

## **Bài 3. Các hệ đếm thường dùng trong tin học và biểu diễn thông tin trong máy tính**

Bài giảng: LẬP TRÌNH CƠ BẢN

## Tài liệu tham khảo

---

- ▶ Giáo trình tin học cơ sở, Hồ Sỹ Đàm, Đào Kiến Quốc, Hồ Đắc Phương. Đại học Sư phạm, 2004 – Chương 4, 6.

# NỘI DUNG

---

- ▶ *Các hệ đếm thường dùng trong tin học*
  - ▶ Hệ đếm
  - ▶ Tìm biểu diễn số trong các hệ đếm
  - ▶ Số học nhị phân
- ▶ *Biểu diễn thông tin trong máy tính*
  - ▶ Dữ liệu kiểu số
  - ▶ Dữ liệu phi số
  - ▶ Biểu diễn thông tin trong máy tính

# HỆ ĐẾM

---

- ▶ Hệ đếm là một tập các ký hiệu (bảng chữ số) để biểu diễn các số và xác định giá trị của các biểu diễn số.
- ▶ Ví dụ: Hệ đếm La mã có bảng chữ là {I,V,X,L,C,D,M} đại diện cho các giá trị là 1, 5,10, 100, 500 và 1000.
  - ▶ Quy tắc biểu diễn số là viết các chữ số cạnh nhau.
  - ▶ Quy tắc tính giá trị là nếu một chữ số có một chữ số bên trái có giá trị nhỏ hơn thì giá trị của cặp số bị tính bằng hiệu hai giá trị. Còn nếu số có giá trị nhỏ hơn đứng phía phải thì giá trị chung bằng tổng hai giá trị

$$MLVI = 1000 + 50 + 5 + 1 = 1056$$

$$MLIV = 1000 + 50 + 5 - 1 = 1054$$

# HỆ ĐẾM

---

VD Hệ đếm thập phân

- ▶ Bảng chữ số  $\{0,1,2,3,4,5,6,7,8,9\}$
- ▶ Quy tắc biểu diễn: ghép các chữ số
- ▶ Quy tắc tính giá trị: mỗi chữ số  $x$  đứng ở hàng thứ  $i$  tính từ bên phải có giá trị là  $x.10^{i-1}$ . Như vậy một đơn vị ở một hàng sẽ có giá trị gấp 10 lần một đơn vị ở hàng kế cận bên phải
- ▶ Giá trị của số là tổng giá trị của các chữ số có tính tới vị trí của nó. Giá trị của 3294,5 là

$$3.10^3 + 2.10^2 + 9.10^1 + 4.10^0 + 5.10^{-1}$$

## HỆ ĐẾM THEO VỊ TRÍ VÀ KHÔNG THEO VỊ TRÍ

---

- ▶ Hệ đếm theo vị trí là hệ đếm mà giá trị của mỗi chữ số không phụ thuộc vào vị trí của nó trong biểu diễn số.
- ▶ Hệ đếm thập phân là hệ đếm theo vị trí
- ▶ Hệ đếm la mã là hệ đếm không theo vị trí

# HỆ ĐẾM THEO VỊ TRÍ CÓ CƠ SỐ BẤT KỲ

---

- ▶ Có thể chọn các hệ đếm với cơ số khác 10.
- ▶ Với một số tự nhiên  $b > 1$ , với mỗi số tự nhiên  $n$  luôn tồn tại một cách phân tích duy nhất  $n$  dưới dạng một đa thức của  $b$  với các hệ số nằm từ 0 đến  $b-1$

$$n = a_k \cdot b^k + a_{k-1} \cdot b^{k-1} + \dots + a_1 b + a_0, \quad 0 \leq a_i \leq b-1$$

Khi đó biểu diễn của  $n$  trong cơ số  $b$  là  $a_k a_{k-1} \dots a_1 a_0$

$$\text{VD } 14 = 1 \cdot 3^2 + 1 \cdot 3^1 + 2 \cdot 3^0 = 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0$$

$$\text{Do đó } 14_{10} = 112_3 = 1110_2$$

# HỆ ĐẾM NHỊ PHÂN

---

- ▶ Hệ nhị phân dùng 2 chữ số là  $\{0,1\}$  và chữ số 1 ở một hàng có giá trị bằng 2 lần chữ số 1 ở hàng kế cận bên phải

$$14,625 = 1.2^3 + 1.2^2 + 1.2^1 + 0.2^0 + 1.2^{-1} + 0.2^{-2} + 1.2^{-3}$$

$$\text{Do đó } 14,625_{10} = 1110,101_2$$

- ▶ Hệ đếm nhị phân là hệ được sử dụng nhiều đối với MTĐT vì MTĐT sử dụng các thành phần vật lý có hai trạng thái để nhớ các bit



# SỐ HỌC NHỊ PHÂN

- ▶ Bảng cộng:  $0+0=0$ ,  $1+0=0+1=1$ ,  $1+1=10$
- ▶ Bảng nhân:  $0 \times 0=0 \times 1=1 \times 0=0$   $1 \times 1=1$
- ▶ Ví dụ  $7+5 = 12$ ,  $12-5 = 7$ ,  $6 \times 5 = 30$ ,  $30:6=5$  được thể hiện trong hệ nhị phân

$$\begin{array}{r} + \quad 111 \\ \quad 101 \\ \hline 1100 \\ 1 \swarrow \searrow \end{array}$$

$$\begin{array}{r} 1100 \\ - \quad 101 \\ \hline 111 \\ 1 \swarrow \searrow \end{array}$$

$$\begin{array}{r} \times \quad 110 \\ \quad 101 \\ \hline \quad 110 \\ + \quad 110 \\ \hline 11110 \end{array}$$

$$\begin{array}{r|l} 11110 & 110 \\ - 110 & \\ \hline \quad 110 & \\ - 110 & \\ \hline \quad 000 & \end{array}$$

# HỆ HEXA (HỆ ĐẾM CƠ SỐ 16)

---

- ▶ Hệ nhị phân tuy tính toán đơn giản nhưng biểu diễn số rất dài. Hệ thập phân thì không thích hợp với máy tính. Người ta thường dùng hệ 16 (hexa) vì biểu diễn số ngắn mà chuyển đổi với hệ nhị phân rất đơn giản
- ▶ Hệ đếm cơ số 16 dùng các chữ số  
 $\{ 0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F \}$
- ▶ Bảng cộng, nhân không hoàn toàn giống như trong hệ thập phân, ví dụ  $5+6 = B$  nhưng cách thực hiện các phép toán số học cũng tương tự như hệ thập phân.

## ĐỔI BIỂU DIỄN SỐ VỚI CÁC CƠ SỐ KHÁC NHAU

---

- ▶ Giả sử có số nguyên  $n$ , trong một hệ đếm cơ số  $p$  nào đó, ta cần tìm biểu diễn của nó trong một hệ đếm cơ số  $b$  và giả sử biểu diễn đó là  $d_k d_{k-1} \dots d_1 a_0$

$$N = d_n \cdot b^n + d_{n-1} \cdot b^{n-1} + \dots + d_1 b^1 + d_0, 0 \leq d_i \leq b-1$$

- ▶ Chia  $n$  cho  $b$  ta được số dư  $d_0$  và thương

$$N_1 = d_n \cdot b^{n-1} + d_{n-1} \cdot b^{n-2} + \dots + d_1 b^1 + d_0$$

- ▶ Chia  $N_1$  cho  $b$  ta được số dư  $d_1$  và thương

$$N_2 = d_n \cdot b^{n-2} + d_{n-1} \cdot b^{n-3} + \dots + d_2 b^1 + d_1$$

- ▶ Như vậy bằng phép chia và tách số dư liên tiếp  $n$  cho cơ số  $b$ , ta lần lượt tách ra các số dư chính là các hệ số của biểu diễn số trong cơ số  $b$ . Quá trình sẽ dừng lại khi nào thương bằng 0

# QUY TẮC THỰC HÀNH ĐỔI BIỂU DIỄN SỐ VỚI PHẦN NGUYÊN

$$23_{10} = ?_2$$

$$\begin{array}{r|l} 23 & 2 \\ \hline 1 & 11 \\ & 2 \\ & 1 \\ & 5 \\ & 2 \\ & 1 \\ & 2 \\ & 2 \\ & 0 \\ & 1 \\ & 2 \\ & 1 \\ & 0 \end{array}$$

$$923_{10} = ?_{16}$$

$$\begin{array}{r|l} 923 & 16 \\ \hline 11 & 57 \\ B & 9 \\ & 16 \\ & 3 \\ & 3 \\ & 16 \\ & 0 \end{array}$$

Lấy các số dư theo  
thứ tự ngược lại

## ĐỔI BIỂU DIỄN SỐ CHO PHẦN LẺ VỚI CÁC CƠ SỐ KHÁC NHAU

---

- ▶ Có số  $x < 1$ , cần đổi ra phần lẻ trong biểu diễn cơ số  $b$

$$x = d_{-1}.b^{-1} + d_{-2}.b^{-2} + \dots + d_{-m}.b^{-m} + \dots$$

- ▶ Nếu nhân  $x$  với  $b$ ,  $d_{-1}$  sẽ chuyển sang phần nguyên và phần lẻ sẽ là

$$x_2 = d_{-2}.b^{-1} + d_{-3}.b^{-2} + \dots + d_{-m}.b^{-m+1} + \dots$$

- ▶ Nếu nhân  $x_2$  với  $b$ ,  $d_{-2}$  sẽ chuyển sang phần nguyên và phần lẻ sẽ là

$$x_3 = d_{-3}.b^{-1} + d_{-4}.b^{-2} + \dots + d_{-m}.b^{-m+2} + \dots$$

- ▶ Do đó có thể tách các số chữ số bằng nhân liên tiếp phần lẻ với  $b$  và tách lấy phần nguyên

## QUY TẮC THỰC HÀNH ĐỔI BIỂU DIỄN SỐ VỚI PHẦN LẺ

►  $0,427_{10} = 0,?_2$

0. 427 x 2

0. 854 x 2

1. 708 x 2

1. 416 x 2

0. 832 ....

●  $0,42_{10} = 0,6B85..._{16}$

0. 42 x 16

6. 72 x 16

11. 52 x 16

8. 32 x 16

5. 12 ....

Một số hữu hạn ở một cơ số này có thể  
là một số vô hạn trong một cơ số khác

## ĐỔI BIỂU DIỄN SỐ VỚI CÁC CƠ SỐ KHÁC NHAU

---

- ▶ Cách đổi như đã nêu trên được sử dụng để đổi một số trong hệ thập phân sang một hệ đếm bất kỳ
- ▶ Để đổi từ một hệ đếm bất kỳ sang hệ thập phân có thể tính trực tiếp giá trị của đa thức

$$P = a_k \cdot b^k + a_{k-1} \cdot b^{k-1} + \dots + a_1 b^1 + a_0 \dots$$

Cách tính tiết kiệm là sử dụng lược đồ Horner

$$P = a_0 + b(a_1 + b(a_2 + b(\dots))))$$

## ĐỔI BIỂU DIỄN SỐ VỚI CÁC CƠ SỐ KHÁC NHAU

---

- ▶ Để đổi một số có cả phần nguyên và phần lẻ thì đổi riêng phần nguyên và phần lẻ rồi ghép lại
- ▶ Để đổi một số âm thì đổi giá trị tuyệt đối sau đó thêm dấu
- ▶ Điều khó khăn đối với hai cơ số bất kỳ khác 10 là ta không quen tính các phép tính số học trong hệ đếm cơ số khác 10. Vì thế có thể chọn hệ đếm thập phân làm trung gian trong tính toán:

$$X_p \rightarrow Y_{10} \rightarrow Z_q$$



## ĐỔI BIỂU DIỄN SỐ TRONG TRƯỜNG HỢP CƠ SỐ LÀ LŨY THỪA CỦA NHAU

---

- ▶ Nếu đổi  $x_p \rightarrow y_q$  mà  $p=q^k$  thì  $p$  sẽ có biểu diễn là  $100\dots 0$  ( $k$  chữ số 0). Khi đó phép nhân để tách phần nguyên và chia để tách phần dư nói trong phần đổi biểu diễn nói trên thực chất là tách biểu diễn số trong hệ đếm cơ số  $q$  thành các nhóm  $k$  chữ số tính từ dấu phẩy ngăn cách phần nguyên và phần lẻ về hai phía. Mỗi nhóm  $k$  chữ số của hệ đếm cơ số  $q$  cho giá trị của một chữ số trong hệ đếm cơ số  $p$
- ▶ Từ đó có quy tắc thực hành như sau:  
  
Nhóm các chữ số của số trong biểu diễn hệ đếm cơ số  $q$  thành từng nhóm đủ  $k$  chữ số tính từ dấu phẩy. Sau đó thay mỗi nhóm này bằng một chữ số tương ứng của hệ đếm cơ số  $p$

## BẢNG TƯƠNG ỨNG GIÁ TRỊ CỦA CÁC CHỮ SỐ TRONG HỆ 16 TRONG HỆ ĐẾM CƠ SỐ 2

Hệ 10	Hệ 16	Hệ 2	Hệ 10	Hệ 16	Hệ 2
0	0	0000	8	8	1000
1	1	0001	9	9	1001
2	2	0010	10	A	1010
3	3	0011	11	B	1011
4	4	0100	12	C	1100
5	5	0101	13	D	1101
6	6	0110	14	E	1110
7	7	0111	15	F	1111

## ĐỔI BIỂU DIỄN GIỮA HỆ ĐẾM CƠ SỐ 16 VÀ HỆ ĐẾM CƠ SỐ 2

- ▶ Ví dụ ta cần đổi số 1001101,010011 ra hệ đếm cơ số 16
- ▶ Ta có  $16 = 2^4$ . Để đổi từ hệ đếm cơ số 2 thành hệ đếm cơ số 16, nhóm các chữ số thành các nhóm đủ 4 chữ số, sau đó thay mỗi nhóm đó bằng một chữ số tương ứng

1001101,0100110  $\rightarrow$  01001101,01011100

$\rightarrow$

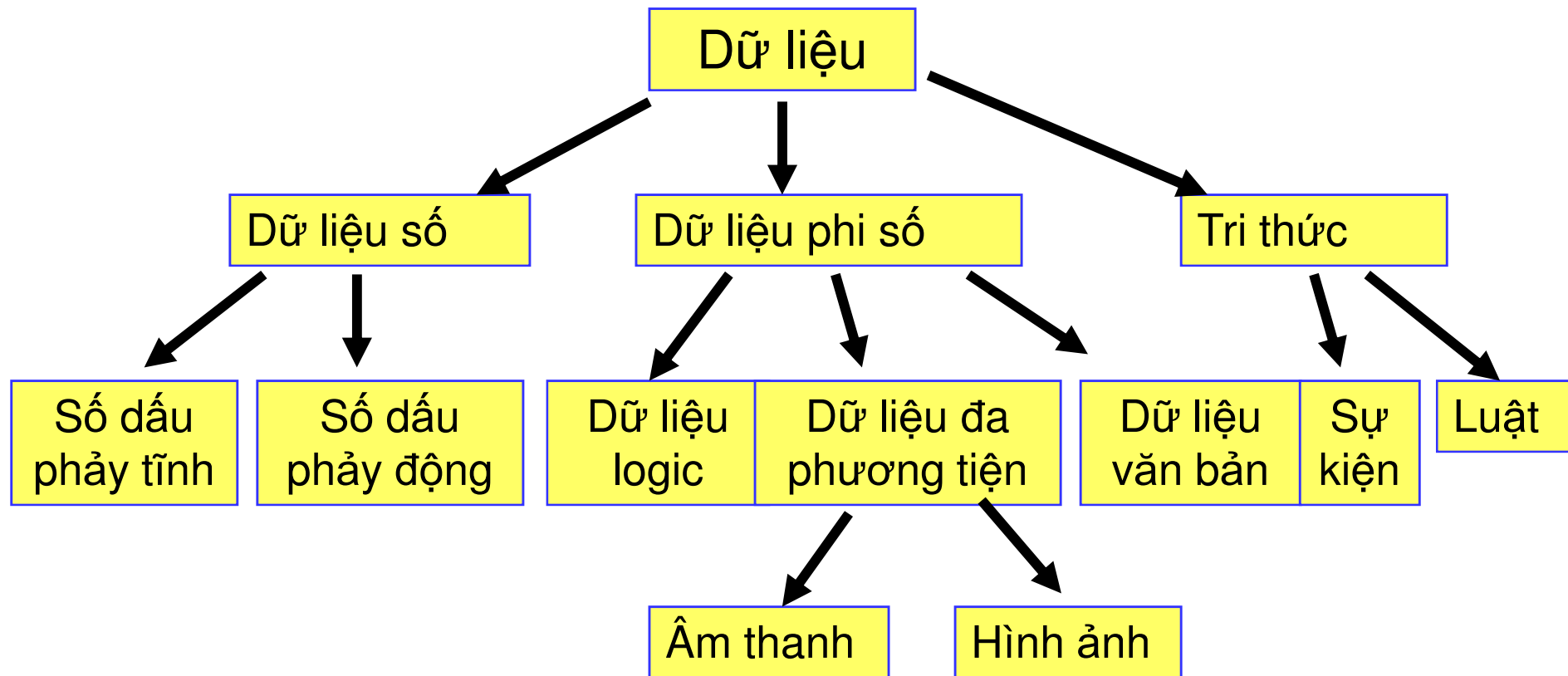
4 D 5 C

- ▶ Ngược lại để đổi một số từ hệ 16 sang hệ 2 chỉ cần thay mỗi chữ số bằng một nhóm 4 đủ chữ số tương ứng

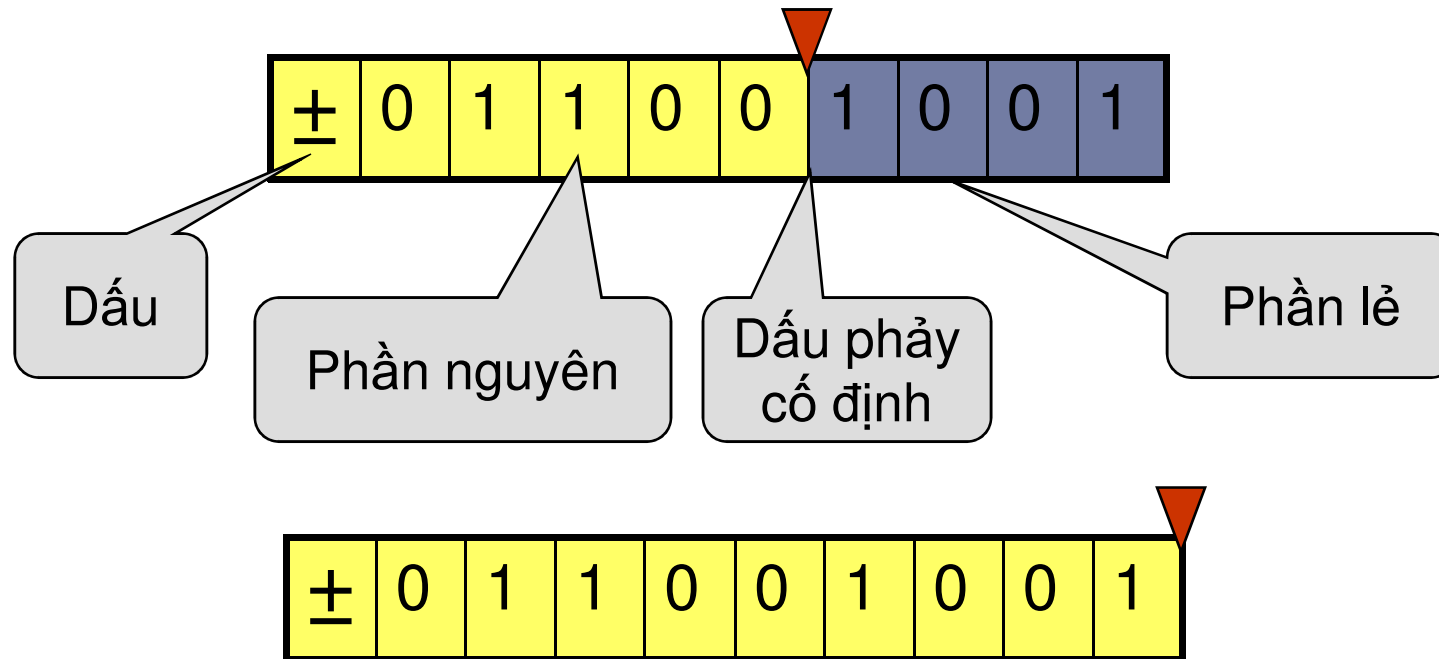
14F,8D  $\rightarrow$  0001 0100 1111, 0111  $\rightarrow$  101001111,0111

# PHÂN LOẠI DỮ LIỆU

---

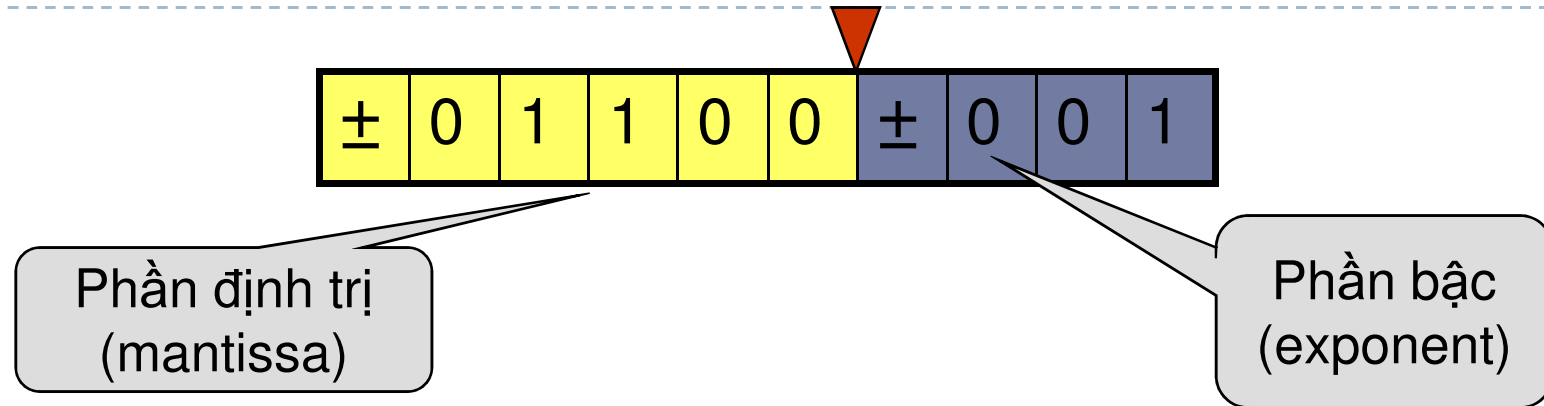


## SỐ DẤU PHẪY TĨNH (fixed point number)



Có một vị trí cố định ngăn cách giữa phần nguyên và phần lẻ -> dấu phẩy tĩnh

## SỐ DẤU PHẪY ĐỘNG ( floating point number)



- Số được biểu diễn dưới dạng nửa logarit  $x = \pm m_x \cdot 10^{\pm P_x}$

- Ví dụ  $3.14 = 0.314 \times 10^2$  hoặc  $-0.0012 = -0.12 \times 10^{-2}$

- Vị trí dấu phẩy trong biểu diễn bình thường do phần bậc định ra trên phần định trị nên gọi là dấu phẩy động. Số dấu phẩy động thường được dùng với tính toán gần đúng. Trong một số ngôn ngữ lập trình nó được khai báo với kiểu là real hay double. Người ta đo tốc độ của các máy tính khoa học kỹ thuật theo **Flops (floating point operations per second) hoặc Gflops**

# SO SÁNH KHOẢNG BIỂU DIỄN

- ▶ Về khả năng biểu diễn số. Với cùng một số ngăn nhớ, số mã khác nhau có thể biểu diễn được hoàn toàn như nhau nhưng khoảng số biểu diễn được khác nhau rất xa. Có thể xem xét qua số dương lớn nhất và số dương nhỏ nhất có thể biểu diễn được. Dưới đây tất cả viết trong hệ đếm cơ số 2.
- ▶ Xét ví dụ với 4 ngăn định trị, 2 ngăn cho bậc và 2 ngăn cho dấu
- ▶ Khoảng biểu diễn được ở chế độ dấu phẩy động là  $0.1 \times 10^{-11}$  đến  $0.1111 \times 10^{11}$  (tổng quát trong trường hợp m ngăn cho định trị và n ngăn cho bậc không kể dấu sẽ là từ  $10(10^{-111..1} - 1)$  đến  $10^{111..1}$ )
- ▶ Với số dấu phẩy tĩnh khoảng biểu diễn chỉ được từ 1 đến  $10^{m+n} - 1$ .
- ▶ Về khoảng biểu diễn, chế độ dấu phẩy động tốt hơn rất nhiều

+	1	1	1	1	+	1	1
+	1	0	0	0	-	1	1
+	1	1	1	1	1	1	1
+	0	0	0	0	0	0	1

# SO SÁNH ĐỘ CHÍNH XÁC

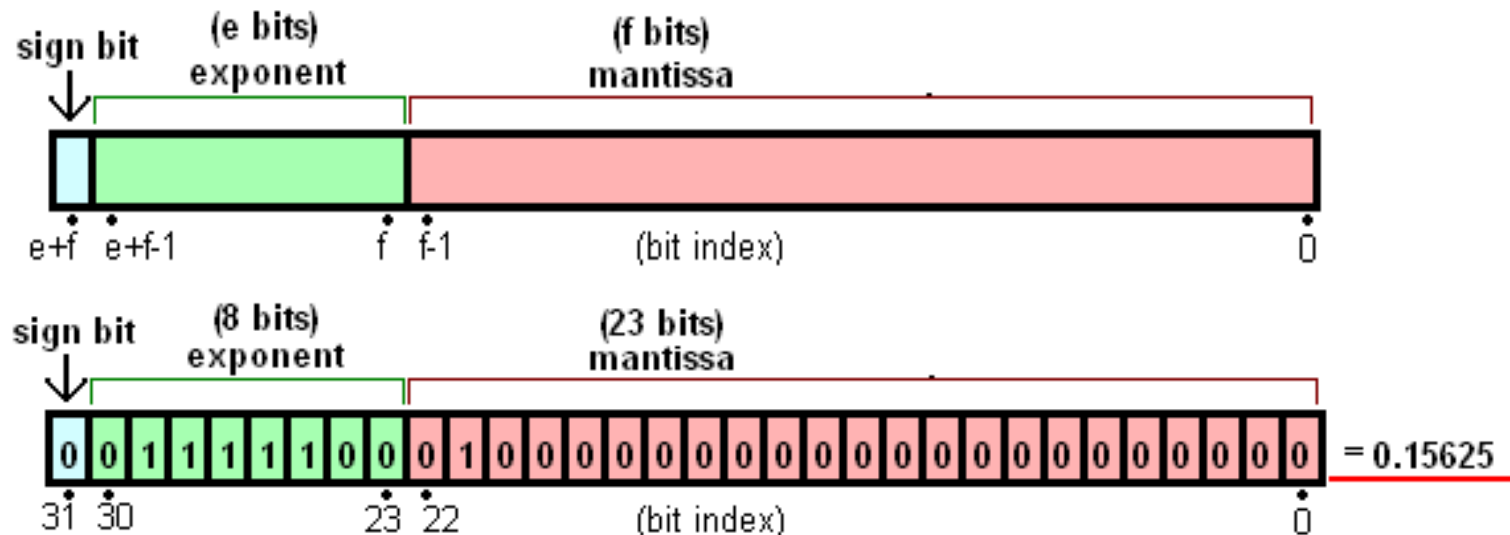
---

- ▶ Do số ngăn của một ô nhớ bị hạn chế nên biểu diễn sẽ mắc sai số làm tròn. Có hai loại sai số: với số  $x$  được xấp xỉ bằng  $x'$  thì  $|x-x'|$  gọi là sai số tuyệt đối, còn  $|(x-x')/x|$  được gọi là sai số tương đối
- ▶ Với dấu phẩy tĩnh trong chế độ số nguyên, sai số tuyệt đối luôn là 1, còn sai số tương đối là có thể lớn tùy theo số nhỏ hay lớn.
- ▶ Với số dấu phẩy động với  $m$  ngăn cho phần định trị và ngăn cho phần bậc sai số tương đối do làm tròn luôn luôn không quá  $10^{-111..1}$  (n so), , còn  $n$  sai số tương đối bị khuếch đại bởi phần bậc có thể lên tới  $10^{10^n-1}$
- ▶ Sai số tuyệt đối có thể lớn nhưng sai số tương đối thì rất tốt. Chính vì vậy trong các bài toán tính toán gần đúng, biểu diễn dấu phẩy động rất phù hợp



# SỐ DẤU PHẪY ĐỘNG CHUẨN IEEE 754

• Chuẩn IEEE 754 là một chuẩn được sử dụng rộng rãi nhất hiện nay cho tính toán dấu phẩy động. Chuẩn này định nghĩa định dạng và cách thực hiện các phép tính trên các số phẩy động trong đó có cả số 0 với dấu âm, các số không chuẩn hoá, các giá trị đặc biệt như vô hạn và giá trị không phải số (NaNs). Chuẩn cũng xác định 4 kiểu làm tròn số và 5 ngoại lệ. Bit cao nhất là dấu của số, sau đó là phần bậc, cuối cùng là phần định trị.



# SỐ DẤU PHẪY ĐỘNG CHUẨN IEEE 754

Kiểu	Phần bậc Exponent	Phần định trị Mantissa
Số 0 (Zeroes)	0	0
Các số không chuẩn hoá (Denormalized numbers)	0	$\neq 0$
Các số chuẩn hoá (Normalized numbers)	1 to $2^e - 2$ (1 -1111...110)	bất kỳ
Vô hạn (Infinities)	$2^e - 1$ (1111...111)	0
Không phải số (NaNs)	$2^e - 1$ (1111...111)	$\neq 0$

# BIỂU DIỄN CHỮ VÀ VĂN BẢN

---

- ▶ Với  $k$  bit, có thể biểu diễn  $2^k$  mã khác nhau. Ta dùng thuật ngữ ký tự (character) để chỉ một biểu diễn cho một ký hiệu phân biệt với chữ (letter) thông thường mà letter cũng chỉ là một loại ký tự giống như chữ số, các dấu chính tả và các dấu đặc biệt khác
- ▶ Bộ mã Mã EBCDIC (Extended Binary Coded Decimal Interchange Code) trong những năm 70 dùng 6 bit có thể mã được 64 ký tự
- ▶ Bộ mã ASCII (American Standard Codes for Information Interchange) dùng 7 bit cho phép biểu diễn 128 ký tự (32 mã đầu tiên dùng cho các mã điều khiển và truyền thông, tiếp theo là các dấu chính tả, các chữ số, các chữ thường, các chữ in và các dấu đặc biệt).
- ▶ Bộ mã ASCII mở rộng dùng 1 byte cho một ký tự nên có khả năng biểu diễn 256 ký tự. 128 chỗ vùng tiếp theo có thể cho chữ của các nước châu Âu, chữ Hy Lạp hoặc bất cứ một bộ chữ nào như tiếng Việt hay ngôn ngữ Slavơ, nhưng không thể đủ cho tiếng Trung Quốc hay Nhật Bản

# BẢNG CHỮ ASCII (128 ký tự đầu)

	000	001	010	011	100	101	110	111
00000	0 NUL	1 SOH	2 STX	3 EXT	4 EOT	5	6	7 BELL
00001	8 BS	9 HT	10 LF	11 VT	12 FF	13 CR	14	15
00010	16	17 DC1	18 DC2	19 DC3	20 DC4	21	22	23
00011	24	25	26	27	28	29	30	31
00100	32	33 !	34 "	35 #	36 \$	37 %	38 &	39 '
00101	40 (	41 )	42 *	43 +	44 ,	45 -	46 .	47 /
00110	48 0	49 1	50 2	51 3	52 4	53 5	54 6	55 7
00111	56 8	57 9	58 :	59 ;	60 <	61 =	62 >	63 ?
01000	64 @	65 A	66 B	67 C	68 D	69 E	70 F	71 G
01001	72 H	73 I	74 J	75 K	76 L	77 M	78 N	79 O
01010	80 P	81 Q	82 R	83 S	84 T	85 U	86 V	87 W
01011	88 X	89 Y	90 Z	91 [	92 \	93 ]	94 ^	95 _
01100	96 `	97 a	98 b	99 c	100 d	101 e	102 f	103 g
01101	104 h	105 i	106 j	107 k	108 l	109 m	110 n	111 o
01110	112 p	113 q	114 r	115 s	116 t	117 u	118 v	119 w
01111	120 x	121 y	122 z	123 {	124	125 }	126 ~	127

# BIỂU DIỄN CHỮ VỚI UNICODE

---

- ▶ Đối với quốc gia có bộ chữ lớn (như Trung quốc, Nhật bản) bộ mã 8 bit không đủ chỗ cho tất cả các chữ. Nhật Bản đã đưa ra một dự án lập bộ chữ cho toàn cầu gọi là UNICODE. Bộ chữ được chia trang cho các quốc gia. Mặt chữ nào của một nước nào đã có sẽ được dùng lại tại các phần mềm khác.
- ▶ Sau này các tổ chức chuẩn chấp nhận UNICODE dưới chuẩn ISO 10646
- ▶ Mỗi quốc gia có thể nhận các trang mã (code page), mỗi ký tự được thể hiện qua mã của trang mã và số thứ tự (code point) của ký tự đó trong trang mã - một số 2 byte). Trong bảng mã UNICODE, chữ “ơ” có điểm mã là 01A1 (so sánh với bảng mã CP1258 của Microsoft, bảng mã 8 bit, chữ “ơ” có điểm mã F5)

# MÃ TIẾNG VIỆT

---

- ▶ Từng tồn tại tới 40 mã tiếng Việt 8 bit dẫn đến tình trạng loạn mã, không chia sẻ được dữ liệu. Có 141 ký tự đặc thù Việt Nam không có chỗ (vùng mở rộng chỉ có 128 chỗ)
- ▶ Năm 1993 xây dựng bộ mã TCVN 5712. Thực chất vẫn là một giải pháp chắp vá với 3 bộ mã khác nhau. Bộ mã 1, chiếm thêm một số chỗ trong vùng mã điều khiển – nguy hiểm cho truyền thông). Bộ mã 2 là bộ mã tổ hợp, dùng một chuỗi ký tự để thể hiện một mã cho các chữ thuần Việt. Bộ mã 3 hy sinh một số ký tự hoa có dấu ví dụ ã. Cả 3 giải pháp đều không giải quyết được triệt để
- ▶ Từ 2001, Bộ KHCN đã ban hành tiêu chuẩn TCVN 6909/2001 về việc sử dụng mã UNICODE có hiệu lực từ 1/1/2003. Các cơ quan nhà nước buộc phải dùng bộ mã này trong trao đổi dữ liệu.
- ▶ TCVN 6909 vẫn chấp nhận cả hai kiểu: mã dựng sẵn (pre-compound) với mỗi ký tự thể hiện bởi một mã 2 byte và kiểu tổ hợp cho phép dùng một chuỗi ký tự 8 bit để thể hiện một ký tự

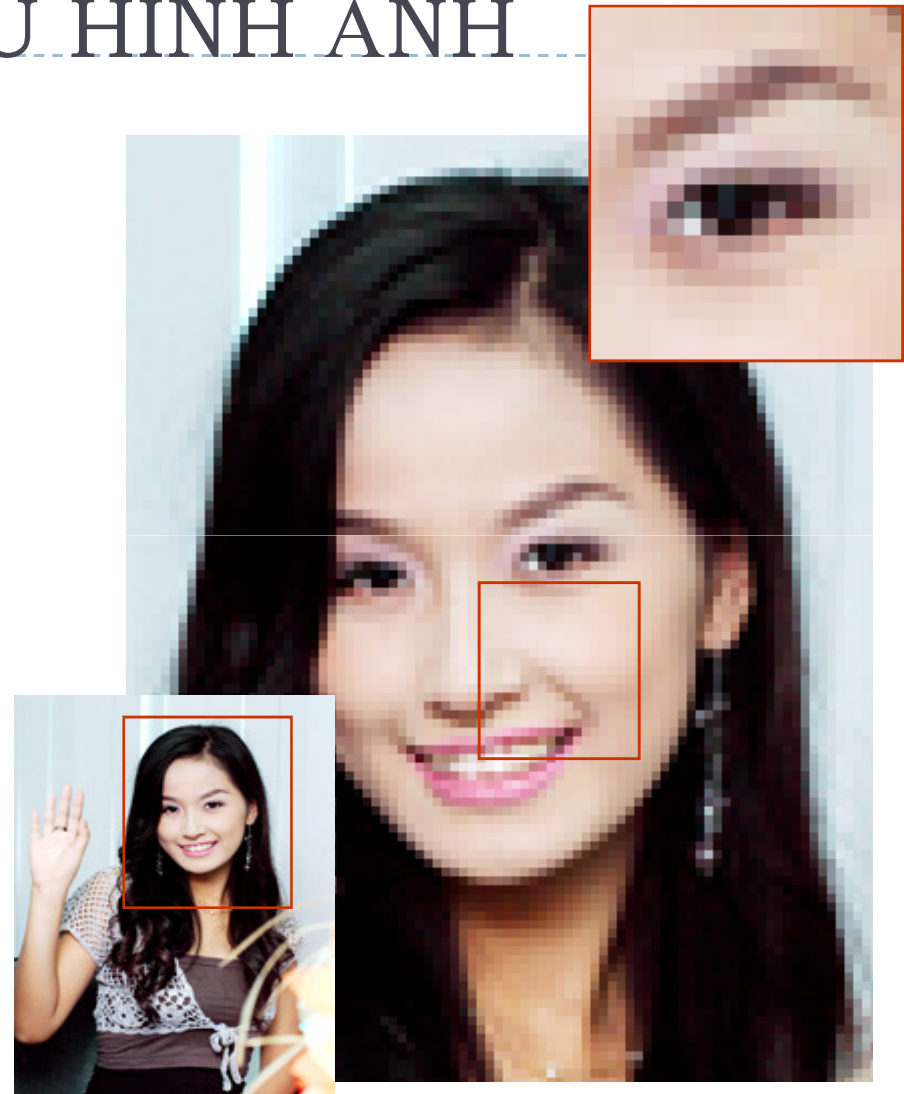
# BIỂU DIỄN CÁC GIÁ TRỊ LOGIC

---

- ▶ Trong đời sống, có các loại thông tin mà giá trị của nó có hai trạng thái đối lập có thể là “có/không”, “đúng/sai”. Dữ liệu loại này gọi là dữ liệu logic
- ▶ Các dữ liệu logic có thể tương tác với nhau thông qua các phép toán logic mệnh đề như “Và”, “hoặc”, “không”
- ▶ Về nguyên tắc có thể mã hoá các đại lượng logic bằng 1 bit (1 là đúng hoặc có, 0 là sai hoặc không có). Tuy nhiên người ta ít khi làm như thế vì đơn vị nhớ cơ sở là byte. Trong cài đặt cụ thể người ta có thể dùng các ký tự như T (true) và F (false) để biểu diễn hai giá trị “đúng” và “sai”

# BIỂU DIỄN DỮ LIỆU HÌNH ẢNH

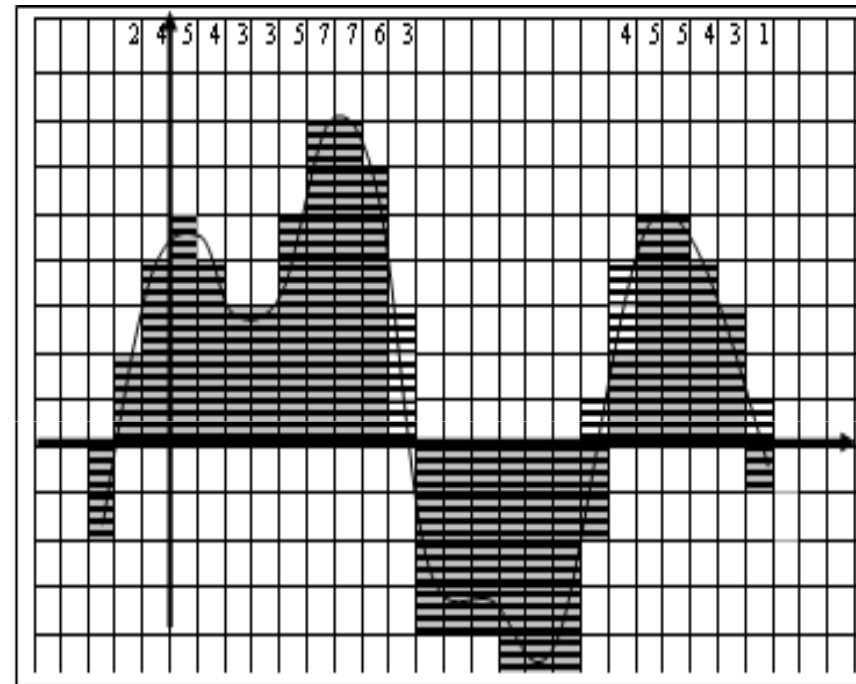
- ▶ Ảnh là một tập hợp các điểm ảnh (pixel), có màu sắc tạo từ 3 màu cơ bản (red, green, blue) với cường độ khác nhau.
- ▶ Ví dụ ảnh màu 24 bit, dùng mỗi byte để mã một màu với các mức từ 0 đến 255. Như vậy sẽ có  $2^{24}$  (khoảng 19 triệu ) sắc độ màu khác nhau.
- ▶ Có các chuẩn ảnh khác chủ yếu khác nhau về việc cấu trúc thông tin ảnh phù hợp với phương pháp nén ảnh và thể hiện ảnh. Một số chuẩn ảnh thông dụng là bitmap, jpeg, gif, tiff
- ▶ Ảnh trực tiếp thể hiện bằng điểm ảnh gọi là ảnh bitmap hay ảnh raster. Còn một kiểu ảnh khác là ảnh vector





# BIỂU DIỄN ÂM THANH

- ▶ Cách đơn giản nhất là mã hoá bằng cách xấp xỉ dao động sóng âm bằng một chuỗi các byte thể hiện biên độ dao động tương ứng theo từng khoảng thời gian bằng nhau.
- ▶ Các đơn vị thời gian này cần phải đủ nhỏ để không làm nghèo âm thanh. Đơn vị thời gian này gọi là chu kỳ lấy mẫu.
- ▶ Khi phát lại, người ta dùng một mạch điện để tái tạo lại âm thanh từ các biên độ dao động của từng chu kỳ lấy mẫu



**Có một số chuẩn định dạng âm thanh như wav, một số chuẩn khác cho phép nén âm thanh cùng với các hình ảnh động**

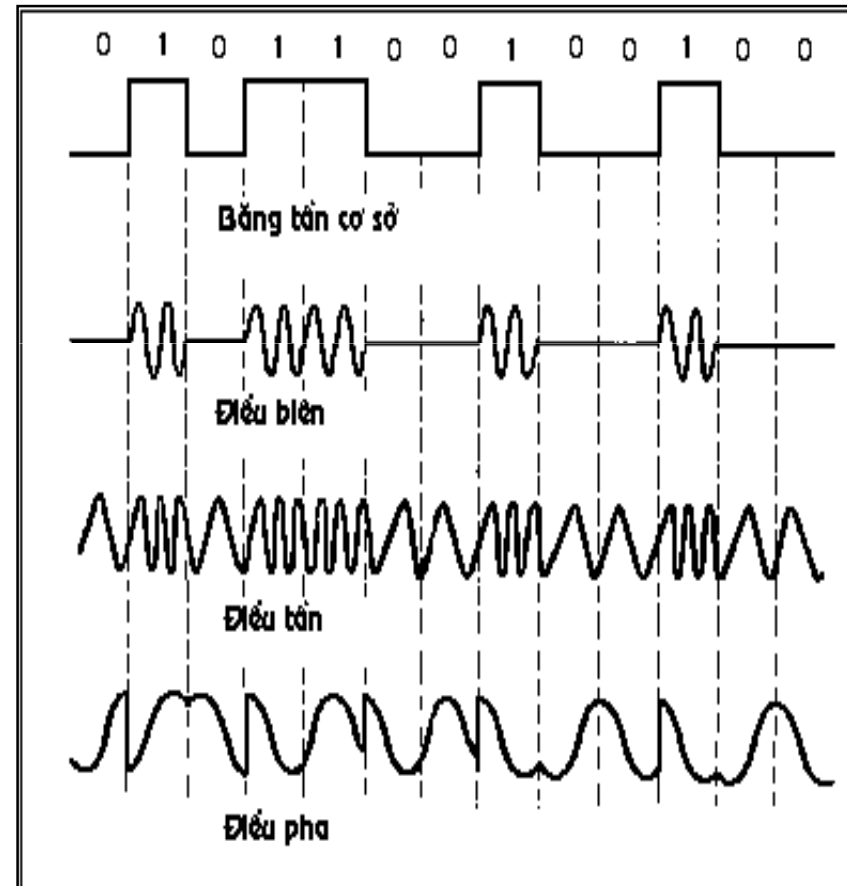
# TRI THỨC = SỰ KIỆN + LUẬT

---

- ▶ Tri thức (knowledge) không chỉ thể hiện bằng các sự kiện (fact) mà ta có thể biểu diễn như các dữ liệu thông thường mà nó còn thể hiện cách suy luận cho bằng các luật (rule)
- ▶ VD quan hệ “Là bố” có thể cho bằng 2 chuỗi ký tự hiểu theo nghĩa tên bố và tên con. Là bố (Hùng, Cường) nghĩa là Hùng là bố của Cường.
- ▶ Quy tắc “**Nếu** (A là bố B) và (B là bố C) **thì** A là ông nội C” cho phép từ một số quan hệ này suy ra một số quan hệ khác
- ▶ Chẳng hạn từ Là bố (Bé, Cường) và Là bố (Cường, Đại) thì theo quy tắc trên sẽ rút ra Bé là ông nội của Đại

# TRUYỀN DỮ LIỆU

- ▶ Dữ liệu được lưu trữ dưới dạng trạng thái nhị phân nhưng truyền đi bằng sóng điện từ
- ▶ Cần điều chế (modulation) tín hiệu trên các sóng mang trong các kênh truyền vật lý.
- ▶ Có thể điều chế theo tần số, biên độ và pha.
- ▶ Đôi khi người ta điều chế bằng cả điều pha và điều biên, cho phép truyền thông với tốc độ cao hơn cả tần số của sóng mang như trong modem 9.6 kb/s với mã hoá kiểu chòm sao (constellation)



# TỔNG KẾT NỘI DUNG

---

- ▶ Trong tin học, người ta thường dùng hệ đếm cơ số 2 và cơ số 16
- ▶ Việc đổi số nguyên có thể thực hiện bằng cách chia liên tiếp cho cơ số mới và tách phần dư liên tiếp sau đó lấy theo chiều ngược lại các số dư
- ▶ Việc đổi phần lẻ có thể thực hiện bằng cách nhân liên tiếp và tách phần nguyên
- ▶ Để chuyển đổi từ hệ đếm cơ số 2 sang 16 chỉ cần nhóm từng cụm đủ 4 chữ số hệ 2 kể từ dấu phẩy về hai phía và thay mỗi cụm này bằng một chữ số hệ 16 tương ứng
- ▶ Ngược lại để đổi một số từ hệ đếm cơ số 16 sang hệ đếm cơ số 2 chỉ cần thay mỗi chữ số của hệ đếm cơ số 16 bởi một nhóm đủ 4 chữ số của hệ đếm cơ số 2.

# Tóm tắt nội dung

---

- ▶ Dữ liệu là cách thể hiện thông tin với mục đích lưu trữ, xử lý và truyền tin
- ▶ Có nhiều loại dữ liệu như số, văn bản, logic, đa phương tiện và tri thức. Mỗi loại có những đặc thù riêng đi kèm với các mã hoá
- ▶ Để truyền dữ liệu, người ta phải điều chế. Đối với tín hiệu điện, thường phải gửi theo sóng mang với cơ chế mã hoá theo kiểu điều tần, điều pha, điều biên hay hỗn hợp.

# THẢO LUẬN

---

- ▶ Biểu diễn dữ liệu trong máy tính, các dạng có thể nén.
- ▶ Biểu diễn dữ liệu hình ảnh.

# CÂU HỎI VÀ BÀI TẬP

---

1. Vì sao người ta sử dụng hệ nhị phân để biểu diễn thông tin cho MTĐT?
2. Hãy đổi các số thập phân sau đây ra hệ nhị phân (chú ý rằng trong tin học ta thường dùng cách viết số theo kiểu Anh, dấu phân cách giữa phần nguyên và phần lẻ là dấu chấm chứ không phải dấu phẩy)

5, 9, 17, 27, 6.625

3. Hãy đổi các số nhị phân sau đây ra hệ thập phân:

11, 111, 1001, 1101, 1011.110

4. Đổi các số nhị phân sau đây ra hệ 16

11001110101, 1010111000101, 1111011101.1100110

5. Đổi các số hệ 16 ra hệ nhị phân

3F8, 35AF, A45

# CÂU HỎI VÀ BÀI TẬP

---

6. Người ta nói dữ liệu là hình thức biểu diễn của thông tin.  
Cũng có người nói dữ liệu là thông tin được xử lý bằng máy tính. Hai cách nói này có mâu thuẫn không.
7. Thế nào là dữ liệu số, thế nào là dữ liệu phi số
8. Tại sao cần các chế độ biểu diễn số khác nhau như chế độ dấu phẩy động và chế độ dấu phẩy tĩnh
9. Nêu các phương pháp điều chế tín hiệu để truyền dữ liệu



---

# HỎI VÀ ĐÁP

---

► Các hệ đếm thường dùng trong tin học và biểu diễn thông tin trong máy tính