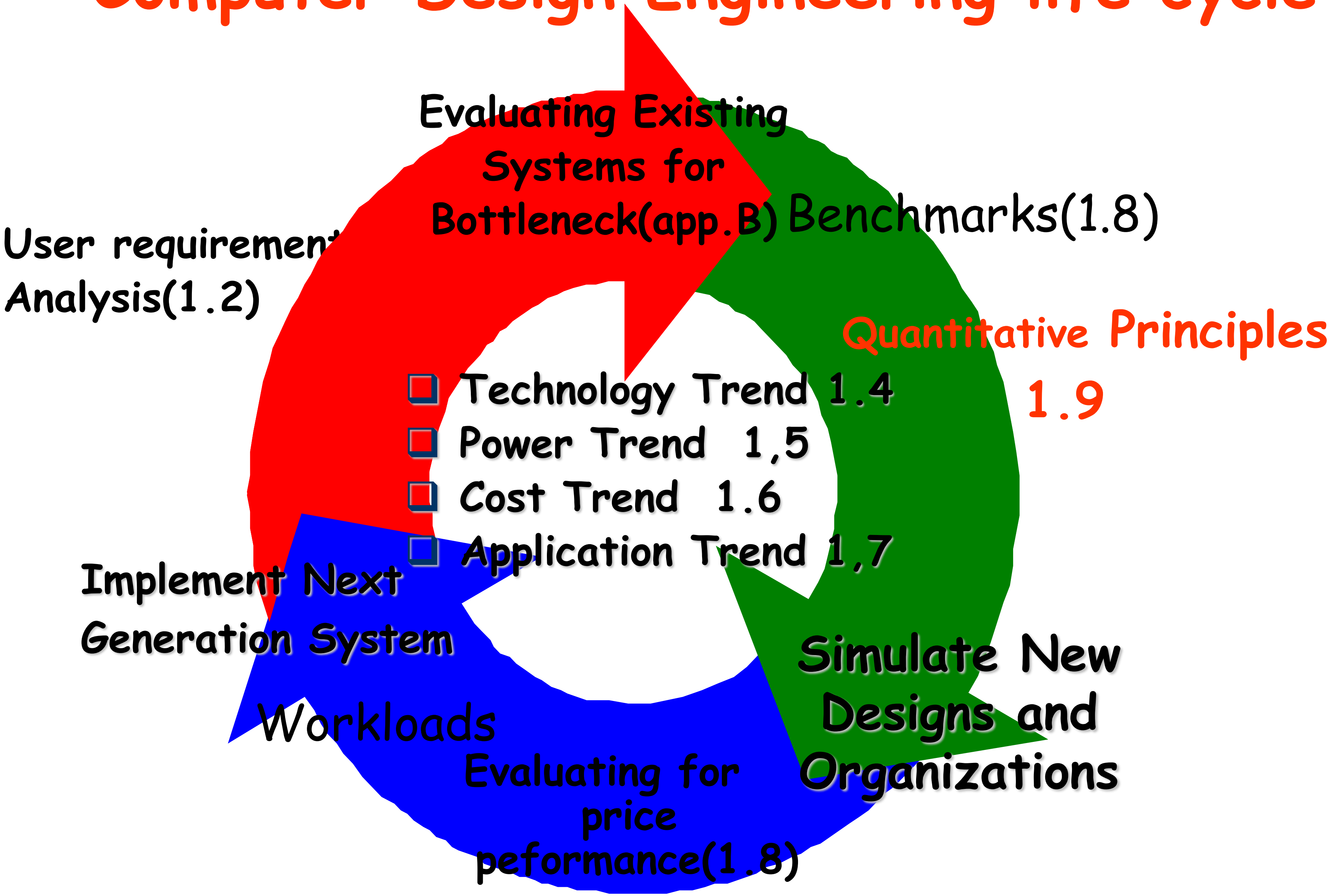
n computer architecture,
on look-ahead, and

Top conferences and Journals

- Top conference:
  - ISCA
  - MICRO,
  - …
- Top journals:
  - IEEE Tran. on Computers IF 2.419
  - ACM Tran. on Computer Systems IF 1.917
  - …

What’s a CA designer’s task ?

Computer Design Engineering life cycle



Topics in Chapter 1

1.1 Introduction
1.2 Classes of computers
1.3 Defining computer architecture and What's the task of computer design?
1.4 Trends in Technology
1.5 Trends in power in Integrated circuits
1.6 Trends in Cost
1.7 Dependability
1.8 Measuring, Reporting and summerizing Perf.
1.9 Quantitative Principles of computer Design
1.10 Putting it altogether

Summary: Task of computer design

- Considerations:
  - functional and non functional requirements
  - implementation complexity
    - Complex designs take longer to complete
    - Complex designs must provide higher performance to be competitive
  - **Technology trends**
    - Not only what's available today, but also what will be available when the system is ready to ship. (more on this later)
  - **Trends in Power in IC**
  - **Trends in cost**
- Arguments
  - Evaluate Existing Systems for Bottlenecks
- Quantitative Principles

Reading Assignments

- Chapter 1

- Homework1 for chapter 1 will be loaded on website

- End !
- Thank you !