



HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

TÊN MÔN HỌC

TS. Nguyễn Đức Minh
minhnd@ptit.edu.vn



Chương 6: Kiến trúc máy tính tiên tiến

6.1 Kiến trúc Intel IA32/64

6.2 Kiến trúc bộ xử lý mảng và song song

6.3 Kiến trúc SIMD và MIMD

6.4 Bộ xử lý đồ họa (GPU) và siêu máy tính



Chương 6: Kiến trúc máy tính tiên tiến

Nhu cầu nâng cao hiệu năng tính toán

- Nhiều lĩnh vực mới như đồ họa máy tính, trí tuệ nhân tạo, xử lý số tín hiệu,... đòi hỏi phải xử lý một khối lượng dữ liệu rất lớn do đó cần phải có những hệ thống máy tính mạnh để thực hiện được những yêu cầu trong thực tế.
- Hầu hết với những bài toán này thì những máy tính xử lý tuần tự (Von Neumann) không đáp ứng được.
- Muốn tăng được khả năng tính toán của các hệ thống máy tính:
 - ✓ Tăng mật độ tích hợp trên chip vi xử lý: Gặp giới hạn vật lý !
 - ✓ Tổ chức kết hợp nhiều bộ xử lý (processor) với nhau → khai thác khả năng xử lý song song của chúng.

Xử lý song song là quá trình xử lý gồm nhiều tiến trình được kích hoạt đồng thời và cùng tham gia giải quyết một vấn đề, thực hiện trên những hệ thống đa bộ xử lý.



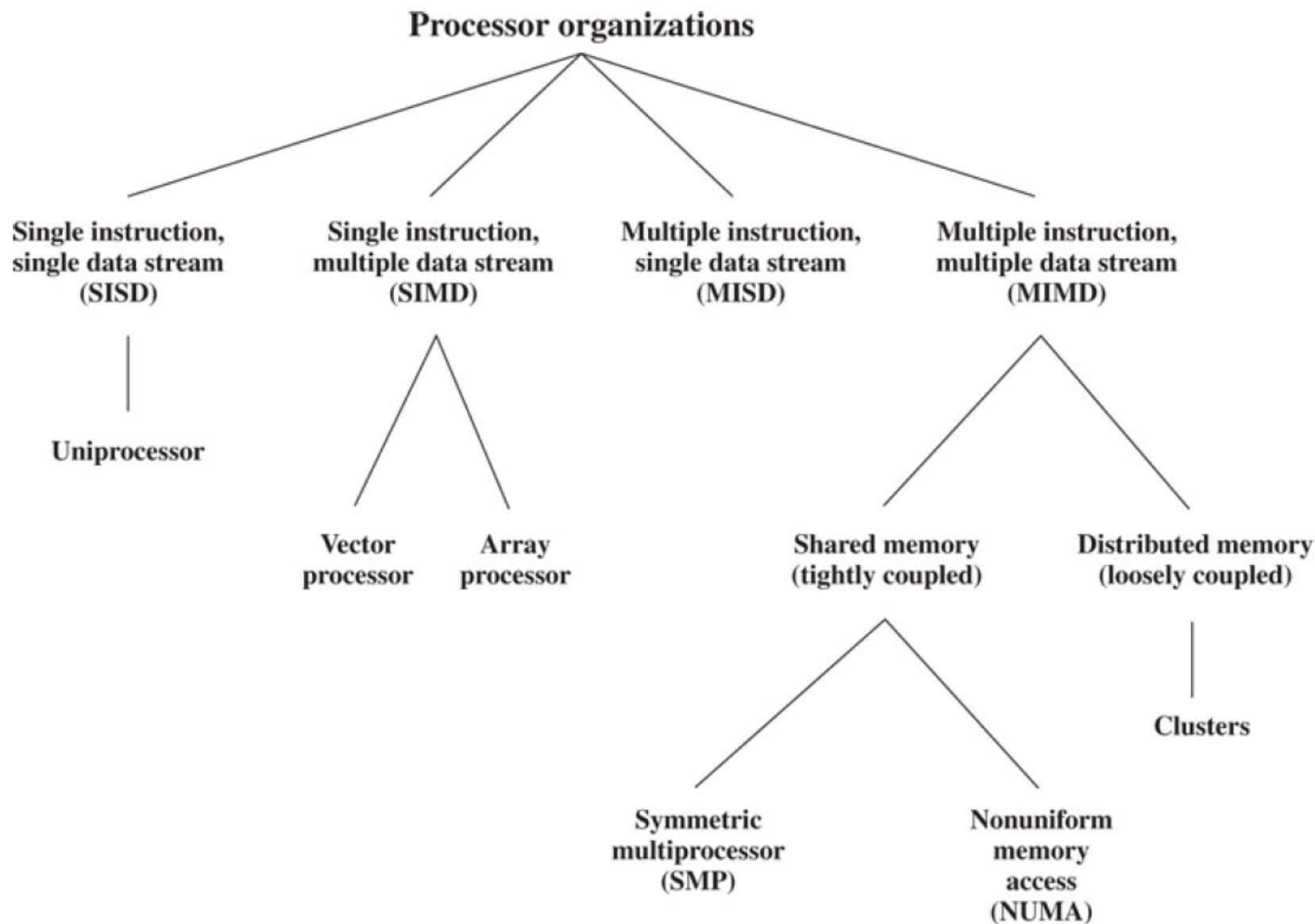
Chương 6: Kiến trúc máy tính tiên tiến

6.1 Phân loại kiến trúc máy tính

- ❖ Có rất nhiều cách phân loại kiến trúc máy tính dựa trên các tiêu chí đánh giá khác nhau:
 - ✓ Flynn's classification (1966) → thông dụng nhất → tìm hiểu theo hướng này !
 - ✓ Kuck (1978)
 - ✓ Hwang and Briggs (1984)
 - ✓ Erlangen (1981)
 - ✓ Giloi (1983)
 - ✓ Skillicorn (1988)
 - ✓ Bell (1992)



Chương 6: Kiến trúc máy tính tiên tiến





Chương 6: Kiến trúc máy tính tiên tiến

Phân loại KTMT của Michael Flynn (1966)

1. Mô hình Đơn dòng lệnh, đơn dòng dữ liệu (SISD: Single Instruction, Single Data Stream).

Đây thực chất chính là kiến trúc Von Neumann (Uniprocessor).

2. Mô hình Đơn dòng lệnh, đa dòng dữ liệu (SIMD: Single Instruction, Multiple Data Stream).

- Một lệnh máy duy nhất điều khiển việc thực thi đồng thời một số phần tử xử lý
- Mỗi phần tử xử lý có một bộ nhớ dữ liệu được liên kết, do đó, các lệnh được thực thi trên các tập dữ liệu khác nhau bởi các bộ xử lý khác nhau.
- Các bộ xử lý kiểu vector (vector processor) và mảng (array processor) đều thuộc loại mô hình này.



Chương 6: Kiến trúc máy tính tiên tiến

3. Mô hình Đa dòng lệnh, đơn dòng dữ liệu (MISD: Multiple Instruction, Single Data Stream).

- Một chuỗi dữ liệu được truyền đến một tập các bộ xử lý, mỗi bộ xử lý thực hiện một chuỗi lệnh khác nhau. Cấu trúc này khó thực hiện và không được phổ biến rộng rãi.

4. Mô hình Đa dòng lệnh, đa dòng dữ liệu (MIMD: Multiple Instruction, Multiple Data Stream).

- Một tập các bộ xử lý thực hiện đồng thời các chuỗi lệnh khác nhau trên các tập dữ liệu khác nhau.
- Các hệ thống SMP (đa xử lý đối xứng), cụm và NUMA (truy nhập bộ nhớ không đồng nhất) phù hợp với kiểu mô hình này.

		Data	Stream
		Single	Multiple
Instruction	Single	SISD	SIMD
Stream	Multiple	MISD	MIMD

Phân loại của Flynn



Chương 6: Kiến trúc máy tính tiên tiến

6.1.1 Mô hình Đơn luồng lệnh, Đơn luồng dữ liệu (SISD)

- Máy tính loại SISD chỉ có một đơn vị xử lý (PU), ở mỗi thời điểm thực hiện một chỉ lệnh và chỉ đọc, ghi một mục dữ liệu.
- Tất cả các máy tính SISD chỉ có một thanh ghi được gọi là bộ đếm chương trình (program counter) được sử dụng để nạp địa chỉ của lệnh tiếp theo khi xử lý tuần tự và kết quả là thực hiện theo một thứ tự xác định của các câu lệnh.
- Mô hình SISD còn được gọi là SPSD, đơn chương trình và đơn luồng dữ liệu.
- Đây chính là mô hình *máy tính truyền thống kiểu Von Neumann* !

Ví dụ: IBM 704, VAX 11/780, CRAY-1



Mô hình máy tính SISD



Chương 6: Kiến trúc máy tính tiên tiến

6.1.2 Mô hình Đơn luồng lệnh, Đa luồng dữ liệu (SIMD)

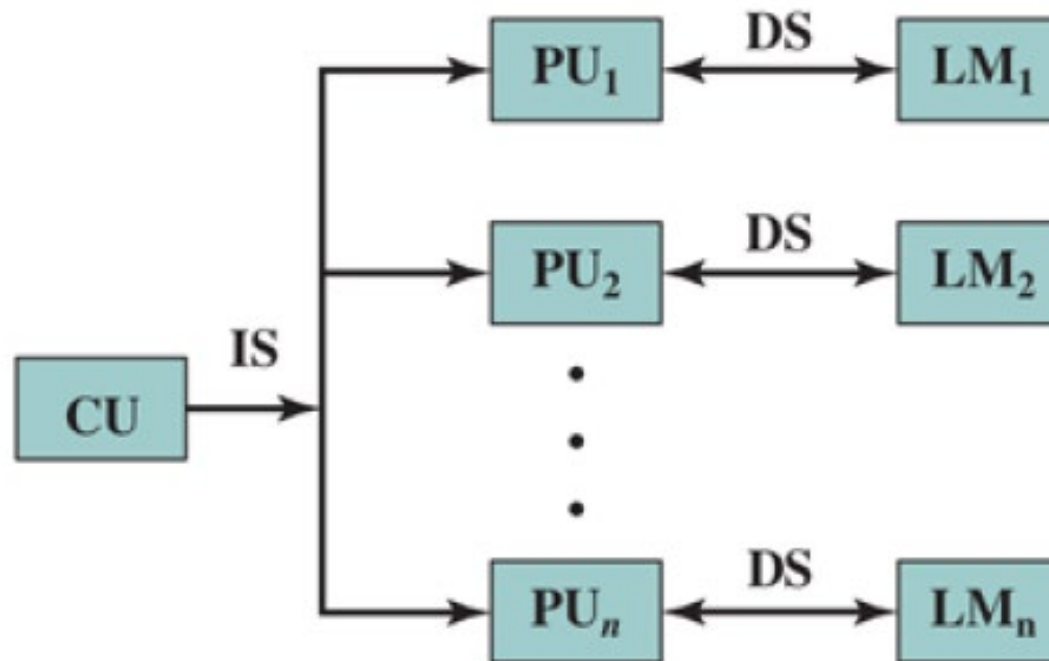
- Máy tính loại SIMD (đơn chương trình, đa luồng dữ liệu), có một đơn vị điều khiển (CU) để điều khiển nhiều đơn vị xử lý (PU) thực hiện theo một luồng các lệnh (IS).
 - CPU đưa luồng lệnh (IS- Instruction Stream) tới tất cả các phần tử xử lý, những bộ xử lý này cùng thực hiện một phép toán trên các mục dữ liệu khác nhau, nghĩa là mỗi bộ VXL có luồng dữ liệu (Data Stream) riêng. Chúng cũng sử dụng các bộ nhớ cục bộ riêng (LM – Local Memory) hoặc 1 bộ nhớ dùng chung (Sharred Memory).
 - Máy tính SIMD có thể hỗ trợ xử lý kiểu vector, trong đó có thể gán các phần tử của dữ liệu vector cho các phần tử xử lý để tính toán đồng thời.
 - *Máy tính vector và các bộ xử lý mảng* là mô hình thuộc loại này.
- Ví dụ: *ILLIAC-IV, MPP, CM-2, STARAN*



Chương 6: Kiến trúc máy tính tiên tiến

61.2 Mô hình Đơn luồng lệnh, Đa luồng dữ liệu (SIMD)

Mô hình máy tính SIMD với bộ nhớ phân tán





Chương 6: Kiến trúc máy tính tiên tiến

6.1.3 Mô hình Đa luồng lệnh, Đơn luồng dữ liệu (MISD)

- Máy tính loại MISD là ngược lại với SIMD. Máy tính MISD có thể thực hiện nhiều chương trình (nhiều lệnh) trên cùng một mục dữ liệu, nên còn được gọi là MPSD (đa chương trình, đơn luồng dữ liệu).

Kiến trúc kiểu này có thể chia thành hai nhóm:

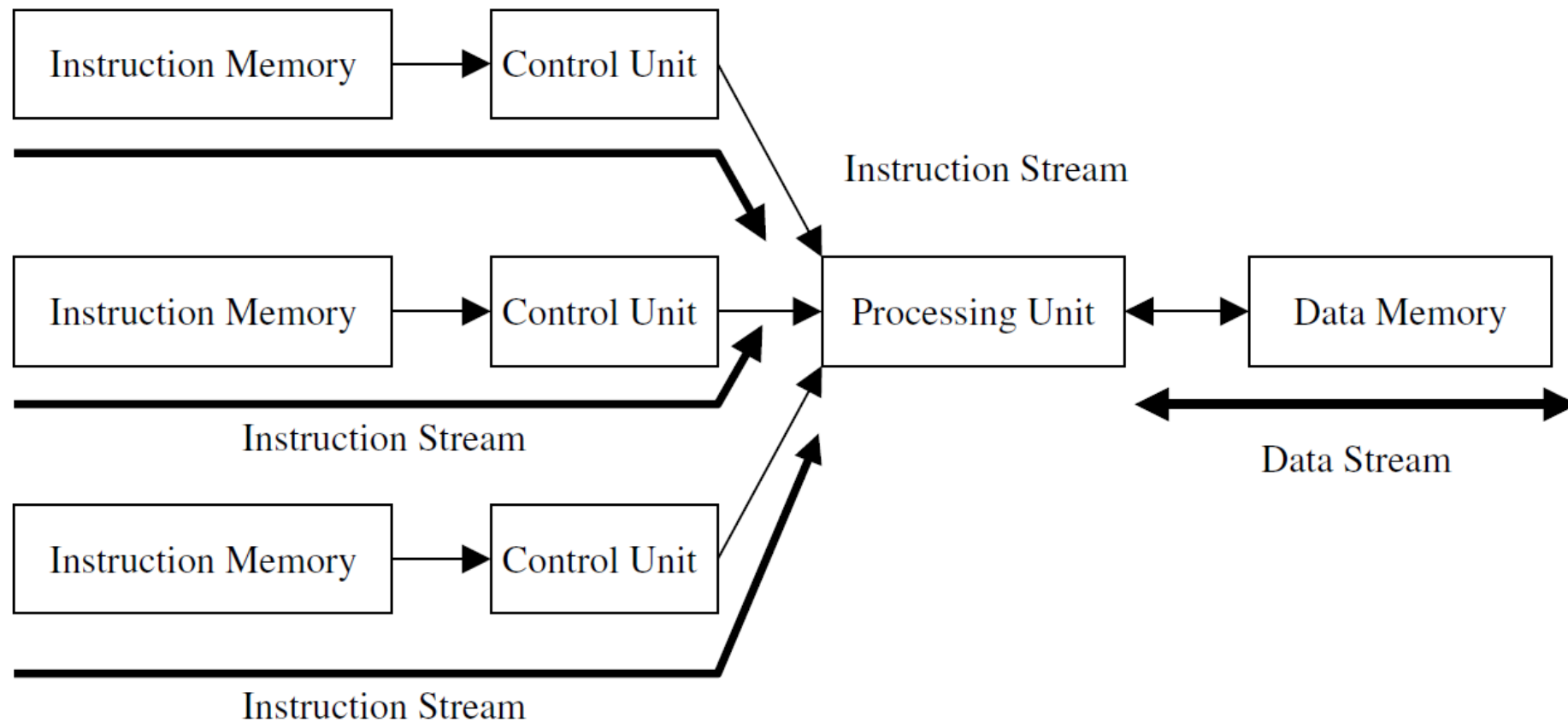
- Lớp các máy tính yêu cầu những đơn vị xử lý khác nhau có thể nhận được những chỉ lệnh khác nhau để thực hiện trên cùng một mục dữ liệu. Đây là kiến trúc khó và hiện nay *chưa có loại máy tính nào được sản xuất theo loại này*.

- Lớp các máy tính có các luồng dữ liệu được chuyển tuần tự theo dãy các CPU liên tiếp. Đây là loại kiến trúc hình ống thực hiện xử lý theo vector thông qua một dãy các bước, trong đó mỗi bước thực hiện một chức năng và sau đó chuyển kết quả cho đơn vị xử lý thực hiện bước tiếp theo.



Chương 6: Kiến trúc máy tính tiên tiến

6.1.3 Mô hình Đa luồng lệnh, Đơn luồng dữ liệu (MISD)





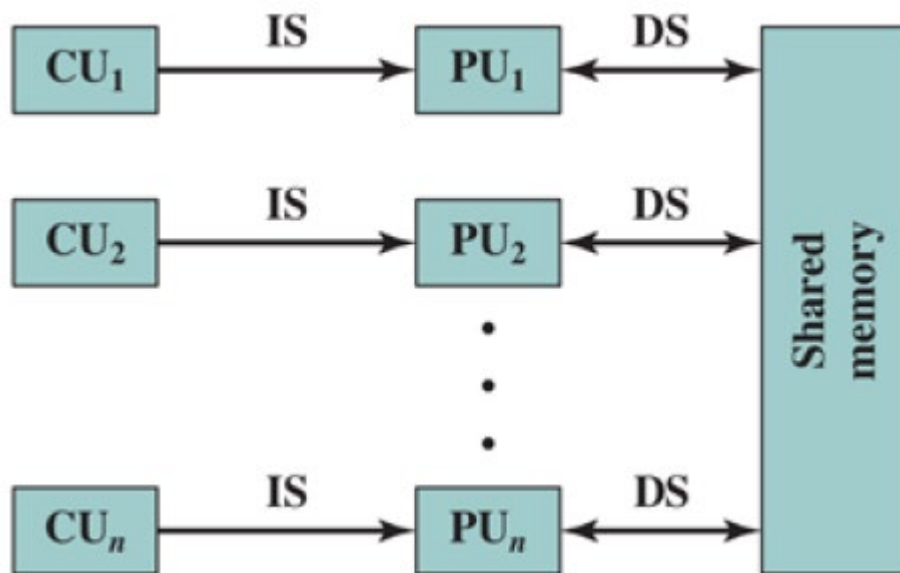
Chương 6: Kiến trúc máy tính tiên tiến

6.1.4 Mô hình Đa luồng lệnh, Đa luồng dữ liệu (MIMD)

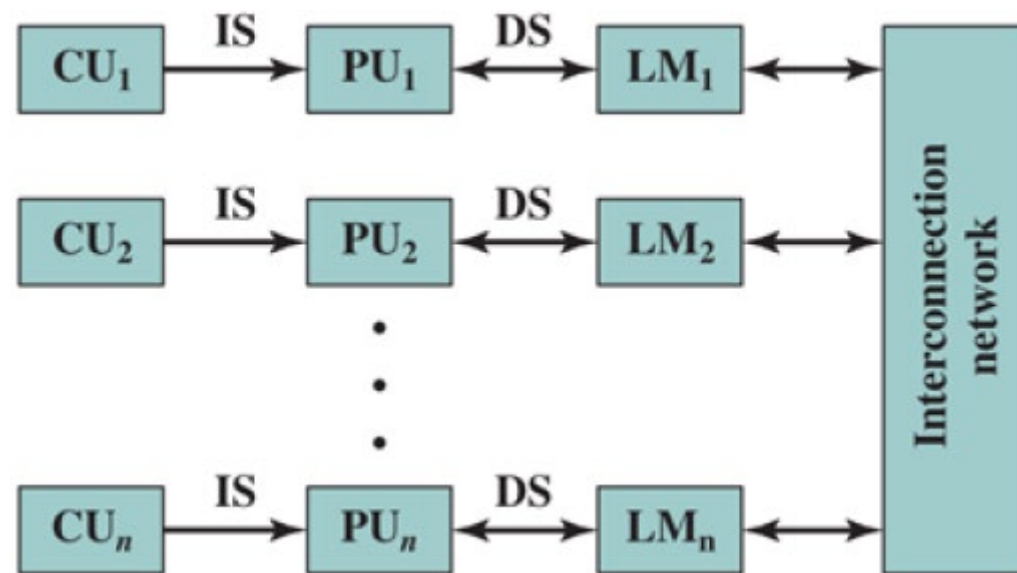
- Máy tính loại MIMD còn gọi là đa bộ xử lý (Processor), trong đó mỗi Processor có thể thực hiện những luồng lệnh khác nhau trên các luồng dữ liệu riêng.
 - Hầu hết các hệ thống MIMD đều có bộ nhớ riêng và cũng có thể truy cập vào được bộ nhớ chung khi cần, do vậy giảm thiểu được sự trao đổi giữa các Processor trong hệ thống.
 - Đây là kiến trúc *phức tạp nhất*, nhưng nó là mô hình *hỗ trợ xử lý song song cao nhất*
- Ví dụ : *BBN Butterfly, Alliant FX, isPC của Intel, ... Cm, CRAY XMP, IBM 370/168M*



Chương 6: Kiến trúc máy tính tiên tiến



Mô hình máy tính MIMD với bộ nhớ chung



Mô hình máy tính MIMD với bộ nhớ phân tán



Chương 6: Kiến trúc máy tính tiên tiến

6.2 Bộ xử lý mảng (Array Processor)

- Bộ xử lý thực hiện tính toán trên một mảng dữ liệu rộng lớn được gọi là bộ xử lý mảng.
- Bộ đa xử lý và bộ xử lý vector là những thuật ngữ khác dành cho bộ xử lý mảng.
- Chỉ thực hiện một lệnh tại một thời điểm trên một mảng dữ liệu.
- Chúng làm việc với các tập dữ liệu khổng lồ để thực hiện tính toán.
- Được sử dụng để nâng cao hiệu suất của máy tính.
- Máy tính **ILLIAC IV** là bộ xử lý mảng SIMD nổi tiếng nhất được thiết kế bởi Tập đoàn Burroughs

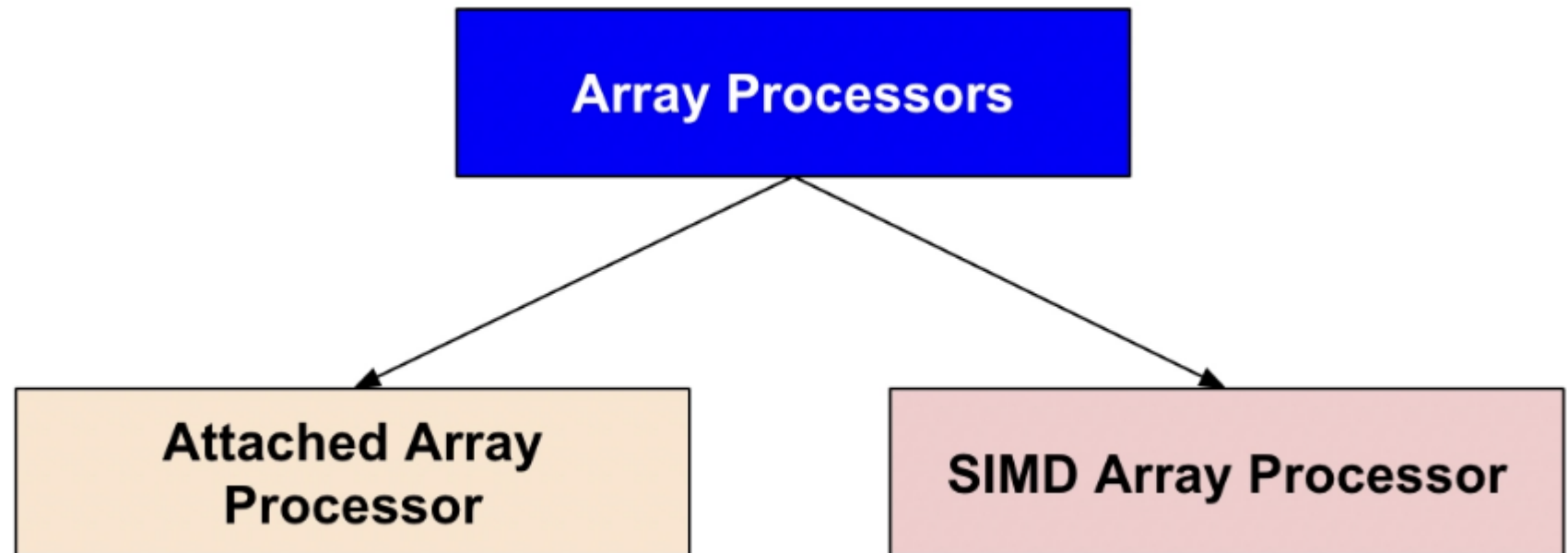


Chương 6: Kiến trúc máy tính tiên tiến

6.2.1 Phân loại bộ xử lý mảng

Bộ xử lý mảng có thể được chia thành hai loại:

1. Bộ xử lý mảng đính kèm
2. Bộ xử lý mảng SIMD (Luồng lệnh đơn, Nhiều luồng dữ liệu)

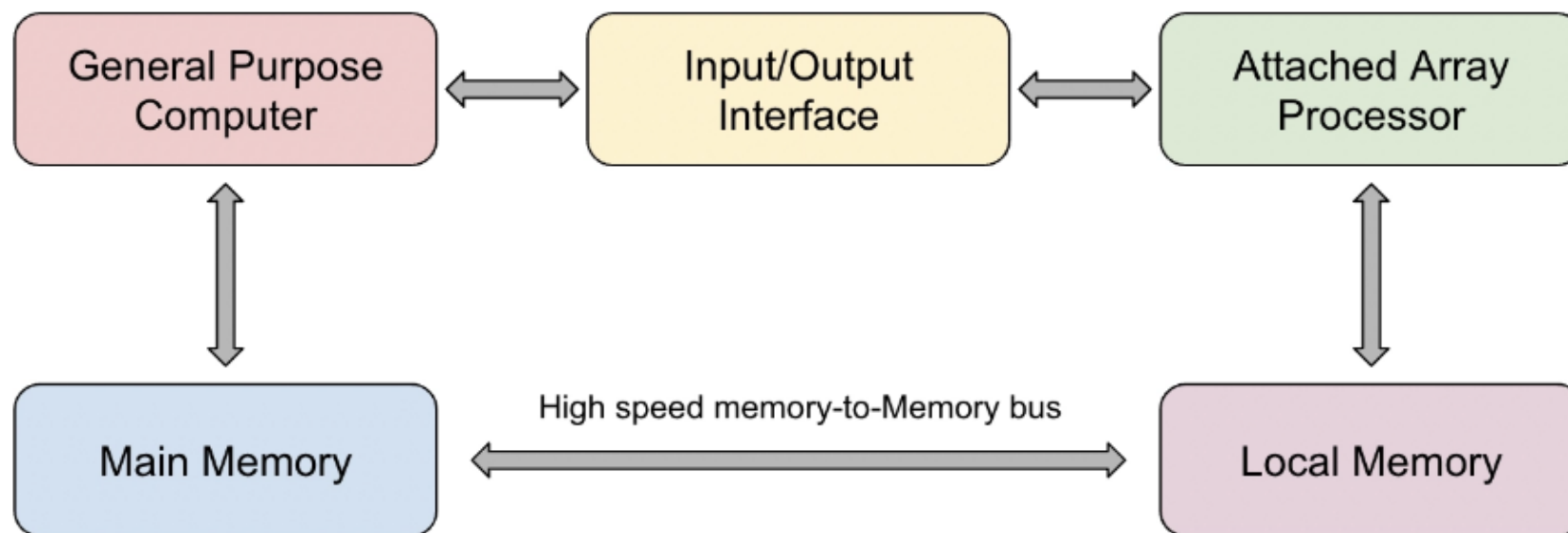




Chương 6: Kiến trúc máy tính tiên tiến

6.2.2 Bộ xử lý mảng đính kèm (Attached Array Processor)

- Bộ xử lý mảng kèm theo là bộ xử lý phụ trợ được kết nối với máy tính đa năng để nâng cao và cải thiện hiệu suất của máy trong các tác vụ tính toán số.
- Cung cấp hiệu suất rất cao bằng cách sử dụng nhiều đơn vị chức năng trong xử lý song song.
- Bộ xử lý mảng đính kèm bao gồm một bộ xử lý chung có giao diện đầu I/O và giao diện bộ nhớ cục bộ.
- Bộ nhớ chính và bộ nhớ cục bộ được liên kết với nhau



Attached Array Processor



Chương 6: Kiến trúc máy tính tiên tiến

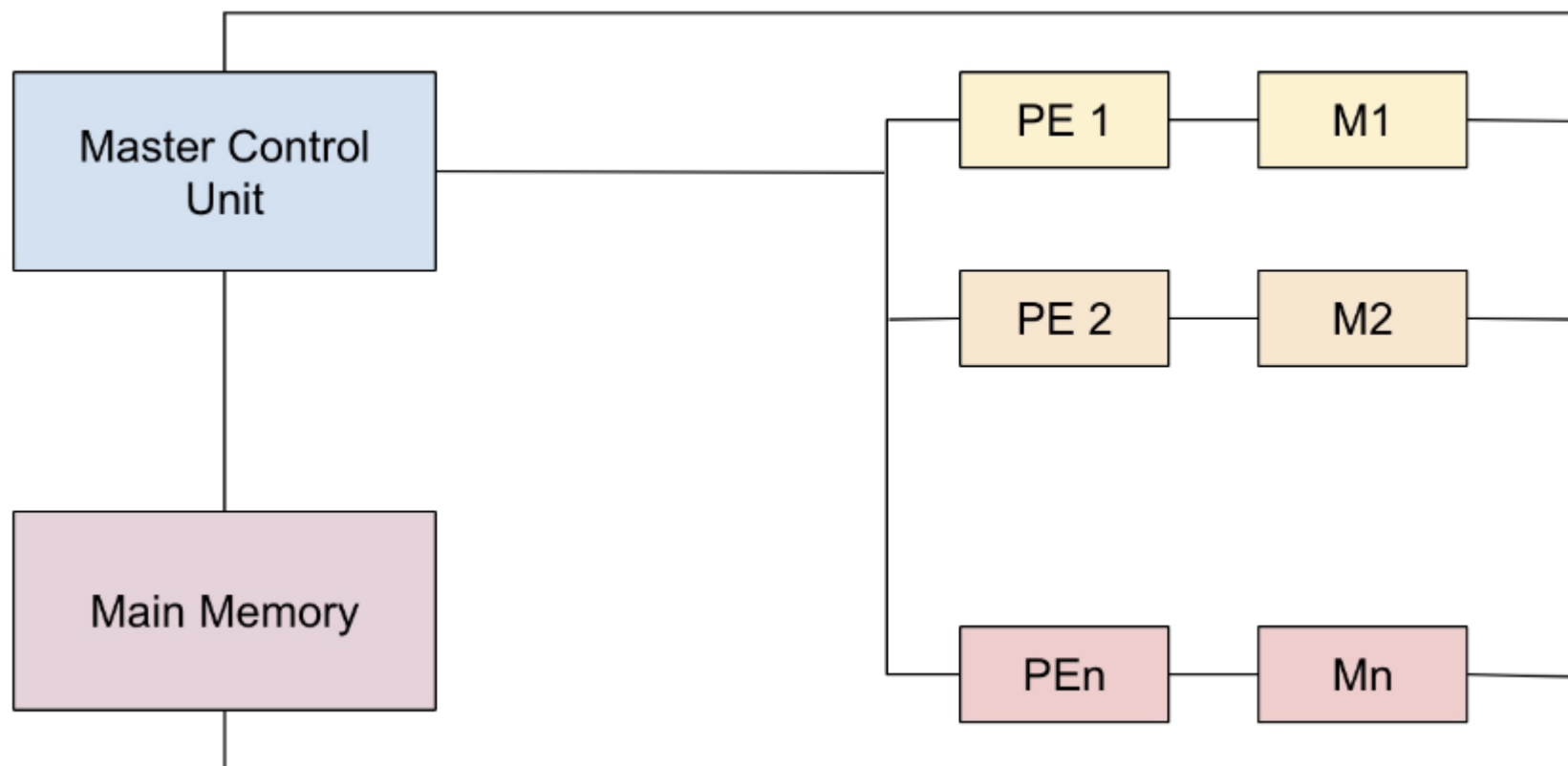
6.2.3 Bộ xử lý mảng SIMD (Single Instruction Multiple Data Stream)

- SIMD đề cập đến việc tổ chức một máy tính có nhiều bộ xử lý song song.
- Các bộ xử lý được thiết kế để hoạt động cùng nhau dưới sự giám sát của một bộ điều khiển duy nhất, tạo ra một luồng lệnh duy nhất và nhiều luồng dữ liệu.
- SIMD bao gồm một số bộ xử lý (PU) giống hệt nhau, mỗi bộ xử lý có bộ nhớ cục bộ LM (Local Memory)
- ALU và các thanh ghi được bao gồm trong mỗi phần tử bộ xử lý.
- Bộ điều khiển chính điều khiển hoạt động của các phần tử xử lý.
- Bộ điều khiển chính giải mã các lệnh và xác định cách thực hiện lệnh.
 - Chương trình được lưu trữ trong bộ nhớ chính.
 - Bộ điều khiển lấy các lệnh vector.
 - Các lệnh vector được gửi đồng thời đến tất cả các PU, kết quả được lưu trong bộ nhớ.



Chương 6: Kiến trúc máy tính tiên tiến

6.2.3 Bộ xử lý mảng SIMD (Single Instruction Multiple Data Stream)



Kiến trúc Bộ xử lý mảng SIMD (Single Instruction Multiple Data Stream)



Chương 6: Kiến trúc máy tính tiên tiến

6.2.3 Bộ xử lý mảng SIMD (Single Instruction Multiple Data Stream)

- Bộ xử lý mảng nâng cao tổng tốc độ xử lý lệnh.
- Thiết kế của hầu hết các bộ xử lý mảng đều tối ưu hóa hiệu suất của nó cho các phép tính số học lặp đi lặp lại, làm cho nó xử lý số học vector nhanh hơn CPU chủ.
- Hầu hết các bộ xử lý mảng chạy không đồng bộ từ CPU chủ nên dung lượng tổng thể của hệ thống nhờ đó được cải thiện.
- Bộ xử lý mảng có bộ nhớ cục bộ riêng, cung cấp thêm bộ nhớ bổ sung cho các hệ thống có bộ nhớ hạn chế. Đây là điều cần cân nhắc đối với các hệ thống có bộ nhớ vật lý hoặc không gian địa chỉ hạn chế.
 - Bộ xử lý mảng cực kỳ hữu ích để xử lý các vấn đề đòi hỏi nhiều sự song song. Tuy nhiên, chúng đòi hỏi phải thay đổi phương pháp lập trình.
 - Việc chuyển đổi các chương trình thông thường (tuần tự) để hỗ trợ các bộ xử lý mảng rất phức tạp và có thể cần các thuật toán (song song) khác nhau để phù hợp với cách tiếp cận song song.



Chương 6: Kiến trúc máy tính tiên tiến

6.2.3 Bộ xử lý mảng SIMD (Single Instruction Multiple Data Stream)

Xử lý mảng được sử dụng ở nhiều nơi khác nhau, bao gồm: -

- Hệ thống radar
- Hệ thống sonar
- Chống nhiễu
- Thăm dò địa chấn
- Truyền thông không dây
 - Ứng dụng y tế
 - Được sử dụng để nâng cao giọng nói
 - Được sử dụng trong các ứng dụng Thiên văn học



Chương 6: Kiến trúc máy tính tiên tiến

6.3 Kiến trúc Intel IA-32

- IA-32 (Intel Architecture 32 bit) đôi khi còn được gọi là i386 (i486, i586, i686) là phiên bản 32 bit của kiến trúc tập lệnh x86 được phát triển lên.
- Triển khai lần đầu tiên trên Intel **80386** (1985) cho tới nay vẫn sản xuất (vi điều khiển Intel Quark)
- Đây là phiên bản đầu tiên của x86 hỗ trợ tính toán 32bit.

Intel IA-32 (x86) Microarchitectures & CPUs	
P5	Mobile
Pentium Pentium MMX	Pentium M Core Duo
P6	Core
Pentium Pro Pentium II Pentium III	Core 2 Duo Core 2 Extreme
NetBurst	Nehalem
Pentium 4 Pentium D Xeon	Core i3 Core i5 Core i7

IA-32 Microarchitectures



Chương 6: Kiến trúc máy tính tiên tiến

Intel Processor	Date Introduced	Micro-architecture	Top-Bin Clock Frequency at Introduction	Transistors	Register Sizes ¹	System Bus Bandwidth	Max. Extern. Addr. Space	On-Die Caches ²
Intel Pentium M Processor 755 ³	2004	Intel Pentium M Processor	2.00 GHz	140 M	GP: 32 FPU: 80 MMX: 64 XMM: 128	3.2 GB/s	4 GB	L1: 64 KB L2: 2 MB
Intel Core Duo Processor T2600 ³	2006	Improved Intel Pentium M Processor Microarchitecture; Dual Core; Intel Smart Cache, Advanced Thermal Manager	2.16 GHz	152M	GP: 32 FPU: 80 MMX: 64 XMM: 128	5.3 GB/s	4 GB	L1: 64 KB L2: 2 MB (2MB Total)
Intel Atom Processor Z5xx series	2008	Intel Atom Microarchitecture; Intel Virtualization Technology.	1.86 GHz - 800 MHz	47M	GP: 32 FPU: 80 MMX: 64 XMM: 128	Up to 4.2 GB/s	4 GB	L1: 56 KB ⁴ L2: 512KB



Chương 6: Kiến trúc máy tính tiên tiến

6.3.1 Đặc trưng kiến trúc IA-32

- ❖ IA-32 có kiến trúc tập lệnh theo kiểu CISC (Complex Instruction set Computing)
- ❖ Các thanh ghi và chế độ định địa chỉ trong kiến trúc IA-32
 - ✓ 80386 mở rộng tất cả các thanh ghi 16 bit (trừ các thanh ghi đoạn) thành 32bit
 - ✓ 80386 có bus dữ liệu 32 bit, bus địa chỉ 32 bit. Có thể làm việc với dữ liệu 8, 16 hoặc 32 bit.
 - ✓ Vẫn có 7 chế độ định địa chỉ
 - ✓ Mã hóa lệnh phức tạp với rất nhiều khuôn dạng lệnh. Các lệnh của kiến trúc IA-32 có thể có chiều dài thay đổi từ 1 byte tới 17 byte.



Chương 6: Kiến trúc máy tính tiên tiến



Tập thanh ghi của Intel 80386

Name	31	0	Use
EAX			GPR 0
ECX			GPR 1
EDX			GPR 2
EBX			GPR 3
ESP			GPR 4
EBP			GPR 5
ESI			GPR 6
EDI			GPR 7
	CS		Code segment pointer
	SS		Stack segment pointer (top of stack)
	DS		Data segment pointer 0
	ES		Data segment pointer 1
	FS		Data segment pointer 2
	GS		Data segment pointer 3
EIP			Instruction pointer (PC)
EFLAGS			Condition codes



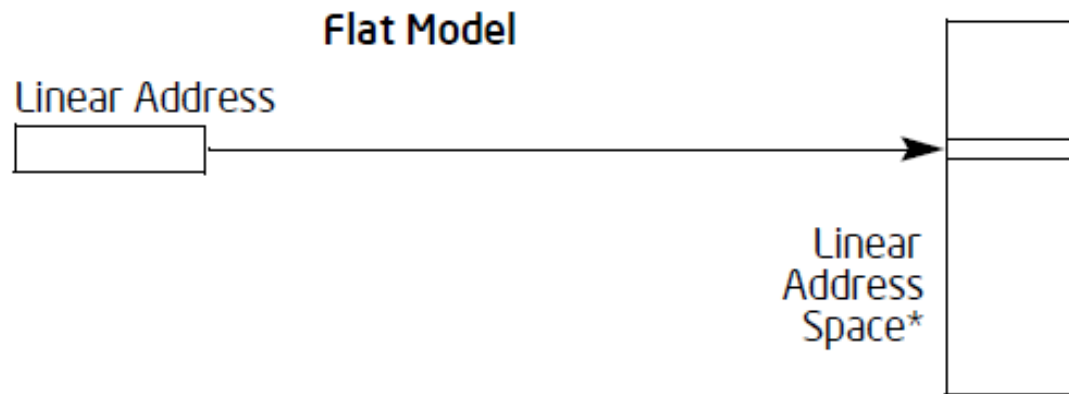
Chương 6: Kiến trúc máy tính tiên tiến

❖ Các mô hình bộ nhớ của kiến trúc IA-32

Khi truy cập bộ nhớ các chương trình sử dụng một trong 3 mô hình bộ nhớ:

1. Flat memory models (mô hình bộ nhớ phẳng):

- Bộ nhớ xuất hiện đối với chương trình dưới dạng một không gian địa chỉ liên tục duy nhất. Không gian này được gọi là không gian địa chỉ tuyến tính.
- Mã, dữ liệu và ngăn xếp đều được chứa trong không gian địa chỉ này. Không gian địa chỉ tuyến tính có thể định địa chỉ theo byte, với các địa chỉ chạy liên tục từ 0 đến $2^{32} - 1$ (nếu không ở chế độ 64 bit).
- Địa chỉ cho bất kỳ byte nào trong không gian địa chỉ tuyến tính được gọi là địa chỉ tuyến tính.





Chương 6: Kiến trúc máy tính tiên tiến

❖ Các mô hình bộ nhớ của kiến trúc IA-32

2. Segmented memory models (mô hình bộ nhớ phân đoạn):

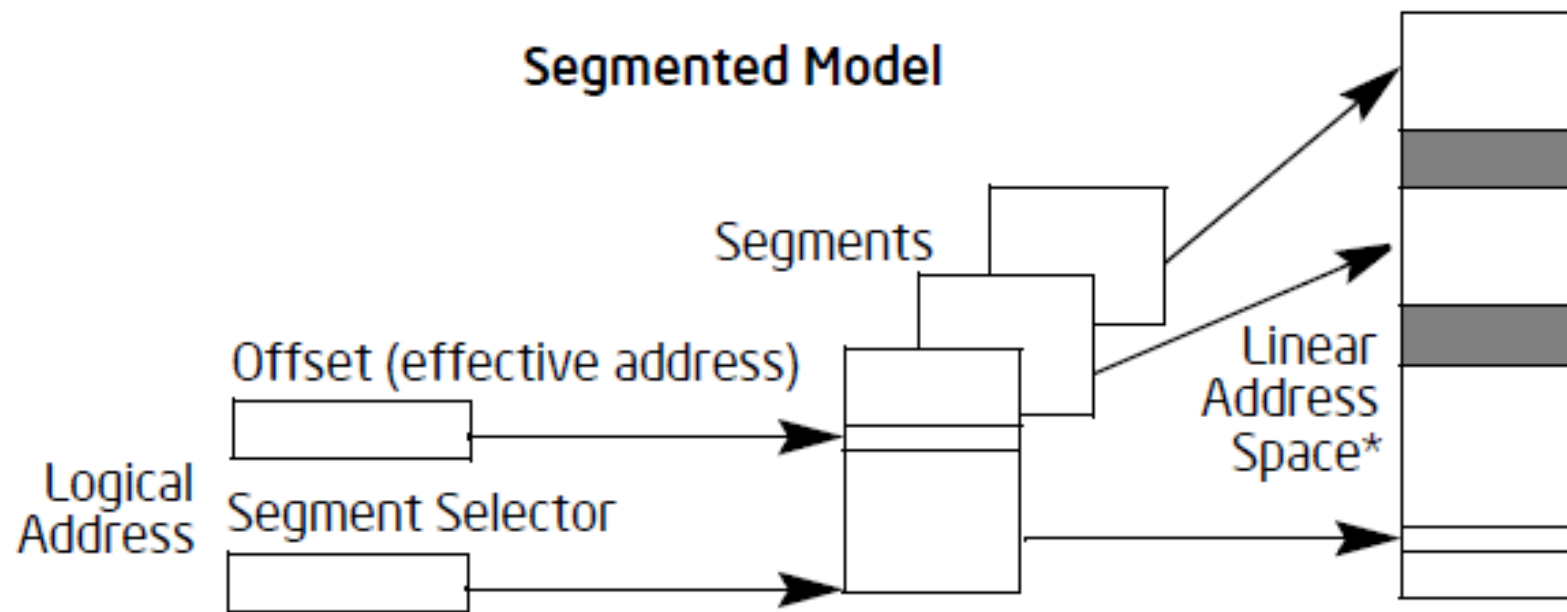
- Bộ nhớ xuất hiện trong chương trình dưới dạng một nhóm các không gian địa chỉ độc lập được gọi là các phân đoạn.
- Mã, dữ liệu và ngăn xếp thường được chứa trong các phân đoạn riêng biệt. Để đánh địa chỉ một byte trong một phân đoạn, chương trình sẽ đưa ra một địa chỉ logic. Địa chỉ này bao gồm một bộ chọn phân đoạn và một offset xác định phân đoạn được truy cập và độ lệch xác định một byte trong không gian địa chỉ của phân đoạn.
- Các chương trình chạy trên bộ xử lý IA-32 có thể xử lý tới 16.383 phân đoạn có kích thước và loại khác nhau và mỗi phân đoạn có thể lớn tới 2^{32} byte.



Chương 6: Kiến trúc máy tính tiên tiến

❖ Các mô hình bộ nhớ của kiến trúc IA-32

2. Segmented memory models (mô hình bộ nhớ phân đoạn):



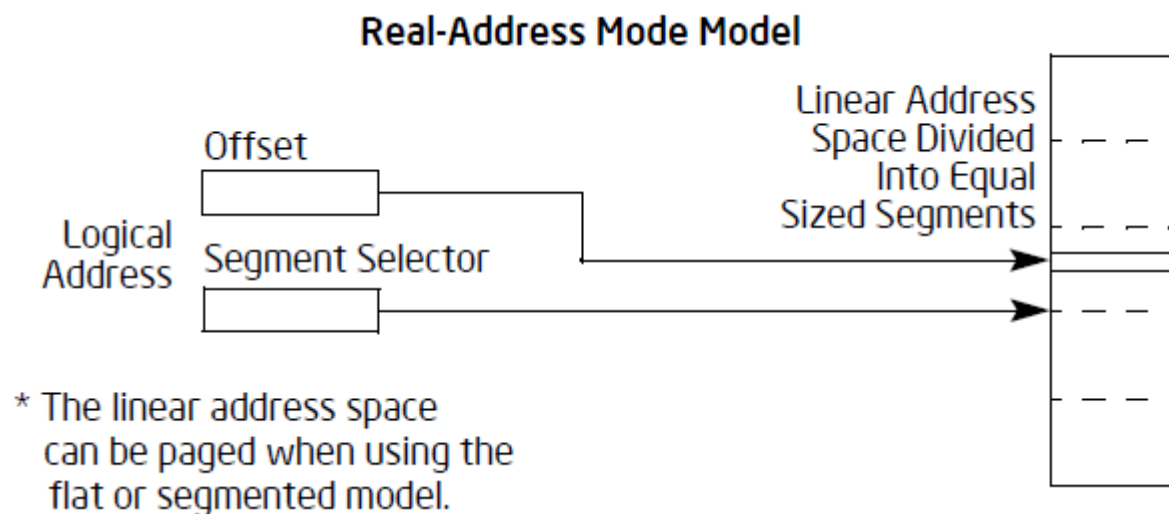


Chương 6: Kiến trúc máy tính tiên tiến

❖ Các mô hình bộ nhớ của kiến trúc IA-32

3. Real-address memory models (mô hình bộ nhớ địa chỉ thực):

- Đây là kiểu bộ nhớ dành cho bộ xử lý Intel 8086. Được hỗ trợ để cung cấp khả năng tương thích với các chương trình hiện có được viết để chạy trên bộ xử lý Intel 8086 (các chương trình cũ!).
- Chế độ địa chỉ thực sử dụng một triển khai cụ thể của bộ nhớ được phân đoạn trong đó không gian địa chỉ tuyến tính cho chương trình và hệ điều hành bao gồm một mảng các phân đoạn có kích thước lên tới 64 KByte mỗi phân đoạn.
- Kích thước tối đa của không gian địa chỉ tuyến tính ở chế độ địa chỉ thực là 2^{20} byte.





Chương 6: Kiến trúc máy tính tiên tiến

❖ Phân trang và bộ nhớ ảo

- Với mô hình bộ nhớ phẳng hoặc phân đoạn, không gian địa chỉ tuyến tính được ánh xạ vào không gian địa chỉ vật lý của bộ xử lý một cách trực tiếp hoặc thông qua phân trang (paging).
- Khi sử dụng ánh xạ trực tiếp (phân trang bị vô hiệu hóa), mỗi địa chỉ tuyến tính có sự tương ứng một-một với địa chỉ vật lý.
- Địa chỉ tuyến tính được gửi đi trên bus địa chỉ của bộ xử lý mà không cần dịch.
- Khi sử dụng cơ chế phân trang của kiến trúc IA-32 (bật chế độ phân trang), không gian địa chỉ tuyến tính được chia thành các trang được ánh xạ tới bộ nhớ ảo.
 - Các trang của bộ nhớ ảo sau đó được ánh xạ khi cần thiết vào bộ nhớ vật lý.



Chương 6: Kiến trúc máy tính tiên tiến

❖ Phân trang và bộ nhớ ảo

- Khi một hệ điều hành sử dụng phân trang, cơ chế phân trang sẽ “trong suốt” đối với chương trình ứng dụng.
- Tất cả những gì ứng dụng nhìn thấy là không gian địa chỉ tuyến tính.
- Ngoài ra, cơ chế phân trang của kiến trúc IA-32 bao gồm các khả năng mở rộng hỗ trợ:
 - Phần mở rộng địa chỉ vật lý (Physical Address Extension - PAE) để giải quyết không gian địa chỉ vật lý lớn hơn 4 GByte.
 - Tiện ích mở rộng kích thước trang (Page Size Extensions - PSE) để ánh xạ địa chỉ tuyến tính thành địa chỉ vật lý trong các trang có kích thước 4 MB.



Chương 6: Kiến trúc máy tính tiên tiến

❖ Các chế độ hoạt động ứng với các mô hình bộ nhớ

(Khi viết mã cho bộ xử lý IA-32 hoặc Intel 64, lập trình viên cần biết chế độ vận hành mà bộ xử lý sẽ sử dụng khi thực thi mã và mô hình bộ nhớ đang được sử dụng). Mỗi quan hệ giữa các chế độ hoạt động và mô hình bộ nhớ như sau:

❑ **Chế độ được bảo vệ (Protected mode)**— Khi ở chế độ được bảo vệ, bộ vi xử lý có thể sử dụng bất kỳ kiểu bộ nhớ nào được. *(Mô hình bộ nhớ ở chế độ địa chỉ thực thường chỉ được sử dụng khi bộ xử lý ở chế độ ảo 8086.)*

- Mô hình bộ nhớ được sử dụng phụ thuộc vào thiết kế của hệ điều hành.
- Khi thực hiện đa nhiệm, các tác vụ riêng lẻ có thể sử dụng các mô hình bộ nhớ khác nhau.



Chương 6: Kiến trúc máy tính tiên tiến

❖ Các chế độ hoạt động ứng với các mô hình bộ nhớ

❑ Chế độ địa chỉ thực (Real address mode)

- Khi ở chế độ địa chỉ thực, bộ xử lý chỉ hỗ trợ mô hình bộ nhớ ở chế độ địa chỉ thực.

❑ Chế độ quản lý hệ thống (System management mode)

- Khi ở SMM, bộ xử lý chuyển sang một không gian địa chỉ riêng, được gọi là RAM quản lý hệ thống (SMRAM).

- Mô hình bộ nhớ được sử dụng để đánh địa chỉ các byte trong không gian địa chỉ này tương tự như mô hình chế độ địa chỉ thực..



Chương 6: Kiến trúc máy tính tiên tiến

❖ Các chế độ hoạt động ứng với các mô hình bộ nhớ

❑ Chế độ tương thích (Compatibility mode)

- Phần mềm cần chạy ở chế độ tương thích phải tuân theo cùng một kiểu bộ nhớ như những phần mềm được thiết kế chạy ở chế độ được bảo vệ 32-bit.
- Tác dụng của việc phân đoạn cũng giống như trong chế độ được bảo vệ 32-bit.

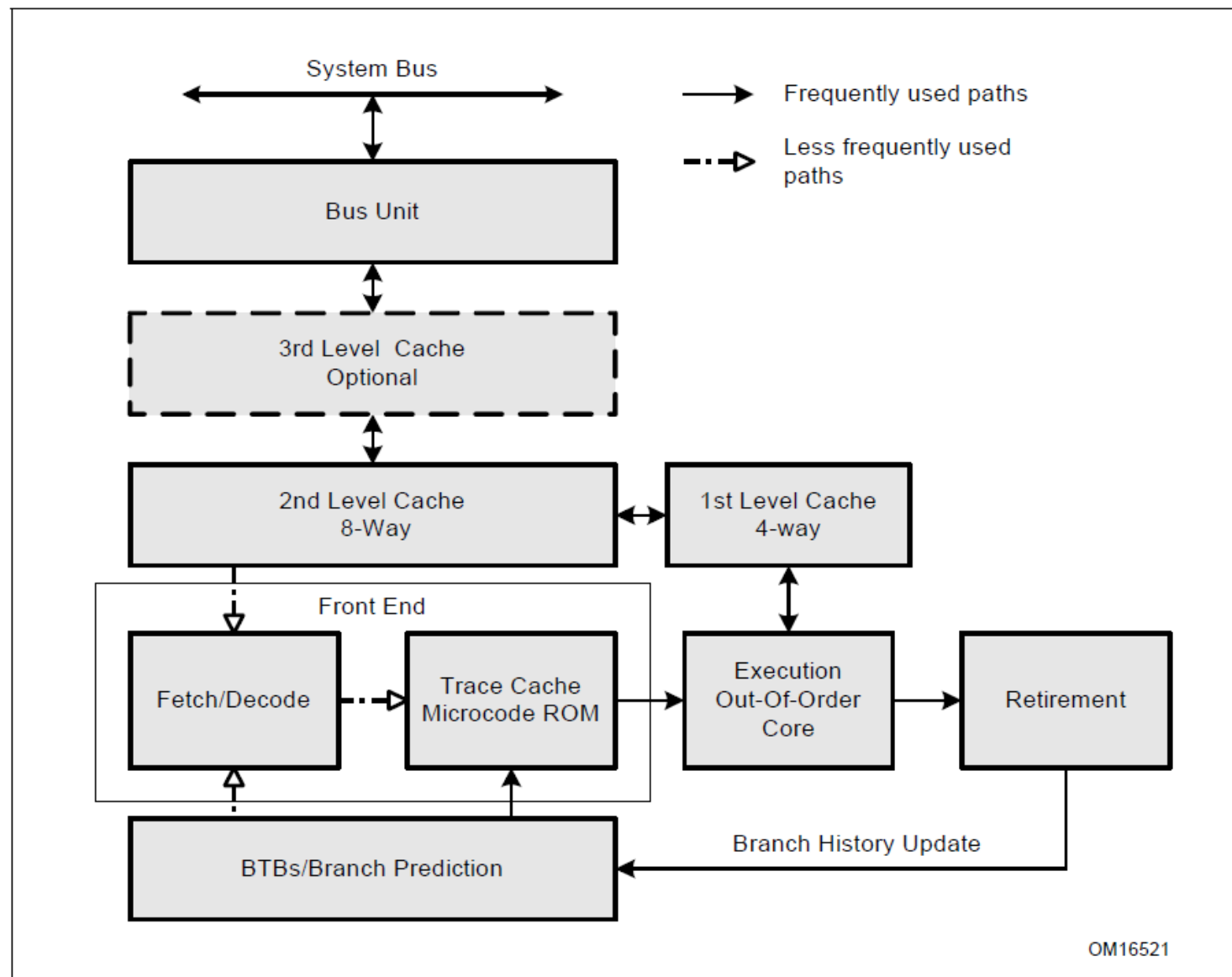
❑ Chế độ 64-bit (64 bit mode)

- Tính năng phân đoạn thường bị vô hiệu hóa (nhưng không hoàn toàn), tạo ra một không gian địa chỉ tuyến tính 64-bit phẳng.
 - Bộ xử lý xử lý cơ sở phân đoạn của CS, DS, ES và SS bằng 0 ở chế độ 64 bit (điều này làm cho địa chỉ tuyến tính bằng địa chỉ hiệu dụng).
 - Chế độ địa chỉ thực và phân đoạn không có sẵn ở chế độ 64 bit.



Chương 6: Kiến trúc máy tính tiên tiến

Vị kiến trúc Intel Netburst





Chương 6: Kiến trúc máy tính tiên tiến

6.3.2 Đặc trưng kiến trúc Intel 64 (*phát triển của IA-32*)

- ❖ Kiến trúc Intel 64 tăng không gian địa chỉ tuyến tính cho phần mềm lên 64 bit và hỗ trợ không gian địa chỉ vật lý lên tới 52 bit.
- ❖ Công nghệ này cũng giới thiệu một chế độ hoạt động mới được gọi là chế độ IA-32e. Chế độ IA-32e hoạt động ở một trong hai chế độ phụ:

(1) chế độ tương thích cho phép hệ điều hành 64-bit chạy hầu hết các phần mềm 32-bit cũ chưa được sửa đổi.

(2) Chế độ 64 bit cho phép hệ điều hành 64 bit chạy các ứng dụng được viết để truy cập không gian địa chỉ 64 bit. Ở chế độ 64 bit, các ứng dụng có thể truy cập:

- Địa chỉ tuyến tính phẳng 64-bit.
- 8 thanh ghi đa năng bổ sung (GPR).
- 8 thanh ghi bổ sung để thực hiện các phần mở rộng SIMD (Intel SSE, SSE2, SSE3 và SSSE3).
- GPR và con trỏ lệnh rộng 64-bit



Chương 6: Kiến trúc máy tính tiên tiến

6.3.2 Đặc trưng kiến trúc Intel 64

- Địa chỉ thanh ghi byte thống nhất.
 - Cơ chế ưu tiên ngắt nhanh.
 - Chế độ đánh địa chỉ tương đối của con trỏ lệnh mới.
- ❖ Bộ xử lý kiến trúc Intel 64 hỗ trợ phần mềm IA-32 hiện có vì nó có thể chạy tất cả các chế độ cũ không phải 64-bit được kiến trúc IA-32 hỗ trợ.



Chương 6: Kiến trúc máy tính tiên tiến

6.3.2 Đặc trưng kiến trúc Intel 64

Intel Processor	Date Introduced	Micro-architecture	Highest Processor Base Frequency at Introduction	Transistors	Register Sizes	System Bus/QPI Link Speed	Max. Extern. Addr. Space	On-Die Caches
Intel Xeon-Processor E3-1280	2011	Intel Turbo Boost Technology, Sandy Bridge microarchitecture; Four core; HyperThreading Technology; Intel 64 Architecture; Intel Virtualization Technology.	3.50 GHz		GP: 32, 64 FPU: 80 MMX: 64 XMM: 128 YMM: 256	DMI: 5 GT/s; Memory: 21 GB/s	1 TB	L1: 64 KB L2: 256KB L3: 8MB
Intel Xeon-Processor E7-8870	2011	Intel Turbo Boost Technology, Westmere microarchitecture; Ten core; HyperThreading Technology; Intel 64 Architecture; Intel Virtualization Technology.	2.40 GHz	2.2B	GP: 32, 64 FPU: 80 MMX: 64 XMM: 128	QPI: 6.4 GT/s; Memory: 102 GB/s	16 TB	L1: 64 KB L2: 256KB L3: 30MB



Chương 6: Kiến trúc máy tính tiên tiến

6.3.2 Đặc trưng kiến trúc Intel 64

Intel Xeon-Processor 5680	2010	Intel Turbo Boost Technology, Westmere microarchitecture; Six core; HyperThreading Technology; Intel 64 Architecture; Intel Virtualization Technology.	3.33 GHz	1.1B	GP: 32, 64 FPU: 80 MMX: 64 XMM: 128	QPI: 6.4 GT/s; 32 GB/s	1 TB	L1: 64 KB L2: 256KB L3: 12MB
Intel Xeon-Processor 7560	2010	Intel Turbo Boost Technology, Nehalem microarchitecture; Eight core; HyperThreading Technology; Intel 64 Architecture; Intel Virtualization Technology.	2.26 GHz	2.3B	GP: 32, 64 FPU: 80 MMX: 64 XMM: 128	QPI: 6.4 GT/s; Memory: 76 GB/s	16 TB	L1: 64 KB L2: 256KB L3: 24MB
Intel Core i7-2600K Processor	2011	Intel Turbo Boost Technology, Sandy Bridge microarchitecture; Four core; HyperThreading Technology; Intel 64 Architecture; Intel Virtualization Technology., Processor graphics, Quicksync Video	3.40 GHz	995M	GP: 32, 64 FPU: 80 MMX: 64 XMM: 128 YMM: 256	DMI: 5 GT/s; Memory: 21 GB/s	64 GB	L1: 64 KB L2: 256KB L3: 8MB



Chương 6: Kiến trúc máy tính tiên tiến

6.4 Kiến trúc Intel IA-64

IA-64 (còn được gọi là **kiến trúc Intel Itanium**) là kiến trúc tập lệnh (ISA) của của họ vi xử lý Intel 64 bit Itanium. Bộ xử lý Itanium đầu tiên, có tên mã *Merced*, được phát hành vào năm 2001 và ngừng sản xuất vào năm 2021.

- Được thiết kế cho các hệ thống máy chủ. Không phải kiến trúc RISC
- Mặc dù các ứng dụng được viết cho máy tính x86 và HP 9000 chạy được trên máy IA-64 nhưng các chương trình phải được biên dịch lại để tận dụng tối đa kiến trúc của IA-64.
- Thay vì các lệnh có độ dài thay đổi như trong x86 (IA-32), IA-64 sử dụng các lệnh bốn byte có độ dài cố định được gói trong bộ ba "từ lệnh dài".
 - IA-64 sử dụng 256 thanh ghi cho các phép tính số nguyên và dấu phẩy động so với 16 thanh ghi trong IA-32.
 - IA-64 cũng sử dụng các cơ chế đường ống và tiên đoán rẽ nhánh.

*Chú ý: phân biệt giữa **Intel 64** và tên mã **IA-64***





Chương 6: Kiến trúc máy tính tiên tiến

6.4 Kiến trúc Intel IA-64

Đặc trưng của Intel Itanium :

- Không phải máy RISC: trình biên dịch tạo ra các khối (bundles), mỗi khối bao gồm 1 nhóm 3 lệnh



Length:
41 bits each =
123 bits

- Không phải máy tính VLIW: Trình biên dịch tạo ra các trường tạm (template field) độ lớn 5 bit chứa đựng các thông tin về cơ chế song song mức lệnh.



- Phương thức hoạt động này được gọi là **EPIC** – *Explicitly Parallel Instruction Computing*



Chương 6: Kiến trúc máy tính tiên tiến

6.4 Kiến trúc Intel IA-64

Đặc trưng của Intel Itanium :

- Tốc độ bộ VXL 1Ghz.... 1.5Ghz
- Có các khối chức năng :
 - ✓ 6 khối Integer
 - ✓ 3 bộ tiên đoán rẽ nhánh
 - ✓ 2 bộ tính toán dấu phẩy động FMA
- Có tập thanh ghi rất lớn:
 - ✓ 128 thanh ghi dấu phẩy động (82 bit)
 - ✓ 128 thanh ghi số nguyên (64 bit)
 - ✓ 64 thanh ghi tiên đoán rẽ nhánh
 - ✓ 64 thanh ghi nhánh
 - ✓ Các thanh ghi quay/ngăn xếp thanh ghi



Chương 6: Kiến trúc máy tính tiên tiến

6.4 Kiến trúc Intel IA-64

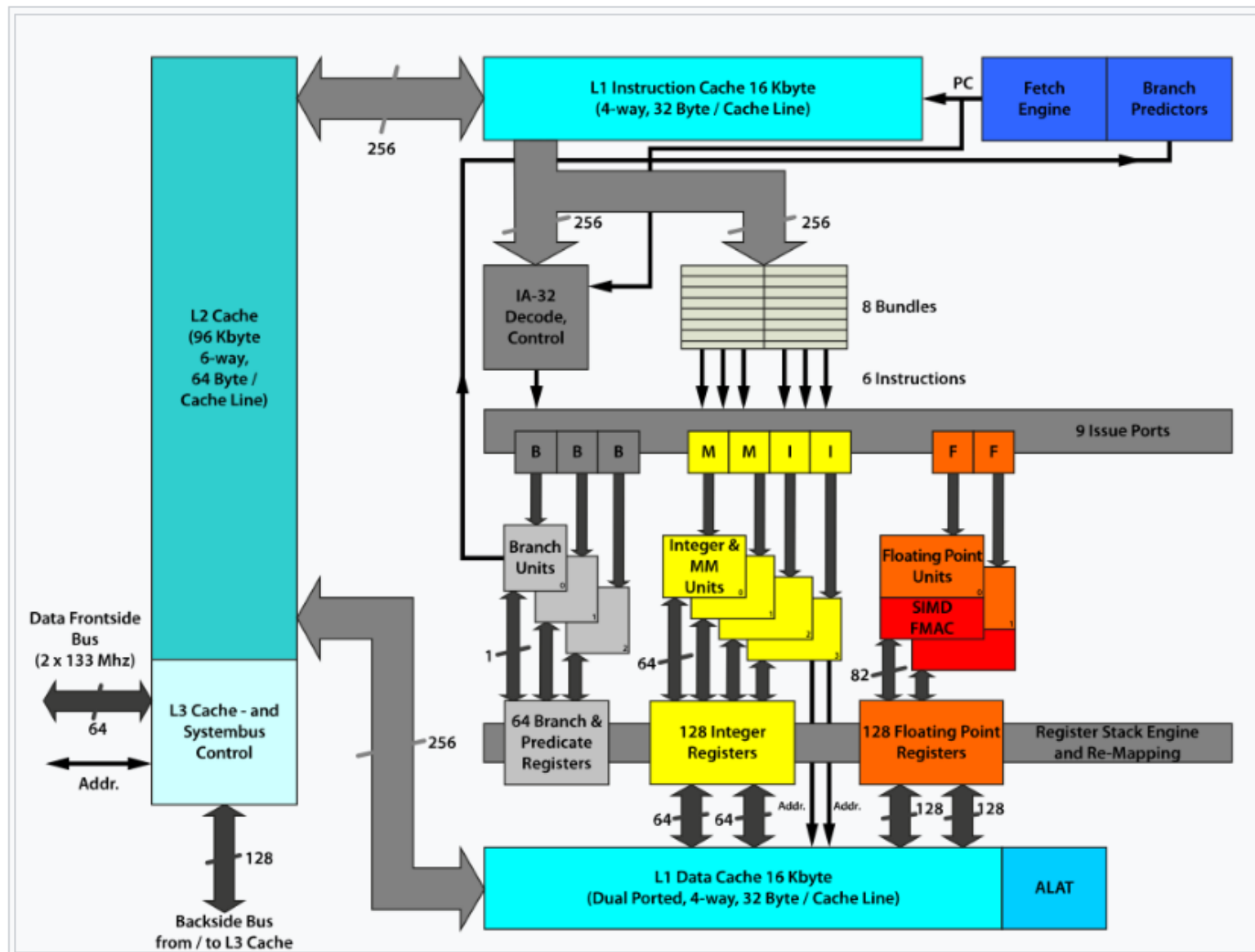
Đặc trưng của Intel Itanium :

- ✓ Có các mức cache L1, L2, L3
- ✓ Bus hệ thống có độ rộng 128 bit, chạy ở tốc độ 400MH
- ✓ Băng thông 6.4Gb/s
- ✓ Bus địa chỉ 50 bit và lên đến 64 bit ảo
- ✓ Phân trang lớn nhất lên đến 4GB



Chương 6: Kiến trúc máy tính tiên tiến

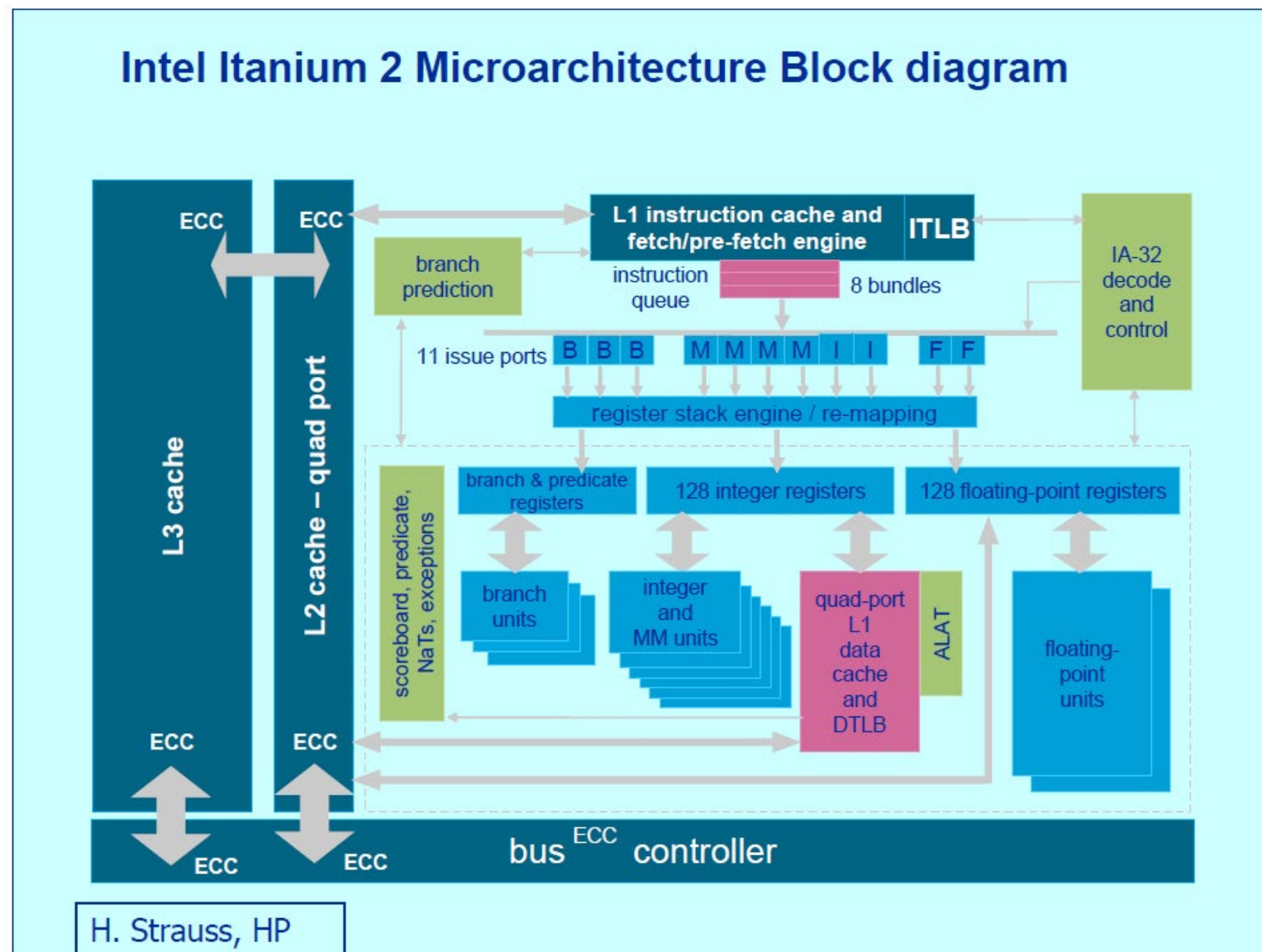
Sơ đồ khối kiến trúc Intel Itanium
(IA-64)





Chương 6: Kiến trúc máy tính tiên tiến

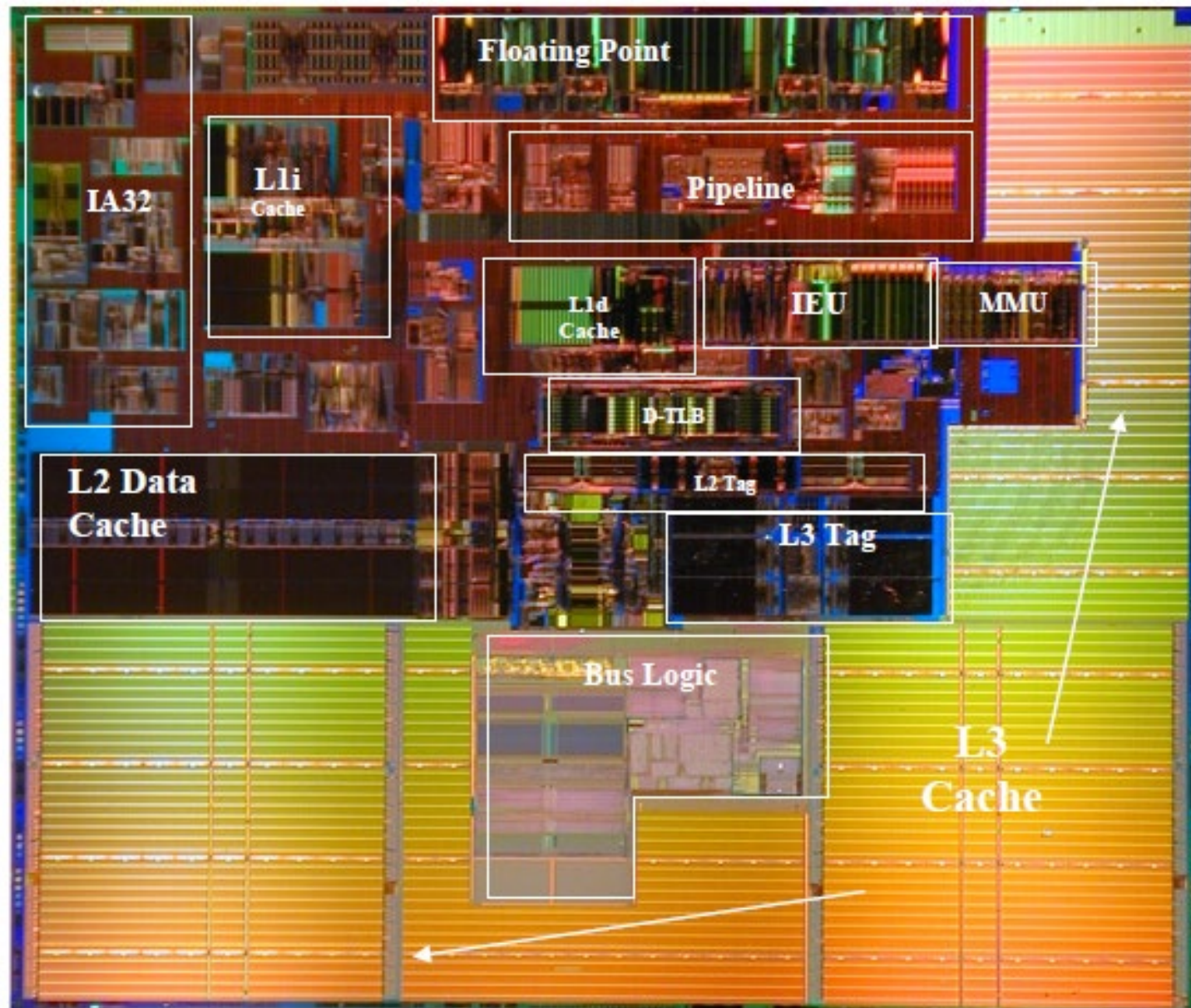
Sơ đồ khối
kiến trúc Intel Itanium 2
(IA-64)





Chương 6: Kiến trúc máy tính tiên tiến

Đế của vi kiến trúc Intel Itanium 2
(IA-64)





Chương 6: Kiến trúc máy tính tiên tiến

6.4 Bộ xử lý đồ họa và siêu máy tính

6.4.1. Bộ xử lý đồ họa GPU

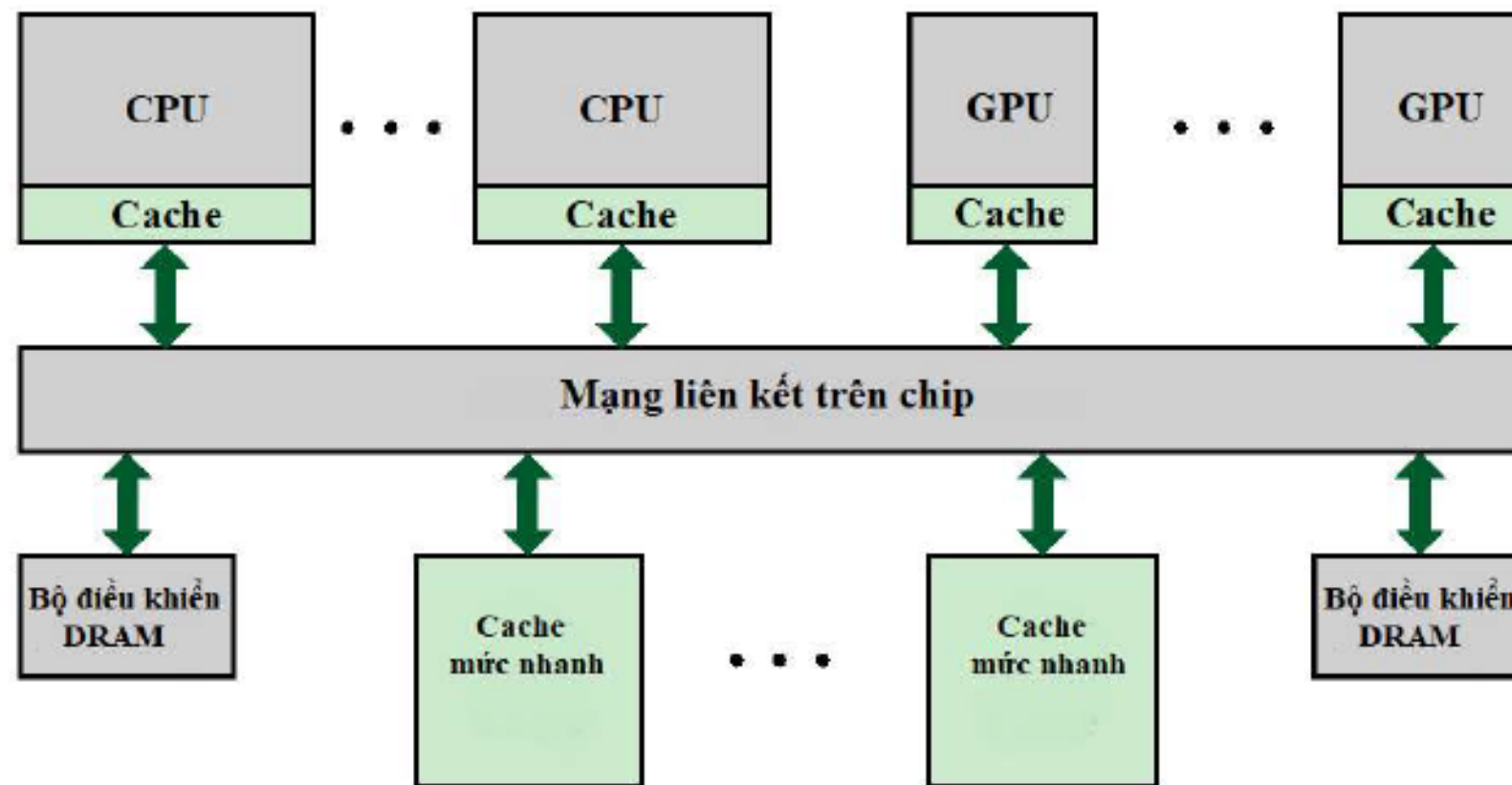
- Một bộ xử lý có thể lập trình chuyên dùng để hiển thị tất cả hình ảnh trên màn hình máy tính.
- Do GPU thực hiện các hoạt động song song trên nhiều bộ dữ liệu (hàng nghìn luồng thực thi song song) nên chúng rất phù hợp cho xử lý dữ liệu vector và ma trận.
- Được sử dụng rộng rãi cho các ứng dụng khoa học và AI cũng như khai thác tiền điện tử.
- Xu hướng tích hợp GPU vào trong các chip VXL đa lõi, tuy nhiên rất phức tạp khi phối hợp phần cứng và lập trình.



Chương 6: Kiến trúc máy tính tiên tiến

6.4.1. Bộ xử lý đồ họa GPU

Tích hợp GPU trên chip
đa lõi không đồng nhất





Chương 6: Kiến trúc máy tính tiên tiến

6.4.1. Bộ xử lý đồ họa GPU

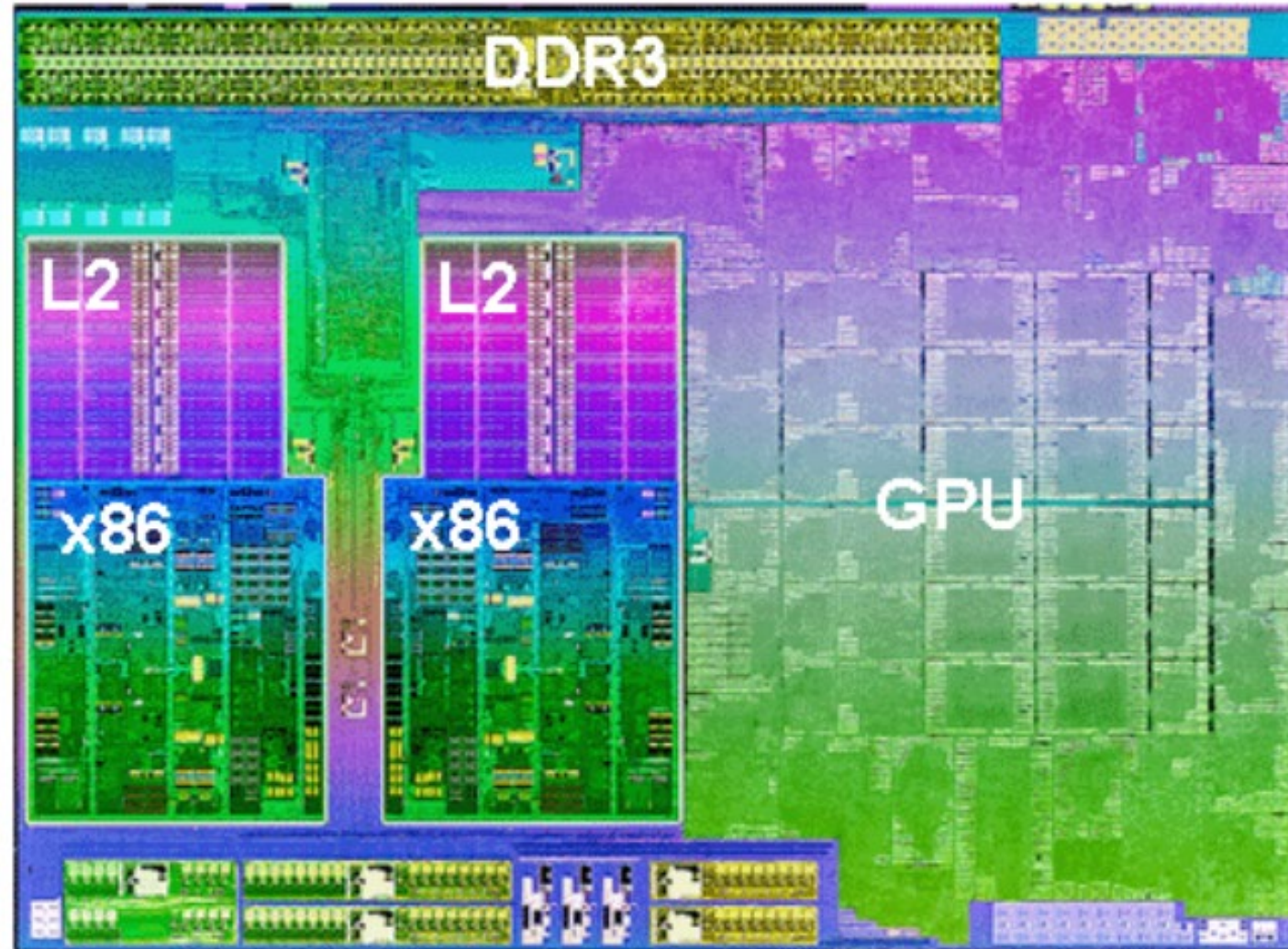
- Cho dù đó là các ứng dụng khoa học hay xử lý đồ họa truyền thống, chìa khóa để tận dụng các bộ xử lý GPU được bổ sung là xem xét thời gian cần thiết để chuyển một khối dữ liệu sang GPU, xử lý nó, sau đó trả kết quả về luồng ứng dụng chính.
- Trong các ứng dụng đã có của các chip tích hợp GPU, bộ nhớ vật lý được phân vùng giữa CPU và GPU. Nếu một luồng ứng dụng đang chạy trên CPU yêu cầu xử lý GPU, CPU sẽ sao chép dữ liệu một cách rõ ràng vào bộ nhớ GPU. GPU hoàn thành quá trình tính toán và sau đó sao chép kết quả trở lại bộ nhớ CPU.
 - Các vấn đề về tính nhất quán của bộ đệm giữa các bộ nhớ đệm CPU và GPU không phát sinh do bộ nhớ được phân vùng.
 - Mặt khác, việc chuyển giao dữ liệu qua lại vật lý cũng làm ảnh hưởng đến hiệu suất.



Chương 6: Kiến trúc máy tính tiên tiến

6.4.1. Bộ xử lý đồ họa GPU

GPU tích hợp →





Chương 6: Kiến trúc máy tính tiên tiến

6.4.1. Bộ xử lý đồ họa GPU

Bo mạch chủ H370 Mining Master hỗ trợ 20 card đồ họa thông qua cáp nối USB (2018- Asus) !





Chương 6: Kiến trúc máy tính tiên tiến

6.4.1. Siêu máy tính

- Siêu máy tính (**Supercomputer**) là một loại máy tính nhưng nó không giống như máy tính thông thường. Các siêu máy tính có kích thước khổng lồ đồng thời có sức mạnh tính toán vượt trội, có khả năng tính toán cao gấp hàng triệu lần máy tính thông thường. **Hoạt động theo cơ chế xử lý song song !**

- Có hai loại siêu máy tính:

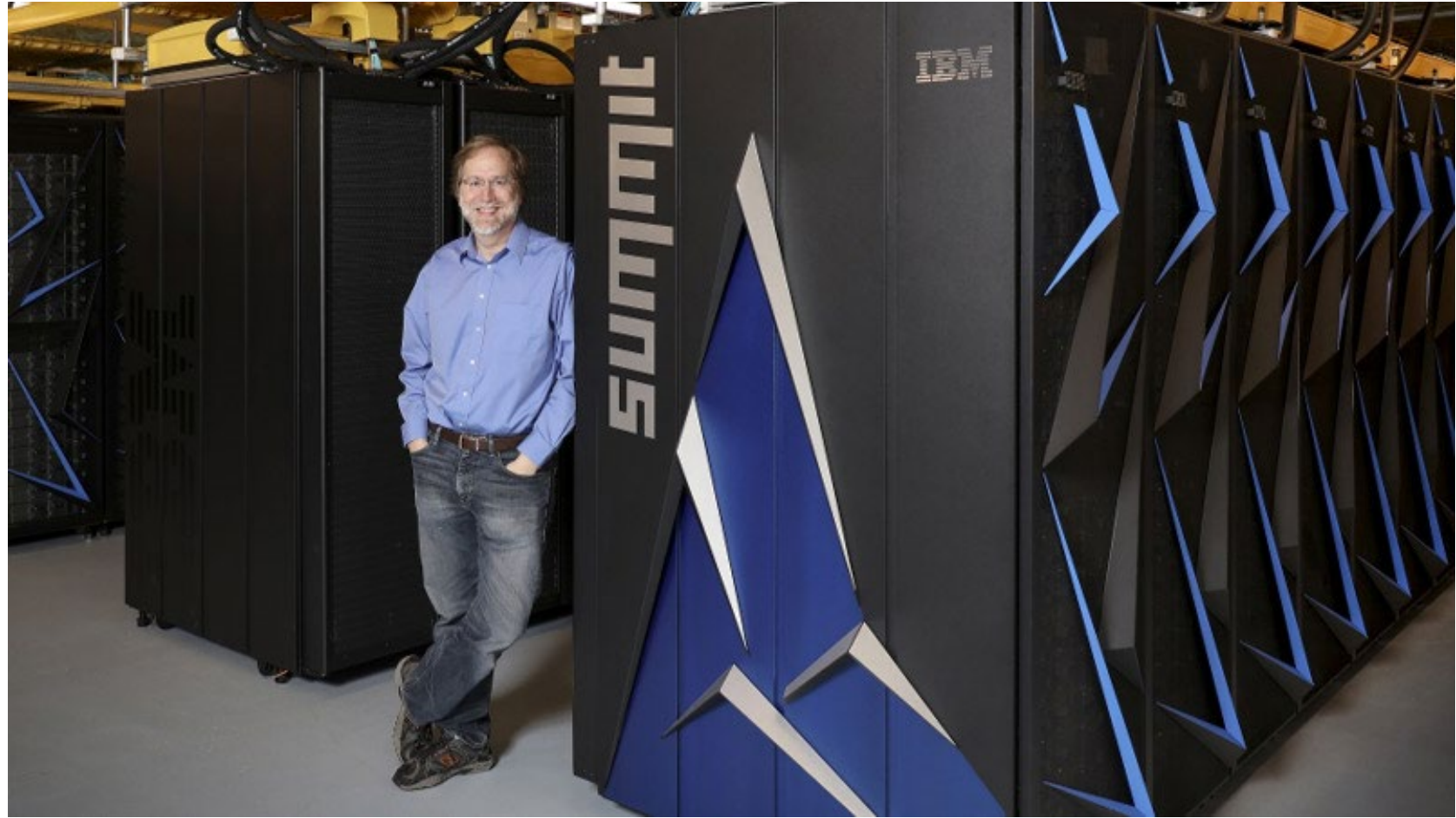
Loại thứ nhất: Kiểu điện toán tập trung (**Computer Cluster**), loại siêu máy tính này bao gồm một hệ thống bao gồm nhiều **CPU** đặt gần nhau. Những CPU này thường nằm trong nhiều máy tính giống nhau, người ta coi hệ thống này như là một siêu máy tính duy nhất. Loại siêu máy tính này chiếm hơn 80% siêu máy tính trên thế giới.

Loại thứ hai: Sử dụng nhiều máy tính nhỏ đặt xa nhau về khoảng cách địa lý, chúng được kết nối với nhau thành một mạng lưới thông qua một máy chính (**Control Node**). Máy chính có nhiệm vụ điều khiển và phân bổ tác vụ cho các máy con xử lý.



Chương 6: Kiến trúc máy tính tiên tiến

6.4.1. Siêu máy tính



Siêu máy tính Summit của IBM



Chương 6: Kiến trúc máy tính tiên tiến

6.4.1. Siêu máy tính

Tên các siêu máy tính mạnh nhất thế giới từ năm 2008 đến nay theo TOP500

Năm	Tên siêu máy tính	Tốc độ Rmax	Địa điểm
2008	IBM Roadrunner	1.026 PFLOPS	Los Alamos, Mỹ
		1.105 PFLOPS	
2009	Cray Jaguar	1.759 PFLOPS	Oak Ridge, Mỹ
2010	Tianhe-1A	2.566 PFLOPS	Tianjin, Trung Quốc
2011	Fujitsu K computer	10.51 PFLOPS	Kobe, Nhật
2012	IBM Sequoia	16.32 PFLOPS	Livermore, Mỹ
2012	Cray Titan	17.59 PFLOPS	Oak Ridge, Mỹ



Chương 6: Kiến trúc máy tính tiên tiến

6.4.1. Siêu máy tính

Một số đặc điểm của siêu máy tính

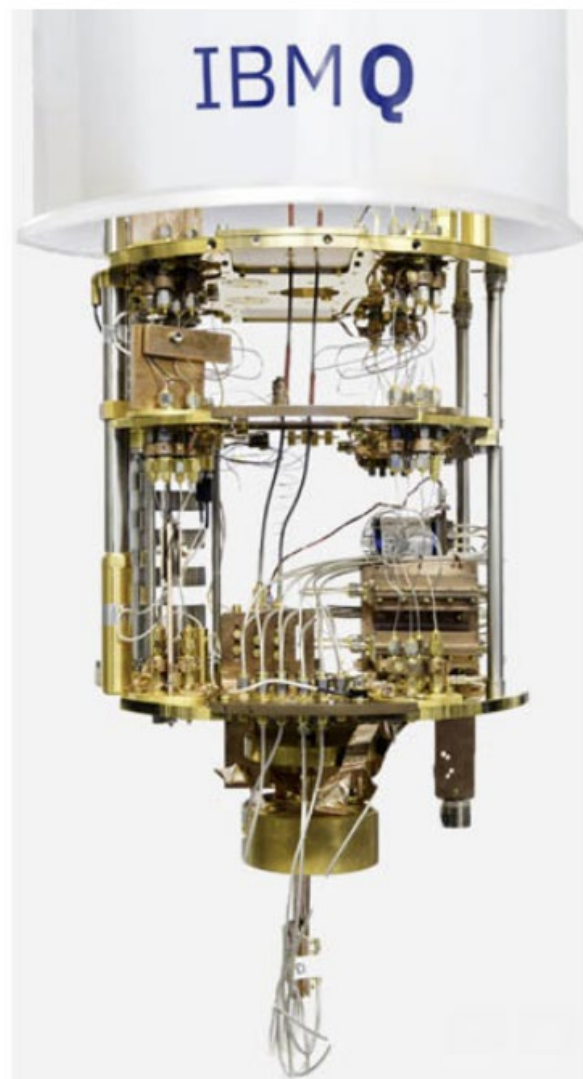
- Kích thước lớn, tiêu thụ nhiều điện năng
- Có cấu tạo gồm nhiều chip vi xử lý với nhiều lõi và bộ xử lý đồ họa GPU tích hợp
- Thích hợp cho tính toán song song
- Hệ điều hành thường là Linux, có phương pháp lập trình riêng.
- Ứng dụng vào các lĩnh vực yêu cầu xử lý số liệu và khối lượng tính toán lớn: thiên văn học, lượng tử học, trí tuệ nhân tạo AI, chơi cờ vua...vv

Sau siêu máy tính là gì ? → Máy tính lượng tử (Qubit, sự vướng víu lượng tử,.. .các cổng lượng tử,...) → *tương lai của các bạn !*



Chương 6: Kiến trúc máy tính tiên tiến

Có khoảng 100 máy tính lượng tử đang hoạt động trên toàn thế giới !



Mô hình cổng Q của IBM