

Computer Architecture Lecture 1: Course Overview

Duy-Hieu Bui, PhD

AloT Laboratory

Email: hieubd@vnu.edu.vn
https://duyhieubui.github.io

Y

Nhập môn

- · Khái niệm chung
 - Kiến trúc và tổ chức
 - Cấu trúc và chức năng
- · Kiến trúc Von Neuman
- · Phân loại kiến trúc
- · Lịch sử phát triển
- Hiệu năng và luật Moore
- · Một số kiến trúc máy tính hiện đại
 - Kiến trúc x86 và x86-64
 - Kiến trúc UltraSparc
- Tổng kết và bài tập

9/10/2024

Duy-Hieu Bui, ITI/SISLAB



Khái niệm chung

- Kiến trúc máy tính
 - Kiến trúc : hướng đến mục tiêu xây dựng sản phẩm từ những thành phần đã có theo một phương thức nào đó
 - Ví dụ: kiến trúc chip, kiến trúc mạng Internet, kiến trúc hệ thống email, ...
- Wikipedia:
 - "computer architecture is the practical art of selecting and interconnecting hardware components to create computers that meet functional, performance and cost goals and the formal modelling of those systems".
 - "The coordination of abstract levels of a processor under changing forces, involving design, measurement and evaluation. It also includes the overall fundamental working principle of the internal logical structure of a computer system"...

9/10/2024

Duy-Hieu Bui, ITI/SISLAB

3



Architecture & Organization

- Architecture: thiết kế logic của máy tính, đặc tả bởi những thuộc tính mà người lập trình có thể sử dụng được
 - Tập lệnh, cơ chế vào/ra, số bits dùng để biểu diễn dữ liêu...
 - VD: Có lệnh Nhân hay không?
- Organization: thiết kế vật lý của máy tính, đặc tả bởi những cách thức cài đặt các chức năng, thao tác
 - Tín hiệu điều khiển, giao diện, công nghệ nhớ.
 - VD: Bộ nhân đã được cài đặt trực tiếp hay thông qua một dãy các phép cộng?

Mathematics
Algorithms
Applications
Operating Systems
Architecture
Organization
Digital Logic
VLSI Design
Semiconductor Manf.
Physics
·



Kiến trúc và tổ chức ...

- Họ Intel x86-64 chia sẻ cùng một kiến trúc cơ bản
- Họ Ultrasparc cũng chia sẻ cùng một kiến trúc cơ bản
- Tuy nhiên, tổ chức của chúng lại khác nhau giữa các phiên bản







9/10/2024

Duy-Hieu Bui, ITI/SISLAB

5



Vai trò của kiến trúc máy tính

- Computer Architecture cho phép:
 - Khai thác những thế mạnh công nghệ
 - Sản xuất các thiết bị nhanh hơn, bé hơn, rẻ hơn, xạch hơn (tiêu thụ ít năng lượng)....
 - Cải thiện được hiệu năng của hệ thống (latency, throughput).
 - Mang đến những ứng dụng mới, khả năng mới
- Sự phát triển của kiến trúc máy tính được xem là mấu chốt cho sự phát triển của những lĩnh vực tính toán khác!

9/10/2024

Duy-Hieu Bui, ITI/SISLAB

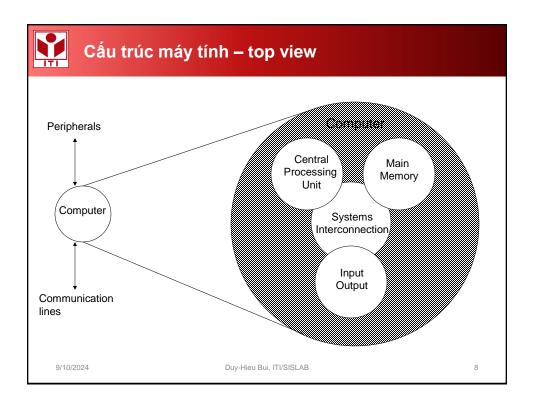


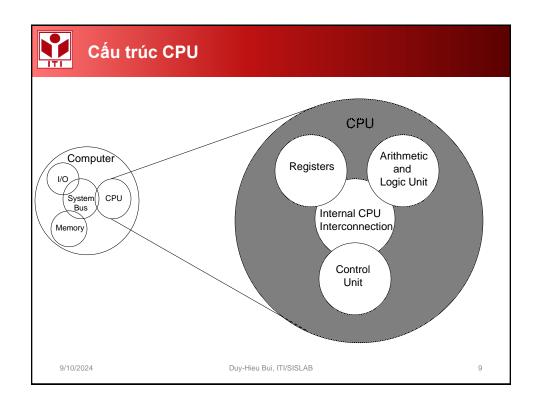
Structure & Function

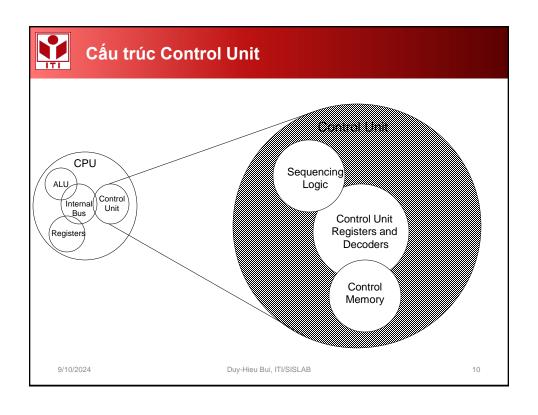
- Cấu trúc: cách để các thành phần liên kết với nhau
- Chức năng: hoạt động của các thành phần cấu trúc
- Chức năng của mọi máy tính:
 - Data processing Xử lý dữ liệu
 - Data Storage Lưu giữ dữ liệu
 - Data Movement Chuyển dữ liệu
 - Control Điều khiển

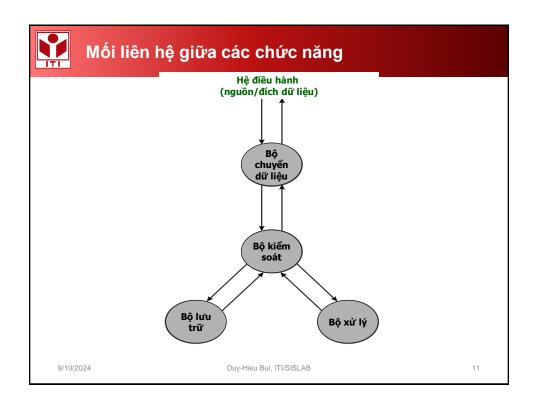
9/10/2024

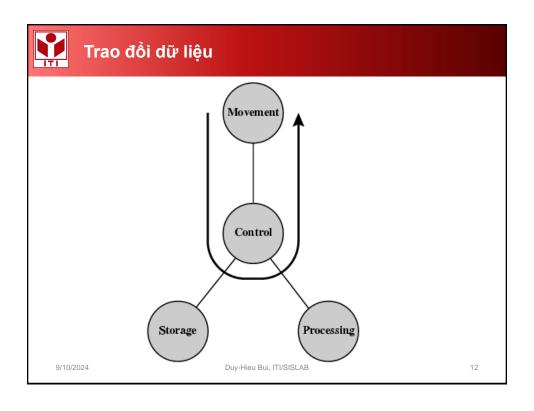
Duy-Hieu Bui, ITI/SISLAB

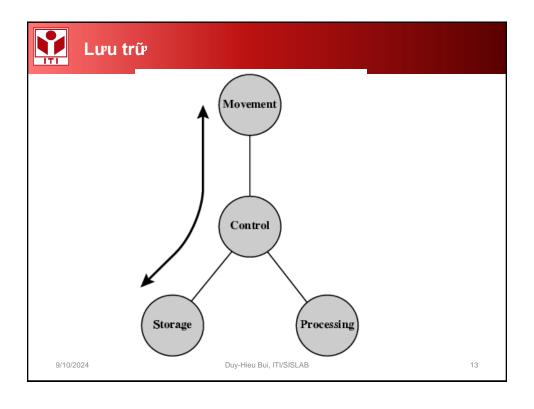


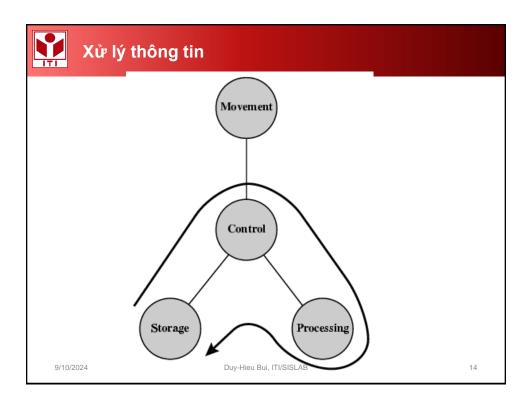


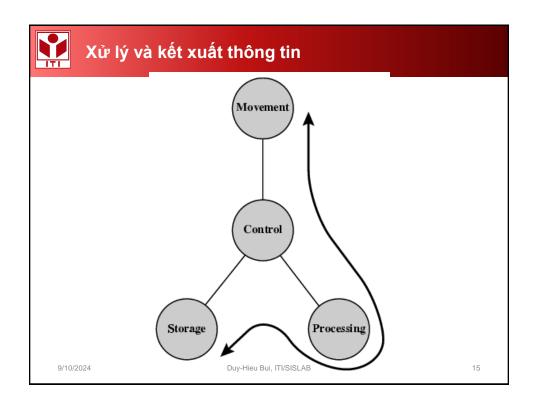




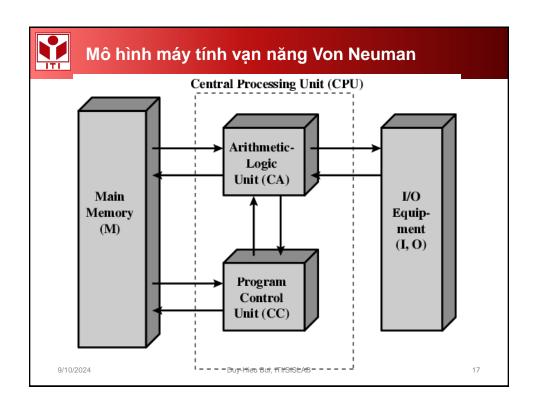


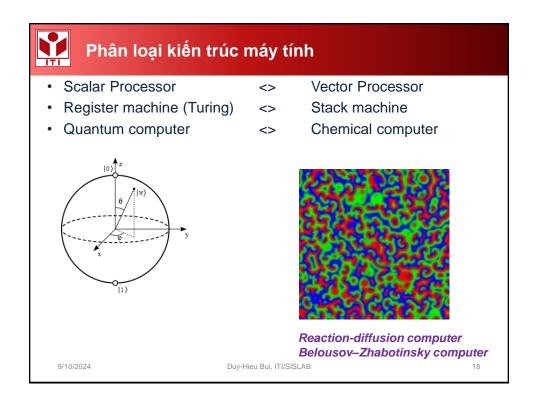


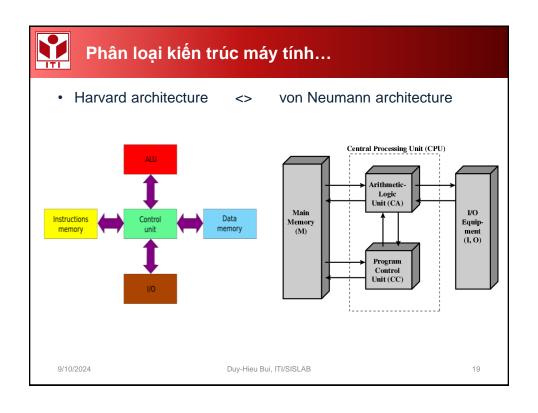


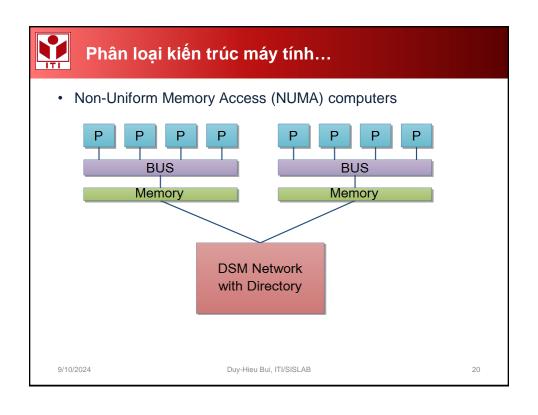


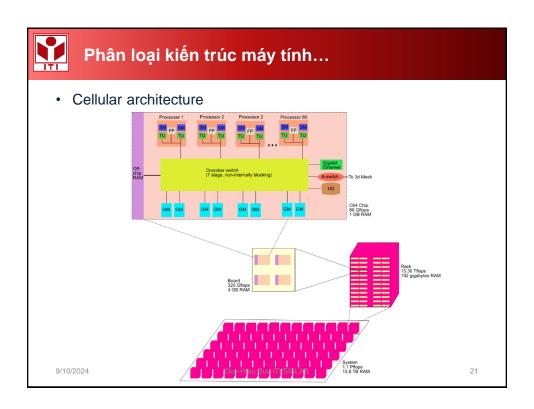














Lịch sử

- Trước thế kỷ 19: automates
 - Hộp chơi nhạc
 - Máy dệt
 - **–** ..
- Thế kỷ 19: Charles Babbage đã xây dựng một máy tính – calculator
 - Lập lại những chuỗi thao tác
 - Chọn trạng thái tính
- Thế kỷ 20:
 - 1946: J. Von Neumann miêu tả một mô hình máy tính vạn năng
 - Máy tính hiện nay:
 - · Dựa trên mô hình đó
 - Được xếp vào thế hệ thứ 5



9/10/2024

Duy-Hieu Bui, ITI/SISLAB



Thế hệ

	Thời gian	Công nghệ	Tốc độ (operations/s)
1	1946-1957	Bóng chân không	40.000
2	1958-1964	Transistor	200.000
3	1965-1971	SSI/MSI	1.000.000
4	1972-1977	LSI	10.000.000
5	1978-1991	VLSI	100.000.000
6	1991-nay	ULSI	>100.000.000



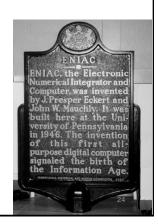
1946-1957

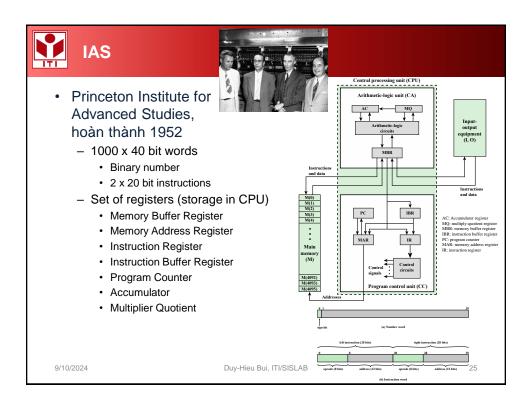
- Máy tính chuyên dụng, mẫu duy nhất
 - Cồng kềnh, vận hành kém
 - Công nghệ bóng chân không, rơle, điện trở
 - Lập trình bằng thẻ đục lỗ
- Tiêu biểu: ENIAC Electronic Numerical Integretor And Calculator dành để nghiên cứu về bom H
 - 1943-1946, được sử dụng đến 1955
 - 18.000 đèn điện tử, 1.500 rơle, 30 tấn, 140KW, 15.000 square feet
 - 5.000 phép cộng/s
 - 20 thanh ghi, mỗi thanh chứa được
 1 số thập phân 10 chữ số

9/10/2024

Duy-Hieu Bui, ITI/SISLAB









1958-1964

- Sử dụng Transistor
 - Transistor: John Barden, Walter Brattain và William Shockley sáng chế ra ở Bell Lab, 1947
- Ngôn ngữ lập trình đầu tiên : COBOL, FORTRAN, LISP
- Tiêu biểu: máy tính mini DEC PDP-1 (1957)
 - 4K từ 18bit
 - Chu kỳ 5μs



9/10/2024

Duy-Hieu Bui, ITI/SISLAB





1972-1977

- Công nghệ LSI (large SI)
 - 10⁷ phần tử logics
 - 3.000-100.000 devices/chip
- Mạng máy tính lên ngôi
- Xử lý phân tán
- 1971 : microprocessor 4004 đầu tiên của INTEL
 - Tất các các thành phần của CPU được tập trung trên cùng một chíp



Duy-Hieu Bui, ITI/SISLAB





1978-

- Công nghệ VL/WSI (very large, wafer)
 - $> 10^8$ phần tử logics
 - 100.000-100.000.000 devices/chip
- · Hệ phân tán tương tác
 - Multimedia,
 - Xử lý dữ liệu không phải dạng số (text, images, speech)
 - Song song,
 - Client-server
- Kiểu:
 - MicroComputer PC
 - MiniComputer
 - SuperMini
 - MainFrame
 - SuperComputer
- Tham khảo: http://www.computerhistory.org/

9/10/2024

Duy-Hieu Bui, ITI/SISLAB

Top 10 siêu máy tính (6/2018)						
Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	DOE/SC/Oak Ridge National Laboratory United States	Summit - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband	2,282,544	122,300. 0	187,659.3	8,806
2	National Supercomputing Center in Wuxi China	Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway NRCPC	10,649,600	93,014.6	125,435.9	15,371
3	DOE/NNSA/LLNL United States	Sierra - IBM Power System S922LC, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband IBM	1,572,480	71,610.0	119,193.6	
4	National Super Computer Center in Guangzhou China	Tianhe-2A - TH-IVB-FEP Cluster, Intel Xeon E5- 2692v2 12C 2.2GHz, TH Express-2, Matrix-2000 NUDT	4,981,760	61,444.5	100,678.7	18,482
5	National Institute of Advanced Industrial Science and Technology (AIST) Japan	Al Bridging Cloud Infrastructure (ABCI) - PRIMERGY CX2550 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR Fujitsu	391,680	19,880.0	32,576.6	1,649
6	Swiss National Supercomputing Centre (CSCS) Switzerland	Piz Daint - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 Cray Inc.	361,760	19,590.0	25,326.3	2,272
7	DOE/SC/Oak Ridge National Laboratory United States	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
8	DOE/NNSA/LLNL United States	Sequoia - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM	1,572,864	17,173.2	20,132.7	7,890
9	DOE/NNSA/LANL/SNL United States	Trinity - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect Cray Inc.	979,968	14,137.3	43,902.6	3,844
10	DOE/SC/LBNL/NERSC United States	Cori - Cray XC40, Intel Xeon Phi 7250/68Q 1AGHz, Aries interconnect Cray Inc.	622,336	14,014.7	27,880.7	3,939



Một số siêu máy tính...



- Summit IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband – N0 1 2018
- Cores: 2,282,544; Memory: 2,801,664 GB; Processor: IBM POWER9 22C 3.07GHz; Interconnect: Dual-rail Mellanox EDR Infiniband
- Linpack Performance (Rmax): 122,300 TFlop/s; Theoretical Peak (Rpeak) 187,659 TFlop/s; Nmax 13,989,888; HPCG [TFlop/s] 2,925.75
- Operating System: RHEL 7.4; Compiler: XLC 13.1, nvcc 9.2; Math Library: ESSL, UBLAS 9.2; MPI: Spectrum MPI

9/10/2024 Duy-Hieu Bui, ITI/SISLAB

31

32



Một số siêu máy tính...



- Số 1 năm 2016, Sunway TaihuLight đạt 125,436 petaflop:
 - Manufacturer: NRCPC; Cores: 10,649,600Linpack Performance (Rmax) 93,014.6 TFlop/s
 - Theoretical Peak (Rpeak) 125,436 TFlop/s
 - Nmax 12,288,000Power: 15,371.00 kW
 - Memory: 1,310,720 GB
 - Processor: Sunway SW26010 260C 1.45GHz
 - Interconnect: Sunway
 - Operating System: Sunway RaiseOS 2.0.5

9/10/2024 Duy-Hieu Bui, ITI/SISLAB



Một số siêu máy tính...



Sequoia

- Bộ năng lượng Mỹ với 1,57 triệu nhân xử lý và đạt tốc độ 16,32 petaflop (16,32 triệu tỷ phép tính mỗi giây)
- IBM sản xuất
- Hai hệ thống nữa của Mỹ trong Top 10 Supercomputer 2012 là "báo gấm" Cray Jaguar (đứng thứ 6) và FERMI (thứ 7).

9/10/2024

Duy-Hieu Bui, ITI/SISLAB

33



Một số siêu máy tính...



Fujitsu K

- 2011, có 705.024 lõi và đạt 10,51 petaflop.
- 830 megaflop mỗi watt

9/10/2024

Duy-Hieu Bui, ITI/SISLAB



Một số siêu máy tính...



 SuperMUC, Đức không chỉ ít tốn năng lượng hơn Tianhe-1a mà còn đánh bại cỗ máy tính của Trung Quốc về hiệu suất với tốc độ 2,9 petaflop. Nó phục vụ cho công việc nghiên cứu và giảng dạy tại một số trường đại học ở Munich. Ngoài ra, Đức còn sở hữu JuQueen (xếp thứ 8) dù hệ thống do IBM sản xuất.

9/10/2024

Duy-Hieu Bui, ITI/SISLAB

35



Một số siêu máy tính...



• Pháp: Curie tốc độ 1,36 petaflop giúp Pháp được nhắc đến trong danh sách 10 siêu máy tính.

9/10/2024

Duy-Hieu Bui, ITI/SISLAB



Hiệu năng - Performance

- Đánh giá hiệu năng máy tính dựa chủ yếu vào 2 tham số
 - Latency (how long to do X)
 - · Còn được gọi là thời gian thực thi / thời gian đáp ứng
 - Throughput (how often can it do X)
- Hiệu năng phụ thuộc vào kiến trúc
 - Tần số xung đồng hồ
- Với CPU, hay sử dụng tham số số lệnh thực hiện được trong 1 giây
 - MIPS: Millions Instructions Per Second
 - MFLOPS: Millions FLOating Point Instructions Per Second
- Hiệu năng toàn hệ thống được xác định thông qua những chương trình đánh giá chuyên biệt benchmarks

9/10/2024

Duy-Hieu Bui, ITI/SISLAB

37



Hiệu năng...

- CPI & IPC
 - Thời gian thi hành T_{exe} = NI * CPI * T_C

NI: Number of instructions

CPI: Cycle per Instruction

T_C: Hàm công nghệ cho phép xác định thời gian thi hành các bước cơ bản

- IPC = 1/CPI : Instructions per Cycle
- $-N_{\text{MIPS}} = NI \times 10^{-6} / T_{\text{exe}} = F \times IPC (F tần số được đo bằng MHz)$
- CPI thể hiện hiệu năng của các kiến trúc cứng: cho phép so sánh hiệu năng của các phần cứng được thực thi từ cùng môt kiến trúc mềm,
- CPI cho phép so sánh các kiến trúc của bộ vi xử lý (chẳng hạn RISC và CISC).

9/10/2024

Duy-Hieu Bui, ITI/SISLAB

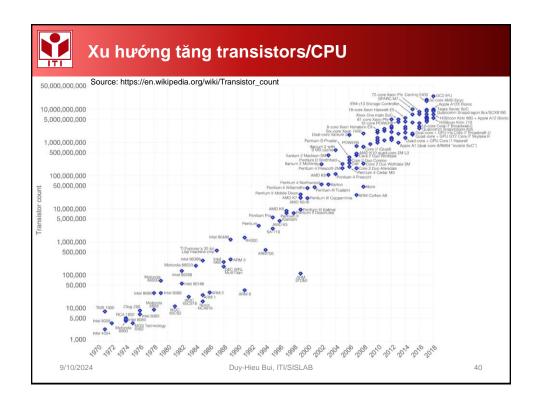
39



Luật Moore - 1965

- Số transistors trên một đơn vị diện tích sẽ tăng gấp đôi sau mỗi 24 tháng
 - <1970, gấp đôi mỗi năm
 - David House: hiệu năng sẽ tăng gấp đôi sau mỗi 18 tháng
- Giá thành chip không đối
- Mật độ cao hơn → hiệu năng cao hơn
- Kích thước bé hơn → mức độ linh hoạt cao hơn (flexibility)
- · Giảm công suất tiêu thụ và cần có các thiết bị làm lạnh
- Giảm hệ thống liên kết giữa các phần tử, tăng độ tin cậy của chip
- Một số khuynh hướng dẫn xuất:
 - Processor performance: Twice as fast after ~18 months
 - Memory capacity: Twice as much in <2 years
- http://www.intel.com/research/silicon/mooreslaw.htm

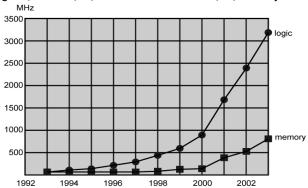
9/10/2024 Duy-Hieu Bui, ITI/SISLAB





Các yếu tố tác động đến hiệu năng

- Tốc độ bộ vi xử lý
- Dung lượng bộ nhớ
 - Ràng buộc: tốc độ bộ nhớ luôn đi sau tốc độ bộ vi xử lý



Hiệu năng truy cập dữ liệu từ thiết bị vào/ra

9/10/2024

Duy-Hieu Bui, ITI/SISLAB

41

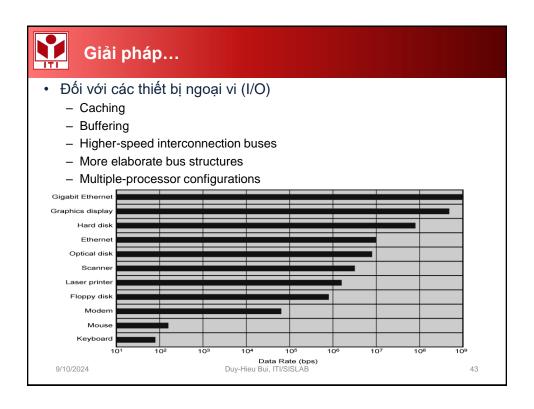


Giải pháp tăng hiệu năng

- Tăng hiệu năng các thành phần cấu thành máy tính
- Tăng số bits thao tác tại mỗi thời điểm (8-16-32-64, ...)
 - Tạo DRAM "rộng hơn" thay vì "nhanh hơn"
- Thay đổi tổ chức/kiến trúc bộ nhớ
 - Phân cấp bộ nhớ, sử dụng bộ nhớ đệm cache
- Giảm tần xuất truy cập bộ nhớ
 - Cache phức tạp hơn và cache ngay trong chip
- Tăng thông lượng liên kết
 - Bus tốc độ cao
 - Phân cấp bus

9/10/2024

Duy-Hieu Bui, ITI/SISLAB



Thực tế Cân đối những yếu tố quyết định đến hiệu năng Processor components Main memory I/O devices Interconnection structures



Thực tế ...

- Tăng tốc độ phần cứng của vi xử lý
 - Giảm kích thước các cổng logic
 - Tăng xung nhịp đồng hồ
 - Giảm thời gian lan truyền tín hiệu giữa các cổng logic
- Tăng dung lượng và tốc độ caches
 - Đưa một phần cache vào bên trong chip → giảm thời gian truy cập cache
 - Phân cấp caches : thường có 2-3 mức cache giữa bộ nhớ và bộ vi xử lý
 - Tăng mật độ nhớ
- Thay đổi tổ chức và kiến trúc vi xử lý
 - → Tăng tốc độ thực thi
 - →song song hoá

9/10/2024

Duy-Hieu Bui, ITI/SISLAB

45



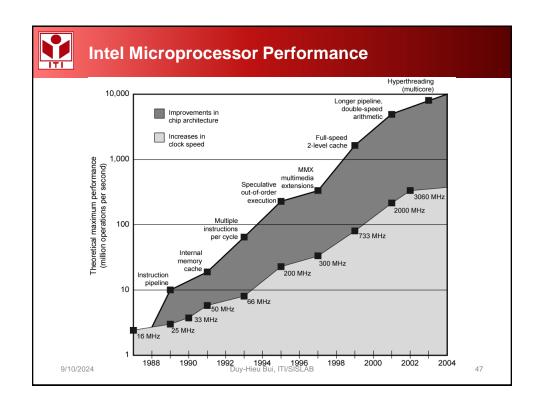
Tuy nhiên...

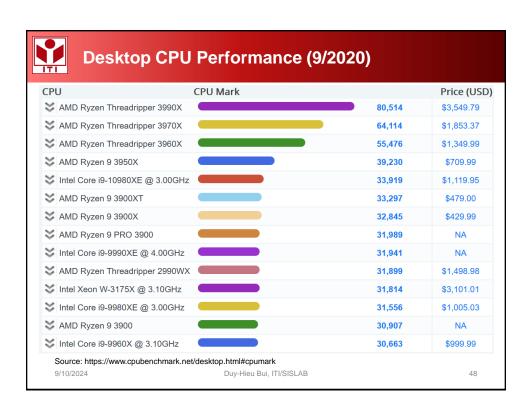
Hiệu năng còn phụ thuộc vào

- Xung đồng hồ
- Mật độ các cổng logics
- · Điện năng
 - Tăng cùng với mật độ các cổng logics và tốc độ xung clock
 - Mất nhiệt năng
- Độ trễ trở/dung kháng RC delay
 - Tốc độ dòng electrons giới hạn bởi trở kháng/dung kháng của vật liệu sử dung
 - Độ trễ tăng khi trở kháng tăng
 - Mạng liên kết càng bé, trở kháng càng tăng
 - Các cổng logics càng gần nhau, dung kháng càng tăng
- Memory latency
 - Tốc độ bộ nhớ cản trở tốc độ bộ vi xử lý
- Giải pháp:
 - Các cách tiếp cận mới về tổ chức và kiến trúc máy tính

9/10/2024

Duy-Hieu Bui, ITI/SISLAB





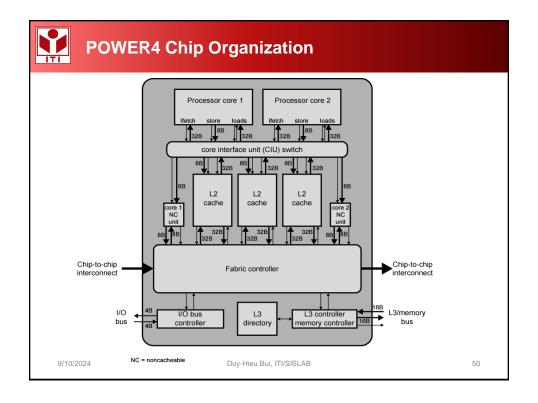


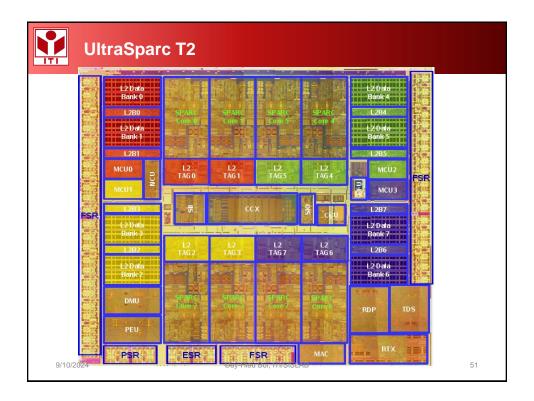
Đa lõi - Multiple Cores

- · Kiến trúc đa lõi trên cùng một chip
 - Chia sẻ bộ nhớ cache lớn
- Why?
 - Nếu phần mềm có thể tận dụng đa VXL, tăng đôi VXL đồng nghĩa với khả năng tăng đôi hiệu năng
 - Với nhiều lõi, bộ nhớ cache lớn phát huy tốt ưu điểm
 - Mức tiêu thụ năng lượng của các phần tử nhớ thấp hơn các phần tử xử lý logic
- Ví du:
 - IBM POWER4: two cores based on PowerPC
 - INTEL: DualCore, Core 2 Duo, QuadCore, ...

9/10/2024

Duy-Hieu Bui, ITI/SISLAB







Intel Processor Evolution (1)

- · 4-bit Processor
 - Intel 4004, 4040
- 8-bit Processor
 - 8080
 - first general purpose microprocessor
 - 8 bit data path
 - Used in first personal computer Altair
- 16-bit processors: origin of x86
 - 8086: much more powerful, instruction cache, prefetch few instructions
 - 8088 (8 bit external bus) used in first IBM PC
 - 80286: 16 Mbyte memory addressable, up from 1Mb
- 32-bit processors:
 - 80386: Support for multitasking
 - 80486: sophisticated powerful cache and instruction pipelining, built in maths co-processor

9/10/2024

Duy-Hieu Bui, ITI/SISLAB



Intel Processor Evolution(2)

- 32-bit processors: P5 microarchitecture
 - Pentium: Superscalar, Multiple instructions executed in parallel
 - Pentium with MMX technology
- 32-bit processors: P6/Pentium M microarchitecture
 - Pentium Pro: Increased superscalar organization, Aggressive register renaming, branch prediction, data flow analysis, speculative execution
 - Pentium II: MMX technology, graphics, video & audio processing
 - Celeron
 - Pentium III: Additional floating point instructions for 3D graphics
 - Pentium II and III Xeon
 - Pentium M
 - Celeron M
 - Intel Core
 - Dual-Core Xeon LV
- · 32-bit processors: NetBurst microarchitecture
 - Pentium 4: Further floating point and multimedia enhancements
 - Xeon
 - Mobile Pentium 4-M
 - Pentium 4 EE, 4E, 4F
- 32-bit processors: Intel 32 Intel Atom

9/10/2024

Duy-Hieu Bui, ITI/SISLAB

53

Intel Processor Evolution (3)

- 64-bit processors: IA-64
 - Itanium 1, 2: Hardware enhancements to increase speed
- 64-bit processors: Intel 64 NetBurst microarchitecture
 - Pentium 4F, Pentium D
 - Pentium Extreme Edition
 - Xeon
- 64-bit processors: Intel 64 Core microarchitecture
 - Xeon
 - Intel Core 2
 - Pentium Dual Core (Wolfdale-3M 45 nm)
 - Celeron, Celeron M
 - 64-bit processors: Intel 64 Nehalem microarchitecture





PowerPC

- 1975, 801 minicomputer project (IBM) RISC
- · Berkeley RISC I processor
- · 1986, IBM commercial RISC workstation product, RT P
 - Not commercial success
 - Many rivals with comparable or better performance
- 1990, IBM RISC System/6000
 - RISC-like superscalar machine
 - POWER architecture
- IBM alliance with Motorola (68000 microprocessors), and Apple, (used 68000 in Macintosh)
- · Result is PowerPC architecture
 - Derived from the POWER architecture
 - Superscalar RISC
 - Apple Macintosh
 - Embedded chip applications

9/10/2024

Duy-Hieu Bui, ITI/SISLAB

55

PPC601FD-080-2



PowerPC Family (1)

- 601:
 - Quickly to market. 32-bit machine
- 603:
 - Low-end desktop and portable
 - 32-bit
 - Comparable performance with 601
 - Lower cost and more efficient implementation
- 604:
 - Desktop and low-end servers
 - 32-bit machine
 - Much more advanced superscalar design
 - Greater performance
- 620:
 - High-end servers
 - 64-bit architecture

9/10/2024

Duy-Hieu Bui, ITI/SISLAB



PowerPC Family (2)

- 740/750:
 - Also known as G3
 - Two levels of cache on chip
- G4 (7xxx 1999):
 - Increases parallelism and internal speed
- 970 G5 (2003):
 - Improvements in parallelism and internal speed
 - 64-bit organization
- Current: e200, e300, e500, e600, e5500, POWER6, POWER7, PPC4xx, PPC7xx, PPC A2, Xenon (Xbox), Cell(PS3), Broadway...

9/10/2024

Duy-Hieu Bui, ITI/SISLAB

57



ARM Evolution

Family	Notable Features	Cache	Typical MIPS @ MHz
ARM1	32-bit RISC	None	
ARM2	Multiply and swap instructions; Integrated memory management unit, graphics and I/O processor	None	7 MIPS @ 12 MHz
ARM3	First use of processor cache	4 kB unified	12 MIPS @ 25 MHz
ARM6	First to support 32-bit addresses; floating-point unit	4 kB unified	28 MIPS @ 33 MHz
ARM7	Integrated SoC	8 kB unified	60 MIPS @ 60 MHz
ARM8	5-stage pipeline; static branch prediction	8 kB unified	84 MIPS @ 72 MHz
ARM9		16 kB/16 kB	300 MIPS @ 300 MHz
ARM9E	Enhanced DSP instructions	16 kB/16 kB	220 MIPS @ 200 MHz
ARM10E	6-stage pipeline	32 kB/32 kB	
ARM11	9-stage pipeline	Variable	740 MIPS @ 665 MHz
Cortex	13-stage superscalar pipeline	Variable	2000 MIPS @ 1 GHz
XScale	Applications processor; 7-stage pipeline	32 kB/32 kB L1 512 kB L2	1000 MIPS @ 1.25 GHz

9/10/2024

Duy-Hieu Bui, ITI/SISLAB

Tổng kết

- · Một số khái niệm cơ bản
 - Kiến trúc & Tổ chức máy tính
 - Cấu trúc và chức năng
 - Mô hình máy tính Von Neuman
- Sơ lược về quá trình phát triển của máy tính
- Các yếu tố liên quan đến hiệu năng và cách thức nâng cao hiệu năng
- Sơ lược về các dòng chip thông dụng của Intel và IBM

Chuẩn bị cho bài sau: Đọc chương 3 của [1]

9/10/2024

Duy-Hieu Bui, ITI/SISLAB