



ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

KIẾN TRÚC MÁY TÍNH

Computer Architecture

Course ID: IT3030, IT3034, IT3283

Version: CA.RISCV.2025



© Nguyễn Kim Khánh

Nội dung học phần

Chương 1. Giới thiệu chung

Chương 2. Hệ thống máy tính

Chương 3. Kiến trúc tập lệnh

Chương 4. Số học máy tính

Chương 5. Bộ xử lý

Chương 6. Bộ nhớ máy tính

Chương 7. Hệ thống vào-ra

Chương 8. Các kiến trúc song song

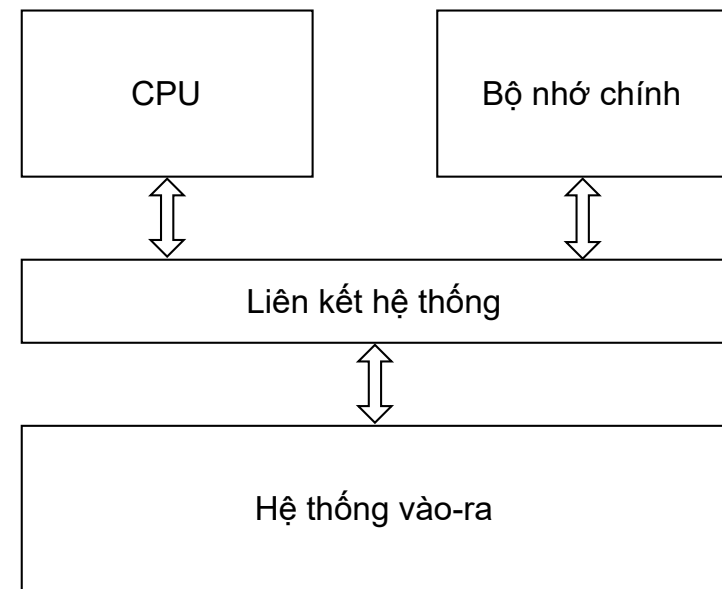


Chương 2 HỆ THỐNG MÁY TÍNH

- 2.1. Các thành phần cơ bản của máy tính
- 2.2. Hoạt động cơ bản của máy tính
- 2.3. Liên kết trong máy tính

2.1. Các thành phần cơ bản của máy tính

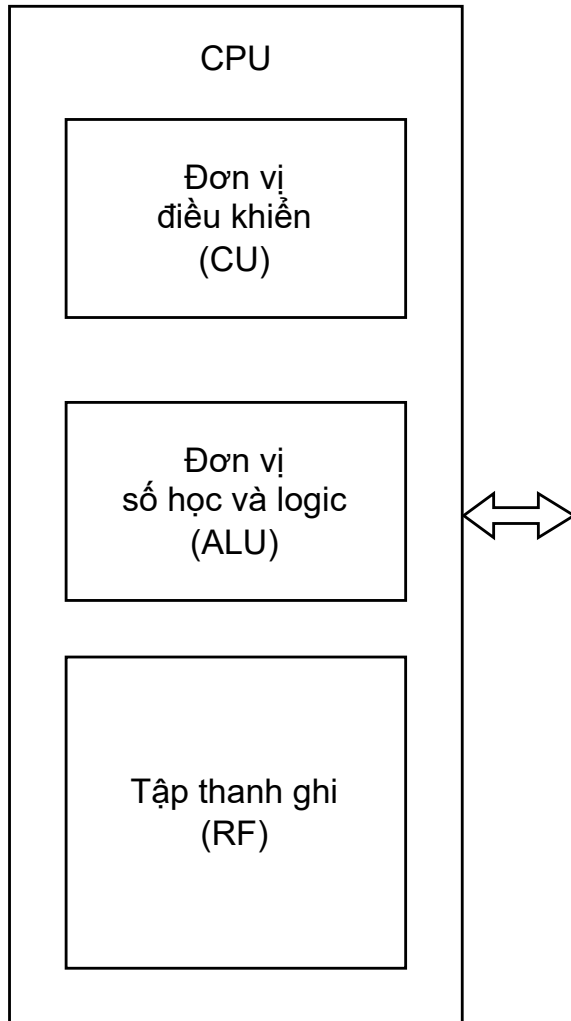
- Bộ xử lý trung tâm
(Central Processing Unit – CPU)
 - Điều khiển hoạt động của máy tính và xử lý dữ liệu
- Bộ nhớ chính (Main Memory)
 - Chứa các chương trình đang thực hiện
- Hệ thống vào-ra (Input-Output)
 - Trao đổi thông tin giữa máy tính với bên ngoài
- Liên kết hệ thống (System interconnection)
 - Kết nối và vận chuyển thông tin
 - Bus: dạng kết nối cơ bản



1. Bộ xử lý trung tâm (CPU)

- Chức năng:
 - điều khiển hoạt động của máy tính
 - xử lý dữ liệu
- Nguyên tắc hoạt động cơ bản:
 - CPU hoạt động theo chương trình nằm trong bộ nhớ chính.
- Là thành phần nhanh nhất trong hệ thống

Các thành phần cơ bản của CPU



- Đơn vị điều khiển

- Control Unit - CU
- Điều khiển hoạt động của máy tính theo chương trình đã định sẵn

- Đơn vị số học và logic

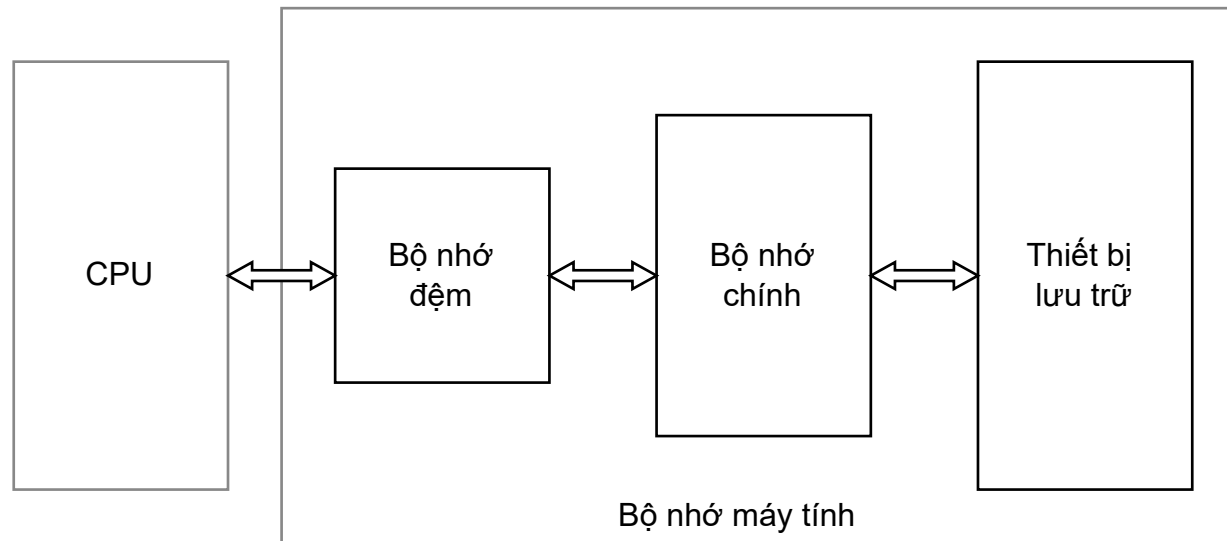
- Arithmetic and Logic Unit - ALU
- Thực hiện các phép toán số học và phép toán logic

- Tập thanh ghi

- Register File - RF
- Gồm các thanh ghi chứa các thông tin phục vụ cho hoạt động của CPU

2. Bộ nhớ máy tính

- Chức năng: nhớ chương trình và dữ liệu (dưới dạng nhị phân)
- Các thao tác cơ bản với bộ nhớ:
 - Thao tác ghi (Write)
 - Thao tác đọc (Read)
- Các thành phần chính:
 - Bộ nhớ chính (Main memory → memory)
 - Bộ nhớ đệm (Cache memory → cache)
 - Thiết bị lưu trữ (Storage Devices → storage)



Bộ nhớ chính (Memory)

- Tồn tại trên mọi máy tính
- Chứa các lệnh và dữ liệu của chương trình đang được thực hiện
- Sử dụng bộ nhớ bán dẫn
- Tổ chức thành các ngăn nhớ được đánh địa chỉ (thường đánh địa chỉ cho từng byte nhớ)
- Nội dung của ngăn nhớ có thể thay đổi, song địa chỉ vật lý của ngăn nhớ luôn cố định
- CPU muốn đọc/ghi ngăn nhớ cần phải biết địa chỉ ngăn nhớ đó

Nội dung	Địa chỉ
0100 1101	00...0000
0101 0101	00...0001
1010 1111	00...0010
0000 1110	00...0011
0111 0100	00...0100
1011 0010	00...0101
0010 1000	00...0110
1110 1111	00...0111
.	
.	
.	
0110 0010	11...1110
0010 0001	11...1111



Bộ nhớ đệm (Cache)

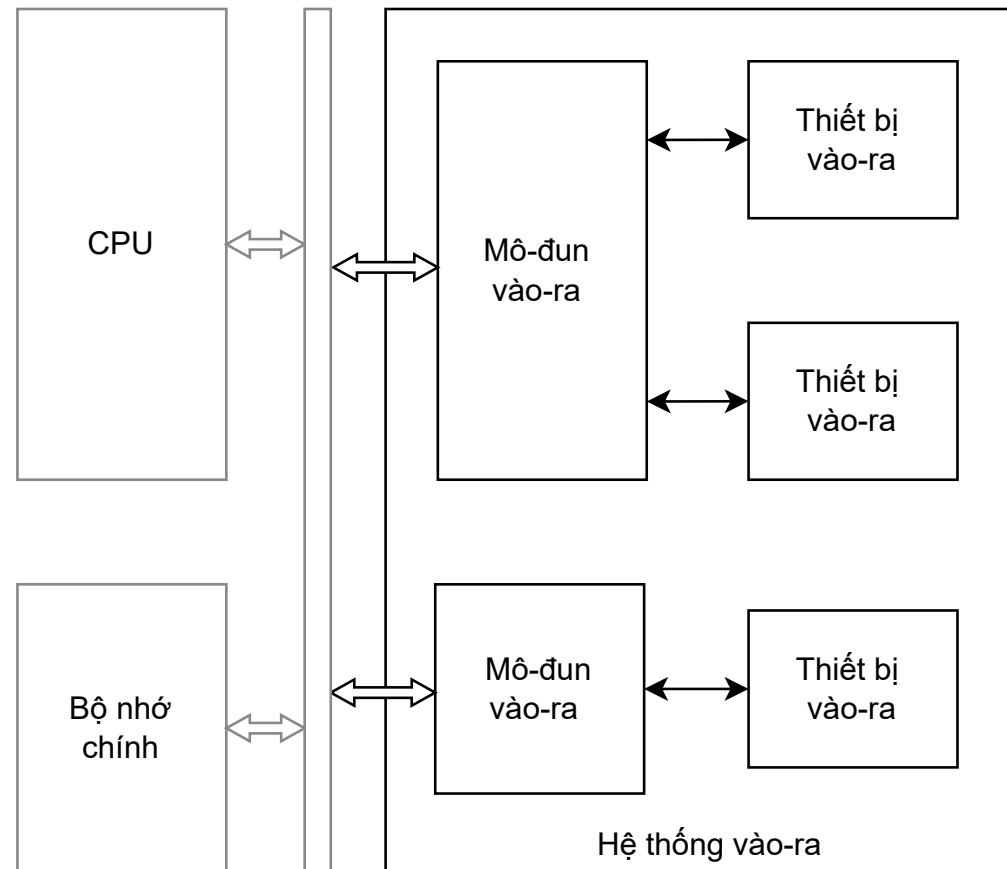
- Bộ nhớ có tốc độ nhanh được đặt đệm giữa CPU và bộ nhớ chính nhằm tăng tốc độ CPU truy cập bộ nhớ
- Dung lượng nhỏ hơn bộ nhớ chính
- Sử dụng bộ nhớ bán dẫn tốc độ nhanh
- Cache thường được chia thành một số mức (L1, L2, L3)
- Cache được tích hợp trên cùng chip bộ xử lý

Thiết bị lưu trữ (Storage)

- Còn được gọi là bộ nhớ ngoài
- Chức năng và đặc điểm
 - Lưu giữ tài nguyên phần mềm của máy tính
 - Được kết nối với hệ thống dưới dạng các thiết bị vào-ra
 - Dung lượng lớn
 - Tốc độ chậm
- Các loại thiết bị lưu trữ
 - Bộ nhớ từ: ổ đĩa cứng HDD
 - Bộ nhớ bán dẫn: ổ thể rắn SSD, ổ nhớ flash, thẻ nhớ
 - Bộ nhớ quang: CD, DVD (ít dùng)

3. Hệ thống vào-ra

- Chức năng: Trao đổi thông tin giữa máy tính với thế giới bên ngoài
- Các thao tác cơ bản:
 - Vào dữ liệu (Input)
 - Ra dữ liệu (Output)
- Các thành phần chính:
 - Các thiết bị vào-ra (IO devices)
 - Các mô-đun vào-ra (IO modules)



Các thiết bị vào-ra

- Còn được gọi là thiết bị ngoại vi (Peripherals)
- Chức năng: chuyển đổi dữ liệu giữa bên trong và bên ngoài máy tính
- Các loại thiết bị vào-ra:
 - Thiết bị vào (Input Devices)
 - Thiết bị ra (Output Devices)
 - Thiết bị lưu trữ (Storage Devices)
 - Thiết bị truyền thông (Communication Devices)

Mô-đun vào-ra

- Chức năng: nối ghép các thiết bị vào-ra với máy tính
- Mỗi mô-đun vào-ra có một hoặc một vài cổng vào-ra (I/O Port)
- Mỗi cổng vào-ra được đánh một địa chỉ xác định
- Các thiết bị vào-ra được kết nối và trao đổi dữ liệu với máy tính thông qua các cổng vào-ra
- CPU muốn trao đổi dữ liệu với thiết bị vào-ra, cần phải biết địa chỉ của cổng vào-ra tương ứng

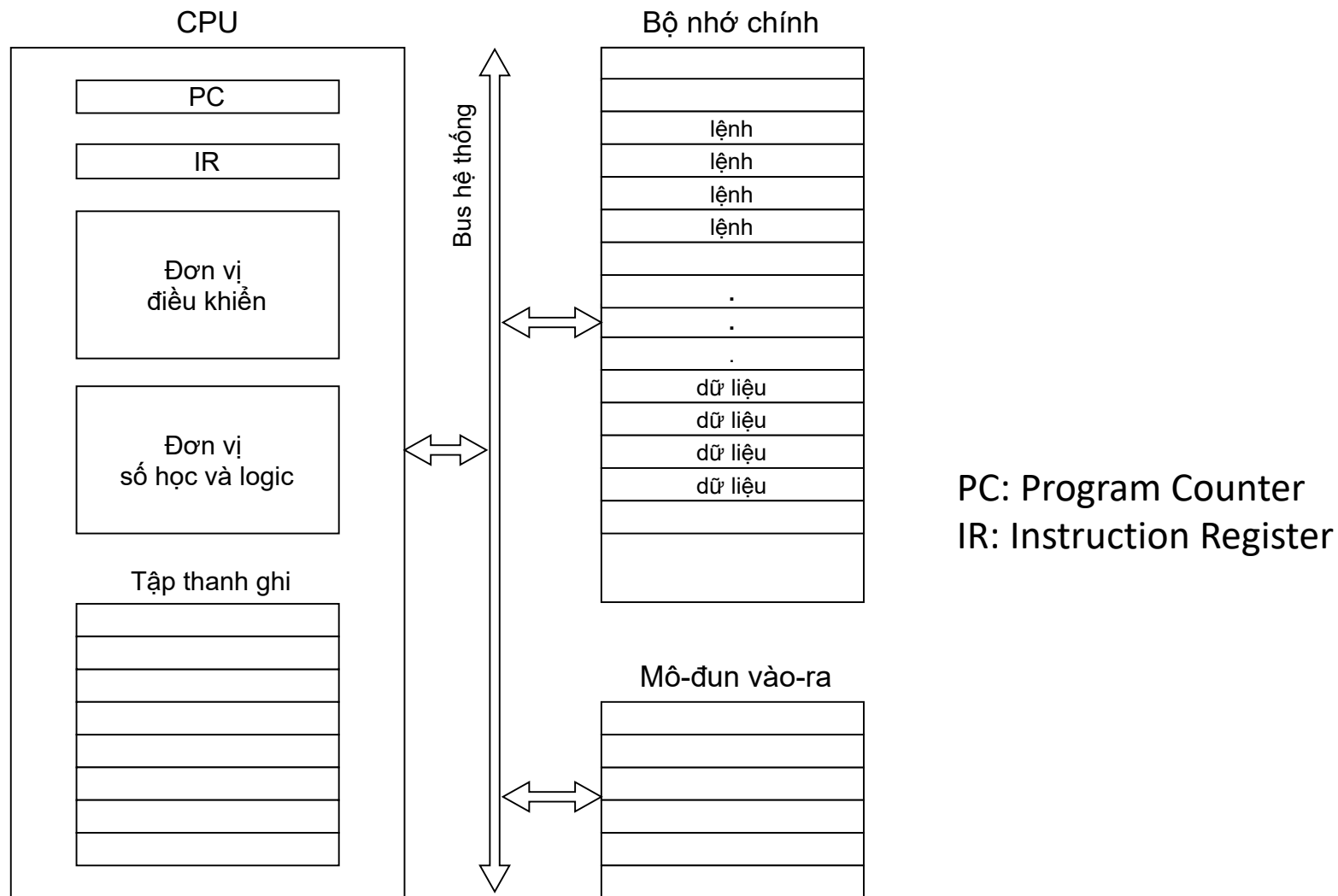
2.2. Hoạt động cơ bản của máy tính

- Thực hiện chương trình
- Hoạt động ngắt
- Hoạt động vào-ra

1. Thực hiện chương trình

- Là hoạt động cơ bản của máy tính
- Máy tính lặp đi lặp lại chu trình lệnh gồm hai bước:
 - Tìm nạp lệnh (Fetch instruction)
 - Thực hiện lệnh (Execute instruction)
- Hoạt động thực hiện chương trình bị dừng nếu:
 - Thực hiện lệnh bị lỗi
 - Gặp lệnh dừng
 - Tắt máy

Sơ đồ mô hình lập trình của máy tính

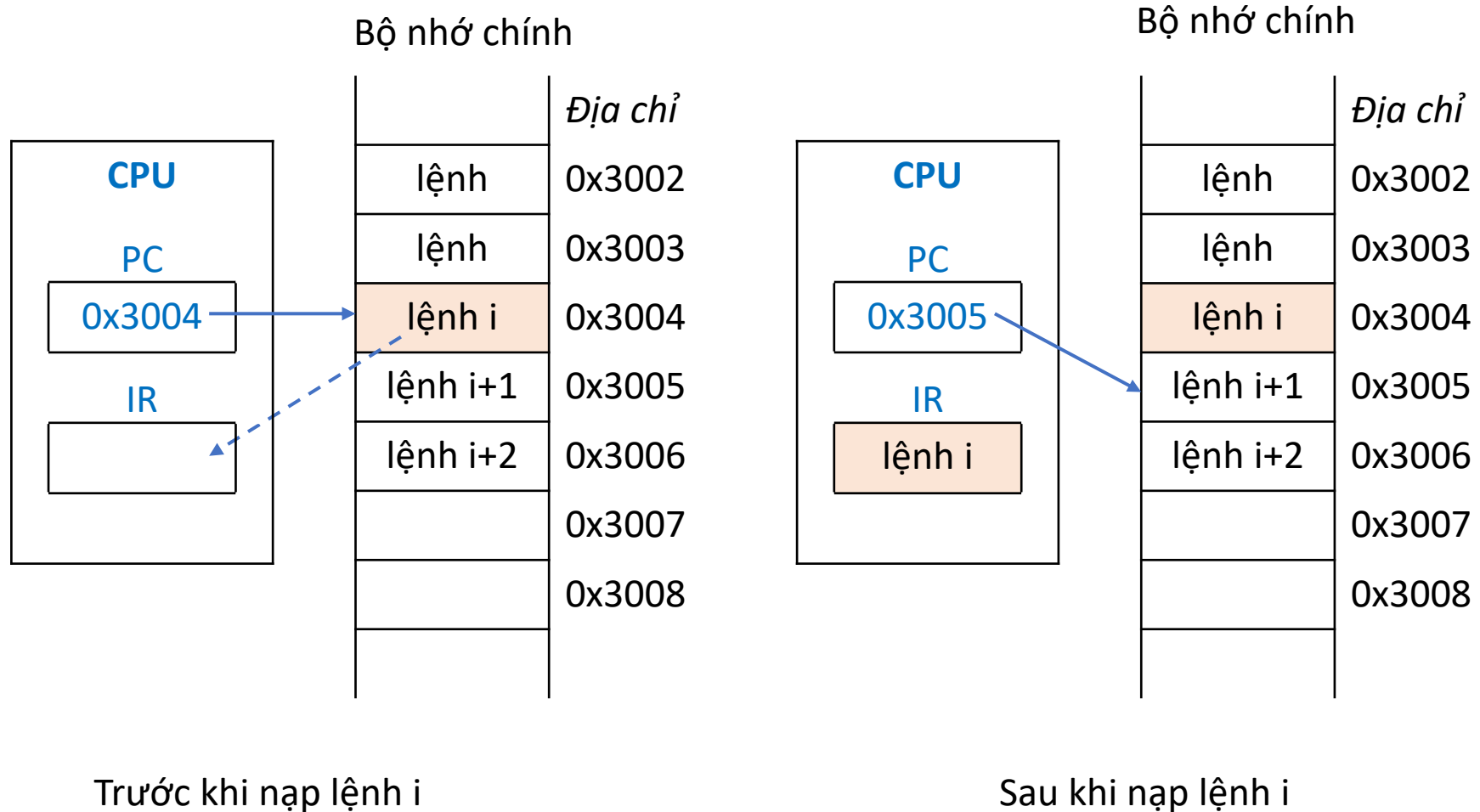


Tìm nạp lệnh

- Bắt đầu mỗi chu trình lệnh, bộ xử lý (CPU) tìm nạp lệnh từ bộ nhớ
- **Bộ đếm chương trình PC** (Program Counter) là thanh ghi của CPU chứa địa chỉ của lệnh sẽ được nạp vào để thực hiện
 - Intel x86: thanh ghi IP (Instruction Pointer)
- CPU phát ra địa chỉ từ bộ đếm chương trình PC đến bộ nhớ để tìm ra ngăn nhớ chứa lệnh
- Lệnh được đọc từ bộ nhớ nạp vào thanh ghi lệnh IR (Instruction Register)
- Sau khi lệnh được nạp vào, trừ một số trường hợp, bộ xử lý tăng nội dung PC để trở đến lệnh kế tiếp.



Minh họa quá trình tìm nạp lệnh



Thảo luận

- Với cấu trúc tuần tự, nội dung PC tăng bao nhiêu ?
- Bản chất của rẽ nhánh (nhảy) trong chương trình ?
- Địa chỉ khởi động ?

Thực hiện lệnh

- Bộ xử lý giải mã lệnh đã được nạp vào và phát tín hiệu điều khiển thực hiện thao tác mà lệnh đó yêu cầu
- Các kiểu thao tác cơ bản của lệnh:
 - Trao đổi dữ liệu giữa bộ xử lý với bộ nhớ chính hoặc với mô-đun vào-ra
 - Thực hiện các phép toán số học hoặc phép toán logic với các dữ liệu
 - Chuyển điều khiển trong chương trình: rẽ nhánh hoặc nhảy đến vị trí khác

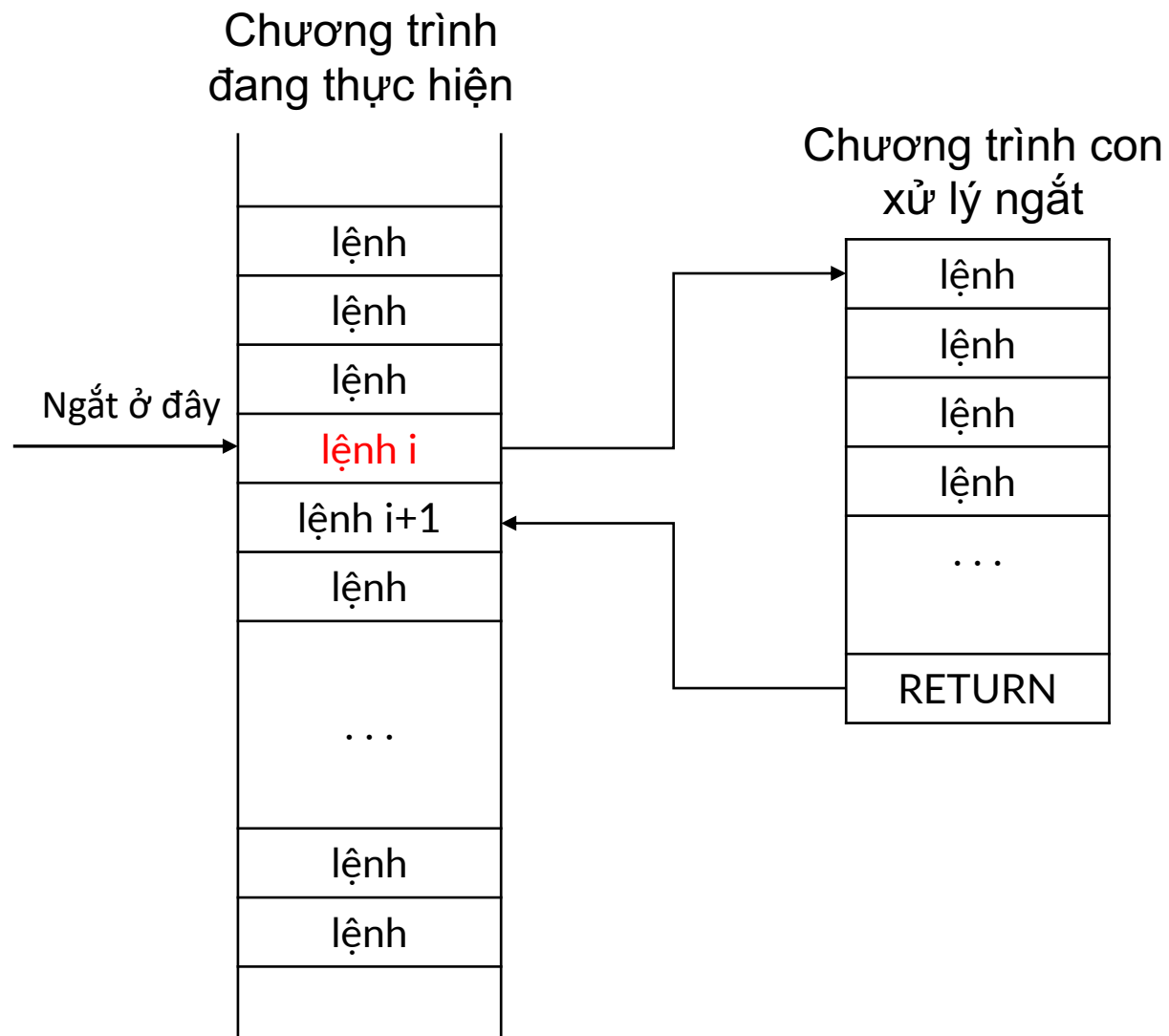
2. Ngắt (Interrupt)

- **Khái niệm chung về ngắt:** Ngắt là cơ chế cho phép CPU tạm dừng chương trình đang thực hiện để chuyển sang thực hiện một chương trình con có sẵn trong bộ nhớ.
 - **Chương trình con xử lý ngắt (Interrupt handlers)**
- **Các loại ngắt :**
 - Do lỗi khi thực hiện lệnh (VD: tràn số, mã lệnh sai, ...)
 - Do lỗi phần cứng
 - Do bộ định thời (Timer)
 - Do bộ điều khiển vào-ra gửi tín hiệu ngắt đến bộ xử lý để yêu cầu trao đổi dữ liệu

Hoạt động với ngắt từ bộ điều khiển vào-ra

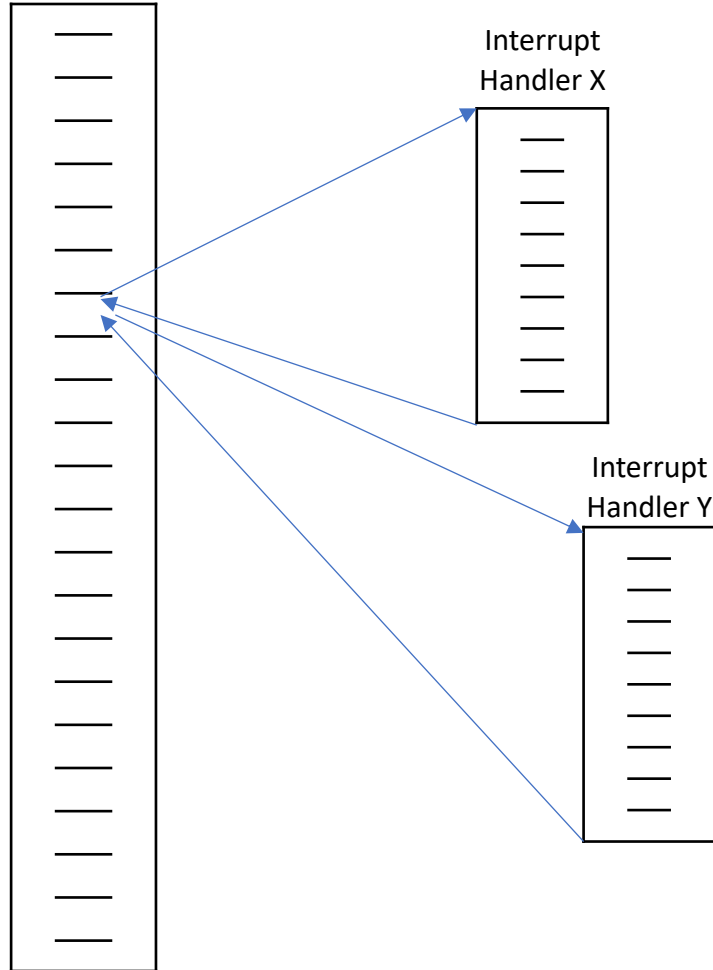
- Sau khi hoàn thành mỗi một lệnh, bộ xử lý kiểm tra tín hiệu ngắt
- Nếu không có ngắt, bộ xử lý tìm nạp lệnh tiếp theo của chương trình hiện tại
- Nếu có tín hiệu ngắt:
 - Tạm dừng (suspend) chương trình đang thực hiện
 - Cắt ngữ cảnh (các thông tin liên quan đến chương trình bị ngắt)
 - Thiết lập bộ đếm chương trình PC trở đến chương trình con xử lý ngắt tương ứng
 - Chuyển sang thực hiện chương trình con xử lý ngắt
 - Khôi phục ngữ cảnh và trở về tiếp tục thực hiện chương trình đang bị tạm dừng

Hoạt động ngắt (tiếp)



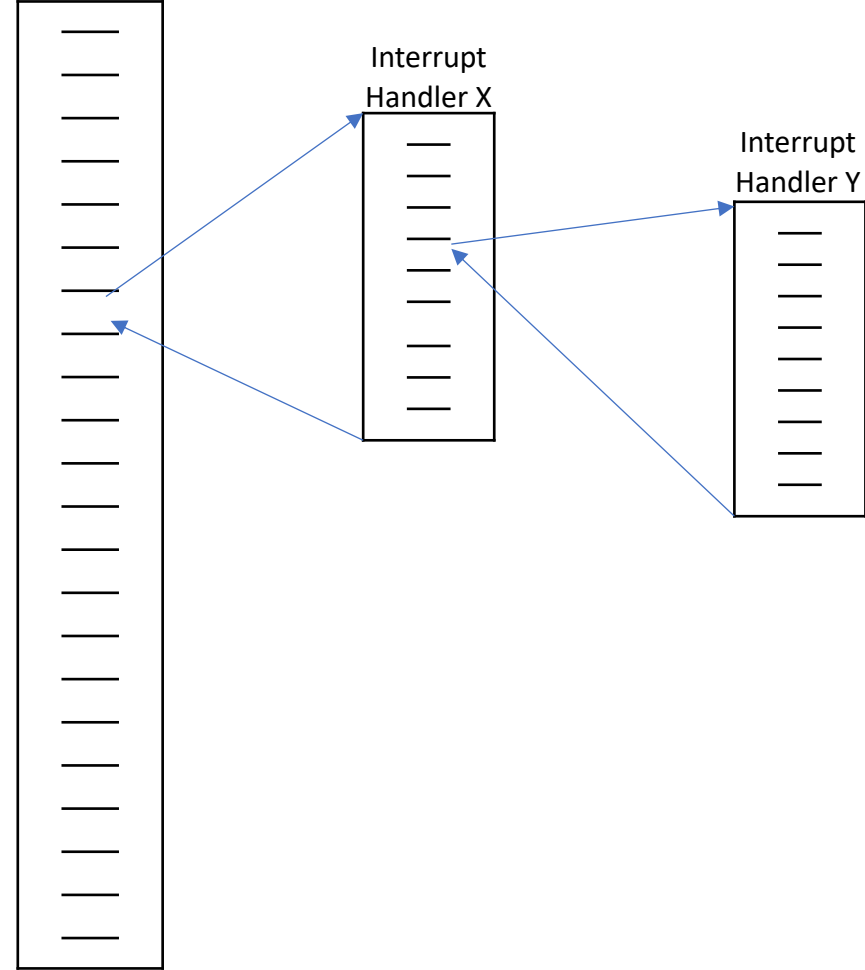
Xử lý với nhiều tín hiệu yêu cầu ngắt

User Program



Xử lý ngắt tuần tự

User Program



Xử lý ngắt ưu tiên



3. Hoạt động vào-ra

- Hoạt động vào-ra: là hoạt động trao đổi dữ liệu giữa mô-đun vào-ra với bên trong máy tính.
- Các kiểu hoạt động vào-ra:
 - Bộ xử lý trao đổi dữ liệu với mô-đun vào-ra bởi lệnh vào-ra trong chương trình
 - Bộ xử lý trao quyền điều khiển cho phép mô-đun vào-ra trao đổi dữ liệu trực tiếp với bộ nhớ chính (DMA - Direct Memory Access).

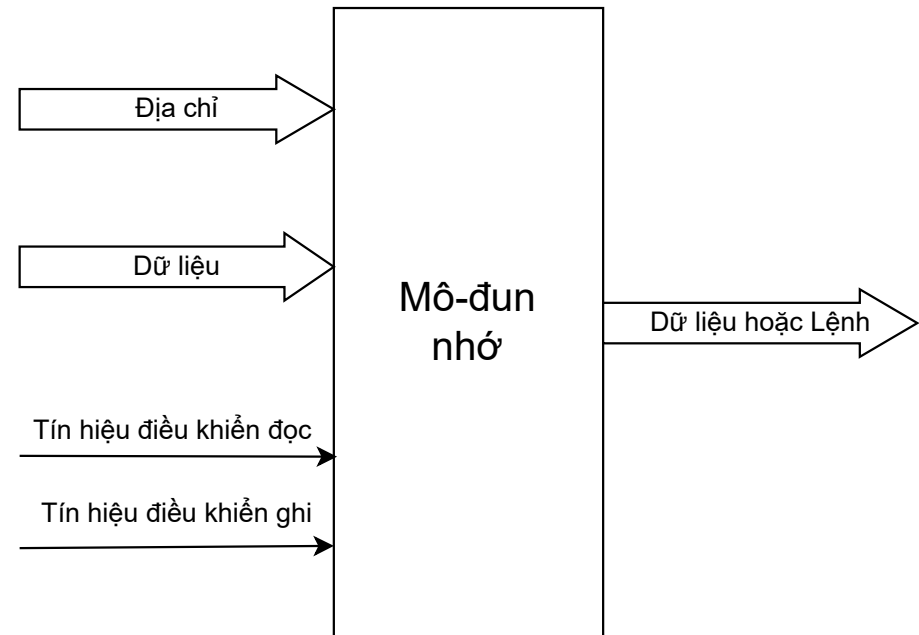
2.3. Liên kết trong máy tính

1. Luồng thông tin trong máy tính

- Các mô-đun trong máy tính:
 - CPU
 - Mô-đun nhớ
 - Mô-đun vào-ra
- cần được kết nối với nhau

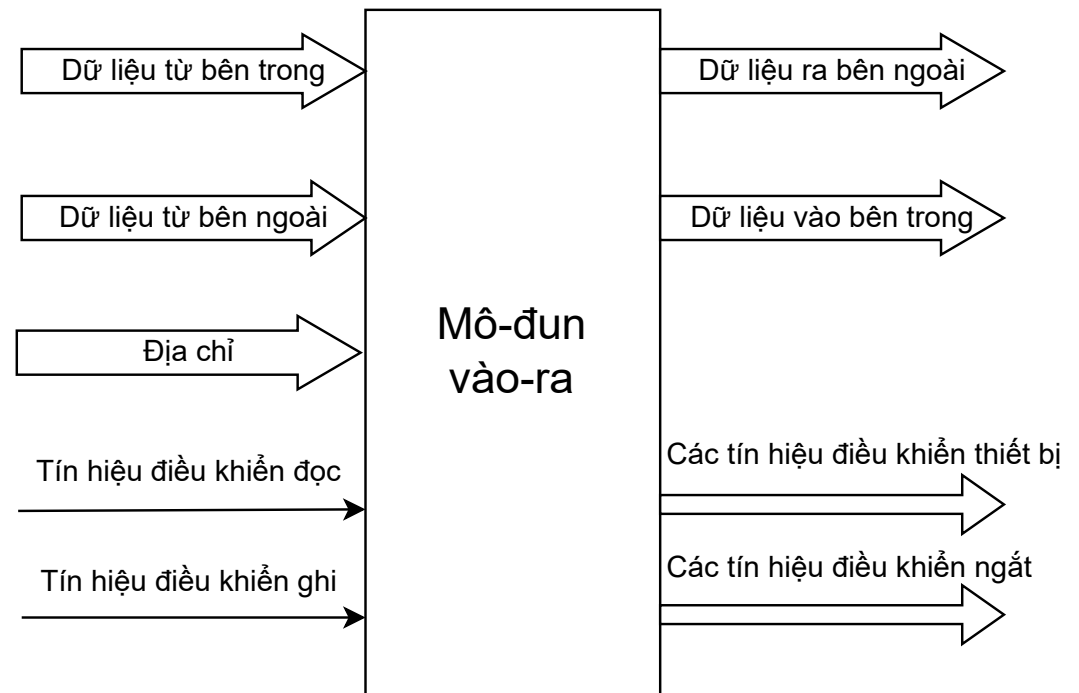
Kết nối mô-đun nhớ

- Địa chỉ đưa đến để xác định ngăn nhớ
- Dữ liệu được đưa đến khi ghi
- Dữ liệu hoặc lệnh được đưa ra khi đọc
 - Bộ nhớ không phân biệt lệnh và dữ liệu
- Nhận các tín hiệu điều khiển:
 - Điều khiển đọc (Read)
 - Điều khiển ghi (Write)



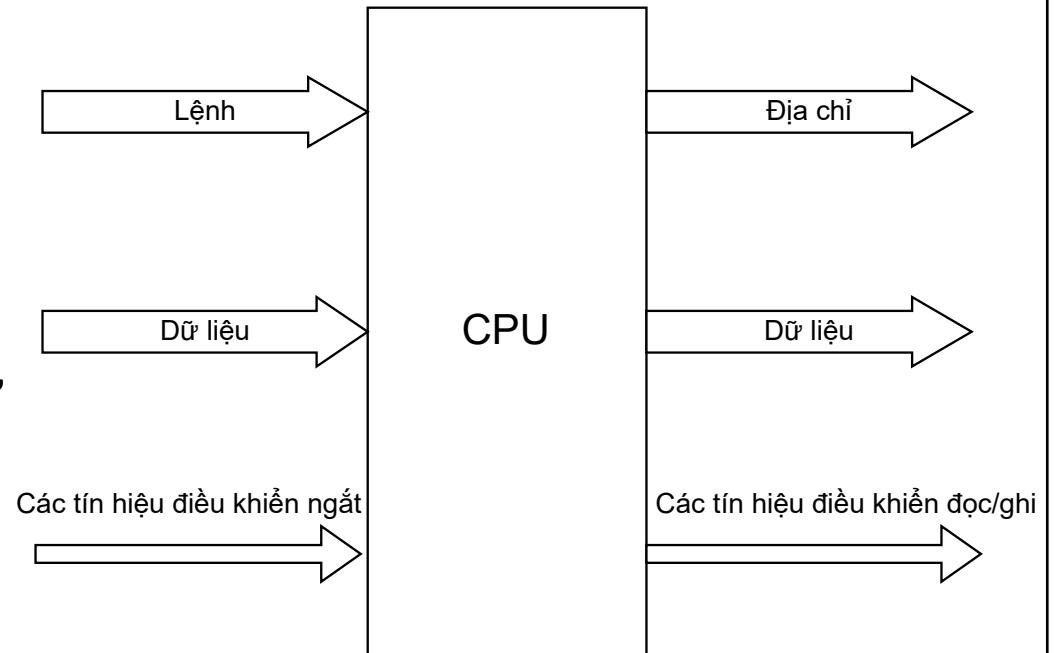
Kết nối mô-đun vào-ra

- Địa chỉ đưa đến để xác định cổng vào-ra
- Ra dữ liệu (Output)
 - Dữ liệu từ bên trong (CPU hoặc bộ nhớ chính) được đưa ra thiết bị vào-ra
- Vào dữ liệu (Input)
 - Dữ liệu từ thiết bị vào-ra được đưa vào bên trong (CPU hoặc bộ nhớ chính)
- Nhận các tín hiệu điều khiển từ CPU
- Phát các tín hiệu điều khiển đến thiết bị vào-ra
- Phát các tín hiệu ngắt đến CPU



Kết nối CPU

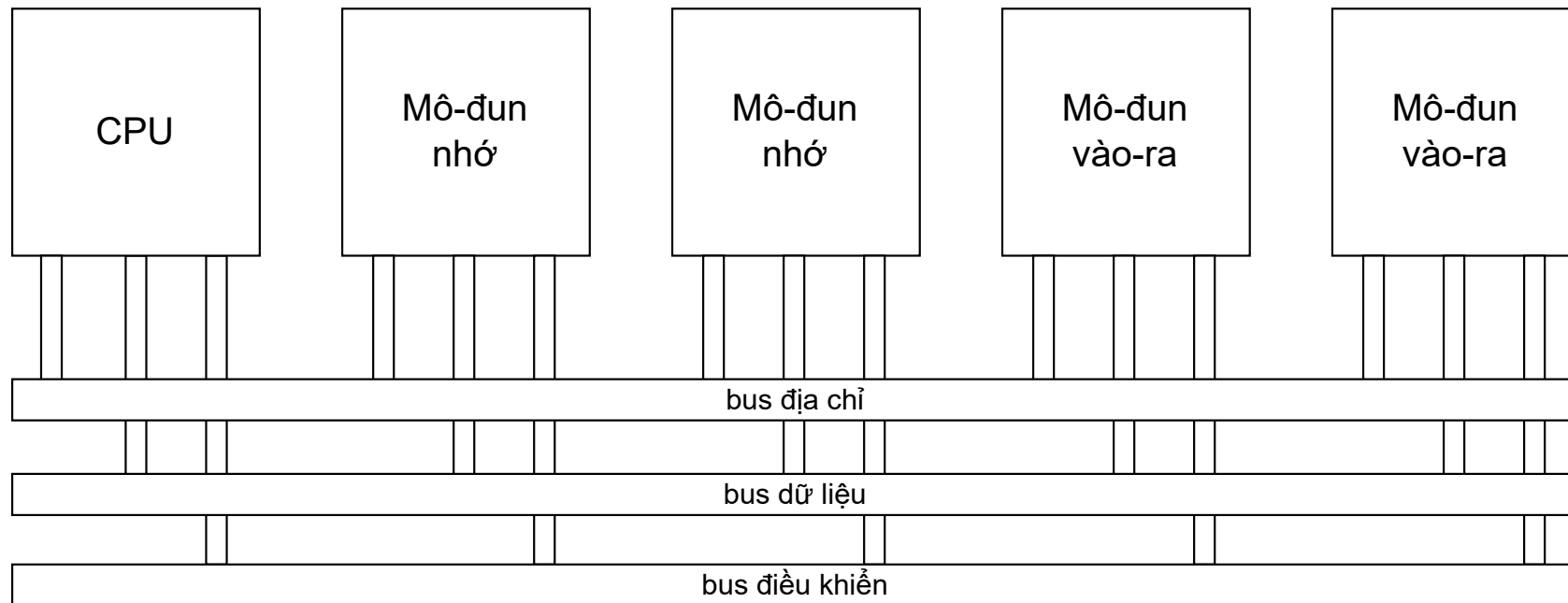
- Phát địa chỉ đến các mô-đun nhớ hay các mô-đun vào-ra
- Đọc lệnh từ bộ nhớ
- Đọc dữ liệu từ bộ nhớ hoặc mô-đun vào-ra
- Đưa dữ liệu ra (sau khi xử lý) đến bộ nhớ hoặc mô-đun vào-ra
- Phát tín hiệu điều khiển đọc/ghi đến các mô-đun nhớ và các mô-đun vào-ra
- Nhận các tín hiệu ngắt



2. Liên kết bus trong máy tính

- **Bus:** tập hợp các đường kết nối để vận chuyển thông tin giữa các mô-đun của máy tính với nhau
- Các bus chức năng:
 - Bus địa chỉ (Address bus)
 - Bus dữ liệu (Data bus)
 - Bus điều khiển (Control bus)
- **Độ rộng bus:** là số đường dây của bus có thể truyền các bit thông tin đồng thời (chỉ dùng cho bus địa chỉ và bus dữ liệu)

Sơ đồ cấu trúc bus cơ bản



Bus địa chỉ

- Chức năng: vận chuyển địa chỉ để xác định vị trí ngăn nhớ hay cổng vào-ra
- Độ rộng bus địa chỉ:
 - N bit: $A_{N-1}, A_{N-2}, \dots, A_2, A_1, A_0$
→ Số lượng địa chỉ tối đa được sử dụng là: 2^N địa chỉ (không gian địa chỉ)
 - Địa chỉ nhỏ nhất: $00 \dots 000_{(2)}$
 - Địa chỉ lớn nhất: $11 \dots 111_{(2)}$
- Ví dụ:
 - Máy tính sử dụng bus địa chỉ 32-bit ($A_{31}-A_0$), bộ nhớ chính được đánh địa chỉ cho từng byte
→ Có khả năng đánh địa chỉ cho 2^{32} bytes nhớ = 4GiB



Bus dữ liệu

- Chức năng:
 - vận chuyển lệnh từ bộ nhớ đến CPU
 - vận chuyển dữ liệu giữa các thành phần của máy tính với nhau
- Độ rộng bus dữ liệu: số bit được truyền đồng thời
 - M bit: $D_{M-1}, D_{M-2}, \dots, D_2, D_1, D_0$
 - M thường là 8, 16, 32, 64 bit
- Ví dụ:
 - Máy tính có bus dữ liệu kết nối CPU với bộ nhớ là 64-bit
→ Có thể trao đổi 8 byte nhớ ở một thời điểm

Bus điều khiển

- Chức năng: vận chuyển các tín hiệu điều khiển
- Các loại tín hiệu điều khiển:
 - Các tín hiệu điều khiển đọc/ghi
 - Các tín hiệu điều khiển ngắt
 - Các tín hiệu điều khiển bus

Một số tín hiệu điều khiển điển hình

- Các tín hiệu (phát ra từ CPU) điều khiển đọc/ghi:
 - Memory Read (MEMR): Tín hiệu điều khiển đọc dữ liệu từ một ngăn nhớ có địa chỉ xác định đưa lên bus dữ liệu (chuyển vào CPU).
 - Memory Write (MEMW): Tín hiệu điều khiển ghi dữ liệu có sẵn trên bus dữ liệu (từ CPU) đến một ngăn nhớ có địa chỉ xác định.
 - I/O Read (IOR): Tín hiệu điều khiển đọc dữ liệu từ một cổng vào-ra có địa chỉ xác định đưa lên bus dữ liệu (chuyển vào CPU).
 - I/O Write (IOW): Tín hiệu điều khiển ghi dữ liệu có sẵn trên bus dữ liệu (từ CPU) ra một cổng có địa chỉ xác định.



Một số tín hiệu điều khiển điển hình (tiếp)

- Các tín hiệu điều khiển ngắt:
 - Interrupt Request (INTR): Tín hiệu từ bộ điều khiển vào-ra gửi đến yêu cầu ngắt CPU để trao đổi vào-ra. Tín hiệu INTR có thể bị che.
 - Interrupt Acknowledge (INTA): Tín hiệu phát ra từ CPU báo cho bộ điều khiển vào-ra biết CPU chấp nhận ngắt để trao đổi vào-ra.
 - Non Maskable Interrupt (NMI): tín hiệu ngắt không che được gửi đến ngắt CPU.
 - Reset: Tín hiệu từ bên ngoài gửi đến CPU và các thành phần khác để khởi động lại máy tính.

Một số tín hiệu điều khiển điển hình (tiếp)

- Các tín hiệu điều khiển bus:
 - Bus Request (BRQ) : Tín hiệu từ mô-đun vào-ra gửi đến yêu cầu CPU chuyển nhượng quyền sử dụng bus.
 - Bus Grant (BGT): Tín hiệu phát ra từ CPU chấp nhận chuyển nhượng quyền sử dụng bus cho mô-đun vào-ra.
 - Lock/ Unlock: Tín hiệu cấm/cho-phép xin chuyển nhượng bus.

Đặc điểm của bus dùng chung (shared bus)

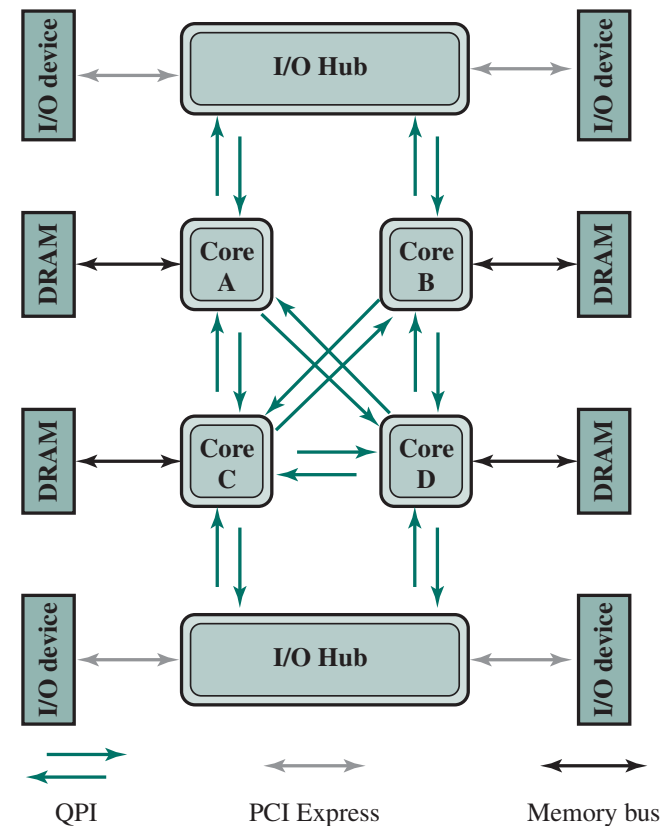
- Nhiều mô-đun kết nối vào bus chung → cần có bộ phân xử bus
- Bus chỉ phục vụ được một yêu cầu trao đổi dữ liệu tại một thời điểm → độ trễ lớn
- Bus phải có tốc độ bằng tốc độ bus của mô-đun nhanh nhất trong hệ thống
- Khắc phục:
 - Đa bus (Multiple bus): chia thành nhiều bus:
 - Bus cho bộ xử lý
 - Bus cho bộ nhớ
 - Bus vào-ra
 - Liên kết điểm-điểm (Point to point interconnection):
 - Sử dụng phổ biến trên các máy tính hiện nay

3. Liên kết điểm-điểm

- Với bus dùng chung, khi tốc độ dữ liệu ngày càng cao, khó thực hiện các chức năng đồng bộ hóa và phân xử bus kịp thời
- Kết nối điểm-điểm có độ trễ nhỏ hơn, tốc độ dữ liệu cao hơn và khả năng mở rộng tốt hơn
- Các loại kết nối điểm-điểm phổ biến:
 - QPI – Quick Path Interconnect
 - PCIe – PCI express

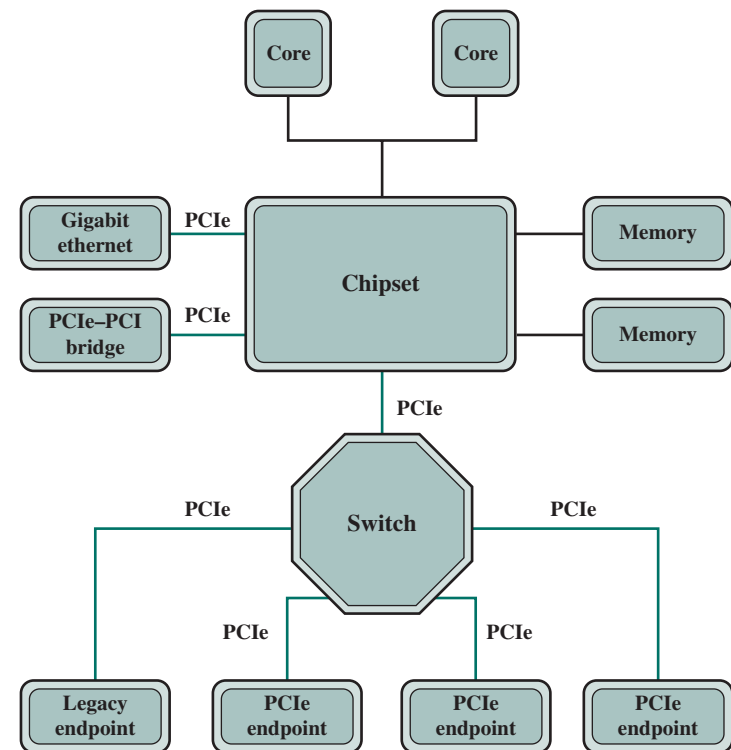
QPI - Quick Path Interconnect

- Được giới thiệu vào năm 2008
- Đa kết nối trực tiếp:
 - Kết nối trực tiếp từng cặp, loại bỏ việc phân xử bus
- Kiến trúc giao thức phân lớp
 - Kết nối các bộ xử lý với nhau và với chipset sử dụng kiến trúc giao thức phân lớp
- Truyền dữ liệu dạng gói

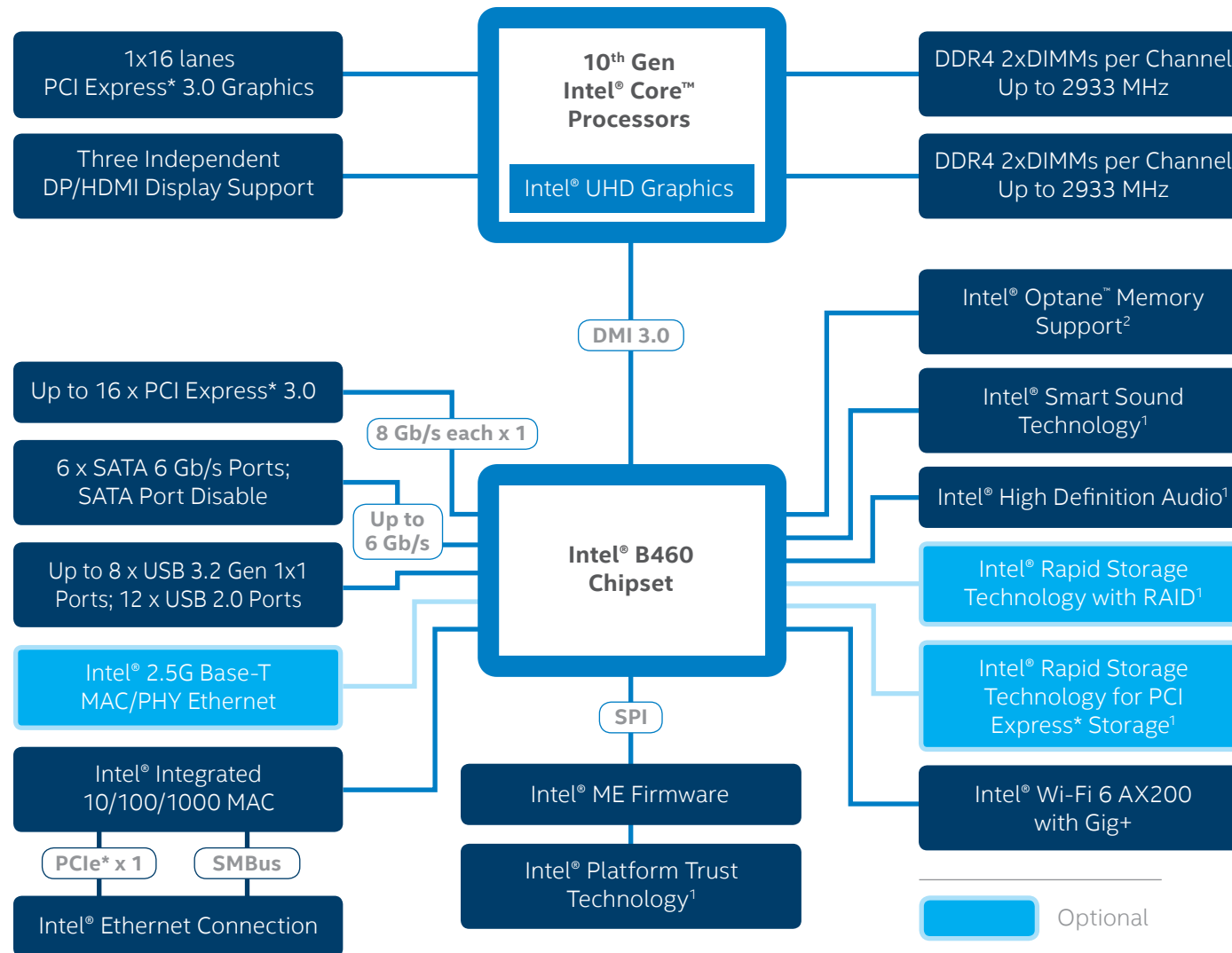


PCI express

- PCI bus: (Peripheral Component Interconnect):
 - Bus vào-ra với băng thông rộng độc lập với bộ xử lý để kết nối với các thiết bị vào-ra
 - Bus dùng chung
- PCI Express (PCIe)
 - Liên kết điểm-điểm thay thế cho PCI bus
 - Kiến trúc giao thức phân lớp
 - Tốc độ cao
 - PCIe 1.0 – 6.0



Ví dụ liên kết trong máy tính Intel



Hết chương 2