

# MULTIMEDIA

Lecturer: Nguyen Duy Nhat Vien

July, 2006.

# TÀI LIỆU THAM KHẢO

- **CMPT 365 Course Contents, Spring 2000, Website:** <http://www.cs.sfu.ca/CourseCentral/365/li/index.html>
- **“Principles of Digital Audio”, Ken C.Pohmanm Fourth Edition McGraw-Hill.**
- **“Digital Video processing”, A. Murat Tekalp, University of Rochester, Prentice Hall PTR.**
- **“Multimedia processing”, Andrew Calway, COMS72200.**
- **“Fundamentals of Digital Image Processing”., Anil.K.Jan, Prentice Hall, 1996.**
- **MPEG Home Page, <http://www.cselt.it/mpeg/>**
- **“Emerging Wireless Multimedia Services and Technologies”, JohnWileySons, Aug 2005**
- **“Multimedia Content and the Semantic Web Standards Methods and Tools”, John Wiley Sons, Jun 2005**
- **“Introduction To Digital Audio Signal Processing”, Davide Rocchesso, 2003**

# NỘI DUNG



- TỔNG QUAN
- KỸ THUẬT AUDIO
- KỸ THUẬT VIDEO

# TỔNG QUAN



# TỔNG QUAN

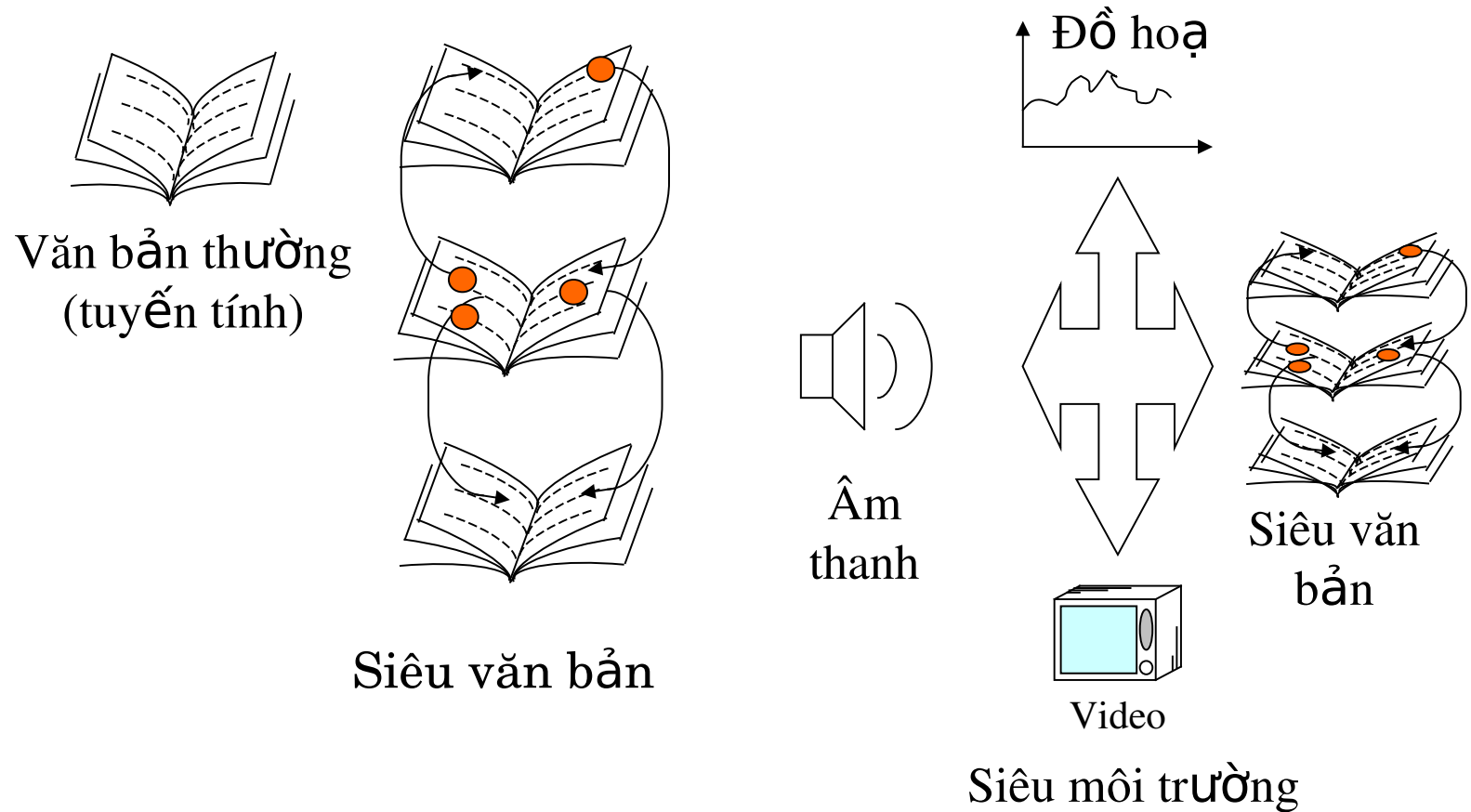


- TỔNG QUAN VỀ MULTIMEDIA
- KHÁI NIỆM CHUNG VỀ AUDIO VÀ VIDEO
- HỆ THỐNG AUDIO-VIDEO
- MỘT SỐ VẤN ĐỀ VỀ TÍN HIỆU
- NÉN DỮ LIỆU

# TỔNG QUAN VỀ MULTIMEDIA

- **Lịch sử phát triển của hệ thống đa môi trường**
  - Báo, tạp chí - môi trường: văn bản, đồ họa và hình ảnh.
  - Cáp đồng: môi trường truyền tải tín hiệu điện.
  - 1895, Guglielmo Marconi phát minh ra máy radio ở Pontecchio – Ý, môi trường chuyển tải tín hiệu audio quảng bá hiện nay.
  - Truyền hình, môi trường truyền thông của thế kỷ 20, truyền hình ảnh và âm thanh đến mọi nơi.
  - Các hệ thống máy tính tích hợp nhiều dạng môi trường số khác nhau, khả năng biểu diễn, tương tác với các dạng thông tin, tiềm năng lớn phục vụ nhu cầu trao đổi thông tin chất lượng cao.
  - Các hệ thống đa môi trường trở nên phong phú, kết hợp các công nghệ khác nhau với khả năng di

# TỔNG QUAN VỀ MULTIMEDIA



Hình 1-1 Hypertext, Hypermedia

# TỔNG QUAN VỀ MULTIMEDIA

- **Siêu môi trường và đa môi trường (hypermedia – multimedia)**
  - Hypertext: Văn bản chứa các liên kết đến các văn bản khác.
  - *“Siêu văn bản là một tài liệu không tuyến tính, bằng cách kích vào một điểm nóng nào đó trên văn bản, nó có thể chuyển đến một tài liệu hay một văn bản khác, rồi có thể quay về, thuận tiện cho người đọc trong việc duyệt văn bản hoặc muốn tổng quan một văn bản từ phần mục lục”*. (Ted Nelson ,1965)
  - Hypermedia: Bao gồm nhiều môi trường truyền thông khác nhau như đồ thị, hình ảnh, đặc biệt là âm thanh, hoạt hình và ảnh động. (Ted Nelson).



# TỔNG QUAN VỀ MULTIMEDIA

- **Multimedia:** thông tin máy tính có thể được mô tả bằng audio, video hay hoạt hình ngoài những môi trường truyền thống (text, graphic, images...).
- **Định nghĩa multimedia:**
  - Multimedia bao gồm sự tích hợp điều khiển bằng máy tính của văn bản, đồ họa, đồ thị, ảnh tĩnh và ảnh động, âm thanh và các môi trường khác, trong đó, mỗi dạng thông tin có thể được biểu diễn, lưu trữ, truyền và xử lý số.
- **Ứng dụng multimedia:**
  - Ứng dụng sử dụng một tập các tài nguyên đa môi trường như văn bản, đồ họa, đồ thị, hoạt hình, video, audio...
- **Hệ thống multimedia:**
  - Hệ thống có khả năng xử lý dữ liệu và ứng dụng multimedia.
  - Được đặc trưng bởi sự lưu trữ, xử lý, tạo, thao tác, diễn xuất thông tin multimedia.

# TỔNG QUAN VỀ MULTIMEDIA



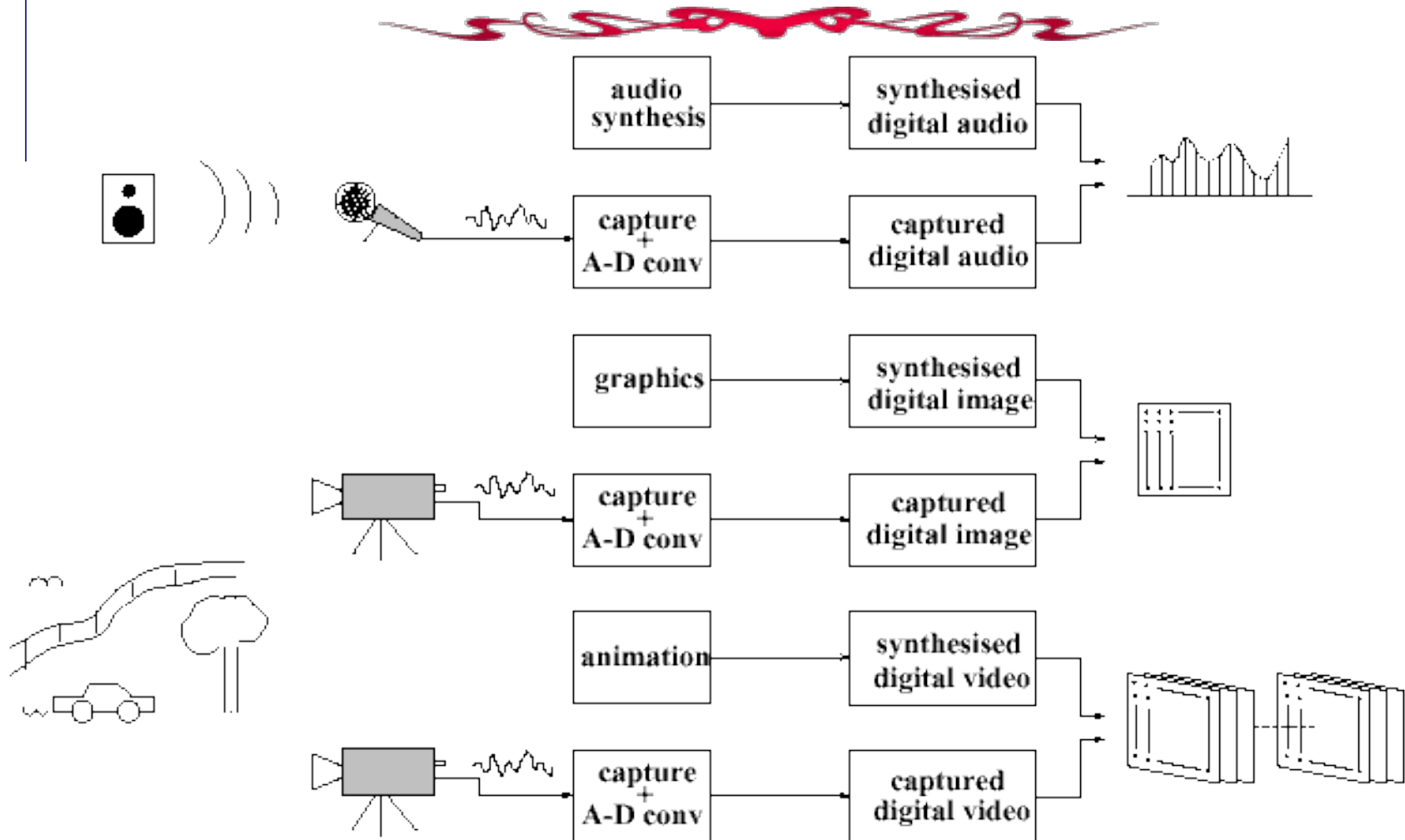
- Ví dụ ứng dụng hypermedia:
  - World Wide Web (WWW) là ví dụ tốt nhất của ứng dụng hypermedia .
  - Powerpoint
  - Word
  - Adobe Acrobat
  - Macromedia Director
  - ...
- Ví dụ ứng dụng multimedia:
  - Hệ thống xây dựng và soạn thảo video số.
  - Tạp chí điện tử.
  - Trò chơi.
  - Thương mại điện tử.
  - Truyền hình tương tác iTV, truyền hình theo yêu cầu.
  - Truyền hình hội nghị.
  - Thực tế ảo.
  - ...

# TỔNG QUAN VỀ MULTIMEDIA



- **Các đặc trưng của hệ thống multimedia:**
  - Hệ thống multimedia phải được điều khiển bởi máy tính.
  - Hệ thống multimedia phải được tích hợp.
  - Thông tin được xử lý được biểu diễn ở dạng số.
  - Giao diện trình diễn cuối cùng của môi trường thường có tính tương tác.

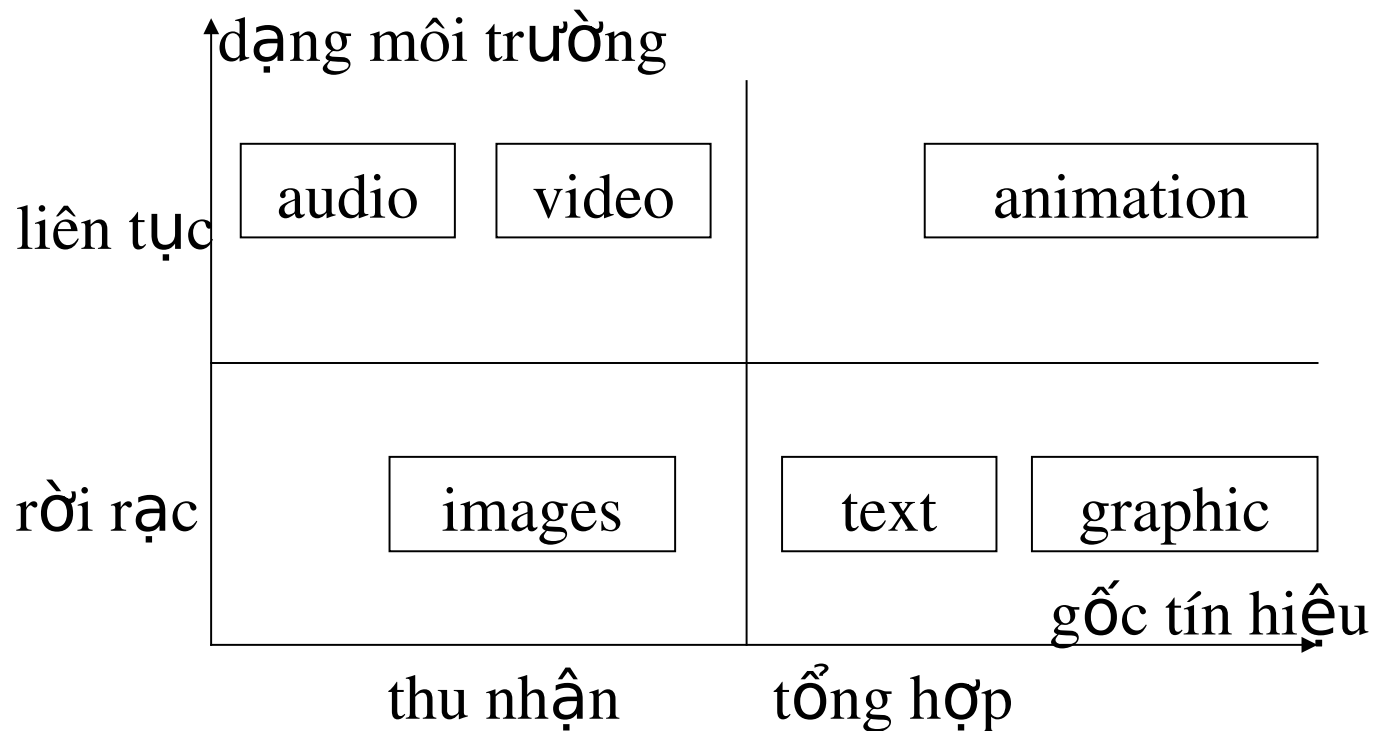
# TỔNG QUAN VỀ MULTIMEDIA



Hình 1-3 Thu nhận và tổng hợp

# TỔNG QUAN VỀ MULTIMEDIA

- Các dạng môi trường và tín hiệu:



Hình 1-2 Dạng môi trường

# TỔNG QUAN VỀ MULTIMEDIA



- **Âm thanh (audio)**
- **Âm thanh:** Là dao động sóng âm gây ra áp lực làm dịch chuyển các hạt vật chất trong môi trường đàn hồi làm tai người cảm nhận được các dao động này. Tai người có thể nghe được trong khoảng tần số từ 20Hz đến 20kHz.
- **Âm thanh tự nhiên:** Là sự kết hợp phức giữa các sóng âm có tần số và dạng sóng khác nhau.
- **Dải động của tai:** Giới hạn bởi ngưỡng nghe thấy (0dB) đến ngưỡng đau (120dB) của người.
- **Ngưỡng nghe tối thiểu:** Mức thấp nhất của biên độ mà tai người có thể cảm nhận được âm thanh tùy thuộc vào từng người, mức áp lực và tần số của âm thanh.

# TỔNG QUAN VỀ MULTIMEDIA

- **Hiệu ứng che khuất âm thanh:** Hiện tượng âm thanh mà tại đó ngưỡng nghe của một âm tăng lên trong khi có mặt của một âm khác (khó nghe hơn). Được sử dụng trong kỹ thuật nén.
- **Hướng âm thanh:** Tai và não có thể giúp ta xác định hướng âm thanh, điều này có thể ứng dụng để tạo các hiệu ứng âm thanh như stereo, surround.
- **Vang và trễ:** Vang là hiện tượng kéo dài âm thanh sau khi nguồn âm đã tắt. Trễ là thời gian âm thanh phản xạ đến đích so với âm thanh trực tiếp. Nếu  $>50\text{ms}$  thì trễ đó gọi là tiếng vọng. Biên độ của âm thanh cứ sau 1 lần phản xạ thì bị suy giảm.
- **Âm nhạc:** Là âm thanh có chu kỳ ở những tần số mà tai người cảm nhận một cách dễ chịu, êm ái, được kết hợp một cách phù hợp. Âm nhạc gồm cao độ, âm sắc và nhịp điệu.

# TỔNG QUAN VỀ MULTIMEDIA



- **Video**

- **Tín hiệu video:** Là sự tái tạo ảnh tự nhiên với những khoảng cách về không gian, thời gian hoặc cả hai.
- **Ảnh tự nhiên:** được tạo nên từ các nguồn sáng mặt trời hay ánh sáng nhân tạo phản xạ lên các vật thể mà ta có thể nhìn thấy được.
- **Ảnh:** Là một ma trận các điểm ảnh mang thông tin về độ chói và màu sắc.
- **Sự lưu ảnh:** Khả năng lưu hình của mắt trong một giây. Mắt có thể lưu được 24 hình trong một giây. Chọn số hình trong một giây của ảnh động phù hợp

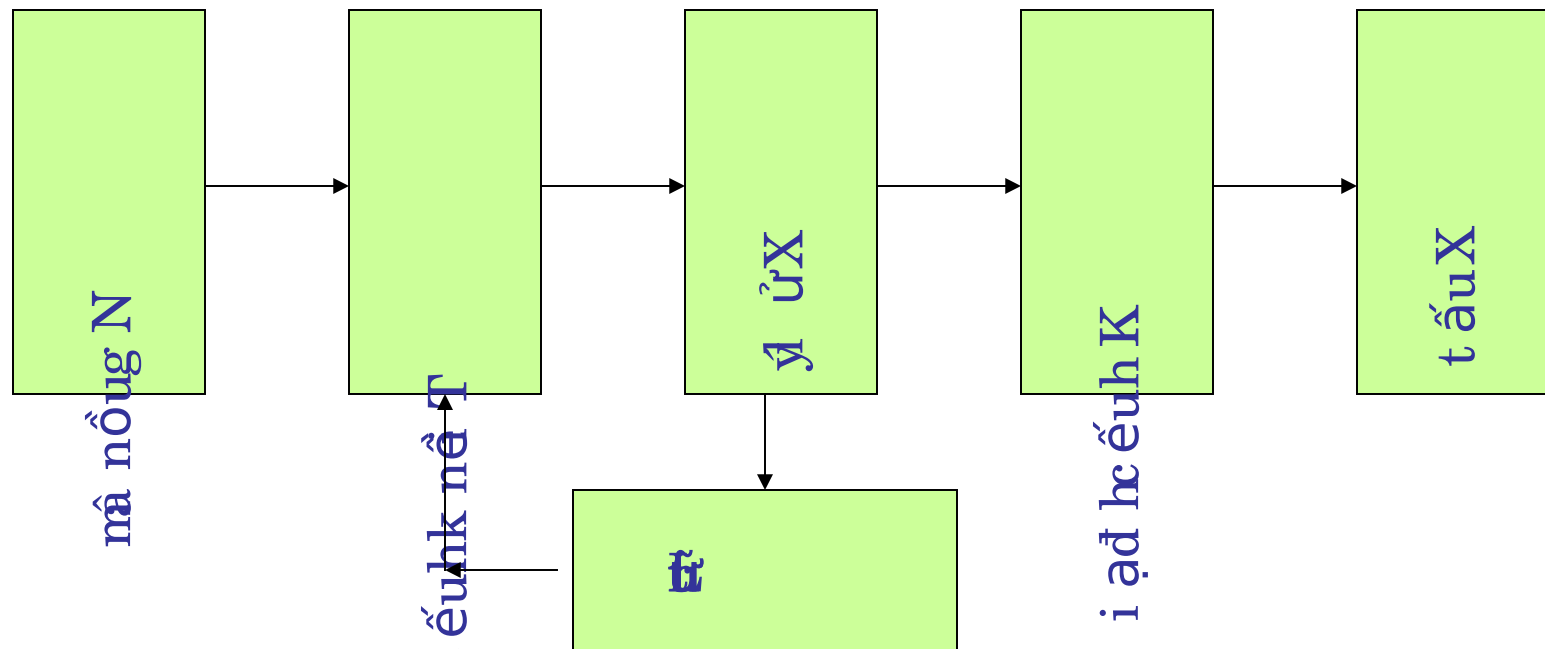


# TỔNG QUAN VỀ MULTIMEDIA

- **Độ chói:** Là biên độ của thành phần trong ảnh (pixel).
  - Ví dụ tín hiệu chói Y được tổng hợp bởi các tín hiệu RGB theo công thức:
    - $E_Y = 0,299E_R + 0,587E_G + 0,114E_B$  (1-2)
  - Thông tin màu được xác định:
    - $E_B - E_Y = 0,587E_G + 0,889E_B + 0,229E_R$
    - $E_R - E_Y = 0,587E_G + 0,114E_B + 0,701E_R$  (1-3)
- **Độ tương phản:** Tỷ số của độ chói thành phần sáng nhất so với độ chói của thành phần tối nhất.

# HỆ THỐNG AUDIO-VIDEO

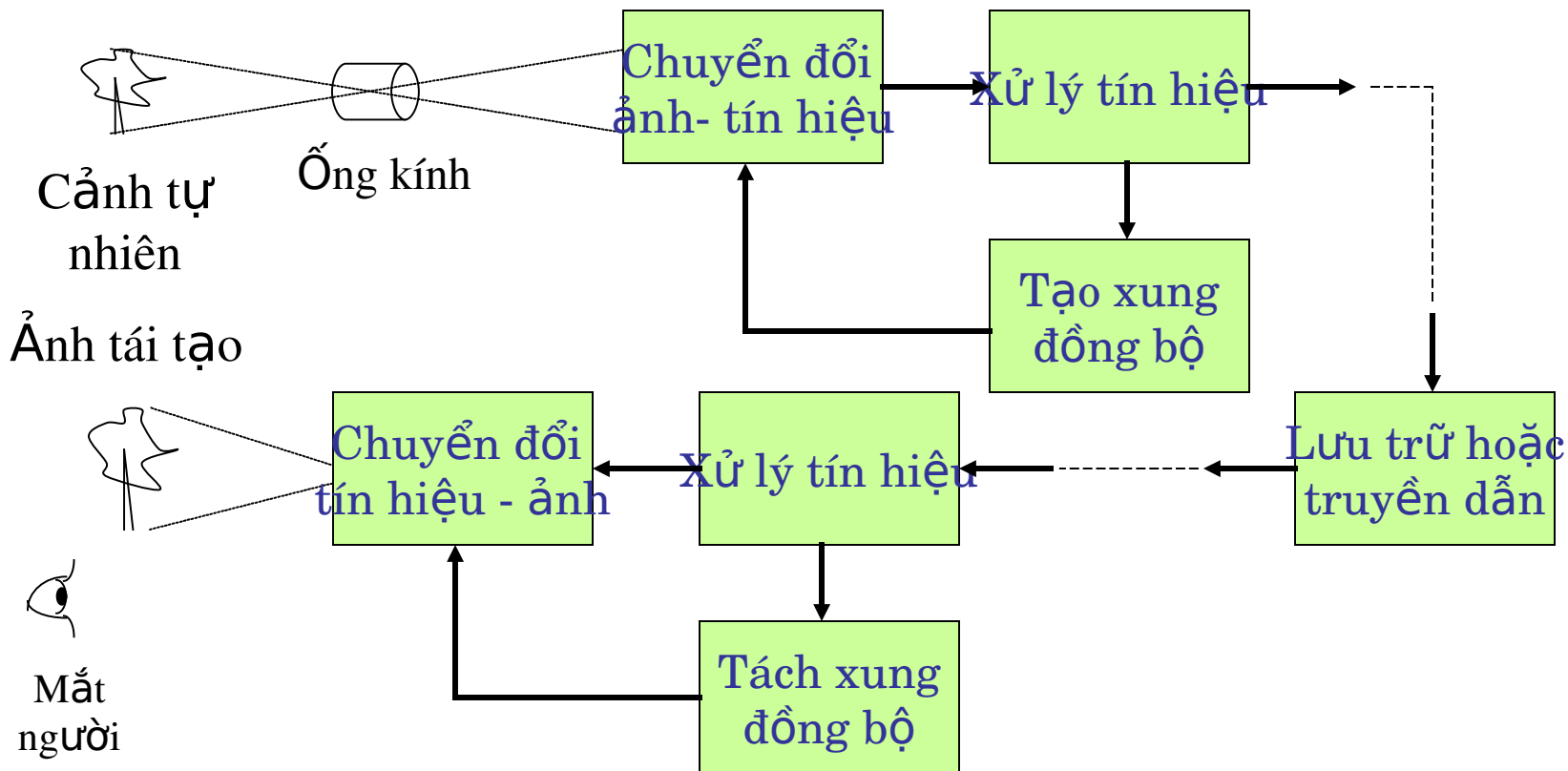
- Hệ thống audio tương tự



Hình 1-4 Hệ thống audio tương tự

# HỆ THỐNG AUDIO-VIDEO

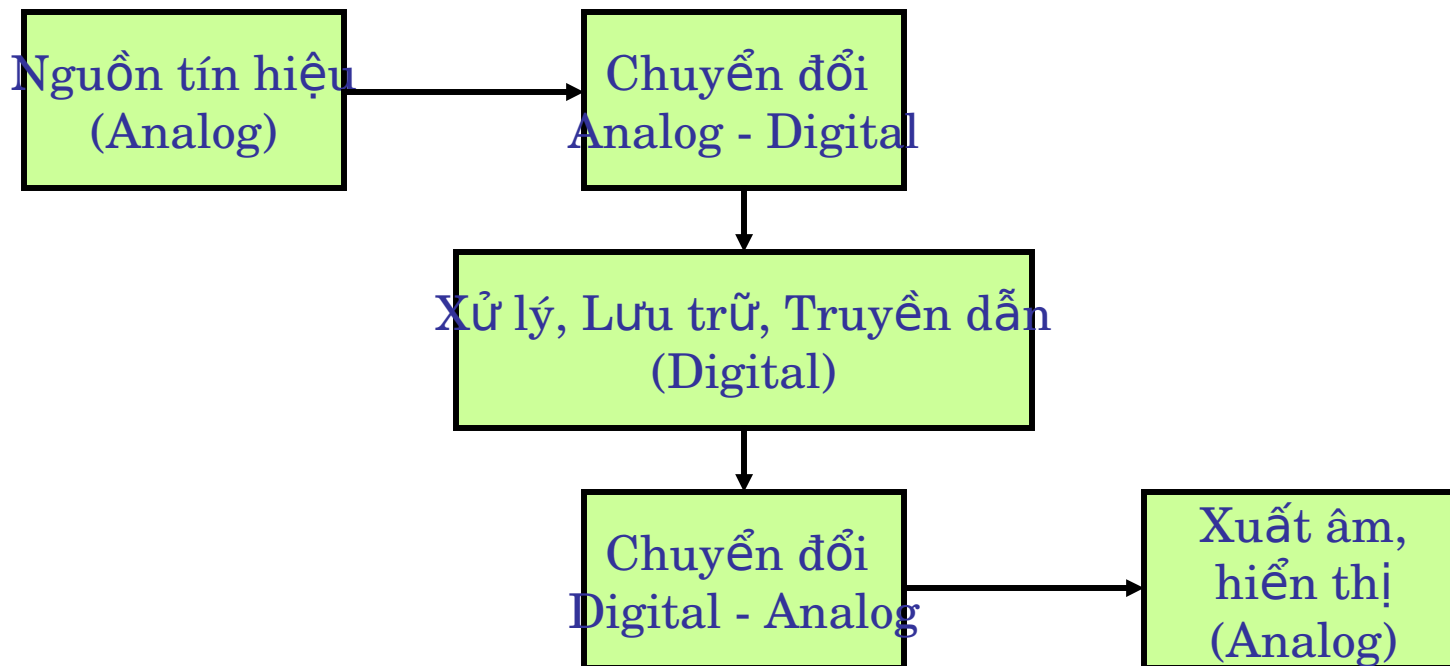
- Hệ thống video tương tự



Hình 1-5 Hệ thống Video tương tự

# HỆ THỐNG AUDIO-VIDEO

- Hệ thống audio-video số:











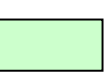




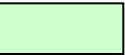
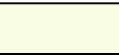



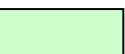
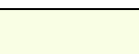



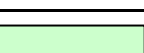
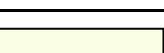

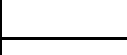
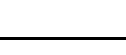
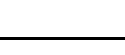
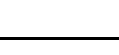



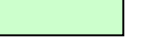
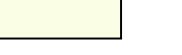





Hình 1-6 Hệ thống audio-video số

# HỆ THỐNG AUDIO-VIDEO

- **Các thành phần của hệ thống:**
  - **Bộ phận thu:** Micro và Camera thu và chuyển tín hiệu (âm thanh hoặc ảnh) sang tín hiệu điện tương tự. Đối với các hệ thống số phải thực hiện việc chuyển đổi tương tự sang số.
  - **Lưu trữ:** Thiết bị lưu trữ là băng từ hoặc đĩa từ. Có thể là các thiết bị riêng biệt sử dụng với mục đích thuận tiện và yêu cầu một chất lượng nào đó.
  - **Xử lý tín hiệu:** Điều chỉnh đặc tuyến tần số, màu sắc, tạo hiệu ứng..
  - **Truyền dẫn:** Truyền tín hiệu từ vị trí này sang vị trí khác với một khoảng cách không gian nào đó qua một môi trường truyền dẫn nào đó.

# HỆ THỐNG AUDIO-VIDEO

- Phân loại các hệ thống Audio-Video:

Lớp hệ thống	Giá	Chất lượng	Dễ sử dụng	Linh hoạt	Khả năng mở rộng	Yếu tố quan trọng nhất
A-V gia đình						Giá, dễ sử dụng
Bán chuyên nghiệp						Chất lượng, giá
Cầu hội thảo						Nén video
Phân phối video						Khả năng lưu trữ
Sản xuất studio						Chất lượng, linh hoạt
Sản xuất chương trình ngoài trời						Cơ động, dễ sử dụng
Sản xuất hậu kỳ						Chất lượng, linh hoạt
Máy tính cá nhân						Toàn bộ

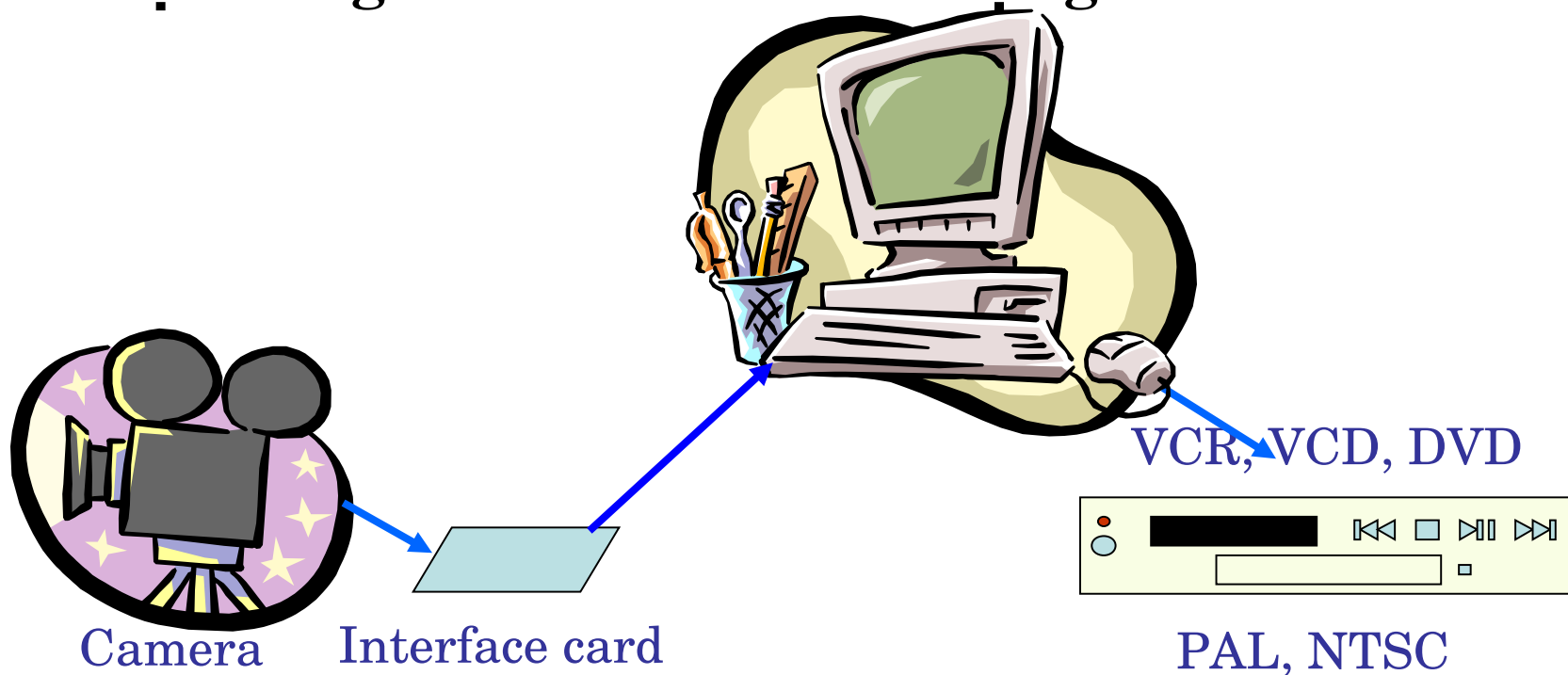
# HỆ THỐNG AUDIO-VIDEO



- **Hệ thống Audio – Video dân dụng:**
  - Xây dựng hoặc tạo lại một số chương trình nhất định
  - Ghi, lưu trữ những sự kiện cá nhân.
  - Hầu hết các chương trình được thu và tạo ra tại chỗ.
  - Hệ thống đáp ứng nhu cầu giá thành thấp, dễ sử dụng để phổ biến rộng rãi.
  - Sử dụng phương pháp sản xuất hậu kỳ với chất lượng giới hạn nhất định.
  - Đa hệ và tương thích với mọi tiêu chuẩn.
  - Mỗi quan tâm của các nhà sản xuất.

# HỆ THỐNG AUDIO-VIDEO

- Hệ thống Audio-Video dân dụng

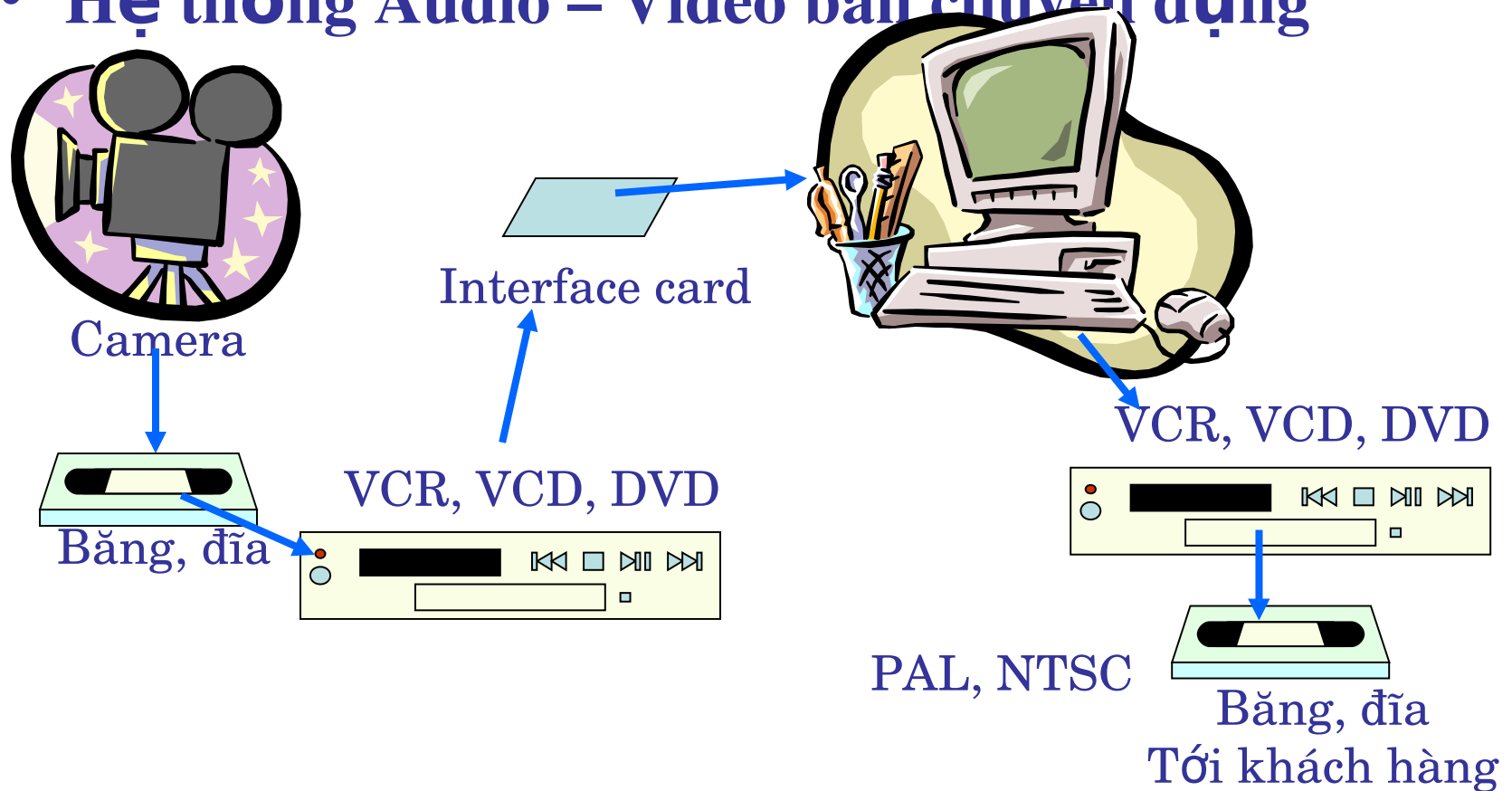


Hình 1-7 Hệ thống Audio – Video dân dụng



# HỆ THỐNG AUDIO-VIDEO

- Hệ thống Audio – Video bán chuyên dụng



Hình 1-8 Hệ thống Audio – Video bán chuyên dụng Page 25

# HỆ THỐNG AUDIO-VIDEO

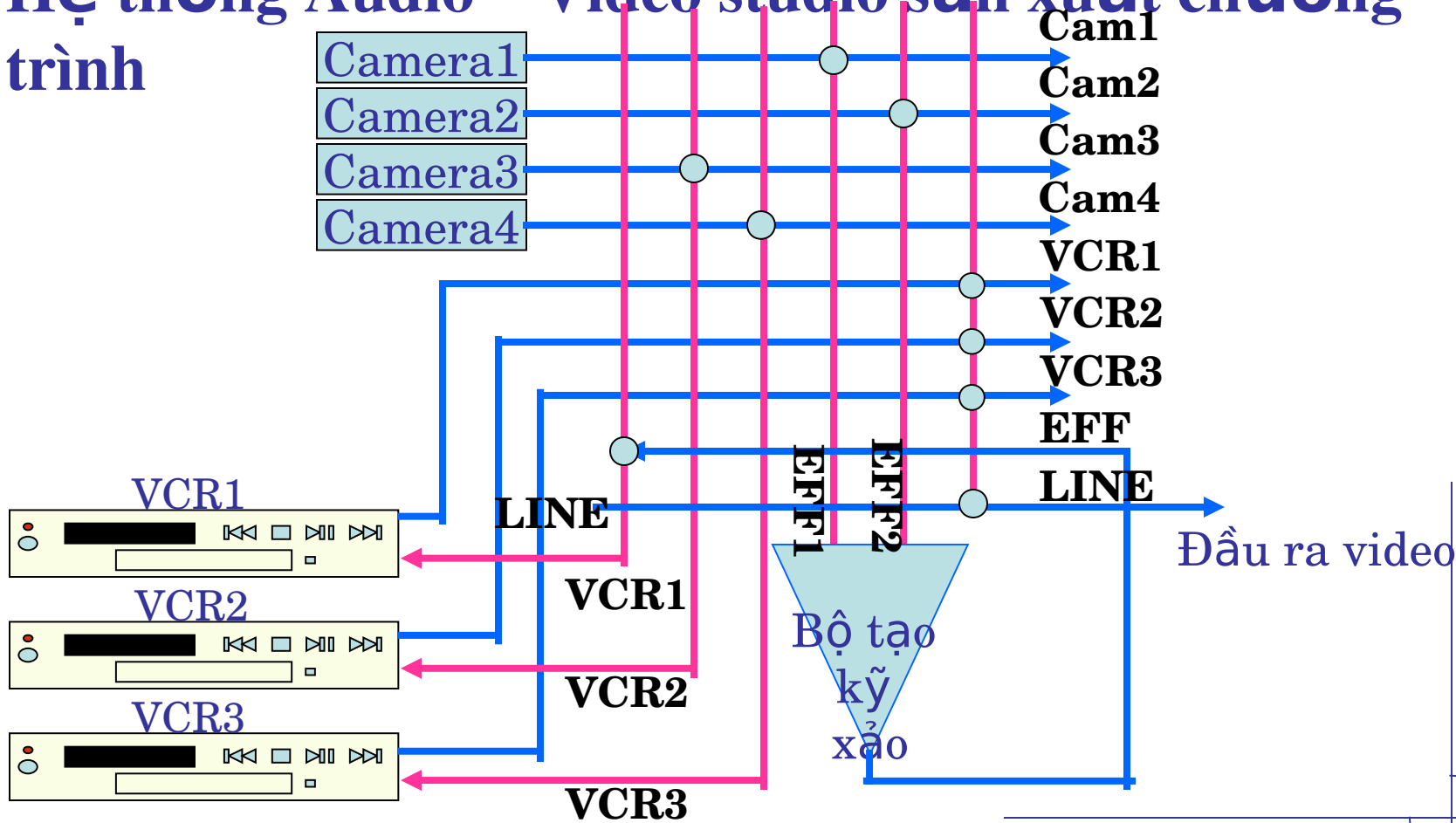


- **Hệ thống phân phối:**

- Tập hợp chương trình thành một dòng dữ liệu để phát quảng bá, truyền hình cáp hay vệ tinh.
- Khả năng chuyển tải đến người xem thông qua máy phát, mạng hay một phương thức nào đó.
- Máy chủ phải đáp ứng khả năng lưu trữ đối với tín hiệu nhằm tạo đường truyền thông suốt giữa các chương trình.
- Yêu cầu tự động cao, giảm chi phí nhân công.
- Truyền hình tương tác yêu cầu khả năng xử lý và chất lượng đường truyền khá cao, đồng thời hệ thống phải có khả năng phát các chương trình khác nhau trong cùng thời điểm.

# HỆ THỐNG AUDIO-VIDEO

- Hệ thống Audio – Video studio sản xuất chương trình



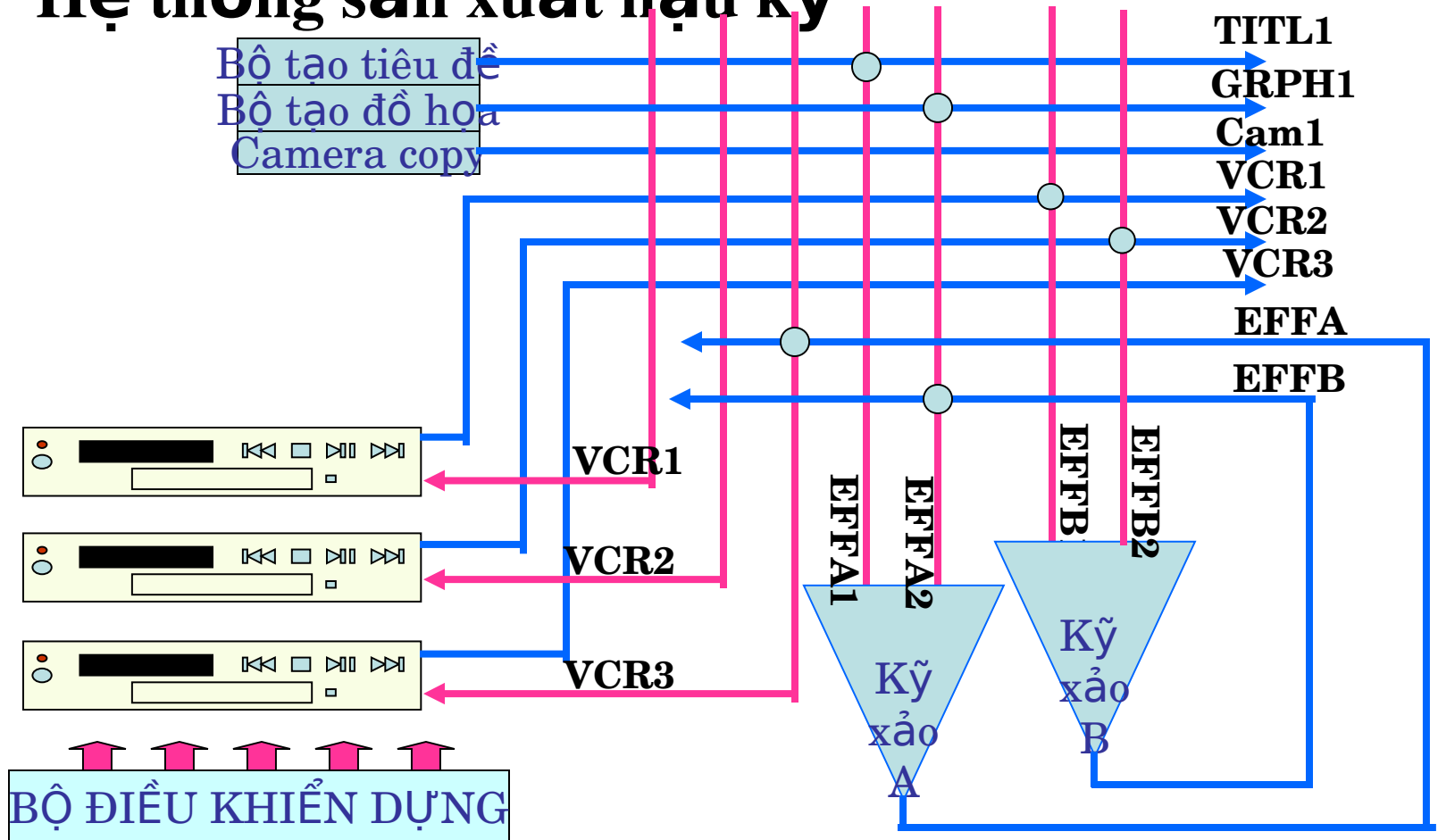
Hình 1-9 Hệ thống Audio – Video studio sản xuất chương trình

# HỆ THỐNG AUDIO-VIDEO

- **Hệ thống sản xuất chương trình ngoài trời:**
  - Được sử dụng để thu các bản tin hay một chương trình nào đó mà không cần nhiều người thực hiện, thường sử dụng các thiết bị cầm tay.
  - Các chương trình truyền trực tiếp thì hệ thống có thể là các hệ thống cố định nhưng với quy mô nhỏ và chất lượng thấp hơn.
  - Yêu cầu tính cơ động cao.
  - Camera được nối với máy ghi riêng mà không sử dụng ma trận chuyển mạch.
  - Máy ghi âm đa đường được sử dụng để thuận tiện trong hậu kỳ âm thanh nhưng phải yêu cầu đồng bộ với hình.

# HỆ THỐNG AUDIO-VIDEO

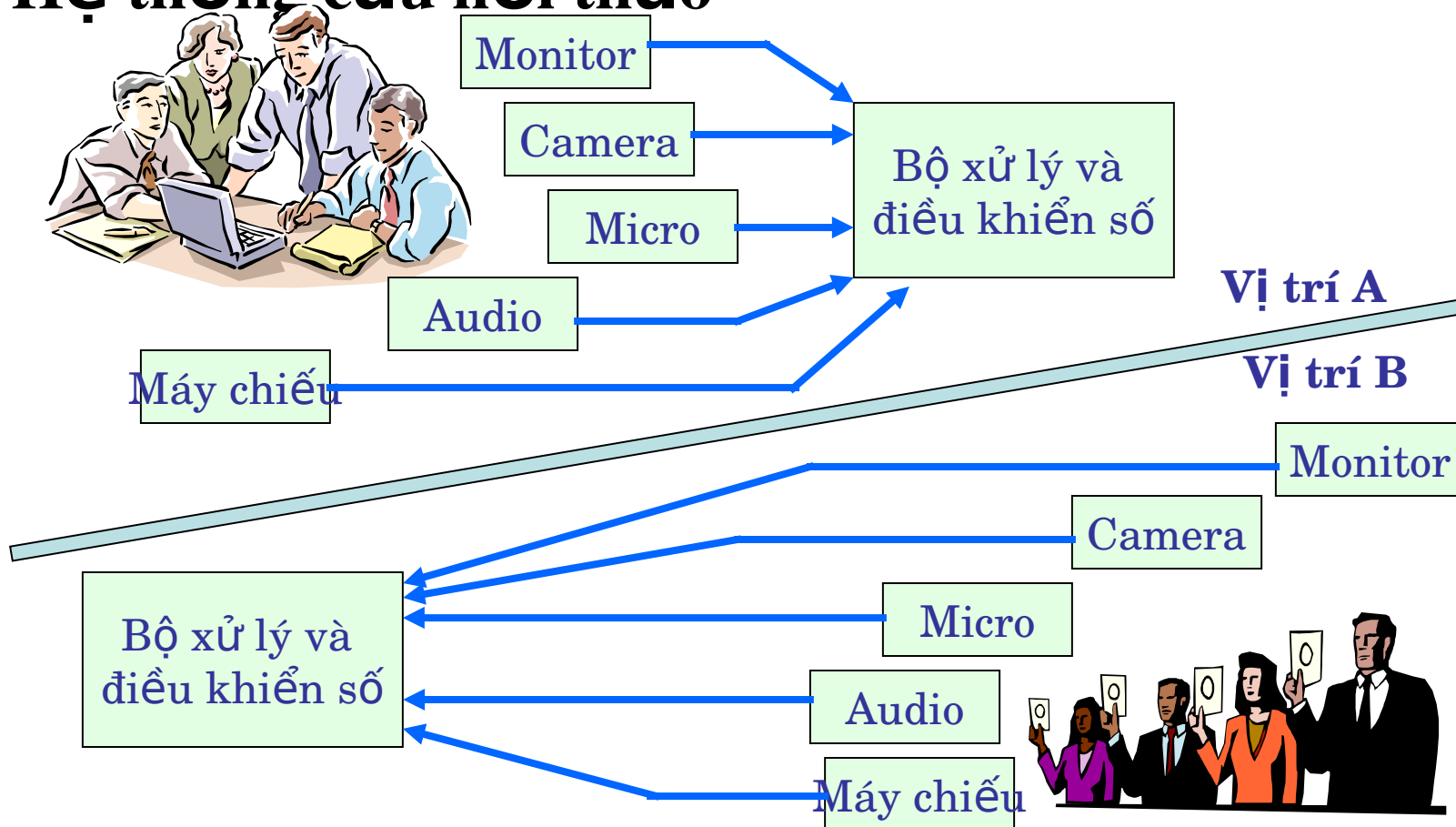
- Hệ thống sản xuất hậu kỳ**



Hình 1-10 Hệ thống Audio - Video sản xuất hậu kỳ Page 29

# HỆ THỐNG AUDIO-VIDEO

- Hệ thống cầu hỏi thảo



Hình 1-11 Hệ thống cầu hỏi thảo

# HỆ THỐNG AUDIO-VIDEO



- **Hệ thống audio video trong PC**

- PC dùng để trình diễn, lưu trữ, xử lý âm thanh, hình ảnh.
- Điều khiển bằng phần mềm chuyên dụng kết hợp với các card đồ họa, xử lý kỹ xảo.
- Đa dạng về tiêu chuẩn dẫn đến khó tương thích.
- Có thể yêu cầu nhiều dạng card thích ứng khác nhau và có thể sử dụng hơn một màn hình để hiển thị.
- Dữ liệu có thể yêu cầu nén và giải nén vì phạm vi ứng dụng khá rộng.

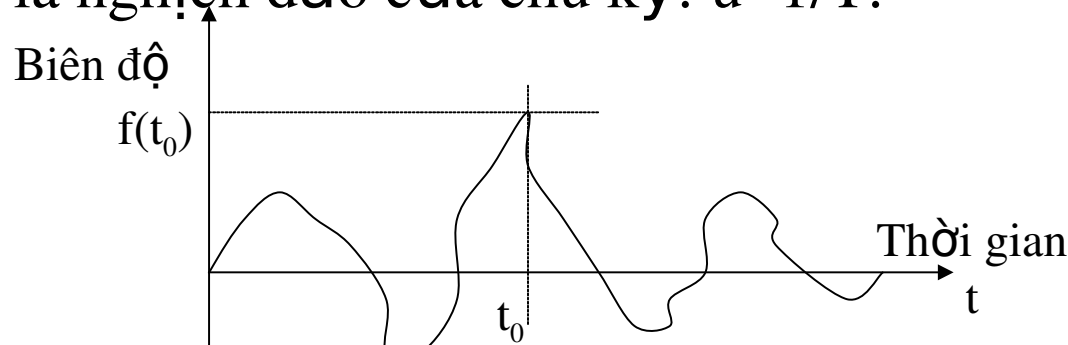
# MỘT SỐ VẤN ĐỀ VỀ TÍN HIỆU

## • Tín hiệu và hàm

- Tín hiệu tương tự là hàm theo thời gian.
- Biên độ âm thanh được biểu diễn bằng mức độ âm thanh tại thời điểm đã cho.
- Tín hiệu được biểu diễn bằng hàm  $f(t)$ .

## • Tín hiệu có chu kỳ

- Sự lặp lại trong một khoảng thời gian ngắn nhất không đổi của tín hiệu gọi là chu kỳ  $T$ .
- Tần số là nghịch đảo của chu kỳ:  $u=1/T$ .



Hình 1-7 Biểu diễn biên độ-thời gian



# MỘT SỐ VẤN ĐỀ VỀ TÍN HIỆU

- **Phân tích Fourier**

- Trong thực tế, rất ít khi ta có được một tín hiệu đơn tần, mà thông thường là các tín hiệu phức tạp, kết hợp bởi nhiều tần số và các hài của nó.
- Việc phân tích Fourier cho kết quả là tổng của các hàm sin và cosin của các tần số khác nhau.

- **Phân tích Fourier một chiều:**

$$F(u) = \int_{-\infty}^{\infty} f(t)e^{-j2\pi ut} dt$$

$$F(u) = F_R(u) + jF_I(u)$$

$$f(t) = \int_{-\infty}^{\infty} F(u)e^{j2\pi ut} du$$

$$|F(u)| = \sqrt{F_R^2(u) + F_I^2(u)}$$

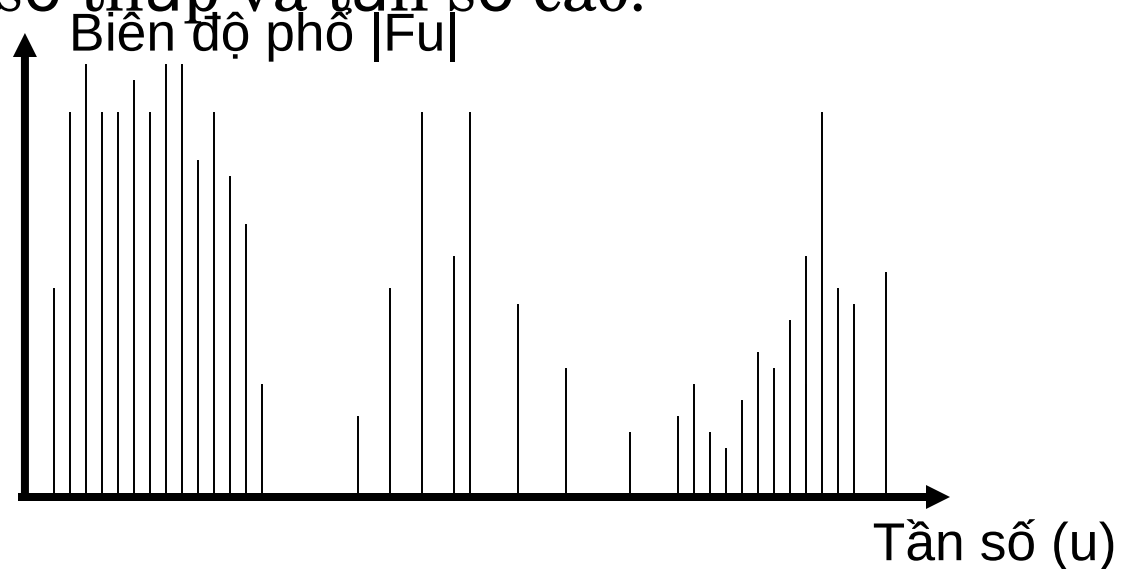
$$\angle F(u) = \arctan\left(\frac{F_I(u)}{F_R(u)}\right)$$

$$F(u) = F_R(u) + jF_I(u) = |F(u)|e^{j\angle F(u)}$$

# MỘT SỐ VẤN ĐỀ VỀ TÍN HIỆU

## • Phổ tần số

- Sự phân bố của  $|F(u)|$  gọi là phổ tần của tín hiệu.
- Tín hiệu biến thiên chậm thì phổ tần tập trung ở tần số thấp và ngược lại. Từ đó hình thành tín hiệu tần số thấp và tần số cao.



# MỘT SỐ VẤN ĐỀ VỀ TÍN HIỆU

## • Tín hiệu Audio và Video

- Tín hiệu âm thanh thường là tín hiệu một chiều.
- Tín hiệu ảnh là tín hiệu hai chiều.
- Tín hiệu Video là tín hiệu 3 chiều.
- Với các chiều khác nhau, ta sẽ có số biến khác nhau tương ứng.

## • Chuyển đổi Fourier 2 chiều

$$F(u, v) = \int \int f(x, y) e^{-j2\pi(ux + vy)} dx dy$$

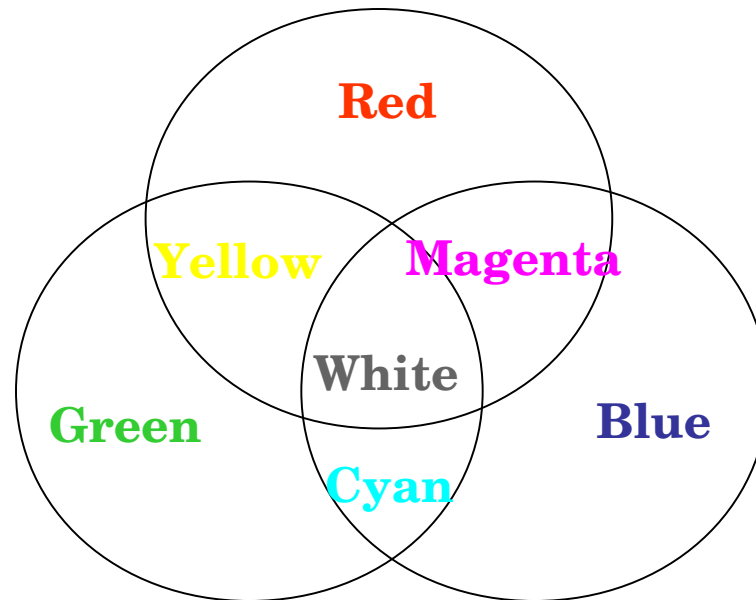
$$= \int e^{-j2\pi ux} dx \int e^{-j2\pi vy} dy$$

$$f(x, y) = \int \int F(u, v) e^{j2\pi(ux + vy)} du dv$$

# MỘT SỐ VẤN ĐỀ VỀ TÍN HIỆU

- **Màu sắc**

- Việc kết hợp các màu khác nhau tạo nên một màu mới. Thông thường, chọn các màu cơ bản để kết hợp, ví dụ RGB

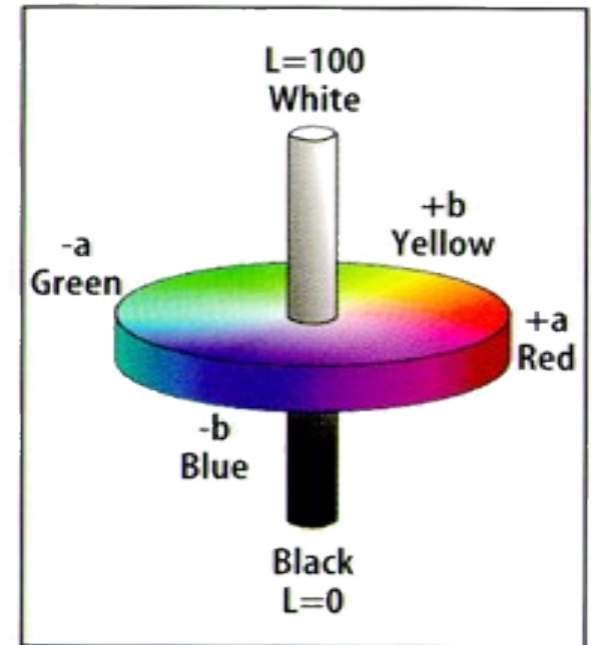


*Hình 1-8 Lý thuyết 3 màu RGB*

# MỘT SỐ VẤN ĐỀ VỀ TÍN HIỆU

- **Không gian cảm quan màu 3 chiều:**

- Con người cảm quan màu sắc ở các khía cạnh sau:
  - brightness: độ sáng như thế nào.
  - hue: màu nào.
  - saturation: sự tinh khiết
- Sự cảm quan này đối với mỗi người là mỗi khác biệt, do đó, không thể so được giữa người này với người kia.



*Lab model*

*Hình 1-9 Cảm quan 3 chiều*

# NÉN DỮ LIỆU

- Đại lượng đo thông tin

- Lượng thông tin trong tín hiệu có thể không bằng lượng dữ liệu của nó mà quan hệ mật thiết với xác suất xuất hiện của nó.

- Tự-thông tin (lượng tin)

- Thông tin được mang bởi một biến cố A có xác suất xuất hiện  $P[A]$

là: 
$$I_A = \log_2 \frac{1}{P[A]} = -\log_2 P[A]$$

- Thông tin không (lượng tin = 0):

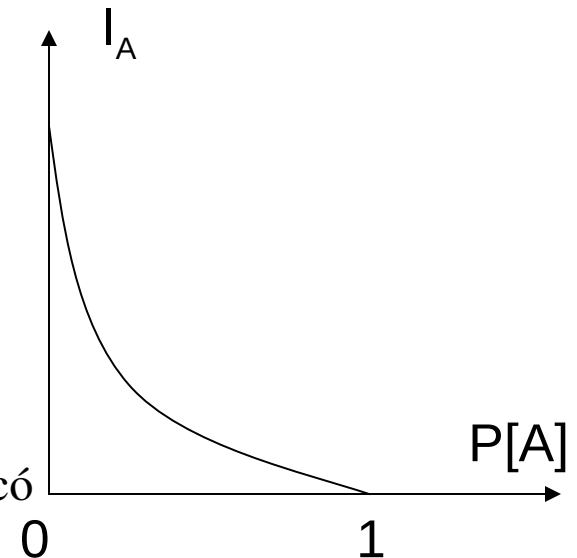
- Mặt trời mọc đằng đông.

- Lượng tin ít

- Máy điện thoại di động trong tương lai đều có khả năng multimedia

- Lượng tin nhiều:

- Trường ĐHK ĐN được xếp hạng nhất trên thế giới về ĐTVT



# NÉN DỮ LIỆU

- **Entropy**

- Lượng tin trung bình của nguồn tin, một cách gần đúng, là số bit trung bình của thông tin yêu cầu để biểu diễn các ký hiệu của nguồn tin.
- Với nguồn N ký hiệu  $X_i$  thì entropy được định nghĩa như sau:

$$H(S) = - \sum_{i=1}^N P[X_i] \log_2 P[X_i]$$

- $H(S) \geq 0$ ; đối với mã hoá nhị phân,  $H(S)$  thể hiện mã hoá với số bit/ký hiệu tối thiểu.
- Ví dụ:

Trong một ảnh phân bố đều ở thang xám (256 mức):  $p_i = 1/256$ , số bit mã hoá cho mức xám là  $\log_2 256 = 8$  bits.  $\rightarrow$  Entropy của ảnh này là

$$H(S) = - \sum p_i \log_2 (1/p_i) = 8 \text{ bits/ký hiệu.}$$

Vậy, trong trường hợp phân bố đều này, mã hoá độ dài cố định sẽ đạt được số bit tối thiểu. Trong trường hợp tổng quan thì mã hóa độ dài cố định sẽ không hiệu quả.

# NÉN DỮ LIỆU

- Mã hoá độ dài cố định FLC (Fixed-Length Code)
  - Đặc điểm:
    - Sử dụng số bit cố định để biểu diễn mọi ký hiệu của nguồn.
    - Đơn giản trong quá trình mã hoá/giải mã.
  - Ví dụ
    - Mã ASCII (American Standard Code for Information Interchange) sử dụng 8 bits để mã hoá các ký tự.
      - Truyền chuỗi: DTVT: 68 84 86 84: 01101000 10000100 10000110 10000100
  - Nhược điểm:
    - Không hiệu quả



# NÉN DỮ LIỆU

- Mã hoá độ dài thay đổi VLC (Variable-Length Code)
  - Đặc điểm
    - Sử dụng số bit khác nhau để biểu diễn các ký tự khác nhau.
    - Các ký tự có xác suất xuất hiện cao được phân bố bởi từ mã ngắn và ngược lại.
    - Hiệu quả trong việc biểu diễn hơn, nén tốt hơn.
  - Ví dụ:
    - Mã Morse.
    - Shannon-Fano.
    - Huffman.
    - Mã hoá loạt dài (RLC).

# NÉN DỮ LIỆU

- Thuật toán Shannon-Fano
  - Ví dụ mô tả thuật toán:

Ký hiệu	A	B	C	D	E
Số lần xuất hiện	15	7	6	5	6

- Mã hoá theo thuật toán Shannon-Fano:
  - Sắp xếp các ký tự theo thứ tự giảm dần của tần suất xuất hiện.
  - Tính xác suất.
  - Đệ quy làm hai phần, mỗi phần có tổng xác suất gần bằng nhau. Mã hoá phần trên bằng bit 0 (hoặc bit 1), phần dưới bằng bit 1 (hoặc bit 0).
  - Vẽ sơ đồ cây.
  - Tính Entropy, số bits mã hoá trung bình và số bit mã hoá thông thường.
  - Nhận xét.

# NÉN DỮ LIỆU

Ký hiệu	Đếm	$P_i$	$\log_2(1/p_i)$	Mã		Tổng bits
A	15	15/39	1.38	0	0	30
B	7	7/39	2.48	0	1	14
C	6	6/39	2.7	1	0	12
E	6	6/39	2.7	1	1	18
D	5	5/39	2.96	1	1	15

- Entropy của nguồn:

$$H(S) = \sum_{i=A} p_i \log_2 \frac{1}{p_i} = \frac{15}{39} \cdot 1.38 + \frac{7}{39} \cdot 2.48 + \frac{6}{39} \cdot 2.7 + \frac{6}{39} \cdot 2.7 + \frac{5}{39} \cdot 2.96$$

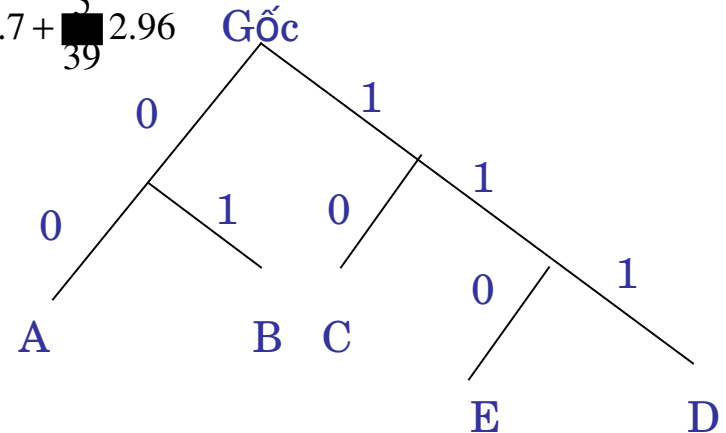
$$H(s) = 2.19.$$

- Số bits sử dụng trung bình:  
(tổng bits / số lần xuất hiện)

$$K = \frac{30+14+12+18+15}{39}; 2.28bits /symbol$$

- Số bit mã hoá thông thường:  
 $\log_2 5 = 3bits/symbol$

- Nhận xét: Số bits sử dụng trung bình gần  $H(S)$  thì bộ mã càng hiệu quả.



# NÉN DỮ LIỆU



- Mã hoá Huffman

- Nguyên tắc:

- Dựa vào mô hình thống kê của dữ liệu gốc, ký tự có xác suất càng cao thì mã hoá với từ mã càng ngắn.

- Thuật toán:

- Tính tần suất xuất hiện trong dữ liệu gốc, sắp xếp theo thứ tự giảm dần.
    - Xét từ dưới lên trên, bắt đầu từ hai ký tự có xác suất bé nhất, quy định mỗi nhánh là 0 (hoặc 1) hợp lại với nhau thành nút có xác suất bằng tổng hai xác suất hợp thành.
    - Lặp lại cho đến hết.

# NÉN DỮ LIỆU

– Xét ví dụ trên

Ký hiệu	Xác suất		Mã	Tổng bit
A	15/39		1	15
B	7/39	0	000	21
C	6/39	1 0 13/39	001	18
E	6/39	0 0 24/39	010	18
D	5/39	1 1 11/39	011	15

- Số bit trung bình:  $87/39=2.23$  /SYMBOL(<2.28)
- Hiệu quả hơn Shannon – Fano.

# NÉN DỮ LIỆU

- Mã hoá loạt dài RLC (Run-Length Coding)

- Nguyên lý

- Mã hoá loạt ký hiệu bằng chiều dài và ký hiệu của loạt đó.

- Đặc điểm

- Mã hoá không tổn hao
- Mã hoá liên ký tự.
- Hiệu quả với một số nguồn tín hiệu, nhất là sau phép chuyển đổi.

- Ví dụ

• 00000000000000000300000000000015: (11,13)3 (9,13)15

• mã: (11,3) (9,15)

Run-length  
Loạt 0 dài

Bit thô

side

số bit cần mã hoá

# NÉN DỮ LIỆU

- Mã hoá Lempel-Zip-Welch:
  - Nén từ điển được Jacob Lampel và Abraham Ziv đề xuất năm 1977, phát triển thành họ LZ, LZ77, LZ78.
  - Năm 1984, Terry Welch cải tiến thành LZW.
  - Nguyên tắc: Dựa vào việc xây dựng một từ điển lưu các chuỗi ký tự có tần suất cao và thay thế bằng một từ mã mới.
  - LZW tổ chức từ điển tốt hơn nên nâng cao tỷ lệ nén.
  - Ví dụ: Xét từ điển có độ lớn bằng 4096 giá trị từ mã, vậy độ dài lớn nhất của từ mã là 12 bits ( $2^{12}=4096$ ).
  - Xét chuỗi vào ABCBCABCABCD.

0	0
...	...
255	255
256	256   Clear Code
257	257   End of Information
258	Chuỗi mới
...	...
4095	Chuỗi mới
<b>256: Mã xóa CC để khắc phục tình trạng mẫu lặp lớn hơn 4096, nếu mẫu lặp lớn hơn 4096 thì gửi CC để xây dựng từ điển cho phần tiếp theo.</b>	
<b>EoI: Báo hiệu hết một phần nén.</b>	

# NÉN DỮ LIỆU

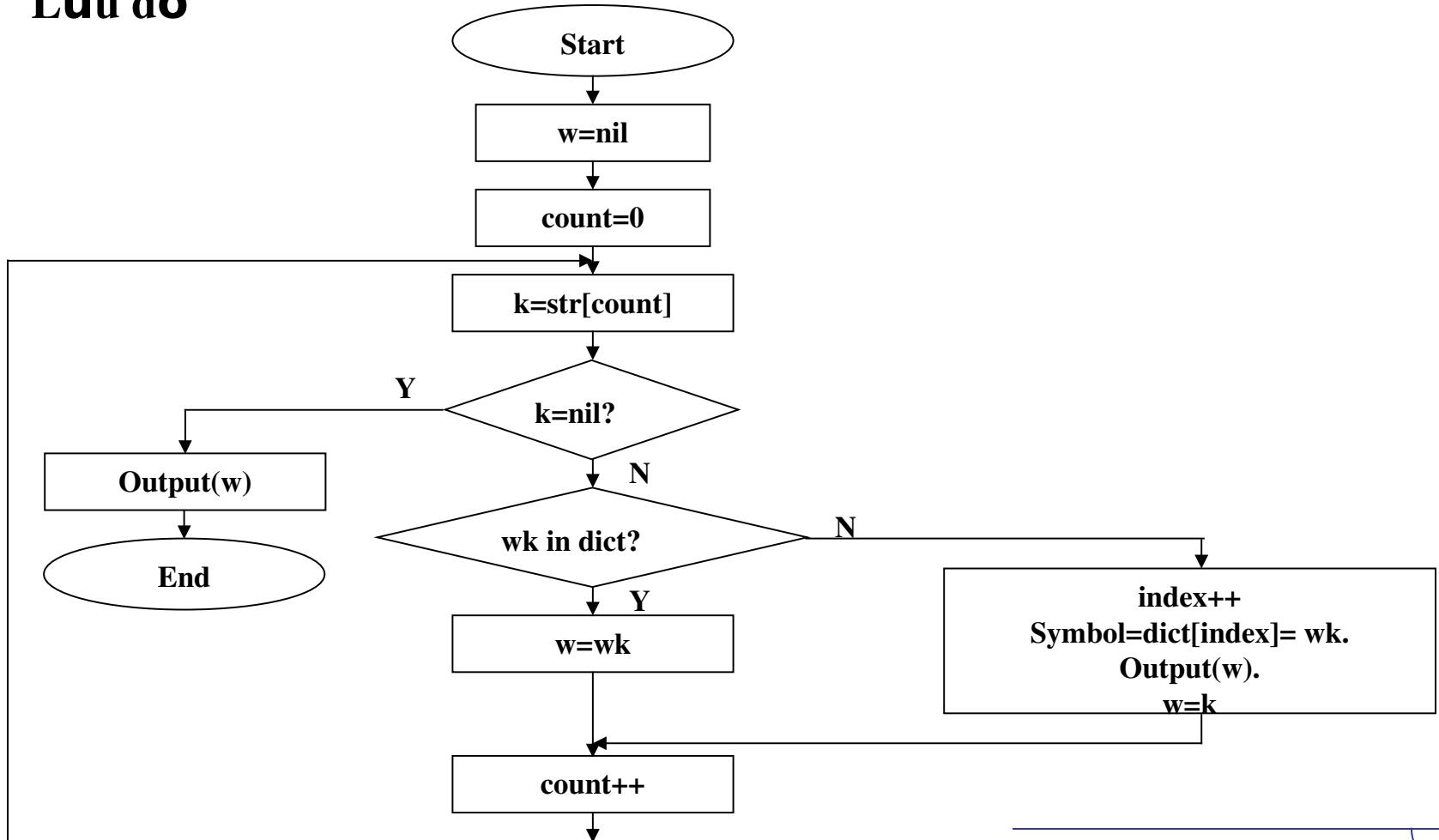
- Thuật toán:
  - $w = \text{NIL}$ ;
  - trong khi đọc được ký tự thứ  $k$  trong chuỗi:
    - nếu  $wk$  đã tồn tại trong từ điển thì  $w = wk$
    - còn không thì thêm  $wk$  vào trong từ điển, mã hoá nó ra cho  $w$ ;  $w = k$ ;
    - $k=k+1$ ;
- Chuỗi ra: 65 66 67 259 258 67 262 68
- Đầu vào  $12 \text{ ktx} 8 \text{ bits} = 96 \text{ bits}$ .
- Đầu ra :  $5 \text{ ktx} 8 + 3 \text{ ktx} 9 = 67 \text{ bits}$ .
- Tỷ lệ nén:  $96/67 = 1.43$

count	w	k	wk	Output	index	Symbol
0	Nil	A	A			
1	A	B	AB	65	258	AB
2	B	C	BC	66	259	BC
3	C	B	CB	67	260	CB
4	B	C	BC			
5	BC	A	BCA	259	261	BCA
6	A	B	AB			
7	AB	C	ABC	258	262	ABC
8	C	A	CA	67	263	CA
9	A	B	AB			
10	AB	C	ABC			
11	ABC	D	ABCD	262	264	ABCD
12	D	nil	D	68		



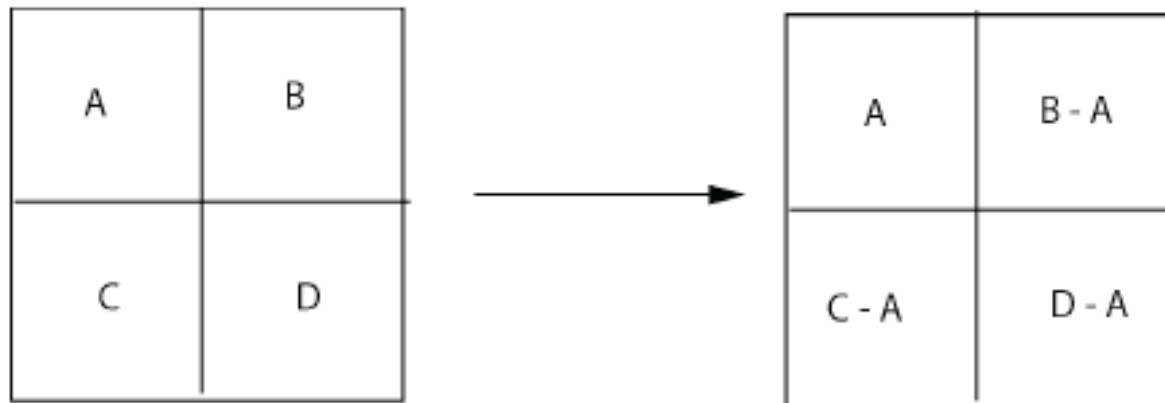
# NÉN DỮ LIỆU

- Lưu đồ



# NÉN DỮ LIỆU

- Mã chuyển đổi đơn giản (Simple Transform Coding)
- Ví dụ: Mã hóa chuyển đổi đơn giản cho block 2x2 pixel đơn sắc với các bước như sau:
  1. Lấy pixel ở góc trên, bên trái làm giá trị gốc cho block (pixel A).
  2. Tính 3 chuyển đổi khác nhau là sự khác biệt giữa các pixel tương ứng và pixel A ( $B-A$ ,  $C-A$ ,  $D-A$ ).
  3. Lưu giá trị gốc và 3 chuyển đổi của các pixel.



# NÉN DỮ LIỆU

- Ta dễ dàng xác định chuyển đổi thuận
  - $X_0=A$ .
  - $X_1=B-A$ .
  - $X_2=C-A$ .
  - $X_3=D-A$ .
- Và chuyển đổi ngược:
  - $A_n=X_0$ .
  - $B_n=X_1+X_0$ .
  - $C_n=X_2+X_0$ .
  - $D_n=X_3+X_0$ .

# NÉN DỮ LIỆU

- **Giảm dư thừa dữ liệu bởi phép chuyển đổi**
  - Độ dư thừa được chuyển đổi thành giá trị Xi.
  - Nén dữ liệu nhờ sử dụng ít bit để biểu diễn sự khác biệt (lượng tử).
- **Ví dụ: Nếu mã hóa đều sử dụng 8 bit để mã hóa cho pixel, ta cần tổng cộng  $8 \times 4 = 32$  bits để mã hóa cho 4 pixel.**
- **Đối với mã hóa chuyển đổi, pixel gốc vẫn dùng 8 bit. Các giá trị chuyển đổi còn lại sử dụng 4 bit để mã hóa. Như vậy, tổng bit dùng là 20bits, số bit mã hóa cho 1 pixel trung bình là 5 bit.**

# NÉN DỮ LIỆU



- **Nhược điểm:**
- **Quá đơn giản.**
- **Cần tổ chức các block lớn (tiêu biểu là 8x8pixel).**
- **Khi mã hóa với sai biệt lớn sẽ là lý do mất mát thông tin.**
- **Ví dụ, đối với các sai biệt, sử dụng 4 bit để mã hóa, nhưng giá trị sai biệt lớn hơn 16 sẽ gây mất mát thông tin.**

# NÉN DỮ LIỆU



- **Mã chuyển đổi sai biệt**
  - Sai biệt được sử dụng trong 1 số thuật toán nén.
    - Nén JPEG.
    - Nén phần tĩnh trong video MPEG.
    - Mã hóa tiếng nói.
    - Hiệu quả trong các đáp ứng liên tục.
    - Kém đối với các tín hiệu có đặc tính biến đổi nhanh.

# NÉN DỮ LIỆU



- **Mã hóa sai biệt:**
  - Mã hóa sai biệt giữa giá trị thực của mẫu và giá trị dự đoán.
  - Còn gọi là mã hóa sai biệt dự đoán hay mã hóa dự đoán.
  - Ví dụ: DPCM, DM và ADPCM.
  - Thích hợp với các dữ liệu liên tiếp không khác nhau nhiều lắm như Video.

# NÉN DỮ LIỆU

- Điều chế xung mã sai biệt (Differential Pulse Code Modulation)
- Dự đoán đơn giản:

$$f_{predict}(t_i) = f_{actual}(t_{i-1})$$

- Mã hóa

$$\Delta f(t_i) = f_{actual}(t_i) - f_{actual}(t_{i-1})$$

- Ví dụ:

- Dữ liệu thật: 9 10 7 6
- Dự đoán: 0 9 10 7
- Mã hóa: +9, +1, -3, -1



# NÉN DỮ LIỆU



- **Điều chế Delta (Delta Modulation)**
  - Là trường hợp của DPCM.
  - Chỉ dùng 1 bit để chỉ thị sự tăng hay giảm của giá trị tiếp theo (ví dụ: 0 tăng, 1 giảm)
  - Không thích hợp đối với các tín hiệu thay đổi liên tục.
- **Điều chế xung mã sai biệt thích nghi ADPCM (Adaptive PCM):**
  - Dữ liệu thực được trừ cho 1 hàm của những giá trị thực trước đó. Ví dụ: giá trị trung bình của những giá trị trước đó.
  - Đặc trưng lấy mẫu tốt hơn.

# Kỹ thuật Audio



# Nội dung



- **GIỚI THIỆU**
- **MÃ HOÁ AUDIO CẢM QUAN**
- **PHÂN TÍCH TÂM LÝ ÂM HỌC**
- **KỸ THUẬT NÉN AUDIO**
- **MPEG-1**
- **MPEG-2**
- **AC-3 (DOLBY DIGITAL)**
- **APT-X100**
- **MÃ HOÁ ÂM THANH NỔI**

# Giới thiệu



- **Âm thanh:**
  - Dạng lan truyền của sóng trong không gian,
  - Sóng âm khi đến tai người nghe, đập vào màng nhĩ, làm cho người đó cảm nhận được sự rung động này
  - Con người có khả năng phân biệt với các âm thanh khác dựa vào một số đặc tính như tần số, nhịp điệu, mức áp lực...
- **Mục đích của các hệ thống audio:**
  - Xử lý, tạo hiệu ứng, nén tín hiệu audio từ tín hiệu thu nhận từ nguồn.
- **Audio số:**
  - Chuỗi các giá trị số được biểu diễn bằng mức âm thanh theo thời gian.

# Ứng dụng

- *Các hệ thống thông tin không dây:*
  - Truyền hình phân giải cao (HighDensity TV HDTV), Âm thanh quảng bá số (Digital Broadcast Audio DBA), Vệ tinh quảng bá trực tiếp (Digital Broadcast Satelite DBS).
- *Các môi trường mạng:*
  - Âm thanh theo yêu cầu (chuyển mạch gói, Internet), Truyền hình cáp (CATV) các tuyến studio.
- *Các ứng dụng đa môi trường:*
  - CD-R, Đĩa đa năng số (DVD).
- *Cinema:*
  - Dolby AC-3 (5 kênh, @384kbps), APT-x100.
- *Lưu trữ khối:*
  - MiniDisc, MP3, MP4.

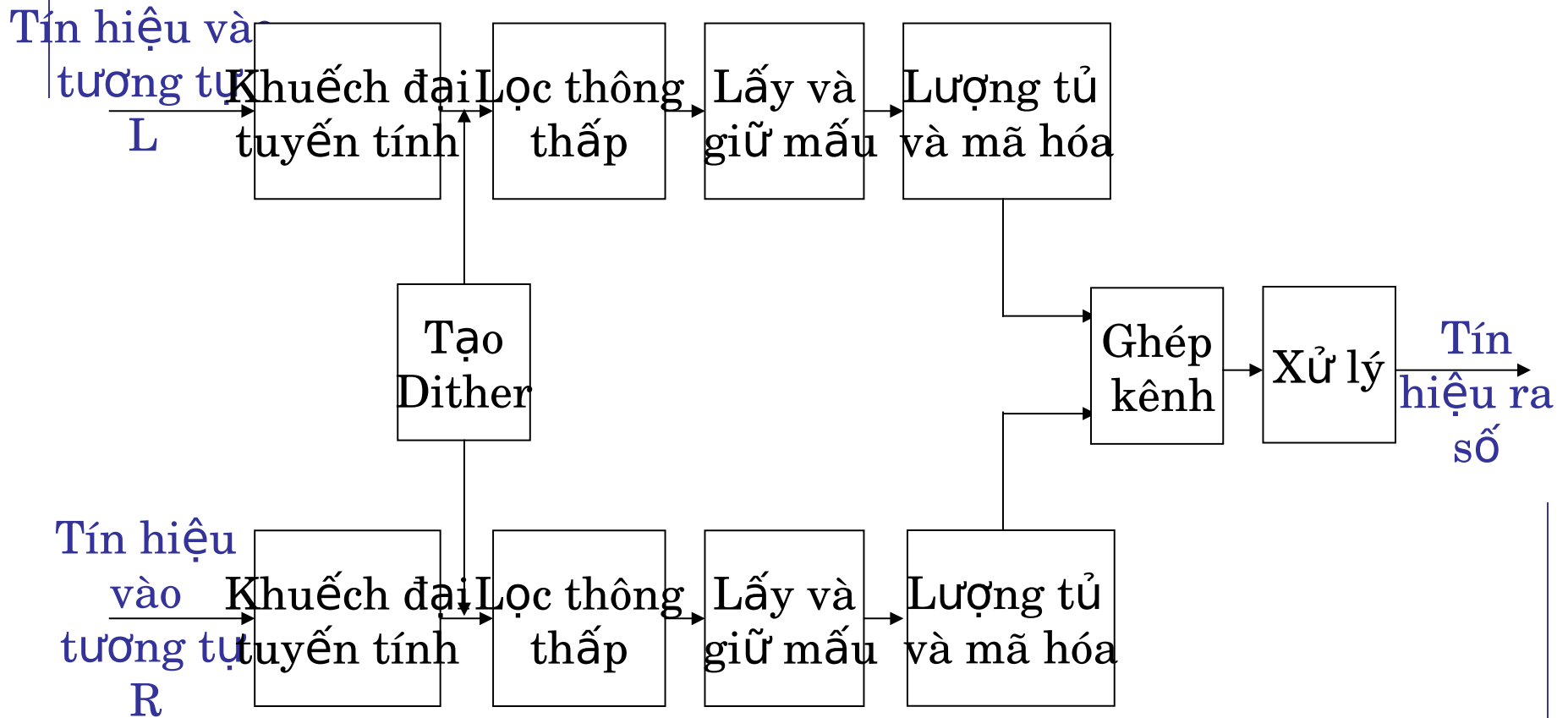
# Audio số



- **Hệ thống audio tương tự và số**

- Hệ thống audio tương tự, việc xử lý tín hiệu gặp phải một số vấn đề như khả năng của linh kiện (về mặt tần số), lưu trữ, phức tạp... từ đó méo phi tuyến cao, SNR bé.
- Hệ thống audio số, việc thu nhận, hiệu chỉnh, xử lý và phát lại trở nên dễ dàng hơn rất nhiều. Các kỹ thuật tổng hợp và nhận dạng phát triển một cách nhanh chóng, tương thích máy tính và con người trở nên phong phú hơn.
- Ưu điểm của audio số là: Độ méo tín hiệu nhỏ (0,01%), Dải động âm thanh lớn gần mức tự nhiên (>90dB), Dải tần rộng hơn (20Hz đến 20kHz), Đáp tuyến tần số bằng phẳng, Cho phép ghi âm nhiều lần mà không giảm chất lượng, Thuận tiện lưu trữ, xử lý, Tăng dung lượng kênh truyền, Khả năng xử lý bằng hệ phi tuyến

# Quá trình thu nhận audio số



Hình 2-1 Quá trình thu nhận âm thanh số

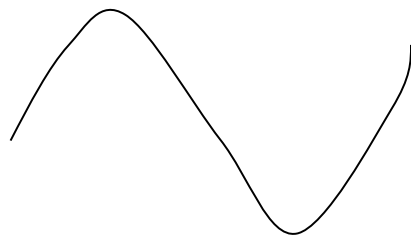
# Lấy mẫu và lượng tử hoá

- **Lấy mẫu và giữ mức:**

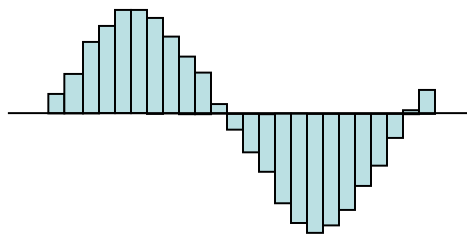
Lấy mẫu là quá trình rời rạc hóa tín hiệu về mặt thời gian và giữ cho biên độ trong khoảng thời gian lấy mẫu không đổi.

- **Lượng tử hóa:**

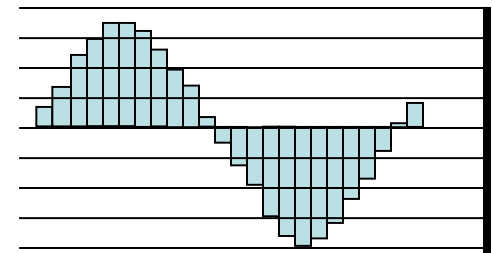
Lượng tử hóa là quá trình rời rạc tín hiệu về mặt biên độ. Tại mỗi mẫu, biên độ được chia thành các mức gọi là các mức lượng tử



Tín hiệu vào



Lấy mẫu và giữ mức



Lượng tử

Hình 2-2 Lấy giữ mẫu và lượng tử

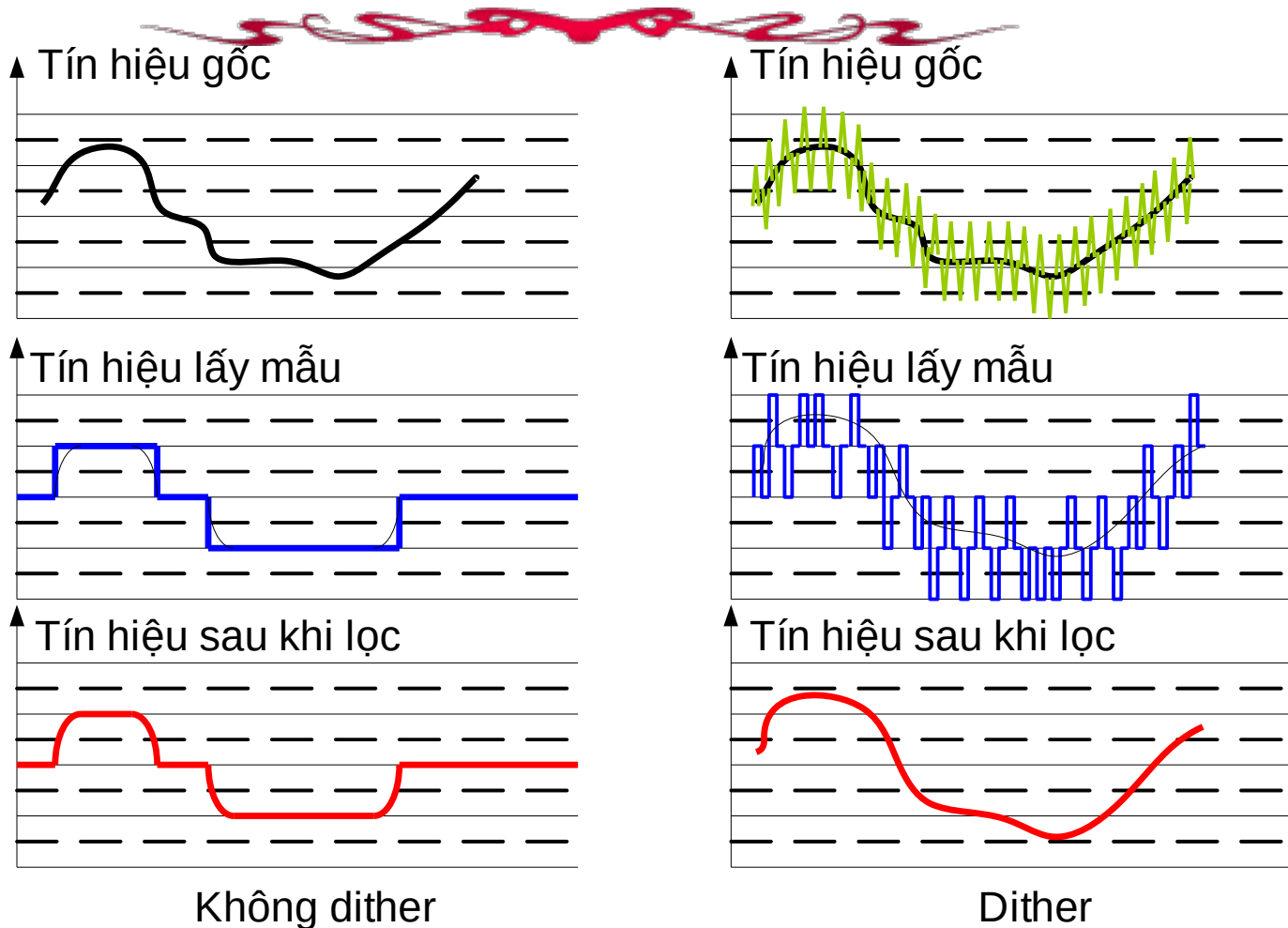


# Dither



- **Nguyên nhân:** Lượng tử hóa  $\rightarrow$  méo. Tín hiệu có biên độ càng nhỏ thì méo lượng tử càng cao.
- **Khắc phục:** Cộng âm thanh trước khi lấy mẫu với một tạp âm tương tự  $\rightarrow$  Ngẫu nhiên hóa các ảnh hưởng méo lượng tử để phân phối đều méo lượng tử thành các lỗi ngẫu nhiên chứ không tập trung nhiều vào phần có biên độ thấp.
- **Khái niệm:** Dither là một nhiễu được cộng vào tín hiệu âm thanh.
- **Mục đích:** Loại bỏ méo lượng tử.
- **Cơ sở:** Dither làm cho tín hiệu âm thanh bị biến đổi giữa các mức lượng tử gần nhau, điều này làm giảm độ tương quan của lượng tử hóa tín hiệu, loại các ảnh hưởng của lỗi và mã hóa các biên độ tín hiệu thấp hơn một mức lượng tử.
- **Nhược điểm:** Cộng nhiễu vào tín hiệu.

# Dither



Không dither

Dither

*Hình 2-3 Dither và không dither*

# Mã hóa và mã hóa kênh



- **Mã hoá**

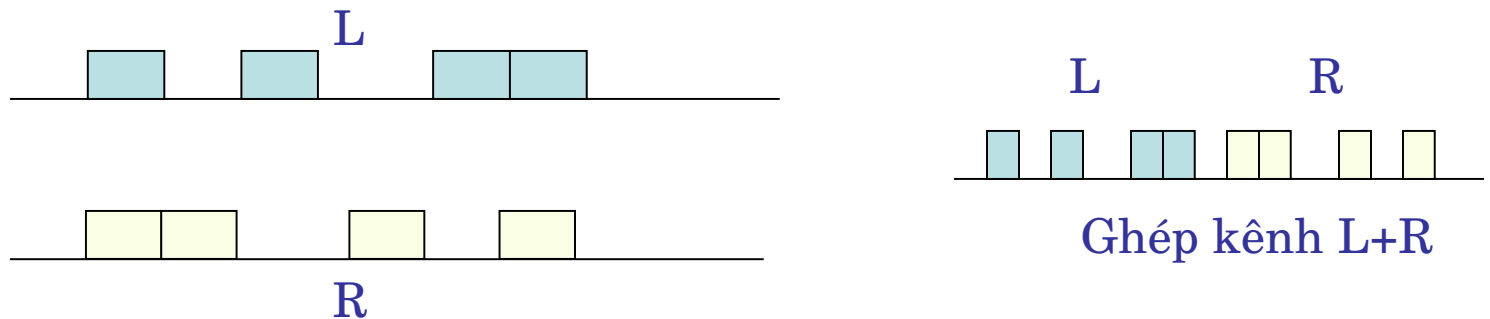
- là quá trình chuyển các mức rời rạc thành một chuỗi các mẫu số nhị phân (hoặc các hệ đếm khác) theo một quy luật nhất định.
- Sau mã hóa nhị phân, ta được tín hiệu điều xung mã PCM.
- Tín hiệu PCM không thích hợp để lưu trữ hoặc truyền dẫn vì vẫn còn tồn tại thành phần một chiều → mã hóa kênh.

- **Mã hóa kênh**

- biến đổi dữ liệu với mục đích đạt được mật độ bit cao trong giới hạn băng thông của kênh truyền.
- Giảm sự tổn hao trong khi truyền hoặc lưu trữ.
- Cải thiện dải thông, dữ liệu truyền dẫn có đặc tính tối ưu.
- Làm cho phổ tín hiệu âm thanh số ít méo.

# Ghép kênh

- Tín hiệu âm thanh số thường bao gồm nhiều kênh, ví dụ hệ thống âm thanh 5.1 gồm các kênh trái, phải, trung tâm, trái vòm, phải vòm và siêu trầm, ngoài ra còn có các tín hiệu mã phụ, mã đồng bộ...
- Nguyên lý ghép kênh có thường được áp dụng đó là ghép kênh phân chia theo thời gian, mỗi kênh sử dụng một khe thời gian được ấn định trước.



Hình 2-4 Ghép kênh tín hiệu audio

# MÃ HOÁ AUDIO CẢM QUAN



- **Giới thiệu**

**Mục đích:**

Biểu diễn chuỗi số ngắn gọn.

Tốc độ bit thấp.

Chất lượng cao

**Động cơ:**

Giảm tốc độ dữ liệu.

Giảm chi phí truyền dẫn (BW).

Giảm các yêu cầu lưu trữ.

**Các yêu cầu:**

Cảm nhận trong suốt.

Độc lập nguồn.

Có khả năng đa kênh.

Độ trễ hợp lý.

# MÃ HOÁ AUDIO CẢM QUAN



Mã hoá nguồn không thực tế với tín hiệu audio, do đó, trong audio, người ta tiến hành

- Khai thác các đặc tính thu được.
- Loại bỏ các thành phần không thích hợp với cảm nhận.
- Giảm các dư thừa thống kê.

	<b>Thoại</b>	<b>Audio</b>
Băng thông	200-3400Hz	20Hz-20kHz
Tốc độ lấy mẫu	8kHz	44.1kHz/48kHz
Số bits trên mẫu	8bits	16+bits
Tốc độ bit thô	64kbps	768kbps
Số kênh	1	1-6+kênh
Mô hình hiệu quả nguồn	Có	Không
Sức chịu đựng nguồn	Có thể yêu cầu	Yêu cầu
Chất lượng mong đợi	Méo do bị giới hạn	Chất lượng “CD”
Tính đa dạng của phổ	Harmonic, V/UV/TR	Không thể phân loại

# MÃ HOÁ AUDIO CẢM QUAN

- Các tiêu chuẩn lấy mẫu theo các yêu cầu cụ thể khác nhau
  - Chất lượng.
  - Độ trễ.
  - Băng thông.
  - Tốc độ

...

Tiêu chuẩn	Tốc độ lấy mẫu	Số bits/mẫu	Mono/stereo	Tốc độ	Băng tần
Điện thoại	8 kHz	8 bits/mẫu	Mono	8 kbytes/s	4Khz
AM radio	11.025 kHz	8 bits/mẫu	Mono	11.0 kbytes/s	5kHz
FM radio	22.05 kHz	16 bits/mẫu	Stereo	88.2 kbytes/s	10kHz
CD	44.1 kHz	16 bits/mẫu	Stereo	176.4 kbytes/s	20kHz
DAT	48 kHz	16 bits/mẫu	Stereo	192.0 kbytes/s	20Khz

# MÃ HOÁ AUDIO CẢM QUAN

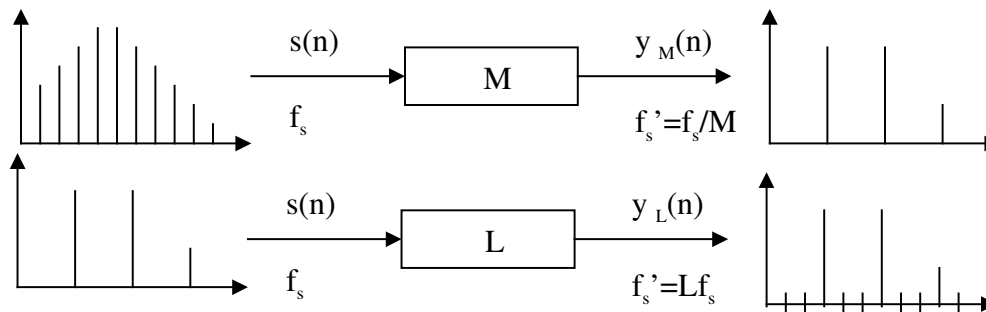
## ❖ Kỹ thuật phân tích thời gian- tần số:

**Mục đích: thuận tiện trong quá trình xử lý.**

## ❖ Biến đổi tần số lấy mẫu

### ❖ Bộ phân chia:

- ❖ Giảm tần số lấy mẫu, chuyển tần số lấy mẫu  $f_s$  về tần số  $f'_s$ , với  $f'_s = f_s/M$ .
- ❖ Tín hiệu ngõ ra có biên độ ở những thời điểm có chu kỳ  $T'_s = 1/f'_s$ .



### ❖ Bộ nội suy:

- ❖ Tăng tần số lấy mẫu, chuyển tần số lấy mẫu  $f_s$  về tần số  $f'_s$ , với  $f'_s = Lf_s$ .
- ❖ Tín hiệu ngõ ra có biên độ của tín hiệu ngõ vào, ngoài ra, nó còn chèn  $L-1$  mẫu có giá trị bằng 0 giữa hai mẫu từ tín hiệu ngõ vào.

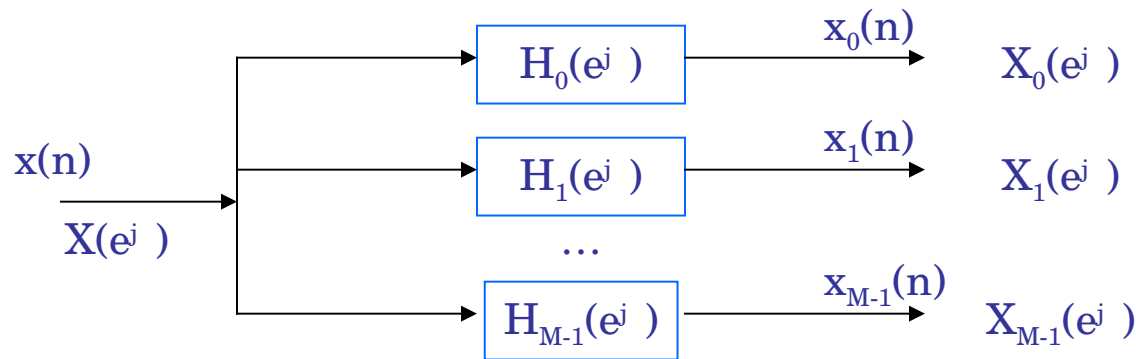


# MÃ HOÁ AUDIO CẢM QUAN

## ❖ *Băng lọc số:*

Băng lọc số là một tập hợp các bộ lọc số có chung đầu vào nhiều đầu ra hoặc chung đầu ra nhiều đầu vào.

Băng lọc số phân tích là tập hợp các bộ lọc số có đáp ứng tần số  $H_k(e^j)$  có chung đầu vào và nhiều đầu ra



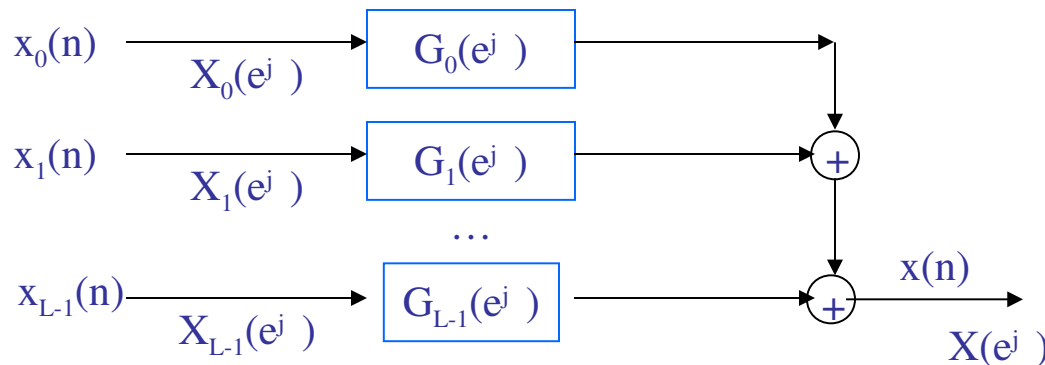
Ngõ ra gồm  $M$  tín hiệu  $X_k(n)$  chiếm dải tần liên tiếp nhau gọi là các tín hiệu băng con (subband).

Các bộ lọc  $H_0(e^j)$ : thông thấp,  $H_{M-1}(e^j)$ : thông cao,  $H_i(e^j)$ : thông dải, với  $i$  từ 1 đến  $M-2$

# MÃ HOÁ AUDIO CẢM QUAN

- **Băng lọc số tổng hợp:**

Là tập hợp các bộ lọc số có đáp ứng tần số  $G_k(e^j)$  có chung đầu ra



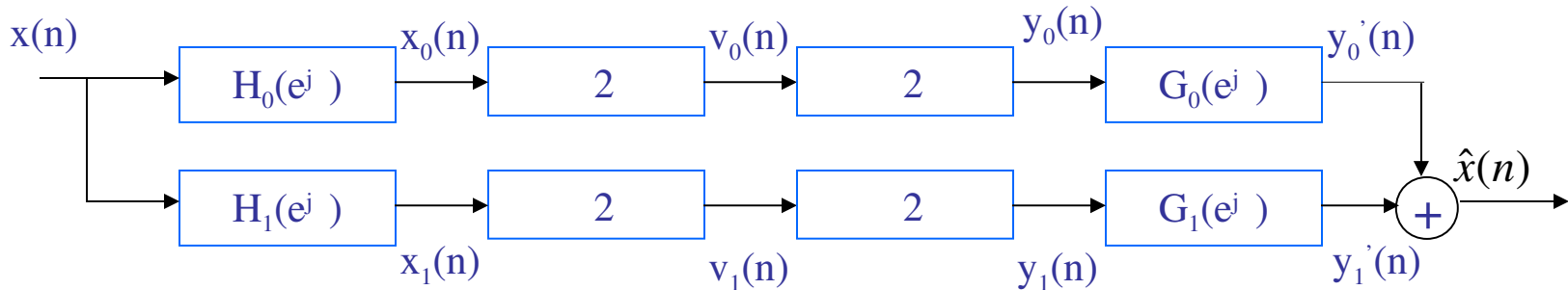
Các bộ lọc :  $G_0(e^j)$ : thông thấp,  $G_{M-1}(e^j)$ : thông cao,  $G_i(e^j)$ : thông dải, với  $i$  từ 1 đến  $L-2$

- **Băng lọc số nhiều nhịp hai kênh và băng lọc gương cầu phương QMF (Quadrature Mirror Filter Bank):**

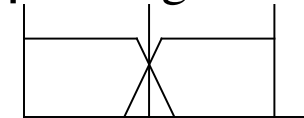
Băng lọc số nhiều nhịp là sự kết hợp của băng lọc số phân tích, băng lọc số tổng hợp với bộ phân chia và bộ nội suy.

Với số bộ lọc của băng lọc phân tích và tổng hợp bằng 2 thì ta có băng lọc số nhiều nhịp hai kênh.

# MÃ HOÁ AUDIO CẢM QUAN



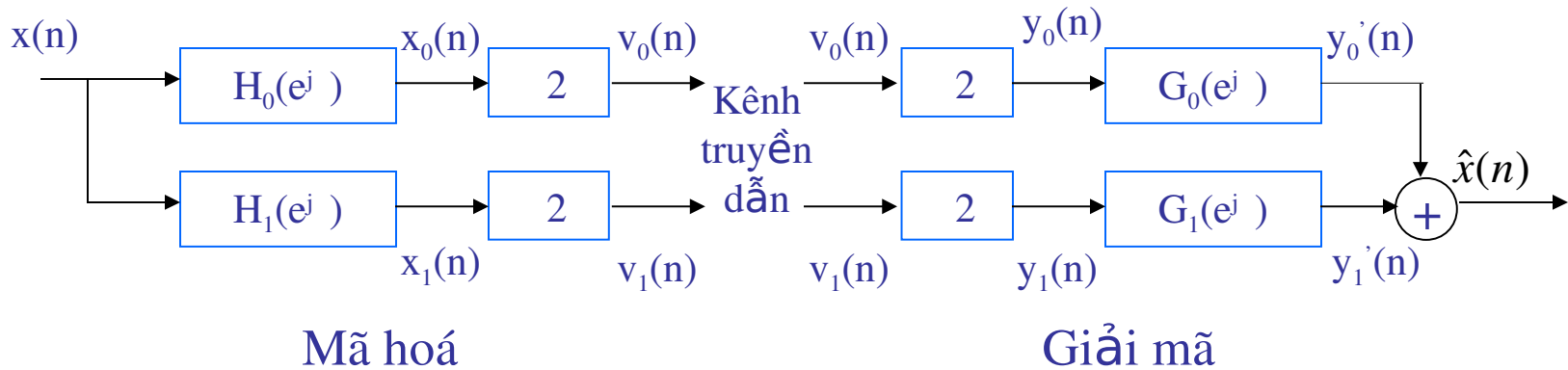
- $H_0(e^j)$ ,  $G_0(e^j)$  :Lọc thông thấp.  $H_1(e^j)$ ,  $G_1(e^j)$  :Lọc thông cao.
- Lý tưởng:  $\hat{x}(n) = x(n)$



- Nếu  $|H_0(e^j)| = |H_1(e^j)|$  và nếu chọn tần số cắt cho 2 bộ lọc là  $\omega_c/2$  thì ta thấy  $|H_0(e^j)|$  là ảnh của  $|H_1(e^j)|$  qua gương đặt ở vị trí  $\omega_c/2$ . Băng lọc nhiều nhịp hai kênh với đặc tính như vậy gọi là băng lọc gương cầu phương  $y(n) = cx(n - n_0)$
- Nếu  $y(n) = cx(n - n_0)$  giống dạng tín hiệu ngõ vào thì ta gọi là băng lọc gương cầu phương khôi phục hoàn hảo PRQMF (Perfect Reconstructive QMF)

# MÃ HOÁ AUDIO CẢM QUAN

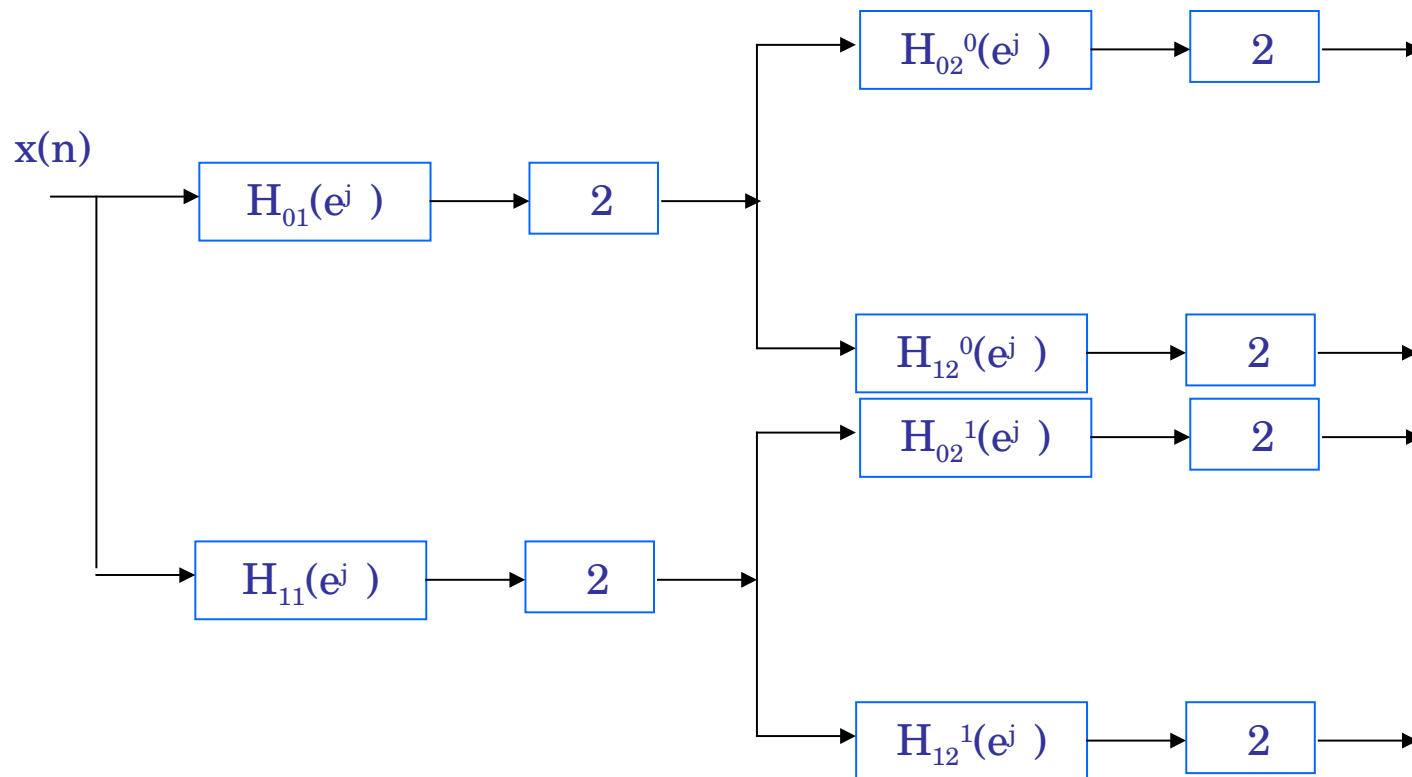
- Mã hoá bằng con và cấu trúc bộ lọc QMF
  - Mã hoá bằng con
    - Sử dụng bộ lọc số để mã hoá bằng con



- Thuận lợi cho việc nén tín hiệu âm thanh vì phổ tập trung không đồng đều. Từ đó, ta có được sự phân bố hợp lý, vừa hiệu quả, vừa đạt chất lượng cao.

# MÃ HOÁ AUDIO CẢM QUAN

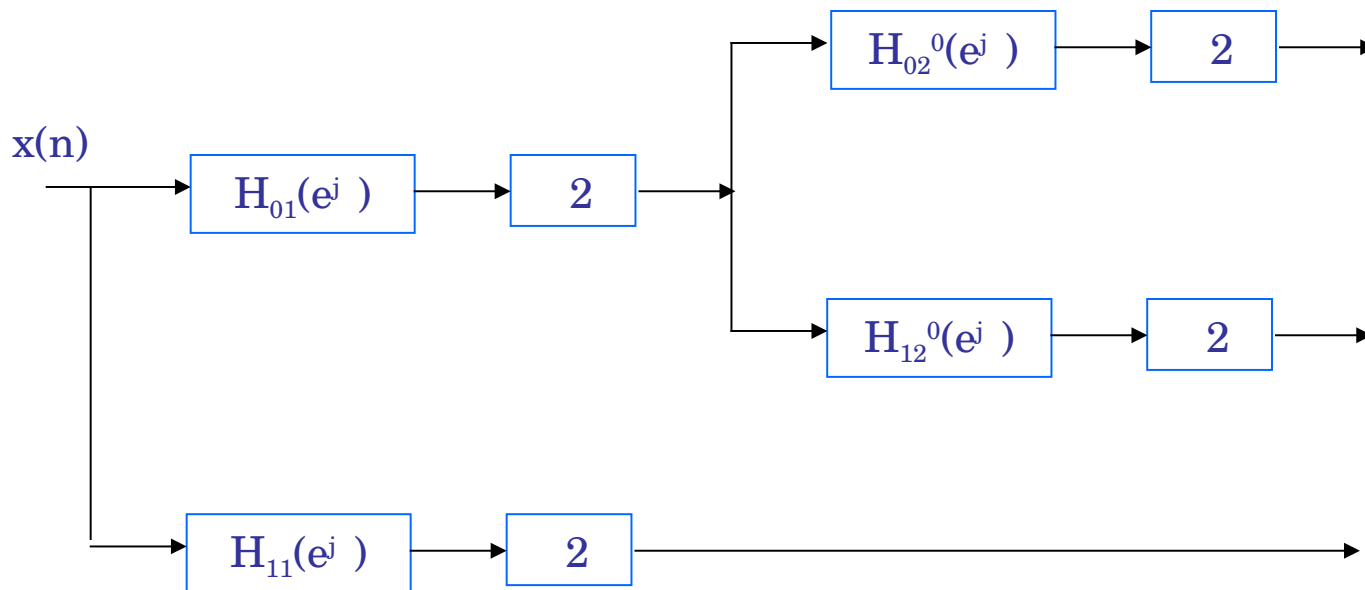
- Cấu trúc dạng cây đơn phân giải



# MÃ HOÁ AUDIO CẢM QUAN

- **Cấu trúc dạng cây đa phân giải**

- Cấu trúc này cho ta lượng bit ngõ ra tối ưu và phụ thuộc vào sự phân bố phổ của tín hiệu.



# MÃ HOÁ AUDIO CẢM QUAN



- Các phương pháp mã hoá chuyển đổi
  - FFT (Fast Fourier Transform)
  - DFT (Discrete Fourier Transform)
  - DCT (Discrete Cosine Transform)
  - MDCT (Modified DCT)
  - Wavelets.

# MÃ HOÁ AUDIO CẢM QUAN

- DCT (Discrete Cosine Transform)

DCT là phép biến đổi trực giao, một thuật toán hiệu quả, cho các đặc tính nén mạnh và giảm độ tương quan.

**Chuyển đổi DCT thuận:**

$$X(k) = \sum_{n=0}^{N-1} x(n) \cos \frac{(2n+1)k}{2N}, \quad k = 0, \dots, N-1$$

$$x(n) = \begin{cases} 1, & n = 0; \\ \frac{2}{N}, & 0 < n < N-1 \end{cases}$$

**Chuyển đổi DCT ngược**

$$x(n) = \sum_{k=0}^{N-1} X(k) \cos \frac{(2n+1)k}{2N}, \quad n = 0, \dots, N-1$$

$$X(k) = \begin{cases} 1, & k = 0; \\ \frac{2}{N}, & 0 < k < N-1 \end{cases}$$



# MÃ HOÁ AUDIO CẢM QUAN

- MDCT (Modified DCT): DCT được hiệu chỉnh.

MDCT là phép biến đổi trực giao tuyến tính được hiệu chỉnh từ DCT.

MDCT thuận:

$$X(k) = \sum_{n=0}^{N-1} p(n)x(n) \cos \left[ \frac{\pi}{2N} (2n+1 + \frac{N}{2})(2k+1) \right], \quad n = 0, \dots, N-1; k = 0, \dots, \frac{N}{2}-1.$$

Và chuyển đổi MDCT ngược (IMDCT):

$$x(n) = \sum_{k=0}^{\frac{N}{2}-1} p(n)X(k) \cos \left[ \frac{\pi}{2N} (2n-1 + \frac{N}{2})(2k-1) \right], \quad n = 0, \dots, N-1; k = 0, \dots, \frac{N}{2}-1.$$

Với  $p(x)$  là hàm cửa sổ sin được xác định:

$$p(x) = \sin\left(\frac{\pi x}{N}\right)$$

- Wavelets

Wavelets có thể được xem như một bộ phân tích băng con, với cây không cân bằng, nghĩa là các tần số được chia một cách không đồng nhất. Vậy, băng lọc tương đồng với dải tới hạn.

# MÃ HOÁ AUDIO CẢM QUAN

•DWT (Discrete Wavelet Transform)

Mở rộng và dịch chuyển “hàm mẹ”  $h(t)$  bằng cách định nghĩa một cơ sở trực giao, wavelet cơ sở:

$$h_{n,m}(t) = 2^{-m/2} h(2^{-m}t - n)$$

Trong đó:  $n$  là tỷ lệ,  $m$  là độ dời và  $t$  là thời gian.

Hệ số tỷ lệ  $n$  chỉ thị độ rộng của các wavelet và hệ số vị trí  $m$  xác định vị trí của nó. Với hàm mẹ  $h(t)$ , ta được một tập hàm wavelet trực giao cơ sở.

Trực giao:

$f(t), g(t)$  là hai vector thuộc không gian  $L2(a,b)$ .  $t \in [a,b]$ .

Hai vector gọi là trực giao khi tích vô hướng của chúng bằng 0.

$$\langle f(t), g(t) \rangle = \int_a^b f(t) g^*(t) dt = 0$$

# MÃ HOÁ AUDIO CẢM QUAN

## • Cơ sở trực giao:

Tập các vector  $\{v_k\} = \{v_1, v_2, \dots, v_n\}$  được gọi là cơ sở trực giao nếu chúng trực giao từng đôi một và có độ dài bằng 1.

$$\langle v_m, v_n \rangle = \delta_{mn}. \text{ Hay: } \int_a^b h_k(t) h_l^*(t) dt = \delta_{kl}.$$

$$\text{• Hàm delta: } \delta_{kl} = \begin{cases} 1, & k = l \\ 0, & k \neq l \end{cases}$$

## Chuyển đổi Wavelet:

$$X(n, m) = \int c_{n,m} x(t) h_{n,m}^*(t) dt$$

## Chuyển đổi wavelet ngược:

$$x(t) = \int c_{n,m} h_{n,m}(t) dt$$

# MÃ HÓA CẢM QUAN

- **Ý tưởng cơ sở:** Khám phá vùng âm thanh con người cảm nhận yếu để nén.
- **Ví dụ:** âm thanh MPEG, Dolby AC



# MÃ HÓA CẢM QUAN



- Âm thanh được sản sinh từ nguồn âm.
- Sự rung động làm chuyển dịch các phần tử trong không khí.
- Sự rung động làm sinh ra áp lực trong không khí: nếu thấp hơn áp lực trung bình thì làm giãn và cao hơn áp lực trung bình thì làm nén không khí lại.
- Sự nén giãn không khí làm sinh ra sóng âm.
- Khi sóng âm va chạm vào bề mặt (trống tai hay microphone) làm rung bề mặt này một cách đồng bộ.
- Lúc này, năng lượng âm thanh được truyền từ nguồn âm đến nơi nhận.

# MÃ HÓA CẢM QUAN



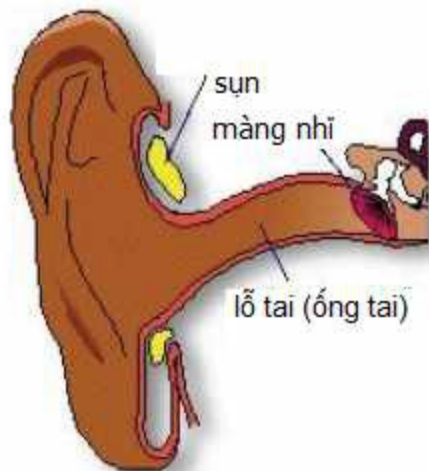
- **Tai người:**

- Khi sóng âm đập vào tai người làm rung động trống tai một cách đồng bộ.
- Qua nhiều cơ chế, năng lượng âm thanh được truyền đến các tế bào thần kinh mà não cảm nhận là âm thanh.
- Tai có thể chia thành 3 phần:
  - Tai ngoài.
  - Tai giữa.
  - Tai trong.

# MÃ HÓA CẢM QUAN

- **Tai ngoài**

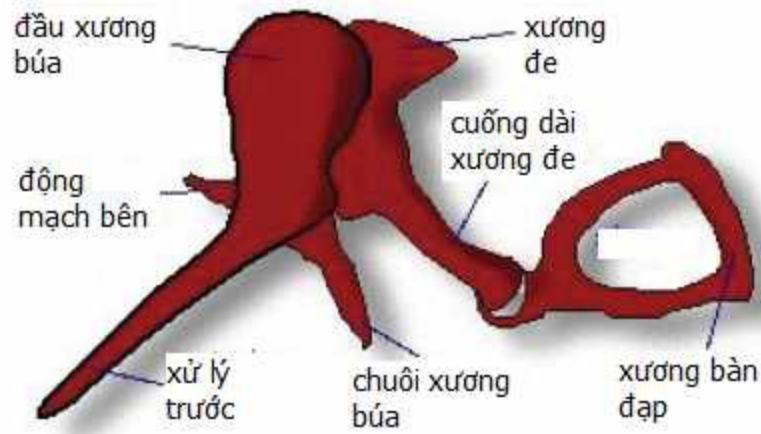
- Ống tai: tập trung năng lượng âm thanh đưa đến.
- Màng nhĩ:
  - Giao tiếp giữa tai ngoài và tai giữa.
  - Âm thanh được chuyển thành rung động cơ học qua tai giữa.
  - Rung động màng nhĩ của trống tai.



# MÃ HÓA CẢM QUAN

- **Tai giữa**

- Gồm 3 xương nhỏ: Xương búa, xương đe và xương bàn đạp.
- Liên kết từ trong tai.
- Các xương này khuếch đại áp lực âm thanh.





# MÃ HÓA CẢM QUAN

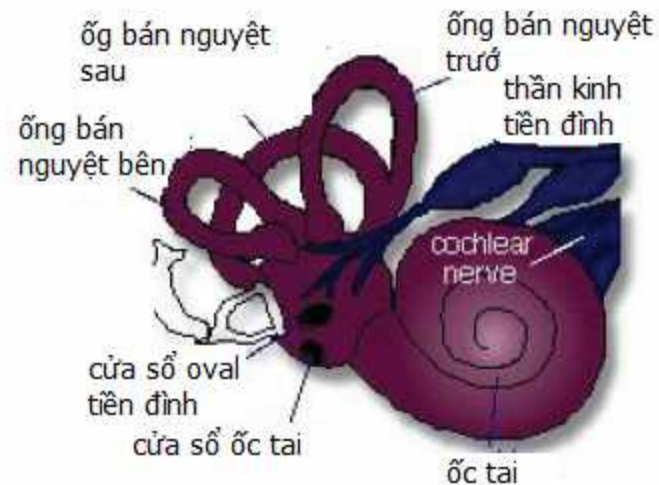
- **Tai trong**

- Ốc tai:

- Chuyển đổi áp lực cơ học các xương nhỏ thành áp lực nước.
    - Ốc tai chứa đầy chất lỏng.
    - Áp lực nước truyền vào ống dẫn ốc tai đến vỏ não.

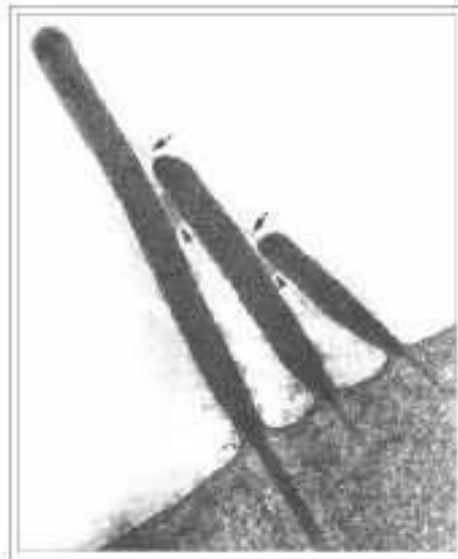
- Các ống bán nguyệt

- Cơ chế cân bằng cơ thể.
    - Cảm nhận âm thanh.



# MÃ HÓA CẢM QUAN

- Áp lực sóng lên màng nhĩ gây sức ép từ cửa sổ oval dùng đột ngột tại màng nhĩ che cửa sổ hình tròn.
- Áp lực này truyền lên tất cả các bộ phận của màng nhĩ.
- Bề mặt bên trong của màng nhĩ có khoảng hơn 20000 tế bào thần kinh hình sợi gọi là stereocilia.



# MÃ HÓA CẢM QUAN



- **Nghe các tần số khác nhau:**
  - Màng nhĩ chặt ở một đầu và lỏng ở đầu kia.
  - Âm thanh tần số cao cực đại ở đầu chặt.
  - Âm thanh tần số thấp ở đầu lỏng.
  - Các tần số cộng hưởng như trên sợi dây cột chặt.
  - Stereo phân biệt từng khoảng nhỏ của độ dài và độ co giãn của chất lỏng mà nó đi qua.

# MÃ HÓA CẢM QUAN



- Âm thanh nén chặt di chuyển vào tai giữa từ màng nhĩ.
- Stereocilia chuyển động.
- Mỗi stereocilia cảm nhận một tần số khác nhau.
- Tế bào stereocilia cộng hưởng với biên độ lớn của rung động.
- Biên độ rung động tăng lên dẫn đến tế bào stereocilia giải phóng xung điện từ dọc theo các tế bào thần kinh thính giác lên não.
- Não cảm nhận âm thanh nhận được nhờ những xung điện từ này.

# PHÂN TÍCH TÂM LÝ ÂM HỌC

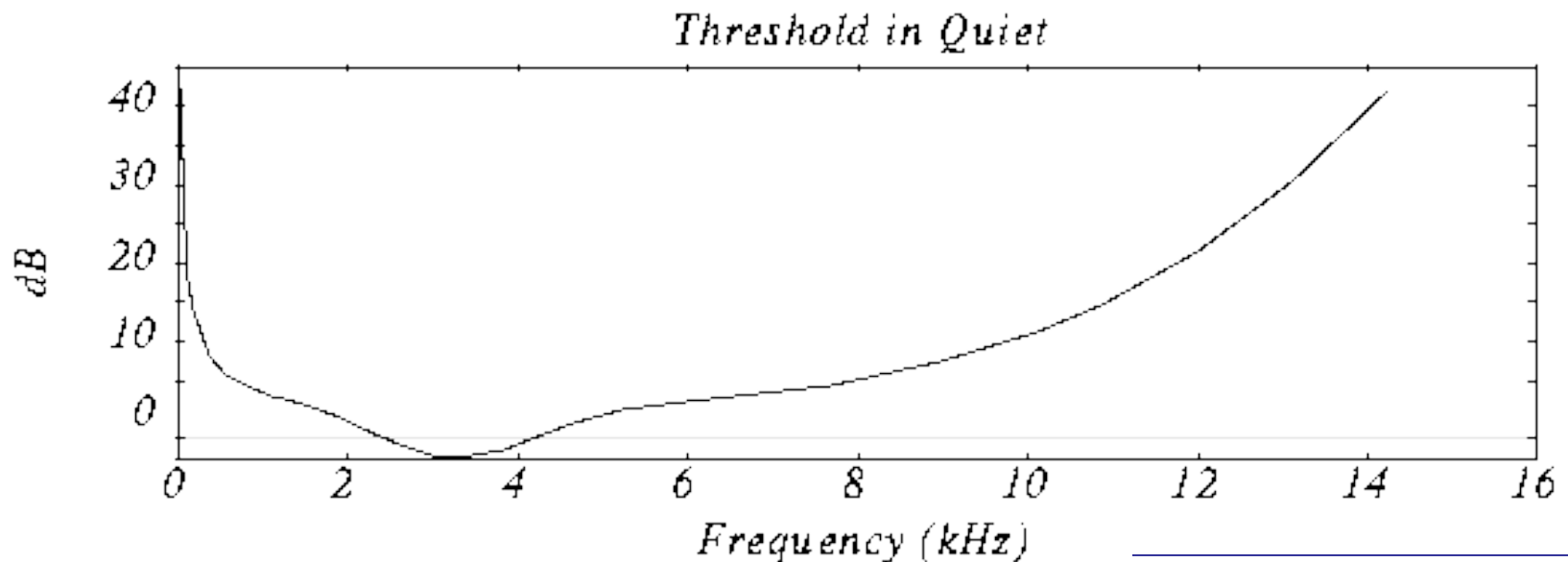
## • Hệ thống thính giác của con người

- Được chú trọng khai thác trong mã hoá audio cảm quan.
- Con người có thể nghe trong dải 20Hz đến 20kHz với khả năng nghe không đồng nhất với các tần số trong dải này. Việc cảm nhận còn phụ thuộc vào mức áp lực và tùy thuộc vào từng người.
- Dải 20Hz – 20kHz được chia thành các dải con không đồng nhất và không tuyến tính. Cảm nhận tốt trong khoảng 2kHz đến 4kHz và ngưỡng nghe đến ngưỡng đau khoảng 96dB.
- Phụ thuộc vào môi trường nghe, với môi trường nhiễu lớn thì hạn chế khả năng nghe và khả năng phân biệt các âm thanh khác nhau.
- Vậy, phân tích tâm lý nghe là xét các vấn đề:
  - Độ nhạy của tai, khả năng đáp ứng của các cường độ khác nhau.
  - Đáp ứng của tai với các tần số khác nhau.
  - Nghe một âm khi có mặt một âm khác.

# PHÂN TÍCH TÂM LÝ ÂM HỌC

- **Ngưỡng nghe tuyệt đối ATH (Absolute Theshold of Hearing)**

- Thí nghiệm:Đặt một người trong phòng kín, im lặng, phát âm kiểm tra (test tone) với tần số xác định (1kHz), tăng mức âm thanh cho đến khi có thể nghe được, ghi lại các giá trị và lặp lại với tần số khác.
- Vẽ đồ thị, ta được ngưỡng nghe tuyệt đối. Thử với người khác, ghi kết quả.



# PHÂN TÍCH TÂM LÝ ÂM HỌC



- **Dải tới hạn (critical bandwidth)**

- Khoảng tần bị che gần đúng theo tần số và biên độ gọi là dải tới hạn.
- Fletcher tiến hành các thử nghiệm và cho thấy việc nghe của con người giống như sử dụng các bộ lọc tâm sinh lý có độ rộng gần bằng một giá trị tới hạn và Fletcher gọi độ rộng của bộ lọc tới hạn là dải tới hạn.
- Dải tới hạn biểu diễn công suất xác định của tai cho các tần số hay dải tần số liên tục.
- Các thí nghiệm cho thấy rằng:
  - Với các tần số che nhỏ hơn 500Hz thì dải tới hạn không đổi với độ rộng khoảng 100Hz.
  - Với các tần số che lớn hơn 500Hz thì dải tới hạn có độ rộng tăng tương đối tuyến tính theo tần số.
- Vậy, thang tần số không tuyến tính → thang bark (Barkhausen)

# PHÂN TÍCH TÂM LÝ ÂM HỌC

- **Dải tới hạn (critical bandwidth)**

- Flecher chia băng thông âm thanh thành 25 dải tới hạn.

<b>Dải</b>	<b>Tần số thấp</b>	<b>Tsố trung tâm</b>	<b>Tần số cao</b>	<b>Băng thông</b>
<b>1</b>	<b>-</b>	<b>50</b>	<b>100</b>	<b>-</b>
<b>2</b>	<b>100</b>	<b>150</b>	<b>200</b>	<b>100</b>
<b>..</b>	<b>..</b>	<b>..</b>	<b>..</b>	<b>..</b>
<b>4</b>	<b>300</b>	<b>350</b>	<b>400</b>	<b>100</b>
<b>5</b>	<b>400</b>	<b>450</b>	<b>510</b>	<b>110</b>
<b>6</b>	<b>510</b>	<b>570</b>	<b>630</b>	<b>120</b>
<b>7</b>	<b>630</b>	<b>700</b>	<b>770</b>	<b>140</b>
<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>
<b>25</b>	<b>15500</b>	<b>18775</b>	<b>22050</b>	<b>6550</b>



# PHÂN TÍCH TÂM LÝ ÂM HỌC

- **Dải tới hạn (critical bandwidth)**

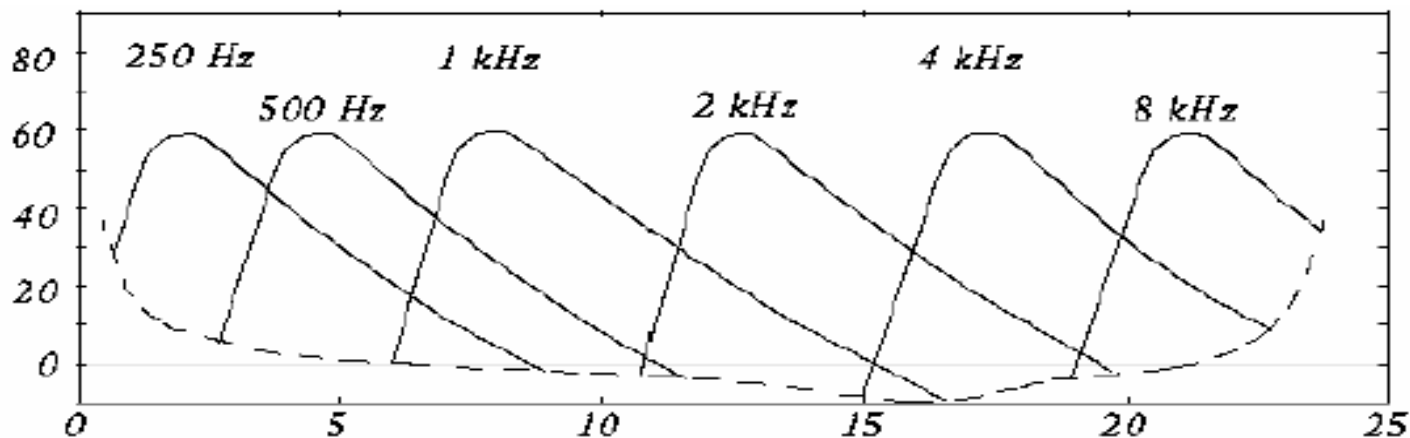
- Bark là đơn vị để biểu diễn một dải tới hạn, 1 bark=1 độ rộng dải tới hạn.

- Công thức chuyển đổi:

$$1 \text{ bark} = \frac{f}{100}, \quad f \leq 500$$
$$1 \text{ bark} = 9.4 \log_2\left(\frac{f}{1000}\right), \quad f > 500$$

- Công thức khác:  $1 \text{ bark} = 13 \arctg(0.76f) + 3.5 \arctg(f/7500)$  [bark]

- Công thức khác nữa:  $1 \text{ bark} = 13 \arctg(0.76f) + 3.5 \arctg(f^2/65.25)$  [bark]



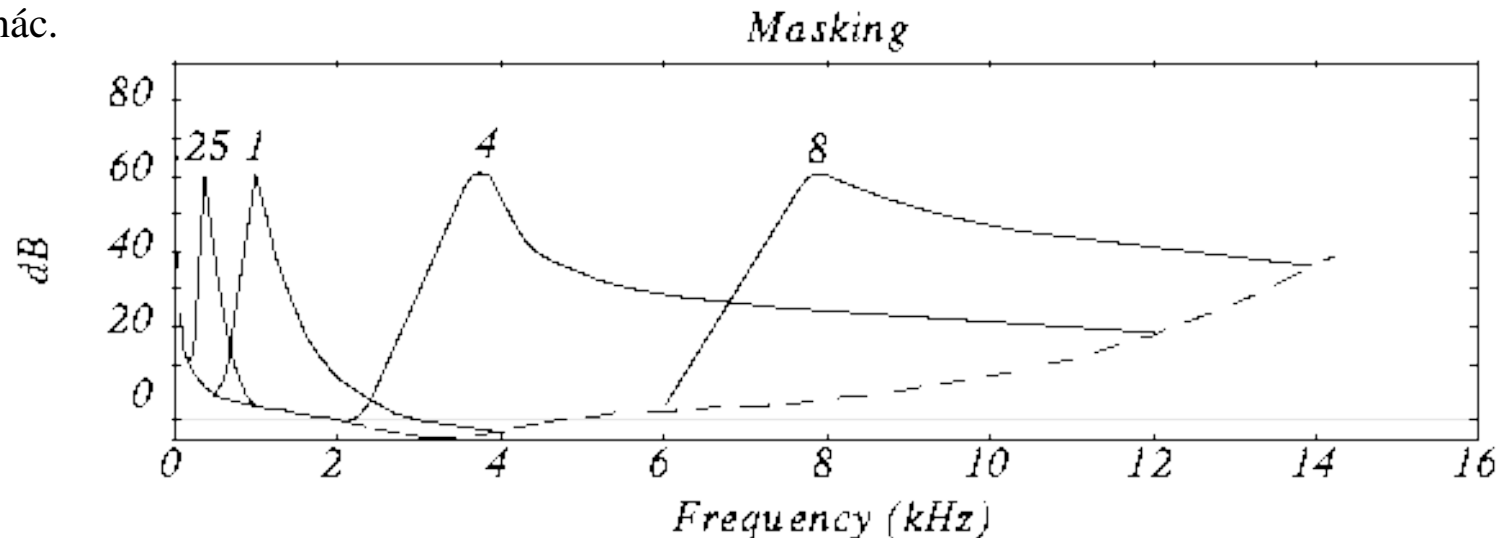
# PHÂN TÍCH TÂM LÝ ÂM HỌC

## • Kỹ thuật che (masking)

Con người khi nghe một âm với sự có mặt của một âm khác sẽ cảm nhận yếu đi khi âm này có tần số gần với âm cần nghe hoặc biên độ lớn.

### – Che tần số (frequency masking)

Thí nghiệm: Đặt một người trong phòng kín, phát ra một âm che (masking tone) với tần số xác định (1kHz) ở một mức nào đó (60dB), sau đó, phát âm kiểm tra (test tone) (1,1kHz), tăng mức đến khi có thể nghe được. Thay đổi âm kiểm tra, vẽ ngưỡng nghe, lặp lại với âm che khác.



# PHÂN TÍCH TÂM LÝ ÂM HỌC

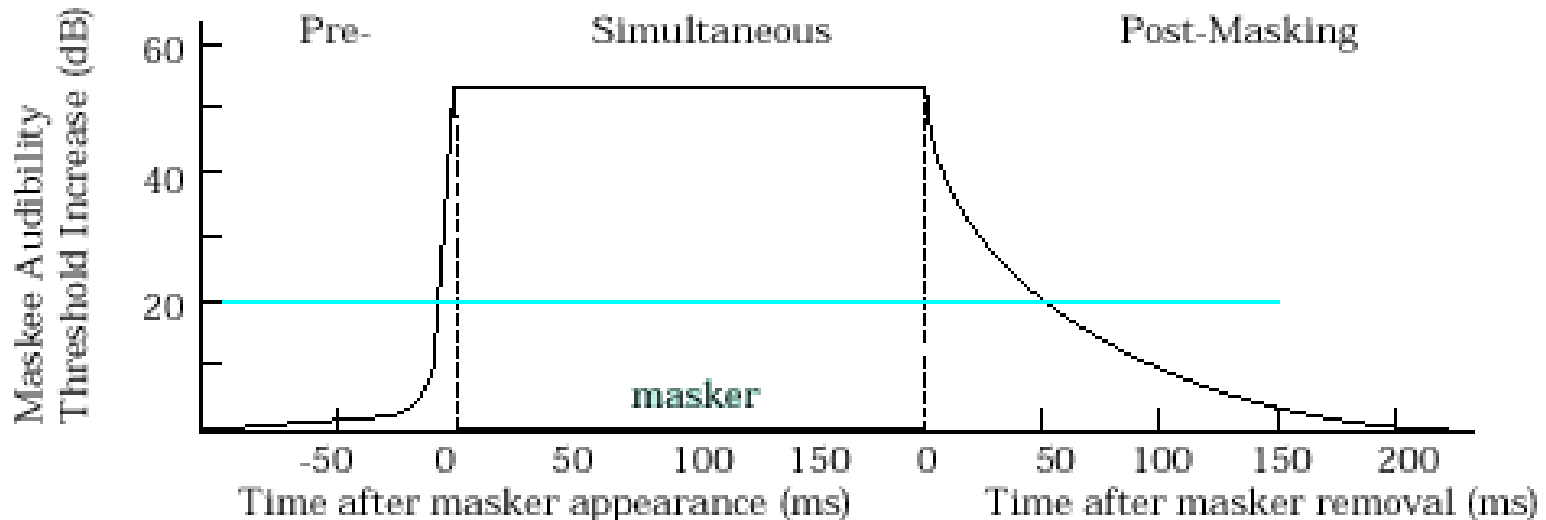
## – Che thời gian (temporal masking)

*Thí nghiệm:*

Phát một âm che ở tần số 1kHz, biên độ 55dB, thêm một âm kiểm tra 1,1kHz, biên độ 20dB trước và sau âm che. Âm kiểm tra không thể nghe được (nó đang bị che).

Lặp lại các mức khác của âm kiểm tra và vẽ.

Với thí nghiệm trên, ta thấy âm 1,1kHz với 20dB bị che trước khoảng 15ms và che sau khoảng 50ms.



# KỸ THUẬT NÉN AUDIO



- **Cơ sở**

Âm thanh trung thực và chất lượng dịch vụ thoả mãn thì tốc độ dòng dữ liệu phải lớn.

Ví dụ : Hệ thống âm thanh đa kênh mã hoá 16 bits, tần số lấy mẫu 48kHz (6 kênh) sẽ có tốc độ:  $48 \times 16 \times 6 = 4.5 \text{ Mbps}$ .

Tốc độ cao → Khó khăn lưu trữ, truyền dẫn và giá thành thiết bị. → Nén.

- **Nén không tổn hao**

Khôi phục đúng thông tin ban đầu sau khi giải nén.

Cơ sở: Loại bỏ dư thừa thống kê, các thông tin xuất hiện trong tín hiệu mà có thể dự báo trước.

Tỷ số nén thấp, khoảng 2:1, phụ thuộc vào độ phức tạp của tín hiệu nguồn.

Thường sử dụng kỹ thuật mã hoá dự đoán trong miền thời gian.

# KỸ THUẬT NÉN AUDIO



## – Thuật toán vi sai:

Tín hiệu âm thanh có đặc tính lặp đi lặp lại nên xuất hiện sự dư thừa số liệu. Thông tin lặp sẽ được loại bỏ trong quá trình mã hoá và được đưa vào lại trong quá trình giải mã sử dụng kỹ thuật DPCM.

Các tín hiệu audio đầu tiên được phân tích thành tập hợp các dải băng con bao gồm một số lượng âm thanh rời rạc, sau đó, DPCM được sử dụng để dự báo các tín hiệu lặp lại theo chu kỳ. Nếu sử dụng ADPCM còn cho kết quả tốt hơn.

## –Mã hoá Entropy:

Tận dụng độ dư thừa trong cách miêu tả các hệ số băng con đã lượng tử hoá nhằm cải thiện tính hiệu quả của quá trình mã hoá. Các hệ số lượng tử được gửi đi theo sự tăng dần của tần số.

Kết quả nhận được là bảng mã tối ưu thống kê các giá trị miền tần số thấp và cao.

Sử dụng mã hoá Huffman, Lempel-Zip để nén.

# KỸ THUẬT NÉN AUDIO



- **Nén tổn hao**

Hệ thống thính giác của con người không thể phân biệt các thành phần phổ có biên độ nhỏ giữa các thành phần phổ có biên độ lớn.

Hệ số nén lớn, khoảng 20:1 phụ thuộc vào quá trình nén và giải nén và chất lượng audio yêu cầu.

- **Các kỹ thuật được sử dụng:**

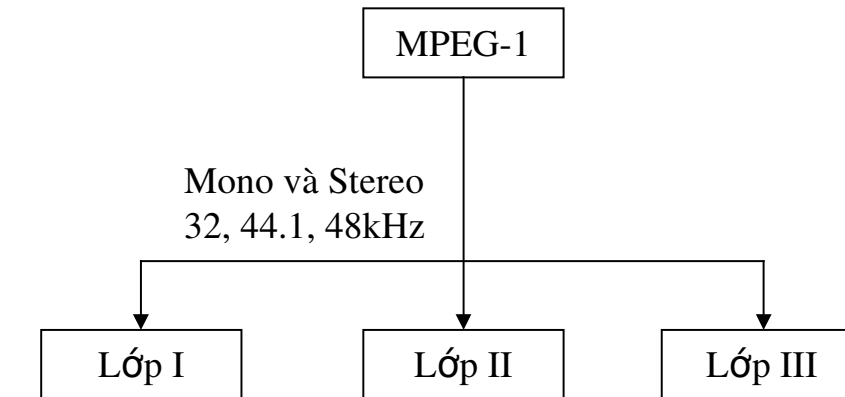
- Kỹ thuật che (masking) đối với các thành phần tín hiệu trong miền thời gian và tần số.

- Che mức tap âm lượng tử cho từng âm độ của tín hiệu âm thanh bằng cách chỉ định số bit vừa đủ để chắc chắn rằng mức nhiễu lượng tử luôn nằm dưới mức giá trị cần che.

- Mã hoá ghép: Khai thác độ dư thừa trong hệ thống audio đa kênh với các thành phần số liệu trong các kênh giống nhau. Mã hoá một phần số liệu chung trên một kênh và chỉ định cho bộ giải mã lặp lại tín hiệu đó trên các kênh còn lại.

# MPEG-1

- Giới thiệu



Được phát triển trên cơ sở phối hợp chuẩn ISO/IEC 11172.

Sử dụng tần số lấy mẫu của CD-DA, với  $f_s=32;44.1;48\text{kHz}$ , mã hoá 16bits/mẫu tín hiệu.

Tốc độ bit: 32 - 768 kbps/channel.

Các kiểu: Mono, dual-mono, dual-stereo, joint-stereo.

Xác định các tham số khác nhau về tốc độ, dòng số sau khi nén, số mẫu trong header cho một kênh, cấu trúc thời gian khung, phương pháp mã hoá dự đoán và các chế độ làm việc.

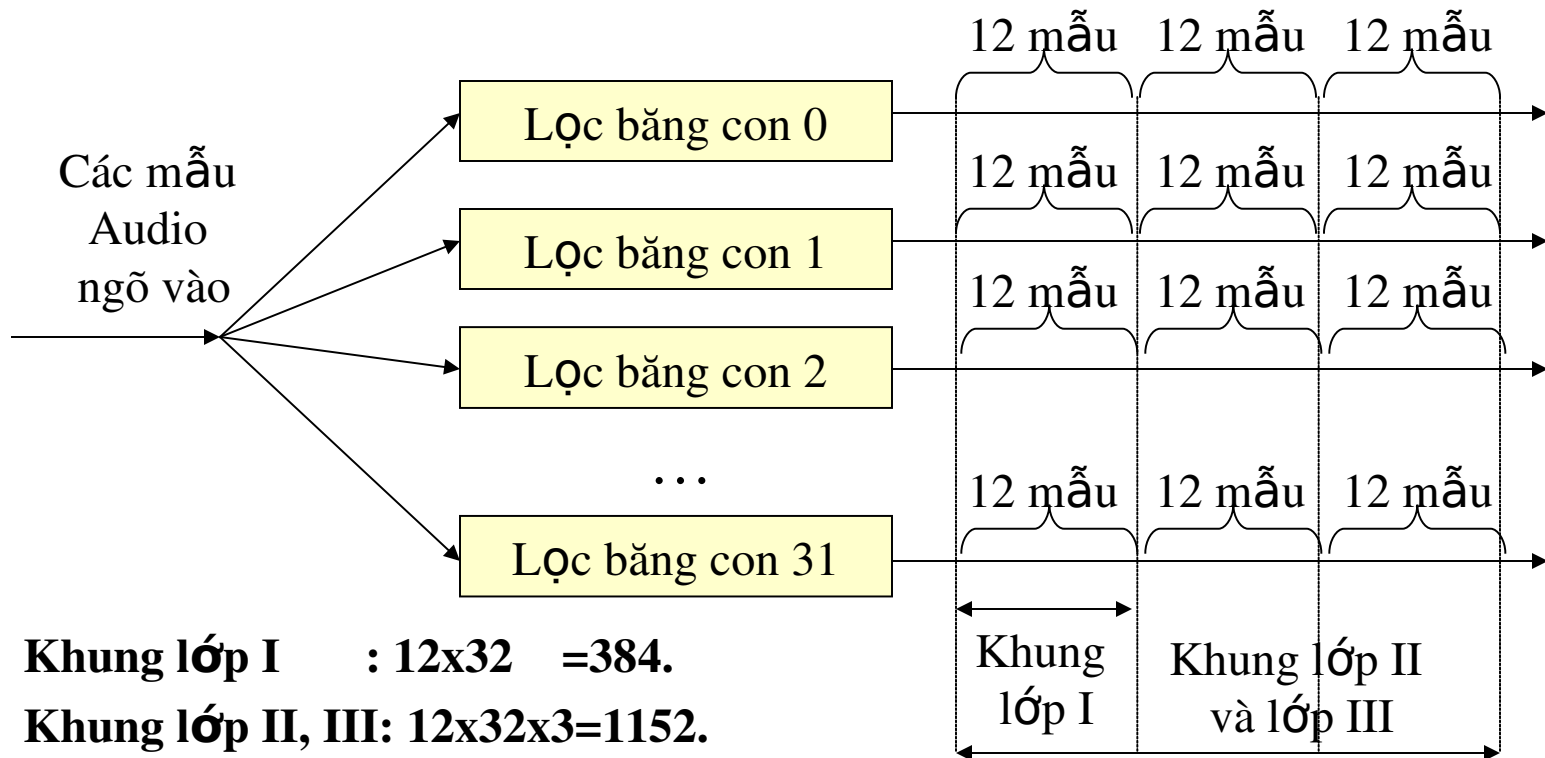
# MPEG-1

- Đặc tính

Lớp I	Lớp II	Lớp III
Dùng cho thiết bị dân dụng	Dùng cho thiết bị chuyên dụng, đa môi trường	Dùng cho thiết bị chuyên dụng, đa môi trường
Tốc độ dòng số liệu từ 32-448kbps	Tốc độ dòng số liệu từ 32-384kbps	Tốc độ dòng số liệu từ 32-320kbps
384mẫu/khung/kênh	1152mẫu/khung/kênh	1152mẫu/khung/kênh
32 băng con đều nhau, mỗi băng con gồm block 12 mẫu	32 băng con đều nhau, mỗi băng con gồm block 36 mẫu	32 băng con tới hạn thành 18 MDCT
Chu kỳ một khung 8ms cho kênh có $f_s=48\text{kHz}$	Chu kỳ một khung 24ms cho kênh có $f_s=48\text{kHz}$	Chu kỳ một khung 24ms cho kênh có $f_s=48\text{kHz}$
Hệ số tỷ lệ 6 bits/băng, phân phối bit theo phương thức ứng trước.	Hệ số tỷ lệ 6 bits/băng, phân phối bit theo phương thức ứng trước.	Hệ số tỷ lệ 6 bits/băng, phân phối bit theo phương thức ứng trước.



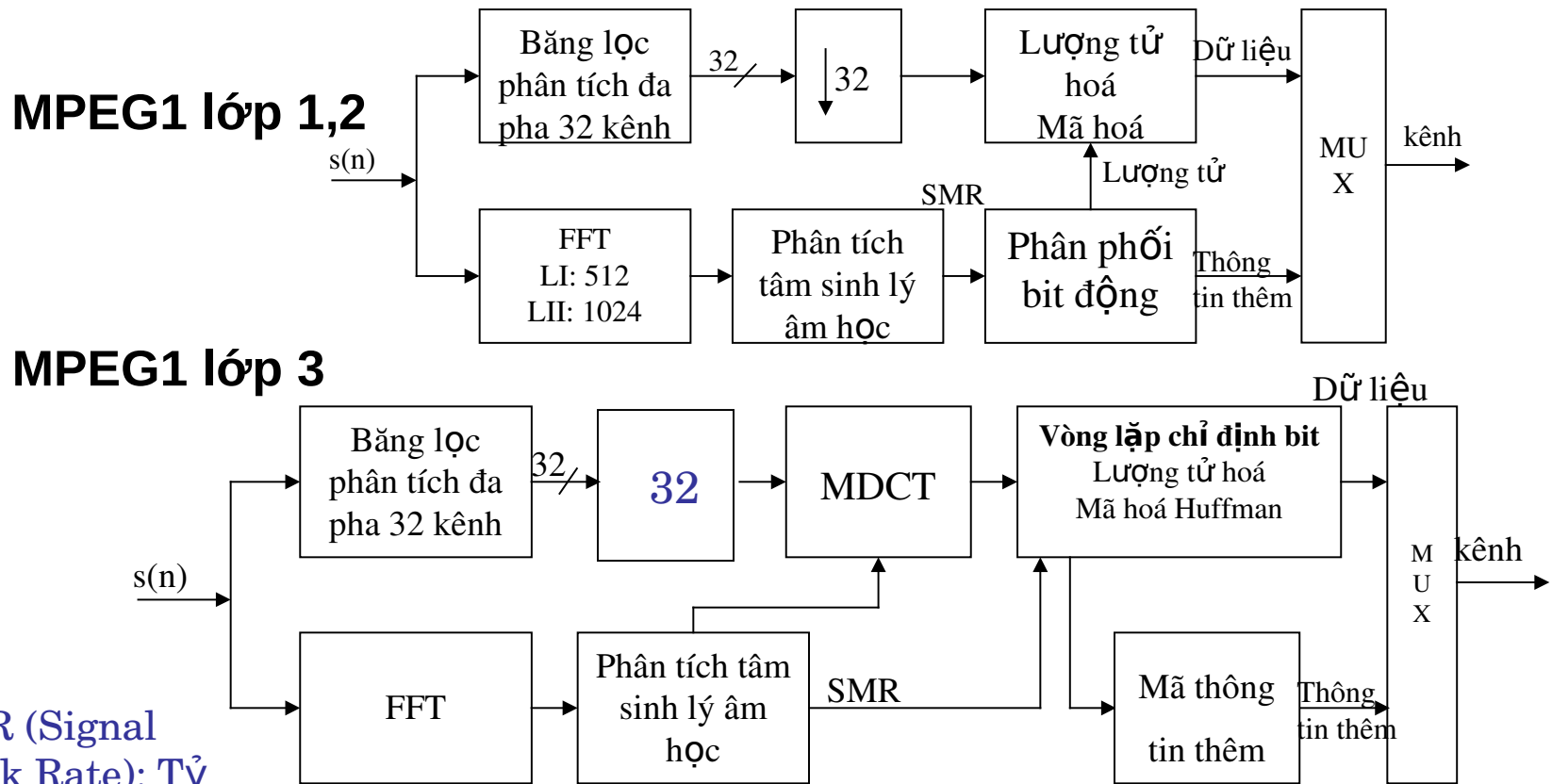
# MPEG-1



- **Khung lớp I** :  $12 \times 32 = 384$ .
- **Khung lớp II, III**:  $12 \times 32 \times 3 = 1152$ .

# MPEG-1

## — Kiến trúc



SMR (Signal Mark Rate): Tỷ số tín hiệu/ngưỡng nghe

# MPEG-1



- **Thuật toán cơ bản**

- Tiến hành chia ngõ vào thành 32 băng con bởi các băng lọc.
  - Lấy 32 mẫu PCM trong cùng một thời điểm, kết quả là 32 hệ số tần số ở ngõ ra.
- Trong MPEG-1 lớp I thì tập 32 giá trị PCM được kết hợp vào trong khối gồm 12 nhóm 32 mẫu này.
- MPEG-1 lớp II và lớp III thì gồm 3 khối 12 nhóm này.
- Phân bố bit đảm bảo rằng mọi nhiễu lượng tử nằm ở dưới các ngưỡng che.
- Với mỗi băng con, xác định mức biên độ và mức nhiễu bằng mô hình tâm sinh lý nghe. SMR (signal-mask rate) được sử dụng để xác định số bit cho quá trình lượng tử hoá đối với mỗi băng con với mục đích giảm thiểu dung lượng.

- **Phân phối bit**

- Là thủ tục xác định số bit cho mỗi băng con.
- Dựa vào thông tin vào từ mô hình tâm sinh lý nghe

# MPEG-1

Ví dụ: Sau khi phân tích, mức của 16 băng con đều là:

Band	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Level (db)	0	8	12	10	6	2	10	60	35	20	15	2	3	5	3	1

Nếu mức của băng con thứ 8 là 60 thì nó che 12 dB ở băng con thứ 7 và 15 dB ở băng con thứ 9.

Băng con 7 có mức 10dB < 12dB: loại. Băng con 9 có mức 35dB > 15dB: gửi đi.

→ Chỉ có các mức lớn hơn mức che là được gửi đi thay vì dùng 6 bits để mã hoá, ta chỉ dùng 5 bits.

MPEG-Layer I: Bộ lọc DCT 1 khung và tần số bằng phẳng trong mỗi băng con. Mô hình tâm sinh lý nghe sử dụng che tần số.

MPEG-Layer II: Có 3 khung trong bộ lọc (trước, hiện tại và kế), tổng là 1125 mẫu. Sử dụng vài bits để che thời gian.

MPEG-Layer III: Sử dụng bộ lọc tối hạn để đáp ứng tốt hơn. Mô hình tâm sinh lý nghe sử dụng che thời gian, che tần số, tính toán độ dư thừa stereo và mã hoá Huffman.

# MPEG-1

- Cấu trúc khung

Layer I	Header (32)	CRC (0-16)	Bit Allocation (128-256)	Scale Factors (0-384)	Samples ( 384)	Ancillary Data	
Layer II	Header (32)	CRC (0-16)	Bit Allocation (128-256)	SCFSI (0-60)	Scale Factors (0-384)	Samples ( 1152)	Ancillary Data
Layer III	Header (32)	CRC (0-16)	Side Information (128-256)	Bit Reservoir (384)	Samples ( 1152)	Ancillary Data	

SCFSI: Scale Factor Selection Information

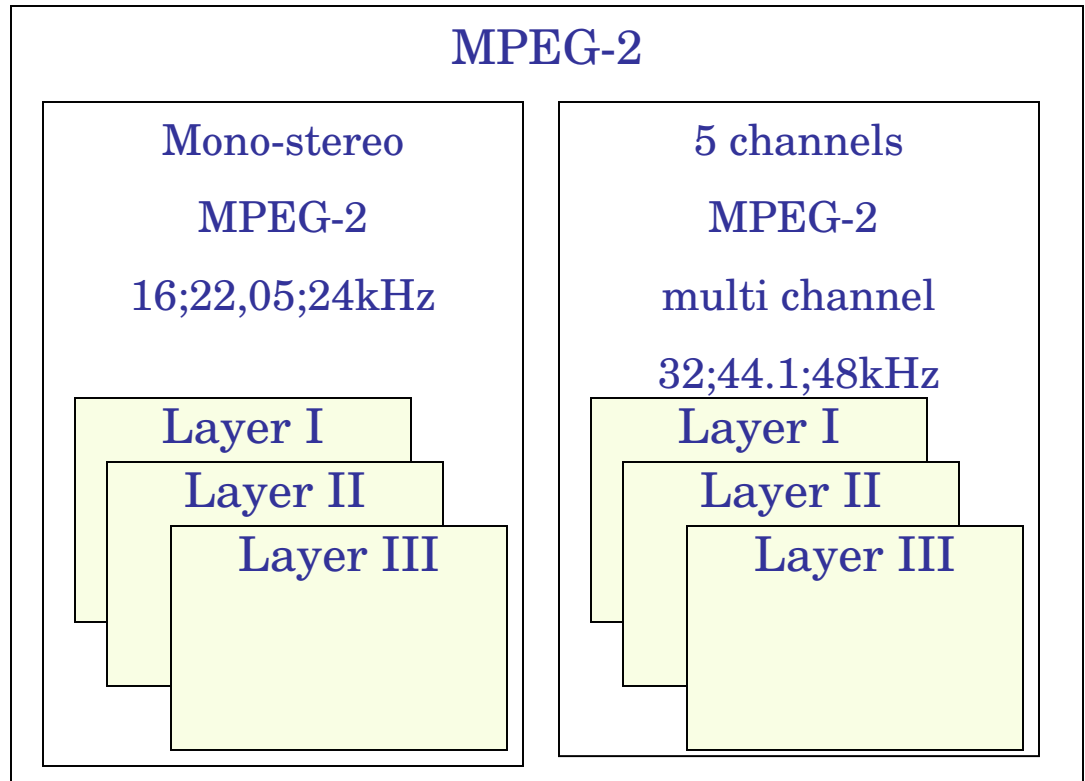
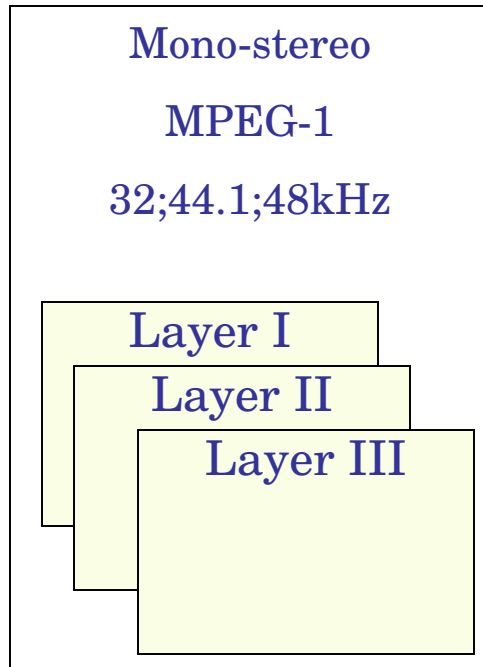
- Header: Gồm 12 bits đồng bộ; 20 bits thông tin hệ thống chỉ thị tốc độ bit
- CRC với đa thức sinh  $x^{16}+x^{15}+x^2+1$ .
- Side Info: Gồm phân bố bit: lớp I với 4 bits tuyến tính cho các băng con, lớp II 4 bits cho các băng con tần thấp, 3 bit tần trung và 2 bits tần cao; hệ số tỷ lệ là 6 bits/băng con kết hợp với phân bố bits và các bits mã hóa cho băng con đó để xác định giá trị, lớp III mã hóa âm thanh nổi.
- Bit Reservoir: Bit cung cấp, các mẫu dữ liệu từ 1 hoặc 2 khung trước.
- Samples: 32x12 mẫu đối với lớp I và 32x36 mẫu đối với lớp II và lớp III
- Ancillary Data: Dữ liệu bổ sung

# MPEG-2



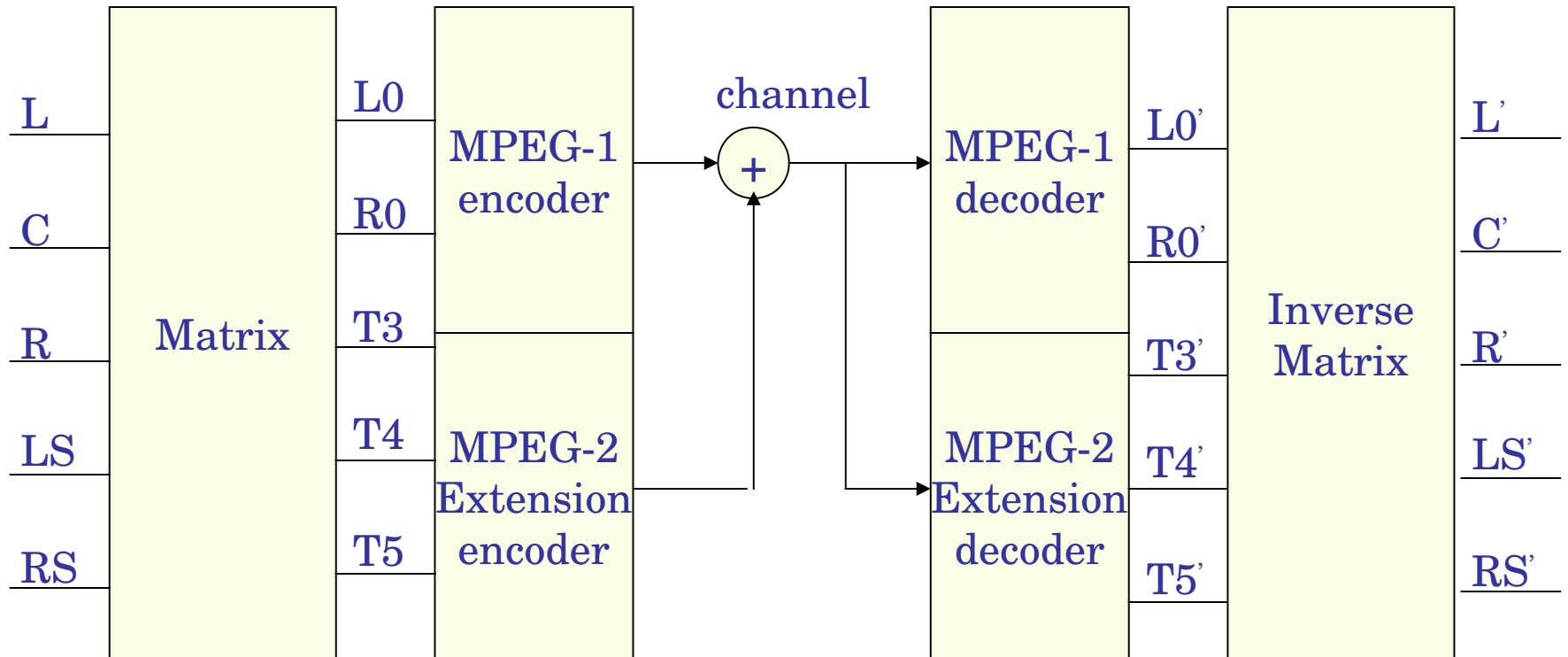
- Mở rộng MPEG-1 cho các ứng dụng mới.
- Có khả năng áp dụng nhiều tốc độ khác nhau, từ 32 đến 1066kbps. Tần số lấy mẫu có thể giảm 1 nửa so với MPEG-1 (16; 22,05; 24kHz).
- Khả năng đa kênh, tốc độ bits mở rộng có thể lên đến 1 Mbps cho các ứng dụng tốc độ cao. Cho phép nén đồng thời nhiều kênh.
- Chất lượng âm thanh tùy thuộc ứng dụng.
- Hỗ trợ khả năng lồng tiếng, bình luận nhiều ngôn ngữ trong phần bits mở rộng (7 kênh).
- MPEG-2 sử dụng mã hoá cường độ cao, giảm xuyên âm, mã hoá dự đoán liên kênh và mã hoá ảo ảnh kênh trung tâm để nhận được tốc độ bit kết hợp 384 kbps.
- Khung MPEG-2 được chia thành 2 phần, phần đầu là MPEG-1 stereo, phần mở rộng MPEG-2 chứa tất cả những dữ liệu surround khác.

# MPEG-2



# MPEG-2

- Mã hoá và giải mã MPEG-2





# MPEG-2



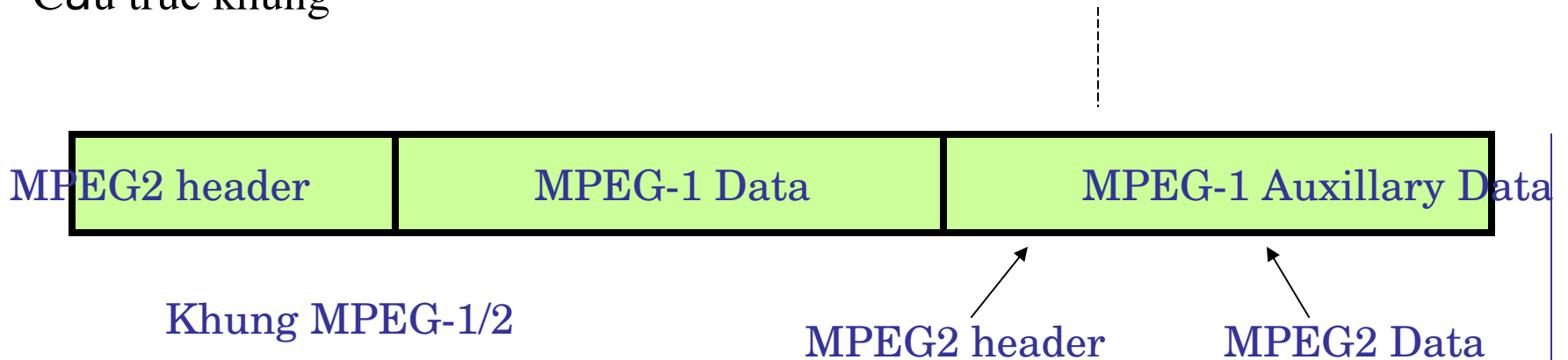
Trong đó:

$L0$  (L C LS) ,  $\begin{matrix} 1 \\ \text{---} \\ 1 \end{matrix} \begin{matrix} 2 \\ \text{---} \\ 2 \end{matrix}$  ;  $\begin{matrix} 1 \\ \text{---} \\ 1 \end{matrix} \begin{matrix} 2 \\ \text{---} \\ 2 \end{matrix}$  or 1; 0  
 $R0$  (R C RS)

Với R: phải, L: trái, C: trung tâm, LS:trái vòm và RS: phải vòm.

Từ đó dễ dàng trong mã hóa thuận nghịch

Cấu trúc khung



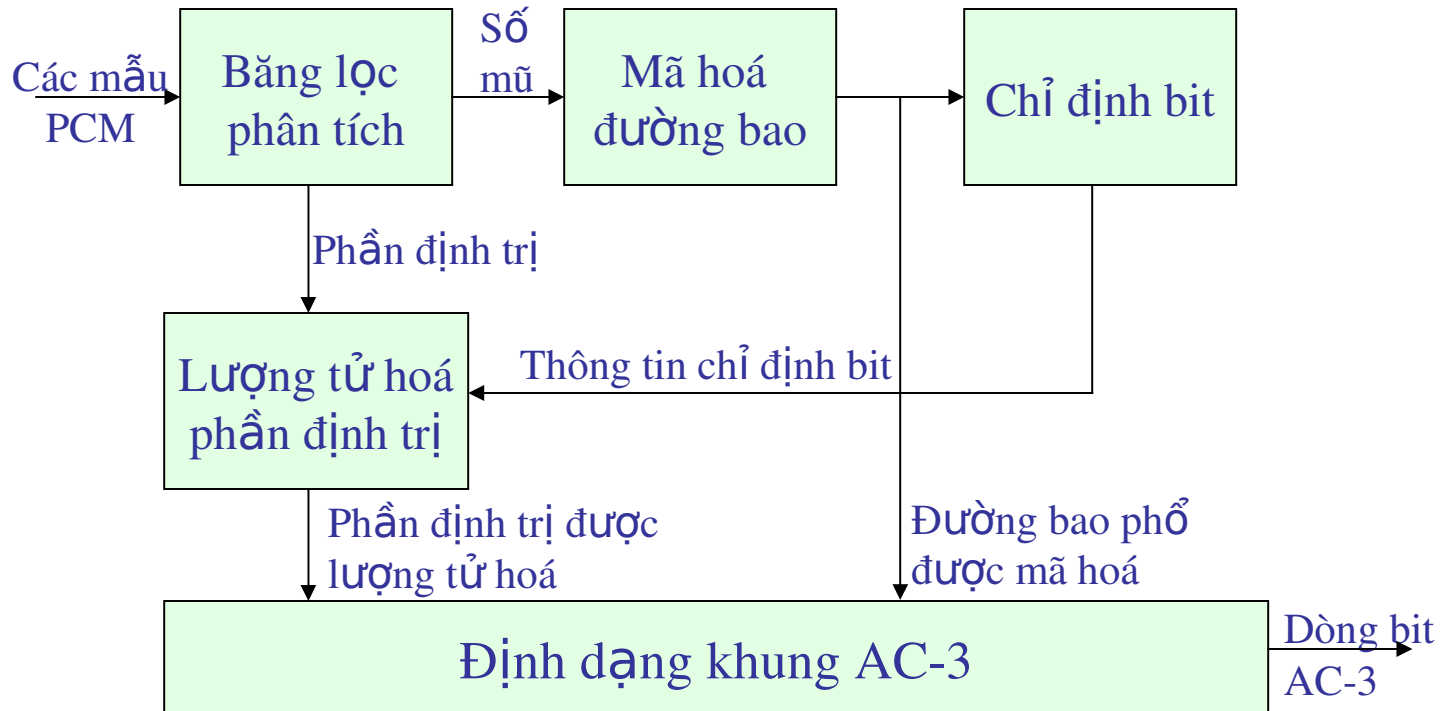
# AC-3 (DOLBY DIGITAL)



- Chuyển tải âm thanh đa kênh trong các ứng dụng như DVD-Video, DTV và DBS.
- Phát triển từ AC-1 và AC-2.
- Mã hoá âm thanh từ 1 đến 6 kênh, thông thường cung cấp âm thanh 5.1 kênh: trái, phải, trung tâm, trái vòm, phải vòm và 1 kênh hiệu ứng tần số thấp.
- 6 kênh yêu cầu  $6 \times 48 \text{kHz} \times 18 \text{bits} = 5.184 \text{Mbps}$  chưa nén có thể được mã hoá tối thiểu với tốc độ 384kbps (tỷ lệ 13:1).
- Tuy nhiên, AC-3 hỗ trợ tốc độ từ 32 đến 640kbps.
- Cung cấp khả năng tự chọn mức âm thanh cho thính giả.
- Cho phép giảm dữ liệu bằng quá trình lượng tử biểu diễn trong miền tần số của tín hiệu âm thanh.
- Bộ mã hoá sử dụng bảng lọc phân tích chuyển các mẫu PCM thành các hệ số trong miền tần số. Mỗi hệ số biểu diễn ký hiệu mũ nhị phân gồm phần số mũ và phần định trị. Các tập số mũ được mã hoá thô qua phổ tín hiệu và xem như là đường bao phổ. Sử dụng phân phối bit xác định số bits cần mã hoá mỗi định trị dựa vào đường bao phổ. Đường bao phổ và các định trị được lượng tử cho 6 khối âm thanh (1536 mẫu âm thanh) được định dạng thành khung rồi chuyển đi.

# AC-3 (DOLBY DIGITAL)

Quá trình giải mã hoàn toàn ngược lại, ở đây chỉ xét quá trình mã hoá



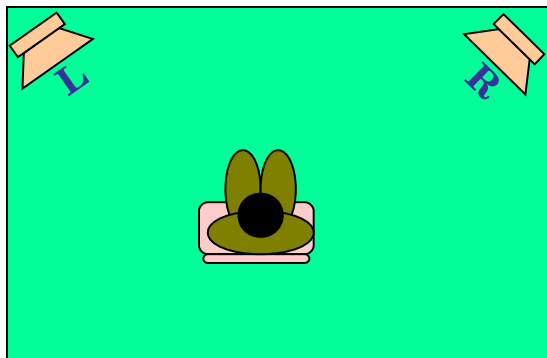
# APT-X100



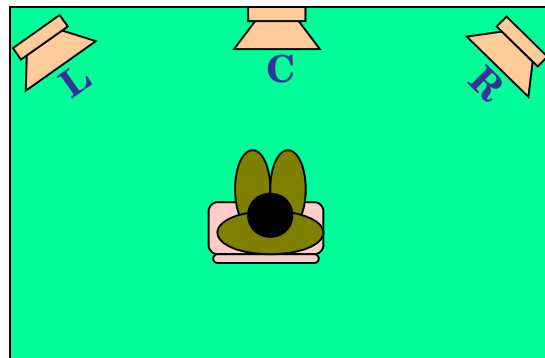
- APT-X100 cho tỷ lệ nén 4:1.
- Sử dụng để truyền dẫn, lưu trữ các tín hiệu audio mono, stereo hay đa kênh chất lượng cao.
- Không hẳn dựa vào mô hình tâm sinh lý nghe, cũng không trực tiếp loại các thành phần không thích hợp trong tín hiệu audio, mà nó ngấm hiểu một mô hình đáp ứng nghe bằng việc phân phối ít bit ở tần số cao.
- Thuật toán hoàn toàn trong miền thời gian sử dụng mã hoá dự đoán tuyến tính trong các băng con.
- Hoạt động với bất kỳ tần số lấy mẫu với ngõ ra 16 bits/từ mẫu.
- Tín hiệu audio được chia thành 4 băng con với băng thông đều nhau sử dụng các bộ lọc QMF.
- Tín hiệu được phân tích trong miền thời gian: sử dụng mã hoá dự đoán tuyến tính ADPCM để lượng tử mỗi băng theo nội dung và loại bỏ độ dư thừa trong các băng con.
- Mã hoá sự khác biệt của mẫu hiện thời và mẫu trước.
- Giải mã được tiến hành một cách ngược lại.

# MÃ HÓA ÂM THANH NỔI

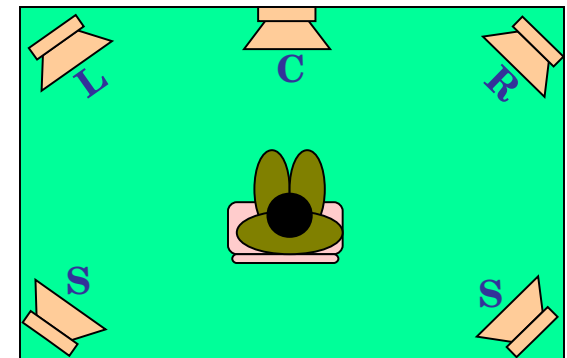
- Âm thanh đa kênh



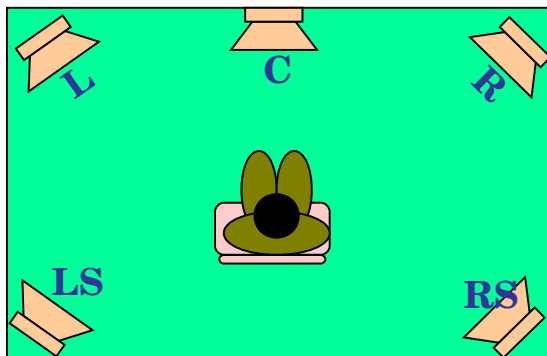
2/0 Stereo



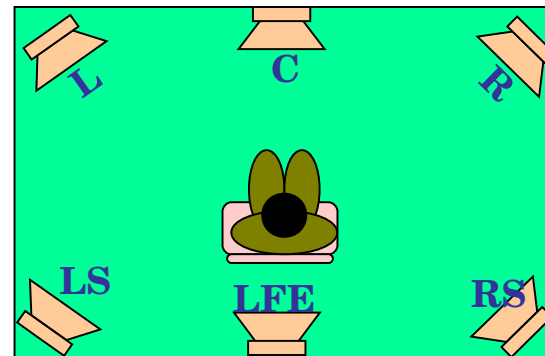
3/0



3/1



3/2



3/2 với siêu trầm (5.1)

S: Surround.  
LFE : Low Frequency  
Enhancement (Woofer)

15 – 120Hz  
Có thể đặt bất kỳ.

# MÃ HOÁ ÂM THANH NỔI

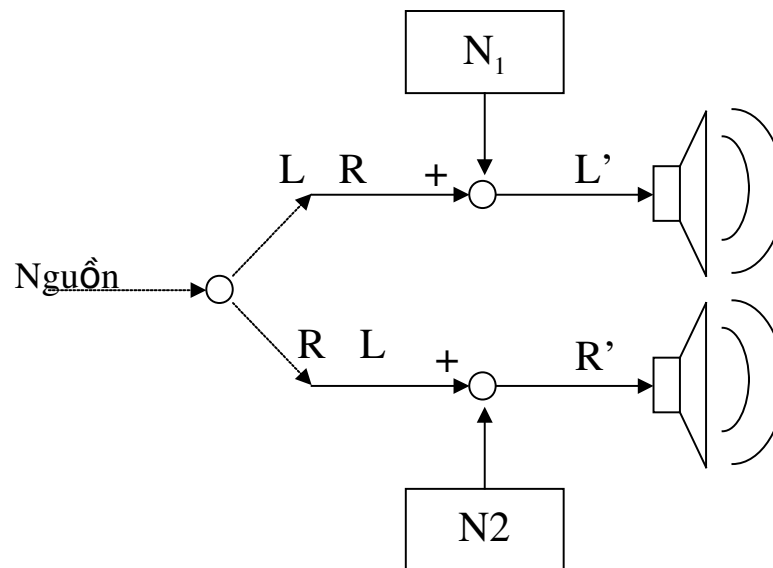
- **Mã hóa Trái phải (LR)**

Không loại bỏ độ dư thừa.

Các kênh riêng biệt được mã hoá độc lập.

Ngưỡng che không liên quan.

Hiệu quả với âm thanh rất khác biệt giữa kênh trái và kênh phải.



# MÃ HOÁ ÂM THANH NỔI

- **Joint Stereo Middle-Side**

Khuynh hướng của Joint Stereo không chỉ là kết hợp các định dạng của chuẩn nén mp3 (MPEG1 lớp III) mà còn kết hợp các dạng khác: MPEG và AAC.

Middle-Side Stereo xét hai kênh dữ liệu theo hai phương diện khác nhau. Thay vì lưu trữ một dữ liệu âm thanh theo hai kênh Left và Right, ta có thể chỉ cần lưu trữ một chuỗi tương tự số trung bình Average (của Left và Right) và sự sai biệt Difference (giữa Left và Right).

Ở dạng Middle-Side, ta có thể lưu trữ các giá trị Middle và Side như sau:

$$\text{Middle}=(L+R)/2, \text{ Side}=(L-R)/2.$$

Dấu của Side là rất quan trọng, nếu Side >0 nghĩa là tín hiệu L lớn hơn tín hiệu R.

Ta hoàn toàn có thể tái tạo hai kênh L, R như sau:

$$L=\text{Middle}+\text{Side}, R=\text{Middle}-\text{Side}.$$

Ưu điểm là sự khác biệt tương đối của các tín hiệu audio của các kênh L và R. Kết quả là kênh Middle lớn hơn nhiều so với kênh Side. Việc mã hoá kênh Side sử dụng ít bit hơn để giải phóng tài nguyên mà có thể được sử dụng để triển khai hữu hiệu hơn trên kênh Middle. Khi tái định dạng lại L, R thì kết quả sẽ thể hiện tín hiệu gốc ngõ vào thực hơn.

# MÃ HOÁ ÂM THANH NỔI

## – Tính chất

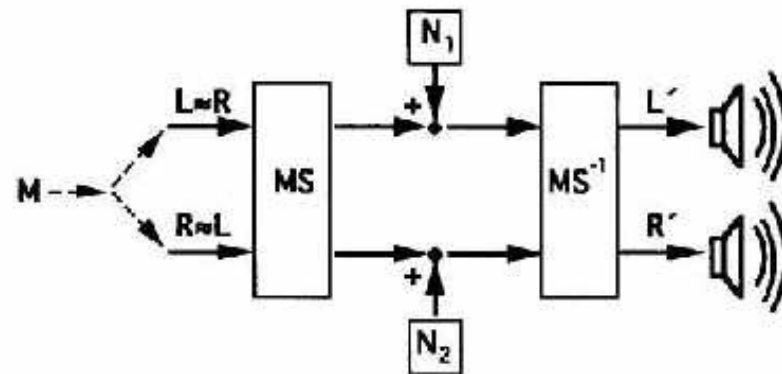
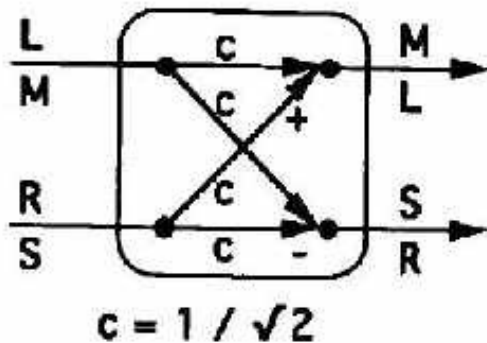
Kênh side sử dụng số bit rất ít.

Loại bỏ độ dư thừa cho tín hiệu mono trong thực tế.

Có thể được áp dụng trong miền thời gian lẫn tần số.

Độ lợi mã hoá cao phụ thuộc tín hiệu.

## – Biến đổi ngược: tổng/hiệu chuẩn hoá.





# MÃ HOÁ ÂM THANH NỔI

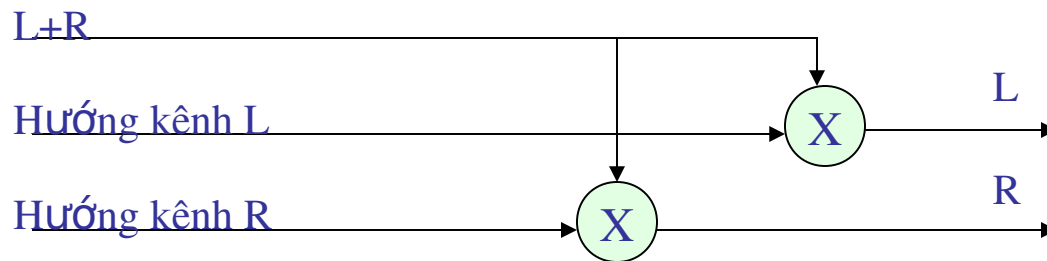
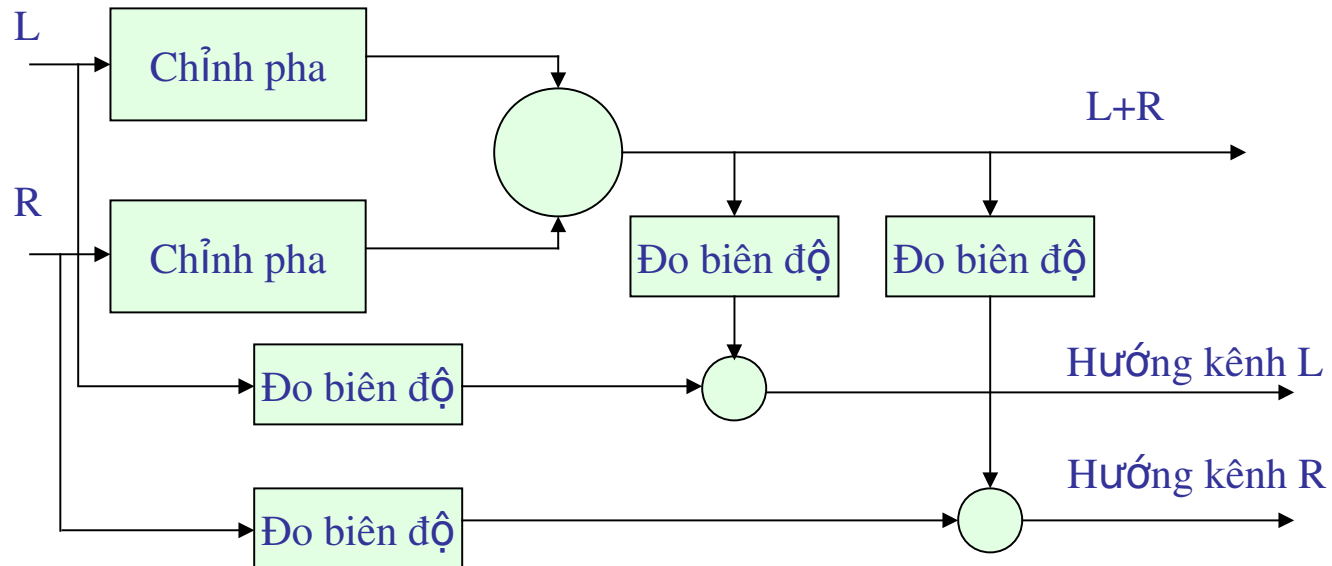


- **Stereo ghép cường độ cao (Intensity Stereo):**

**Mục đích:** Tối thiểu hoá thông tin stereo để nhận được tốc độ bit thấp nhất nếu có thể.

- Mã hoá tín hiệu tổng các kênh + các hướng của các kênh. Truyền đường bao, sau đó tỷ lệ theo các kênh.
- Kiểm chứng dựa trên việc cảm nhận của con người kém đối với tần số trên 3kHz.
- Không liên quan đến kỹ thuật suy giảm.
- Biên độ và pha là không quan trọng.
- Giảm gần 50% dữ liệu.
- Không đảm bảo thông tin về pha của tín hiệu.
- Một vài sự méo có thể nhận biết được.
- Sử dụng trong các ứng dụng tốc độ bit thấp.

# MÃ HOÁ ÂM THANH NỔI





# **VIDEO ENGINEERING**



# NỘI DUNG



- TỔNG QUAN
- KỸ THUẬT AUDIO
- KỸ THUẬT VIDEO

# Khái niệm Ảnh số



# Khái niệm



- **Ảnh số** bao gồm các nguyên tố ảnh gọi là **điểm ảnh - pixel**.
- **Số pixel** xác định **chất lượng của ảnh** (**độ phân giải – resolution**)
- **Chất lượng ảnh** càng cao, **độ phân giải** càng lớn.
- **Bit-map** là kiểu lưu ảnh trong bộ nhớ video theo cách thể hiện ra màn hình.

# Ảnh bit-map đơn sắc

- Mỗi pixel lưu trữ 1 bit với giá trị 0 hoặc 1.
- Một ảnh 600x480 đơn sắc yêu cầu lưu trữ trong 37.5KB.
- Kỹ thuật dither thường được sử dụng để hiển thị ảnh đơn sắc.



# Ảnh thang xám

- **Mỗi pixel lưu trữ 1 byte (8 bits), giá trị từ 0 đến 255.**
- **Một ảnh 600x480 yêu cầu lưu trữ trong 307.2KB.**





# Ảnh 8 bits màu

- Sử dụng 1 byte trong 1 pixel.
- Hỗ trợ 256 màu.
- Yêu cầu bảng tra màu LUT (Colour Look-Up Table).
- Một ảnh 600x480 yêu cầu lưu trữ trong 307.2KB (như thang xám)



# Ảnh 24 bits màu

- Sử dụng 3 byte (tương ứng với 3 màu RGB) trong 1 pixel.
- Hỗ trợ 16 777 216 màu.
- Một ảnh 600x480 yêu cầu lưu trữ trong 921.6KB.
- Thông dụng hơn ảnh 24 bits là ảnh 32 bits với 1 byte thêm vào cho giá trị alpha biểu diễn thông tin hiệu ứng đặc biệt.



# CƠ SỞ VIDEO

- Các dạng của tín hiệu Video màu

- **Tín hiệu video thành phần:**

- Các tín hiệu video thành phần được xử lý riêng lẻ.

- Tập các thành phần RGB: Là các tín hiệu cơ bản video màu được camera cung cấp. Ba tín hiệu màu có cùng độ rộng băng tần.

- Tập các thành phần Y, R-Y, B-Y: Là tổ hợp của các giá trị màu cơ bản. Thông thường tín hiệu Y có băng tần rộng hơn hai tín hiệu R-Y và B-Y.

- Tái tạo hình ảnh tốt nhất nhưng yêu cầu băng thông và đồng bộ tốt cho các thành phần.

- **Tín hiệu video tổng hợp:**

- Là tín hiệu video mà trong đó thông tin độ chói (luminance), màu (chrominance) và đồng bộ (synchronization) được phối hợp với nhau (theo tần số, thời gian và biên độ) để tạo ra một tín hiệu duy nhất.

- Phổ năng lượng tập trung vào hài của tần số tín hiệu quét dòng.

- Cho phép quét cách dòng để nhận được cảm nhận tốt hơn.

- Yêu cầu băng thông nhỏ, không đồng bộ nhưng xử lý khó khăn..

- **S-Video:**

- Gồm 2 đường, một đường cho tín hiệu video tương tự và đường cho sự kết hợp các thành phần chói.

# CƠ SỞ VIDEO

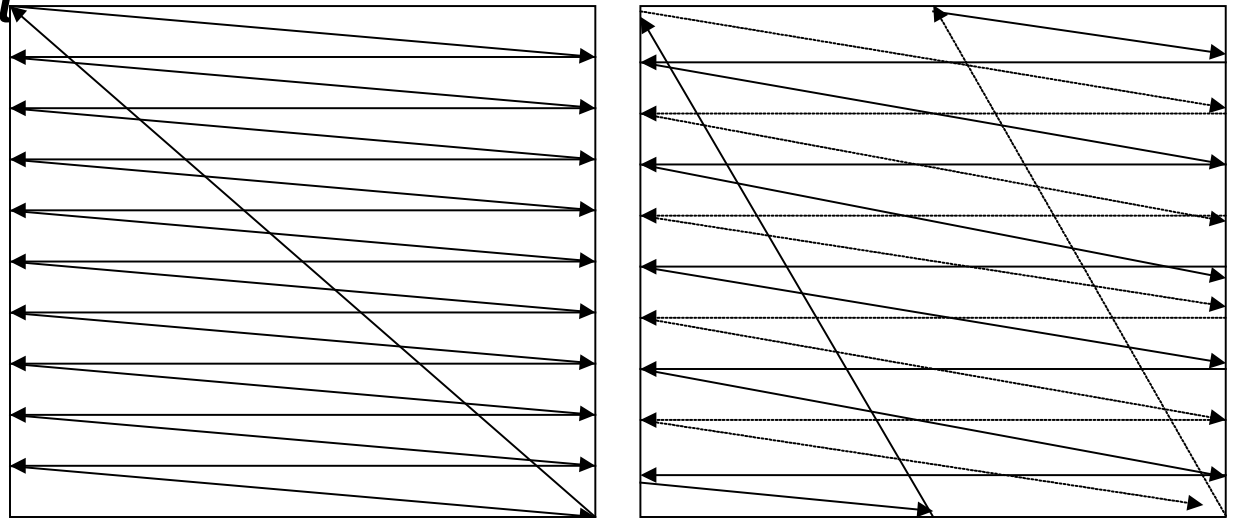
- **Video tương tự**

- Thông tin về cảnh vật truyền đi mang các tính chất về độ chói, màu sắc và sự thay đổi theo thời gian.
- Một tín hiệu video bao gồm các ảnh theo trình tự thời gian, mỗi ảnh bao gồm điểm ảnh. Các điểm ảnh mang thông tin về độ chói và màu sắc của vật thể, cảnh quan.

- **Kỹ thuật quét**

Quét liên tục: Tần số quét lớn.

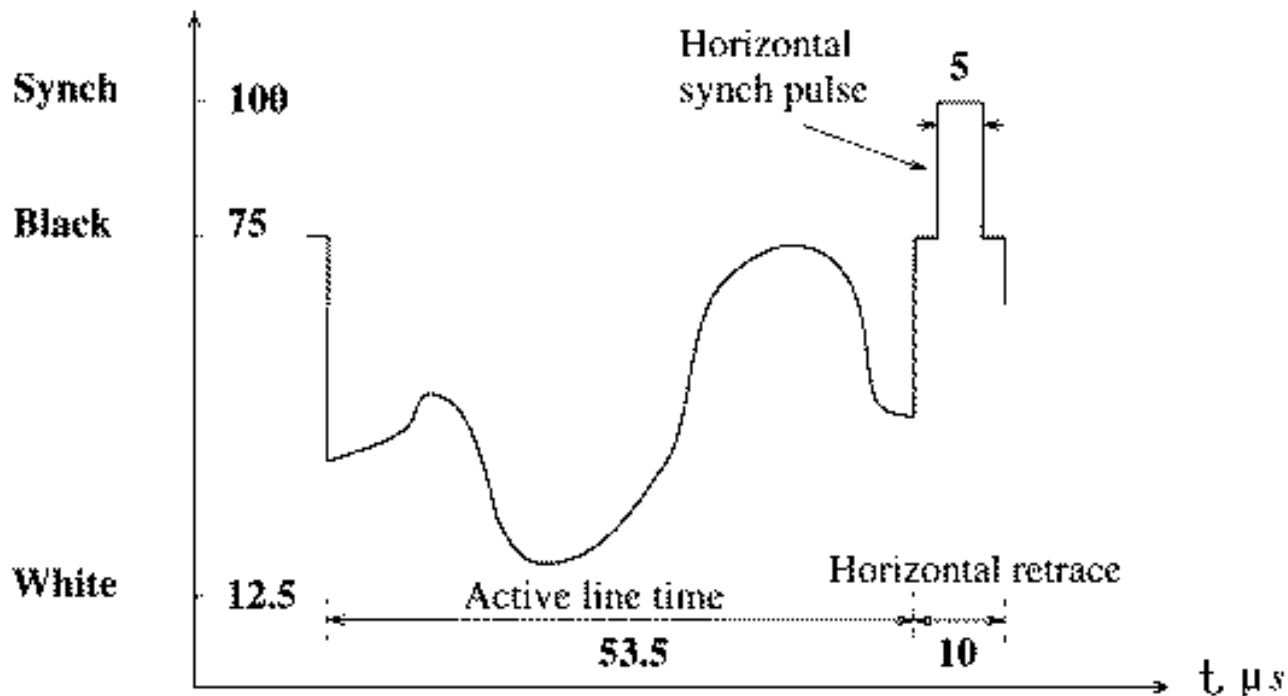
Quét cách dòng: Giảm được tần số quét dòng nhưng vẫn đảm bảo được cảm nhận liên tục, không bị trôi, nhấp nháy



Hình 3-1 Quét liên tục và quét cách dòng

# CƠ SỞ VIDEO

- Đặc điểm của video tương tự
  - Tín hiệu đơn cực, mức 1 chiều (DC=0V) biểu diễn mức đen và 0.7V biểu diễn mức trắng. -25mV biểu thị mức xoá.



Hình 3-2 Tín hiệu video tương tự điều chế âm

# CƠ SỞ VIDEO

- Các tiêu chuẩn video màu

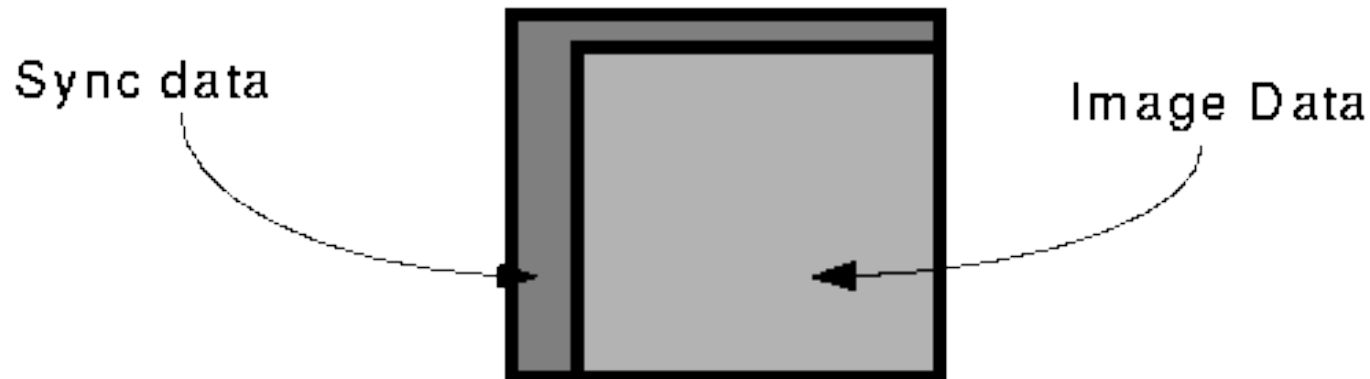
- *NTSC Video:*

- 525 dòng trên một khung, 30 khung hình trong một giây.

- Quét cách dòng. Chia làm hai trường (mỗi trường 262.5 dòng).

- Có 20 dòng dự trữ cho thông tin điều khiển tại thời điểm bắt đầu mỗi trường.

- Phù hợp vì độ phân giải của Laser Disk và S-VHS là 420 và TV thông thường có độ phân giải khoảng 320 dòng.



Hình 3-3 Tín hiệu video với thông tin điều khiển

# CƠ SỞ VIDEO

Biểu diễn màu:

NTSC sử dụng kiểu màu YIQ.

Thành phần tổng hợp  $= Y + I \cos(f_{sc} t) + Q \sin(f_{sc} t)$ .

Trong đó,  $f_{sc}$  là tần số của sóng mang phụ của màu.

– ***PAL Video:***

625 dòng trên một khung, 25 khung hình trong một giây.

Quét cách dòng.

Khung gồm hai trường chẵn lẻ, mỗi trường 312.5 dòng.

Sử dụng kiểu màu YUV.

Dải tần tín hiệu chói Y rộng 5MHz. Tín hiệu U và V được xác định theo công thức:

$$V = 0.877(R - Y) = 0.615R - 0.515G - 0.100B$$

$$U = 0.493(B - Y) = -0.147R - 0.293G + 0.473B$$

# CƠ SỞ VIDEO



- **Video số**

- **Thuận lợi:**

- Truy cập ngẫu nhiên trực tiếp thuận tiện,
    - Việc tạo, lưu trữ và ghi đọc nhiều lần không ảnh hưởng đến chất lượng ảnh.
    - Không cần xung xoá và xung đồng bộ.
    - Xử lý thuận tiện, không gặp trở ngại về giới hạn tần số, băng thông.

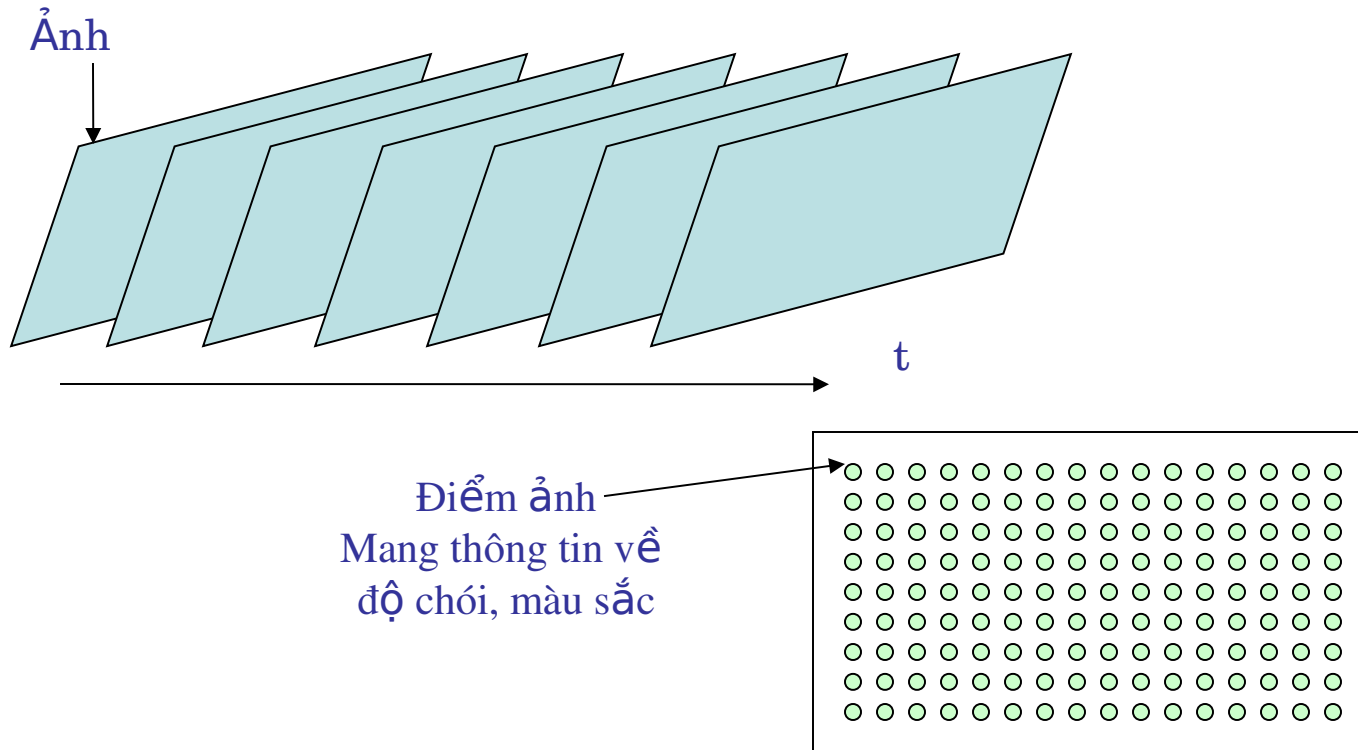
- **Khó khăn**

- Tuy nhiên, tín hiệu số gặp một số trở ngại xoay quanh vấn đề về tính hiệu quả, chẳng hạn bộ lọc số có giá thành tương đối cao.



# CƠ SỞ VIDEO

- Tín hiệu video số



Hình 3-5 Video số

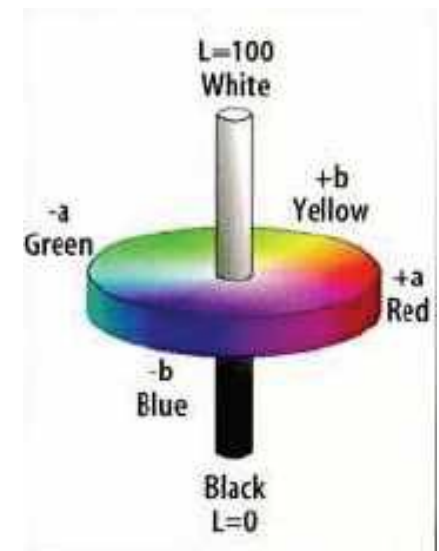
# CƠ SỞ VIDEO

- **Không gian màu**

- Không gian màu được hình thành bởi các thành phần cường độ sắc R, G, B.

- **Mô hình màu LAB**

- Bao gồm:
- L: luminant,
- A: Phạm vi từ màu Green đến Red.
- B: Phạm vi từ Blue đến Yellow.



# CO' SO' VIDEO



Original Color Image



L, A, B Images

# CƠ SỞ VIDEO

- **Mô hình YUV (CCIR601 hoặc YCrCb)**

- Được thiết lập năm 1982 cho các chuẩn video số.
- Tín hiệu video được chia thành 2 trường chẵn và lẻ. Hai trường này kết hợp với nhau thành 1 khung.
- Ứng dụng tốt với NTSC và PAL.
- Sử dụng không gian màu YCrCb hay YUV.

$$Y = 0.299R + 0.587G + 0.114B \quad Cr = R - Y \quad Cb = B - Y$$

- Chuyển đổi YUV:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

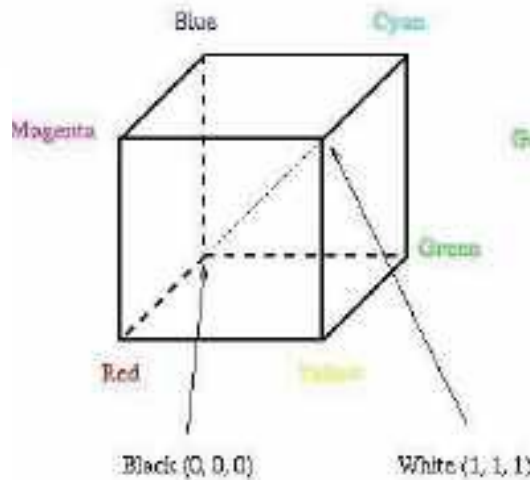
# CƠ SỞ VIDEO

- **Mô hình CMY**

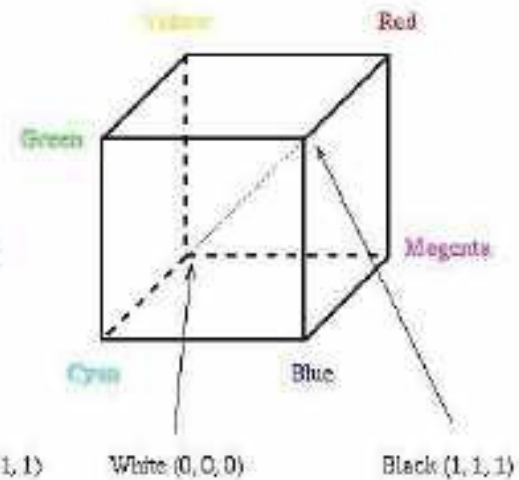
- Cyan, Magenta, Yellow là các màu bù của Red, Green, Blue.
- Thường được sử dụng trong các máy in.

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} C \\ M \\ Y \end{bmatrix}$$



The RGB Cube



The CMY Cube

# CƠ SỞ VIDEO

- **Mô hình CMYK**

- Được sử dụng trong các máy in.
- K: đại diện cho màu đen Black.
- Thường cho các bản in với màu đen đậm hơn mô hình CMY.

$$K = \min(C, M, Y),$$

$$C = C - K,$$

$$M = M - K,$$

$$Y = Y - K.$$

# CƠ SỞ VIDEO



Original Color Image



C, M, Y, K images

# CƠ SỞ VIDEO

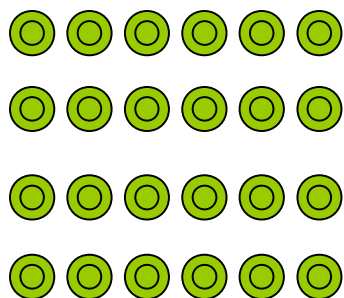


## – Tiêu chuẩn lấy mẫu màu:

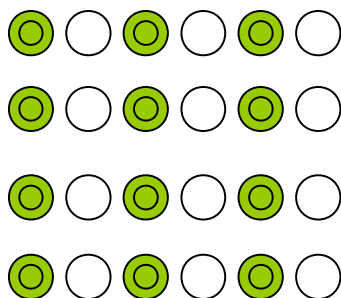
- Thuận lợi trong việc xử lý đối với tín hiệu video thành phần, nhưng băng thông yêu cầu lớn.
- Điểm khác nhau chủ yếu của các tiêu chuẩn lấy mẫu là ở tỷ lệ giữa tần số lấy mẫu và phương pháp lấy mẫu tín hiệu chói và các tín hiệu màu.
- Tần số chuẩn là 3,375MHz. Mẫu tín hiệu được lấy chỉ đối với phần tử tích cực của tín hiệu video. Cấu trúc lấy mẫu là trực giao.



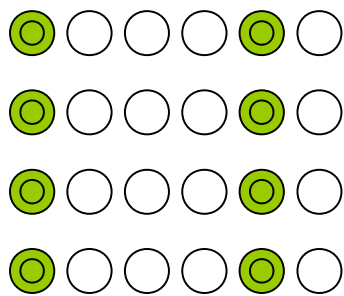
# CƠ SỞ VIDEO



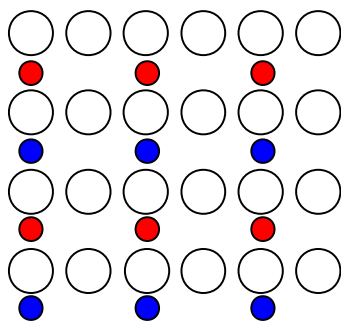
4:4:4



4:2:2



4:1:1



4:2:0

● Lấy mẫu Y, C<sub>R</sub>,  
● Lấy mẫu C<sub>R</sub>

○ Lấy mẫu Y  
● Lấy mẫu C<sub>B</sub>

## 4:4:4:

Tần số lấy mẫu màu cho các thành phần Y, C<sub>R</sub>, C<sub>B</sub> là 13.5MHz.

## 4:2:2:

Y: 13.5MHz, C<sub>R</sub>, C<sub>B</sub> là 6.75MHz.

Khi giải mã, màu điểm ảnh sau được suy từ điểm ảnh trước.

## 4:1:1:

Y:13.5MHz, C<sub>R</sub>, C<sub>B</sub> là 3.375MHz. Lấy mẫu màu liên tục từng dòng

## 4:2:0:

Y:13.5MHz, C<sub>R</sub>, C<sub>B</sub> là 3.375MHz. Lấy mẫu màu xen kẽ cách dòng.

Hình 3-4 Các tiêu chuẩn lấy mẫu màu

# CƠ SỞ VIDEO

- Các tiêu chuẩn video số của CCIR (Consultative Committee for International Radio)

	CCIR 601525/60 NTSC	CCIR 601625/50 PAL/SECAM	CIF	QCIF
Độ phân giải chói	<b>720x480</b>	<b>720x576</b>	<b>352x288</b>	<b>176x144</b>
Độ phân giải sắc	<b>360x480</b>	<b>360x576</b>	<b>176x144</b>	<b>88x72</b>
Lấy mẫu màu	<b>4:2:2</b>	<b>4:2:2</b>	<b>4:2:0</b>	<b>4:2:0</b>
Số trường /s	<b>60</b>	<b>50</b>	<b>30,15,10,7.5</b>	<b>30,15,10,7.5</b>
Cách quét	<b>Cách dòng</b>	<b>Cách dòng</b>	<b>Liên tục</b>	<b>Liên tục</b>

# JPEG

## • Khái niệm

• Chuẩn JPEG mô tả một họ kỹ thuật nén ảnh cho tone liên tục (mức xám hay màu) của ảnh. JPEG khai thác độ dư thừa sinh lý thị giác trong ảnh. JPEG được tiến hành vào tháng 3 năm 1986, đến tháng 1 năm 1988 thì JPEG được nhất trí với giải pháp DCT thích nghi để cải thiện và tăng cường ảnh.

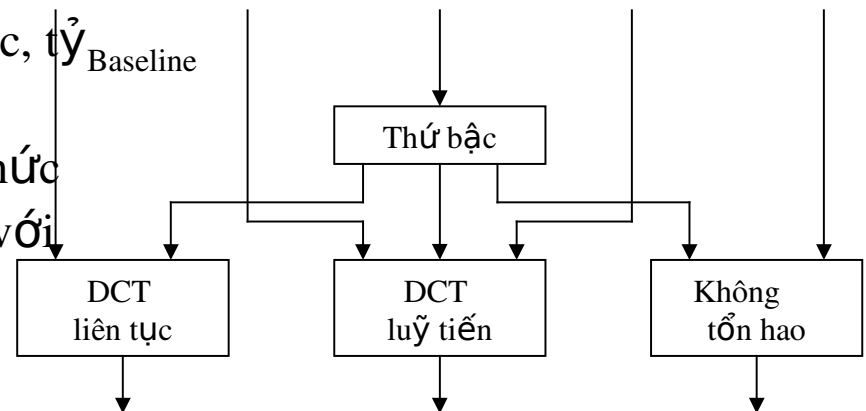
• DCT liên tục: Ảnh được mã hoá từ trái sang phải, từ trên xuống dưới dựa vào DCT.

• DCT luỹ tiến: Ảnh mã hoá quét phức hợp theo chế độ phân giải không gian cho các ứng dụng băng hẹp

• Không tổn hao: Khôi phục chính xác, tỷ lệ nén thấp.

• Thứ bậc (phân cấp): Mã hoá quét phức hợp phân giải không gian, hiệu quả với những ảnh có độ phân giải cao.

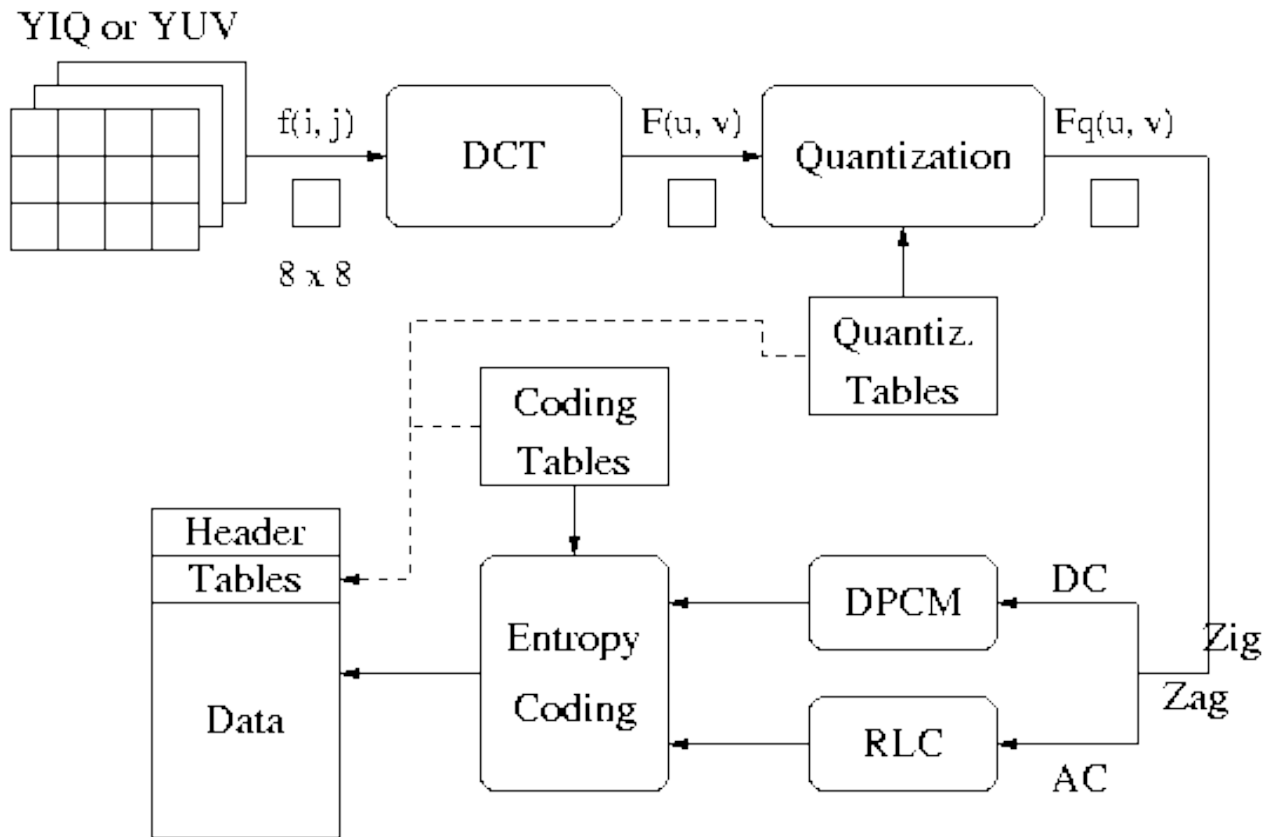
**Tài liệu này chỉ xét DCT liên tục.**



*Hình 3-6 Các kiểu hoạt động của JPEG*

# JPEG

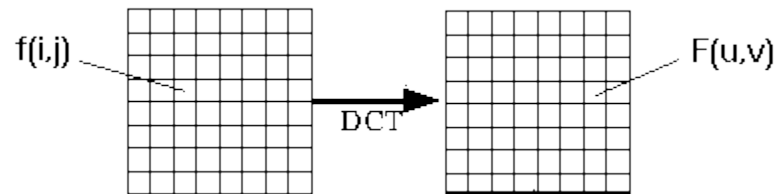
- Mã hoá JPEG



Hình 3-7 Mã hoá JPEG

# JPEG

- Chuyển đổi Cosin rời rạc DCT (Discrete Cosine Transform)



*Hình 3-8 DCT*

- DCT thuận:

$$F(u, v) = \frac{\Lambda(u)\Lambda(v)}{4} \sum_{i=0}^7 \sum_{j=0}^7 \cos \frac{(2i+1) \cdot u\pi}{16} \cdot \cos \frac{(2j+1) \cdot v\pi}{16} \cdot f(i, j)$$
$$\Lambda(\xi) = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } \xi = 0 \\ 1 & \text{otherwise} \end{cases}$$

- DCT ngược:

$$\hat{f}(i, j) = \frac{1}{4} \sum_{u=0}^7 \sum_{v=0}^7 \Lambda(u)\Lambda(v) \cos \frac{(2i+1) \cdot u\pi}{16} \cdot \cos \frac{(2j+1) \cdot v\pi}{16} \cdot F(u, v)$$
$$\Lambda(\xi) = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } \xi = 0 \\ 1 & \text{otherwise} \end{cases}$$

# JPEG



• Ví dụ

98	92	95	80	75	82	68	50
97	91	94	79	74	81	67	49
95	89	92	77	72	79	65	47
93	87	90	75	70	77	63	45
91	85	88	73	68	75	31	43
89	83	88	71	66	73	59	41
87	81	84	69	64	71	57	39
85	79	82	67	62	69	55	37

DC



591	106	-18	28	-34	14	18	3
35	0	0	0	0	0	0	0
-1	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
-1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
-1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

$$F(0,0) = \frac{1}{8} \sum_{i=0}^7 \sum_{j=0}^7 f(i,j) = 591$$

- Thành phần DC:
- Các giá trị còn lại là thành phần AC
- Nhận xét:

DCT làm giảm độ tương quan không gian của thông tin trong khối, → biểu diễn DCT có độ dư thừa thông tin ít hơn, đồng thời, DCT chứa thông tin về nội dung tần số không gian của thông tin trong khối, dựa vào đặc tính sinh lý thị giác ta chỉ mã hoá những hệ số DCT quan trọng, → nén.

# JPEG



- Lượng tử hoá
  - Lượng tử các hệ số  $F(u,v)$  để giảm số bits.
  - Các hệ số tương ứng với các tín hiệu tần số thấp là các giá trị lớn nên phải được lượng tử chính xác.
  - Các hệ số tương ứng với các tín hiệu tần số cao (AC) có giá trị bé nên cho phép sai số, ta có thể chia với một số lớn hơn.
  - Lượng tử hoá thay đổi theo khoảng cách để đạt được hiệu quả nén cao.

$$F_q(u,v) = \text{round} \left( \frac{F(u,v) + \frac{q(u,v)}{2}}{q(u,v)} \right)$$

The diagram illustrates the quantization process. On the left, the original coefficient  $F(u,v)$  is shown with a horizontal bar and a vertical bar, both labeled  $q(u,v)$ , indicating the quantization step. This is followed by the word "round". On the right, the quantized coefficient is shown as  $F(u,v) + \frac{q(u,v)}{2}$  divided by  $q(u,v)$ . The numerator is also shown with a horizontal bar and a vertical bar, both labeled  $q(u,v)$ .

- Trong đó,  $q(u,v)$  là giá trị trong bảng lượng tử hoá 8x8, tùy thuộc vào kênh chói hay kênh sắc.
- Mắt người ít cảm nhận được các nội dung ở tần số cao và càng kém đối với các nội dung tần số cao của kênh sắc

# JPEG



- Ma trận lượng tử kênh chói và kênh sắc:

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	14	16	24	40	57	69	56
14	17	22	29	51	57	69	56
18	12	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Ma trận lượng tử kênh chói

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

Ma trận lượng tử kênh sắc

- Nhận xét:*

Hệ số lượng tử  $q(u,v)$  càng xa thì có giá trị càng lớn,  $\rightarrow$  kết quả của các thành phần được lượng tử có tần số càng cao càng tiến về 0.

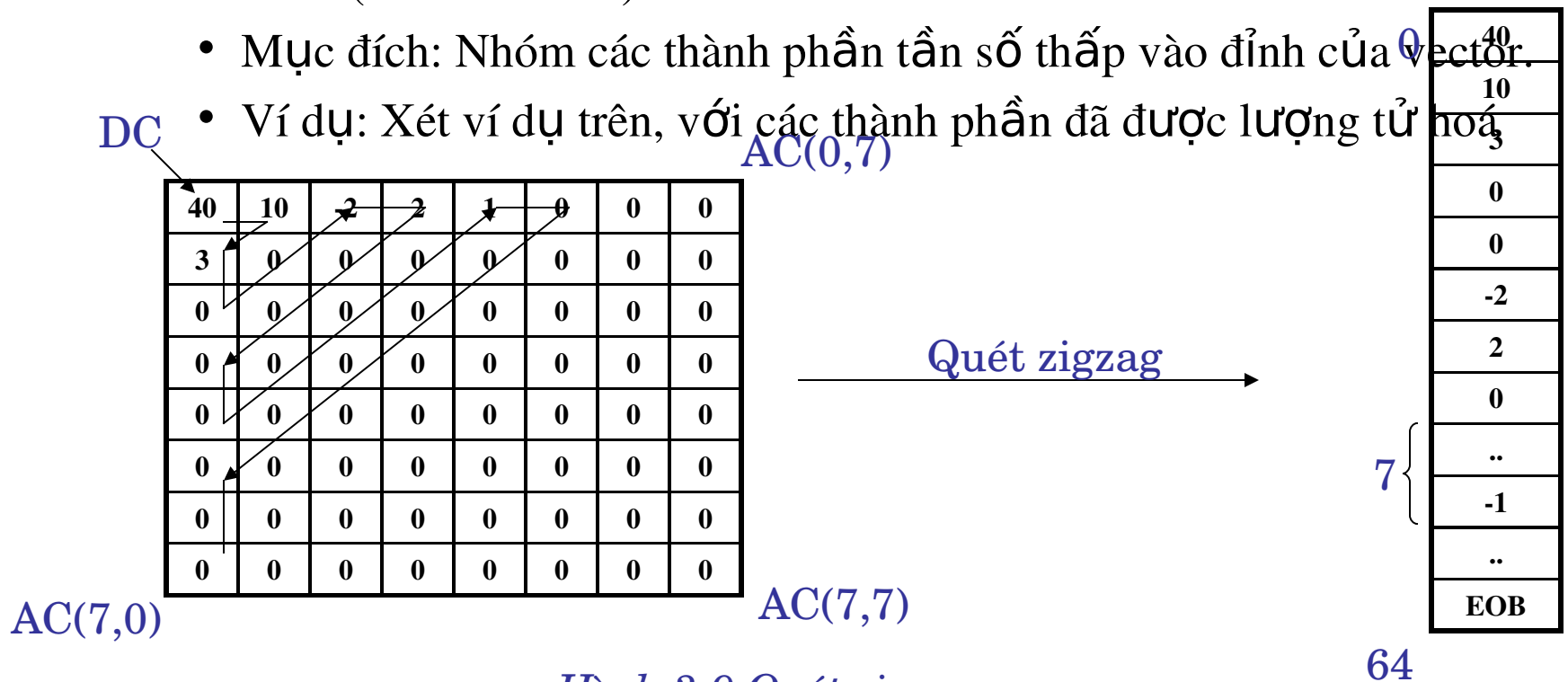
Đây là quá trình tổn hao thông tin duy nhất trong nén ảnh không tổn hao.



# JPEG

- Quét zigzag

- Ánh xạ ma trận 8x8 thành vector 1x64, đáy của vector là giá trị EOB (End of Block).
- Mục đích: Nhóm các thành phần tần số thấp vào đỉnh của vector.
- Ví dụ: Xét ví dụ trên, với các thành phần đã được lượng tử hoá



Hình 3-9 Quét zigzag

# JPEG

- Điều chế xung mã sai biệt trên các thành phần DC

- Thành phần DC là thành phần đầu sau quét zigzag.

- Giá trị của thành phần DC là lớn, thay đổi nhưng gần với giá trị của block trước đó.

- ✕ → Điều chế DPCM cho thành phần DC.

- Mã hoá loạt dài cho thành phần AC

- Trừ thành phần DC, các thành phần còn lại là AC gồm nhiều giá trị 0 liên tiếp.

- Mã hoá loạt dài giá trị 0 sẽ mang lại hiệu quả rất cao.

- Ví dụ: Giả sử block trước có giá trị của

15	DPCM cho tp DC
0,10	Không có bước chạy 0 trước giá trị 10
0,3	
2,-2	Có 2 bước chạy 0 trước giá trị -2
0,2	
7,-1	
EOB	

Kết thúc khối

*Hình 3-10 Mã hoá các thành phần DC và AC*

# JPEG



- Mã hoá Entropy

- Mã hoá Entropy sử dụng kỹ thuật mã hoá Huffman với các bảng mã hoá gồm bảng phân loại và bảng Huffman dựa vào đặc tính thống kê của tín hiệu.

- **Mã hoá Entropy cho thành phần DC:**

Phân loại	Từ mã (chói)
8	1111 110
7	1111 10
6	1111 0
5	1110
4	110
3	101
2	01
1	00
0	100

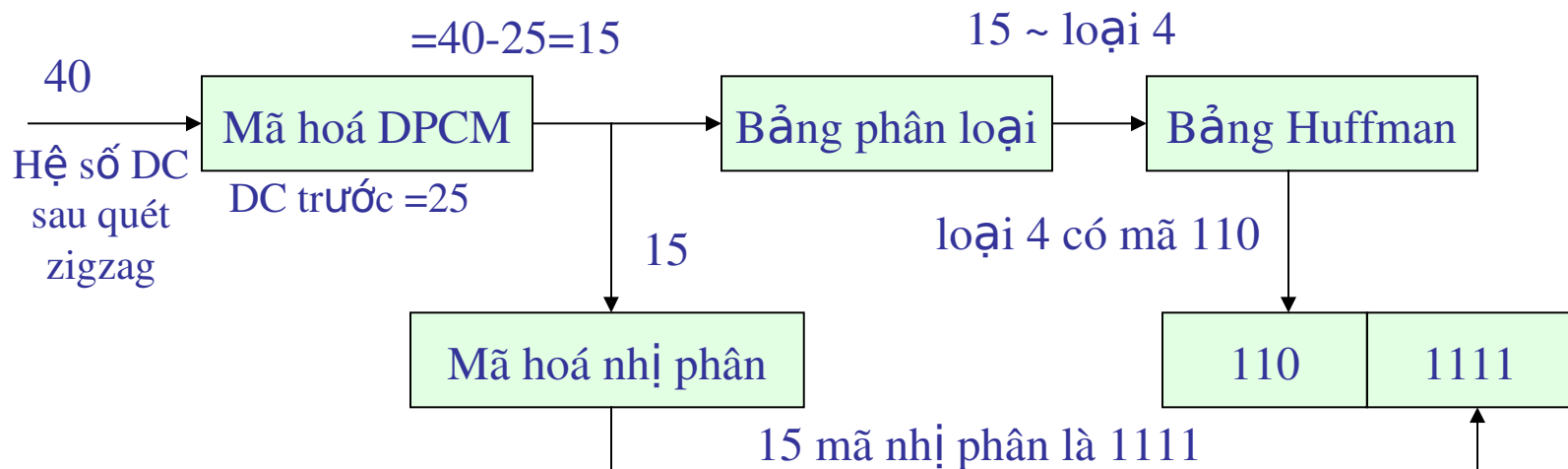
*Hình 3-11 Bảng phân loại và bảng Huffman cho thành phần DC chói*

# Phân loại DC, AC

Các hệ số DC sai lệch	Phân loại
-255...-128; 128...255	8
-127...-64; 64...127	7
-63...-32; 32...63	6
-31...-16; 16...31	5
-15...-8; 8...15	4
-7...-4; 4...7	3
-3;-2; 2;3	2
-1;1	1
0	0

# JPEG

- Sơ đồ mã hoá



- *Hình 3-12 Mã hoá entropy thành phần DC*

## • Mã hoá các thành phần AC

- Các từ mã với độ dài thay đổi có tần suất xuất hiện cao được mã hoá ngắn và ngược lại. Quá trình mã hoá như vậy gọi là mã hoá độ dài từ mã thay đổi VLC.
- Bảng phân loại giống như bảng phân loại của thành phần DC nhưng khác bảng Huffman.

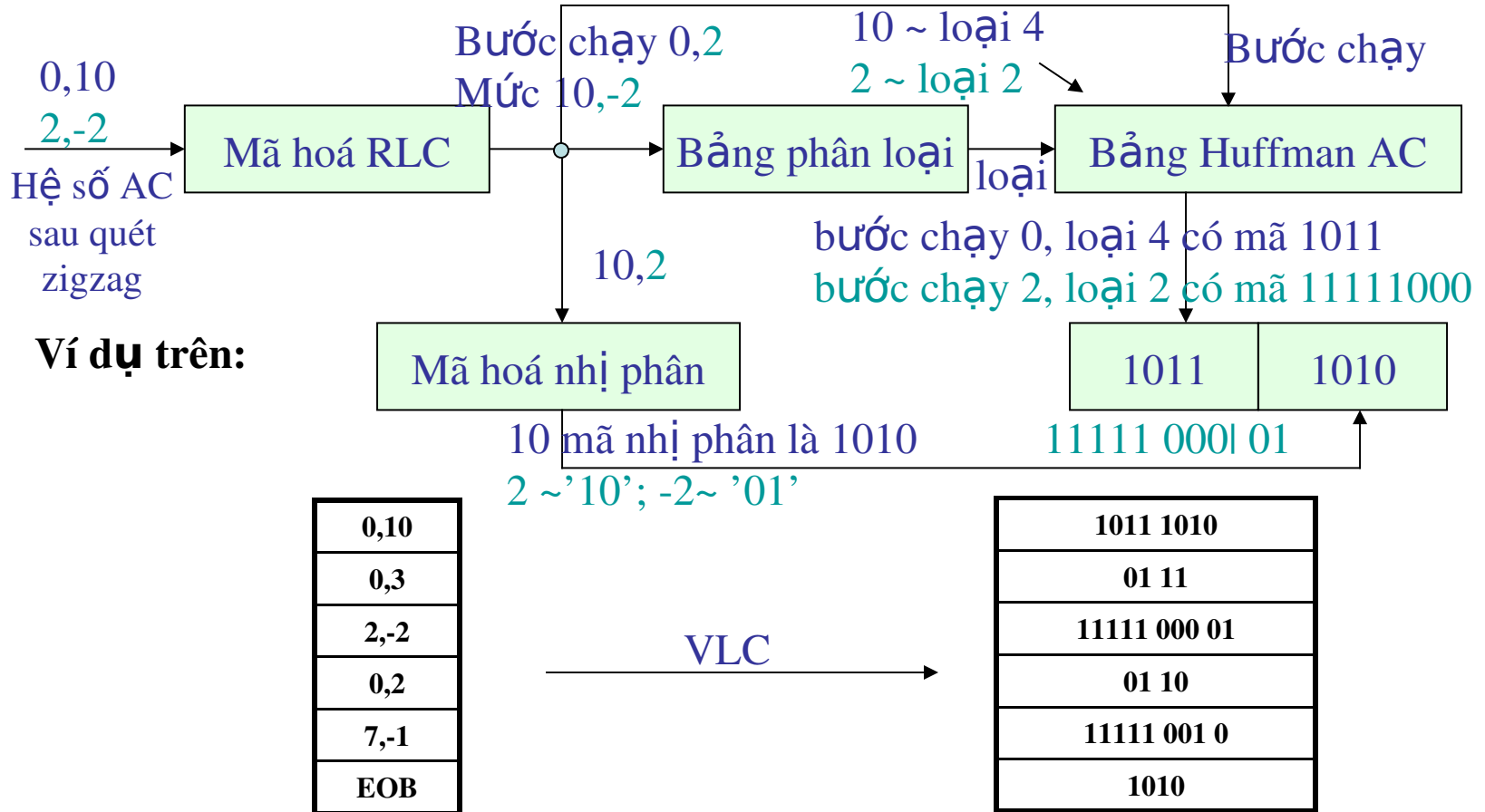
# JPEG



Bước chạy	Phân loại	Độ dài mã	Từ mã
0	1	2	00
0	2	2	01
0	3	3	100
0	4	4	1011
1	1	4	1100
1	2	6	111001
2	1	5	11011
2	2	8	11111 000
3	1	6	111 010
4	1	6	111 011
5	1	7	1111 010
6	1	7	1111 011
7	1	8	11111 001
EOB		4	1010

Hình 3-13 Bảng Huffman các hệ số AC

# JPEG



• Ví dụ trên:

Hình 3-14 Mã hoá entropy thành phần AC

# NÉN VIDEO





# Nén video

- Tín hiệu video nguyên gốc chiếm dung lượng rất lớn, yêu cầu tốc độ bit cao (hàng Gbps) → khó khăn trong lưu trữ và truyền dẫn.
- Ví dụ: HDTV: độ phân giải  $1920 \times 1080$  với 30 khung hình/s và 8bits mã hóa cho 1 kênh YCrCb (hệ PAL) yêu cầu tốc độ 1.5Gbps.
- Các phương thức mã hóa không tổn hao (Huffman, LZW...) không có tỷ lệ nén cao để phù hợp với tín hiệu image và video → mã hóa tổn hao

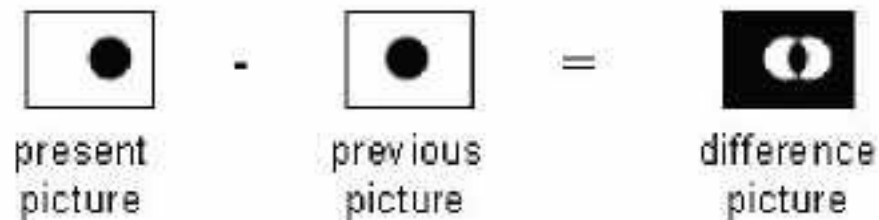
# Ý tưởng nén video



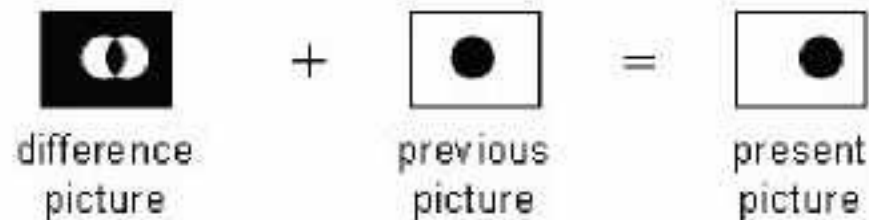
- Mã hóa độ dư thừa thống kê không gian (mã hóa trong khung, giống JPEG).
- Mã hóa trong miền thời gian liên khung: độ nén lớn hơn do có độ tương quan lớn giữa các khung liên tiếp nhau. Đặc biệt lưu ý đến sự khác biệt giữa các khung.
- Ý tưởng chính là: Mã hóa bù chuyển động kết hợp với mã hóa dư thừa thống kê không gian.

# Ví dụ mã hóa bù chuyển động

- Ý tưởng dựa trên kỹ thuật DPCM.
- Xét ảnh chấm bi di chuyển
- Xác định sự khác biệt giữa khung hiện hành và khung trước đó.
- Thực hiện mã hóa và giải mã.



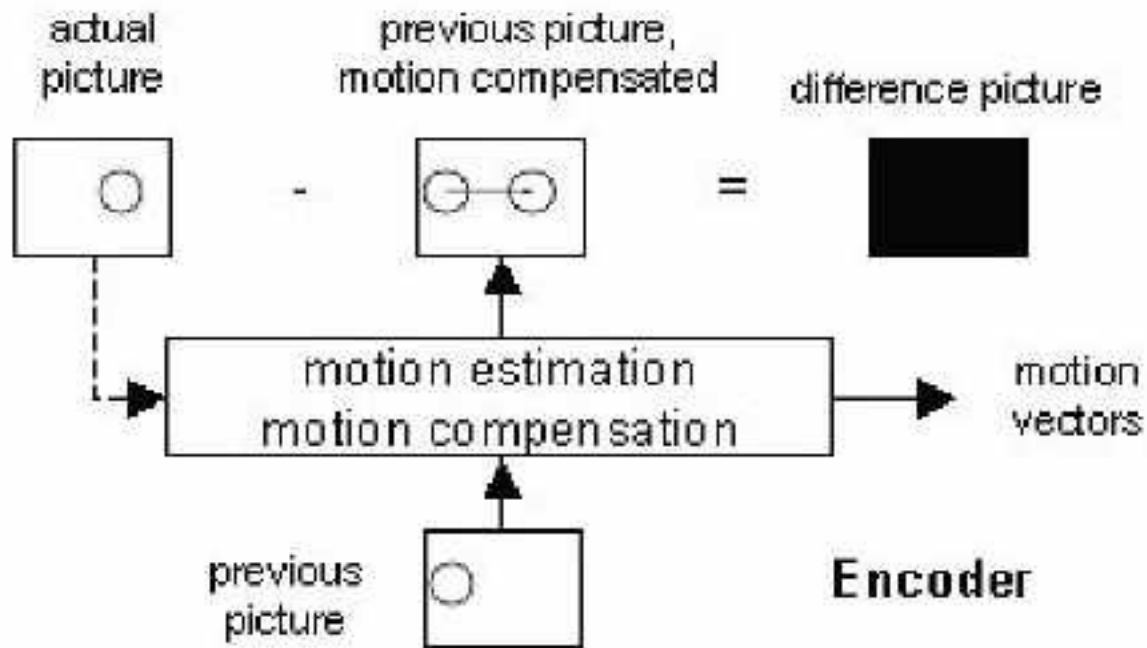
Encoder



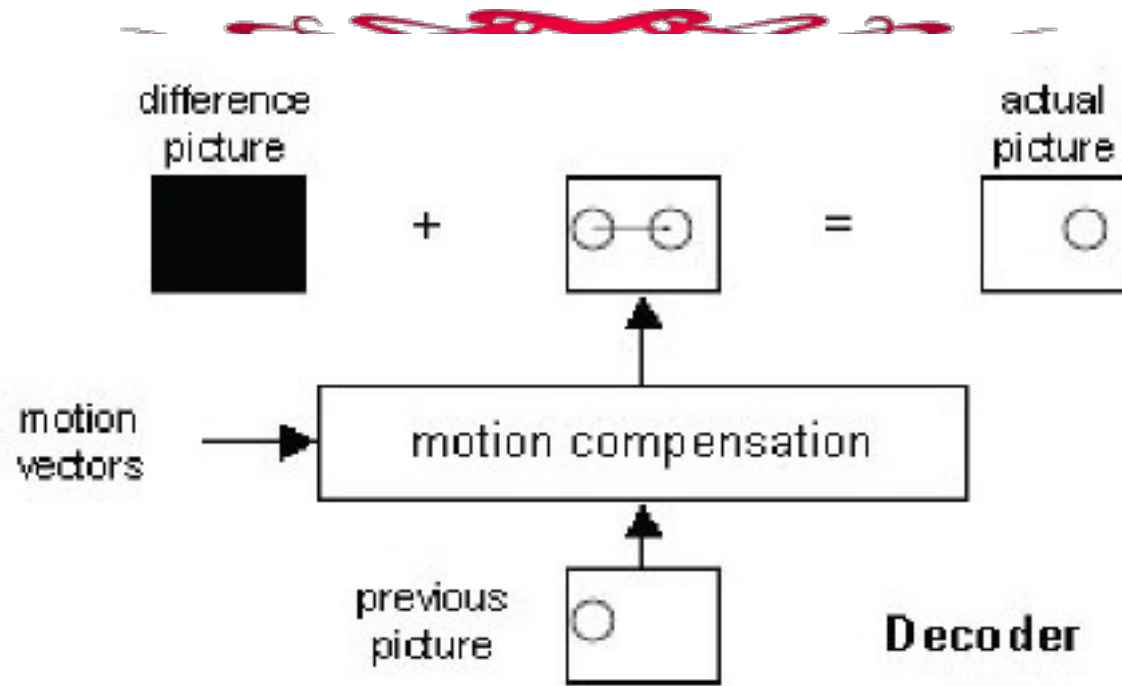
Decoder

# Ví dụ mã hóa bù chuyển động

- Dự đoán chuyển động cho block.



# Ví dụ mã hóa bù chuyển động



- Giải thích tại sao phương pháp này tốt hơn mã hóa sai biệt giữa các khung.

# CÁC CHUẨN NÉN VIDEO



- H.261
- H.263
- MPEG-1
- MPEG-2

# H.261



- **Khái niệm**

- ITU (CCITT) H.261 được phát triển cho dịch vụ truyền hình hội nghị và video phone qua ISDN ở tốc độ  $p \times 64\text{kbps}$  ( $p=1..30$ ).

- Ví dụ: Tốc độ 64kbps truyền 48kbps video, 16kbps audio.

- Truyền hình hội nghị yêu cầu chất lượng ảnh cao hơn,  $p \geq 6$ , tốc độ 384kbps.

- Là cơ sở cho các chuẩn nén sau này như MPEG 1, 2.

- Các đặc tính:

- Trễ mã hoá  $< 150\text{ms}$ ,  $\rightarrow$  truyền hình hội nghị song công, gây ấn tượng tốt cho khán giả.

- Thực hiện trên linh kiện VLSI (giá thành thấp)  $\rightarrow$  mở rộng thị trường ở các dịch vụ video phone, truyền hình hội nghị.

- **Các dạng ảnh ngõ vào**

- Khả năng phối hợp giữa các chuẩn 625 và 525 dòng của TV, H.261 sử dụng dạng thức trung gian chung CIF (Common Intermediate Format) với các tốc độ bit thấp hơn, H.261 sử dụng dạng có tốc độ nhỏ hơn  $\frac{1}{4}$  là QCIF (Quadrature).

- Với tốc độ 30 khung hình/s thì tốc độ dữ liệu của CIF là 37.3Mbps, QCIF là 9.35 Mbps. Tốc độ càng thấp thì càng giảm số khung hình/s

# H.261

- Ghép tín hiệu video

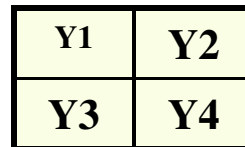
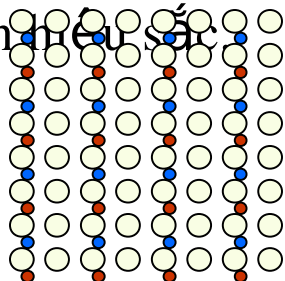
- Mục đích: Định nghĩa cấu trúc dữ liệu để bộ giải mã có thể hiểu được khi nhận dòng bit.

- Tín hiệu video được phân thành các lớp. Mỗi lớp như vậy đều có header để định nghĩa các tham số được sử dụng bởi bộ mã hoá khi tạo ra dòng bits.

- Lớp ảnh được phân thành các nhóm khối GOB (Group of Block).

- Các GOB bao gồm các MB (Macro Block).

- MB là đơn vị nhỏ nhất bao gồm 4 khối 8x8 của tín hiệu Y và hai khối 8x8 của tín hiệu sắc.  $\frac{1}{4}$  MB



Hình 3-15 Macro block



# H.261

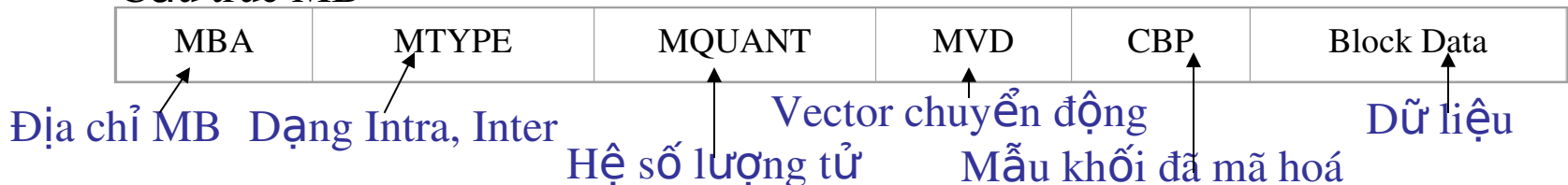
• Lớp GOB luôn được kết hợp của 33 macroblock, hợp thành ma trận 3x11. Chú ý rằng mỗi MB có một header chứa địa chỉ MB và kiểu nén, tiếp theo là dữ liệu của khối.

Kiểu	Số GOB/khung	Số MB/GOB	Tổng MB/khung
CIF	12	33	396
QCIF	3	33	99

•Cuối cùng, lớp ảnh bao gồm header ảnh theo sau bởi dữ liệu cho các GOB. Một header chứa dữ liệu là dạng ảnh (CIF hay QCIF) và số khung. Chú ý rằng ảnh CIF có 12 GOB còn QCIF chỉ có 3 GOB.

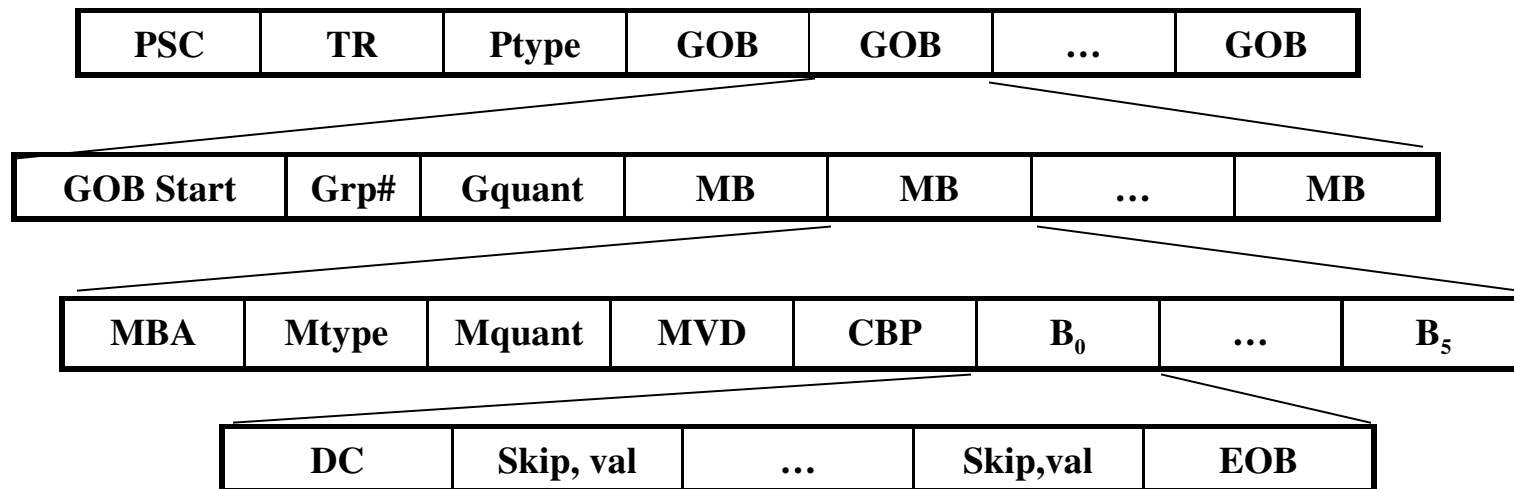
1	2	3	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	20	21	22
23	24	25	26	27	28	29	30	31	32	33

## •Cấu trúc MB



# H.261

- Cấu trúc chuỗi bit H.261

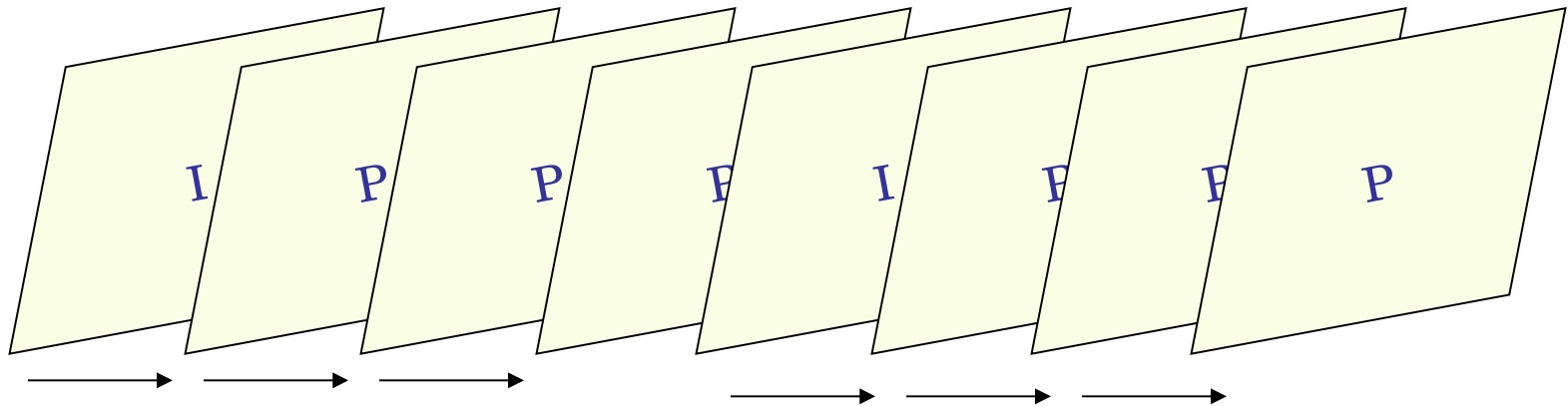


*Hình 3-16 Cấu trúc chuỗi bits H.261*

- PSC: Picture Start Code Bắt đầu một ảnh.
- TR: Temporal Reference Sử dụng để đồng bộ audio
- Ptype: Kiểu ảnh (khung P hay khung I).
- Grp# Số group.
- GQuant: Lượng tử cả nhóm với cùng một hệ số lượng tử.

# H.261

- Các khung

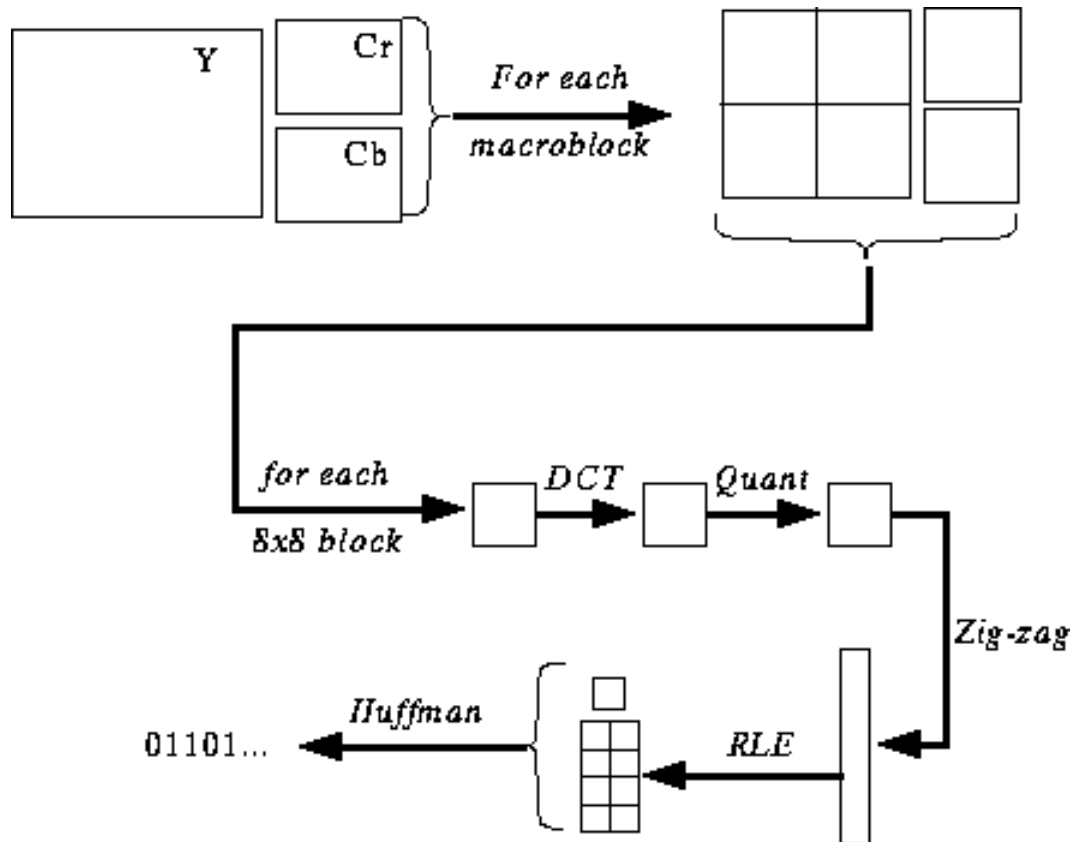


*Hình 3-17 Chuỗi các khung H.261*

- Intra Frame: Trong khung, khung I cho khả năng truy cập ngẫu nhiên, kiểu nén intra tương tự như JPEG dựa trên mã hoá DCT từng MB.
- Inter Frame: Liên khung, liên hệ với các khung I và P trước, MB đầu có thể bù chuyển động MC (Motion Compensation) hoặc không. Mã hoá DCT sai biệt dự đoán. Cung cấp các tùy chọn cho từng MB như lượng tử với cùng hệ số lượng tử, sử dụng bộ lọc bù chuyển động.

# H.261

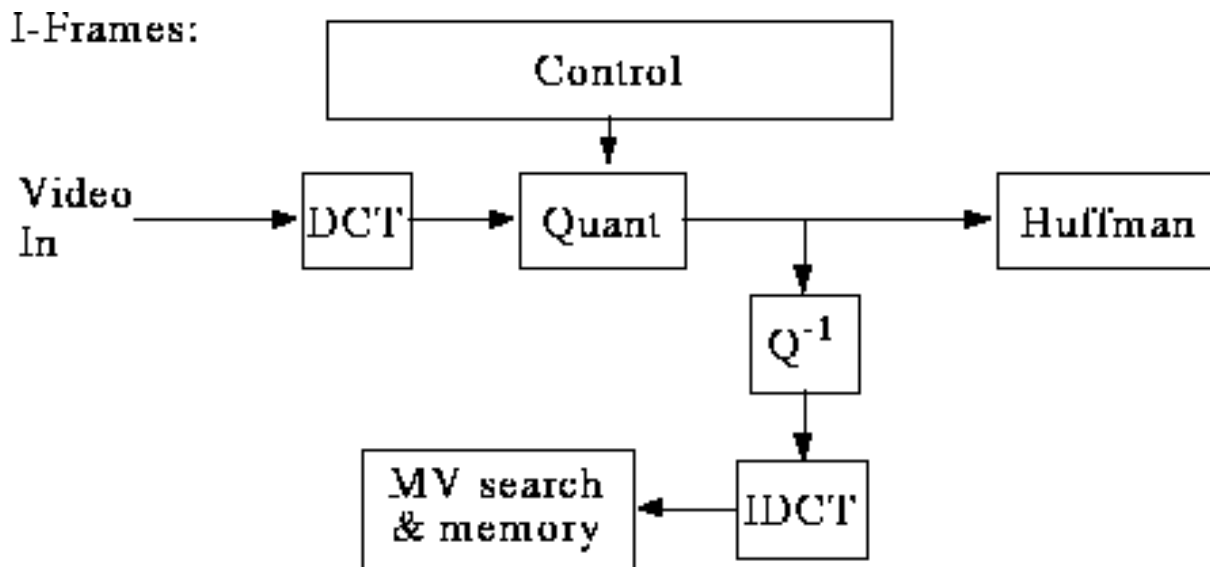
- Mã hoá trong khung



Hình 3-18 Mã hoá trong khung H.261

# H.261

- Sơ đồ mã hoá trong khung



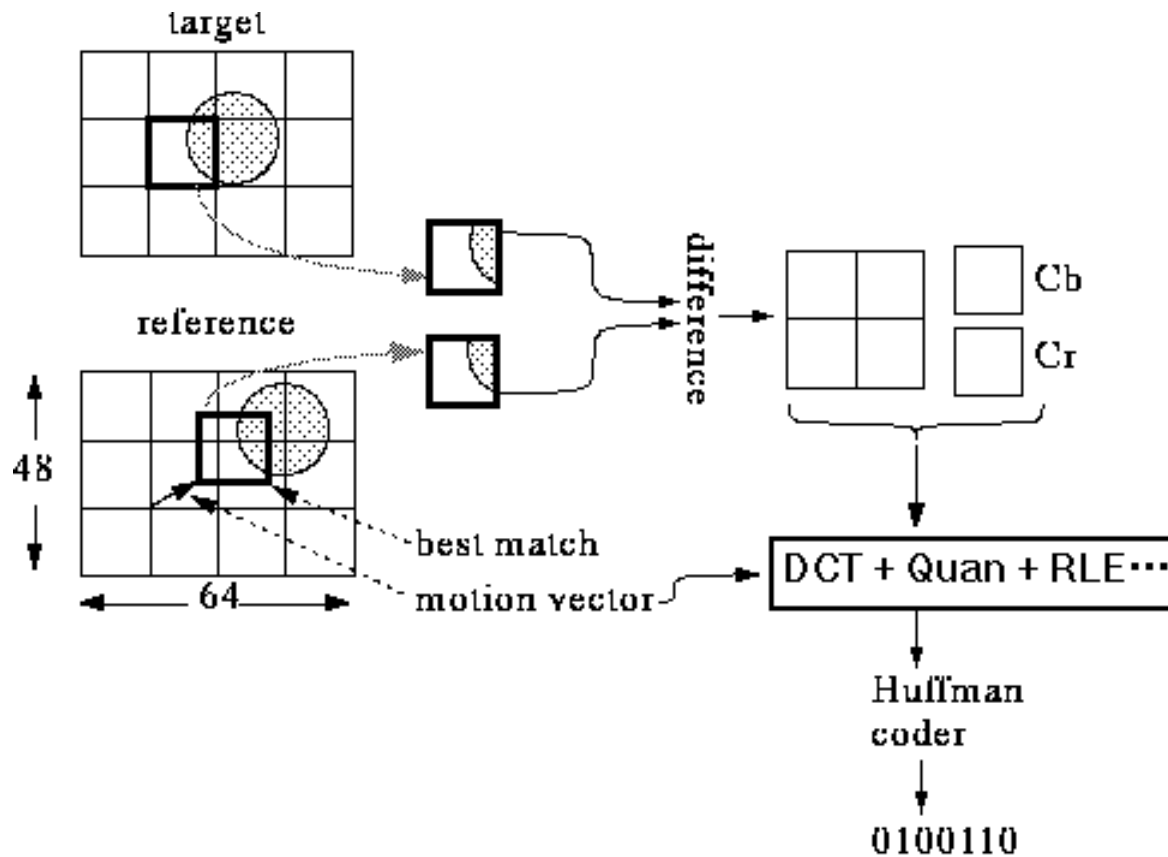
*Hình 3-19 Mã hoá khung I*

“Control” điều khiển tốc độ bit và bộ đệm, nếu bộ đệm đầy thì giảm tốc độ bit.

“Memory” dùng để lưu hình ảnh được khôi phục lại với mục đích tìm vector chuyển động cho Frame P tiếp theo.

# H.261

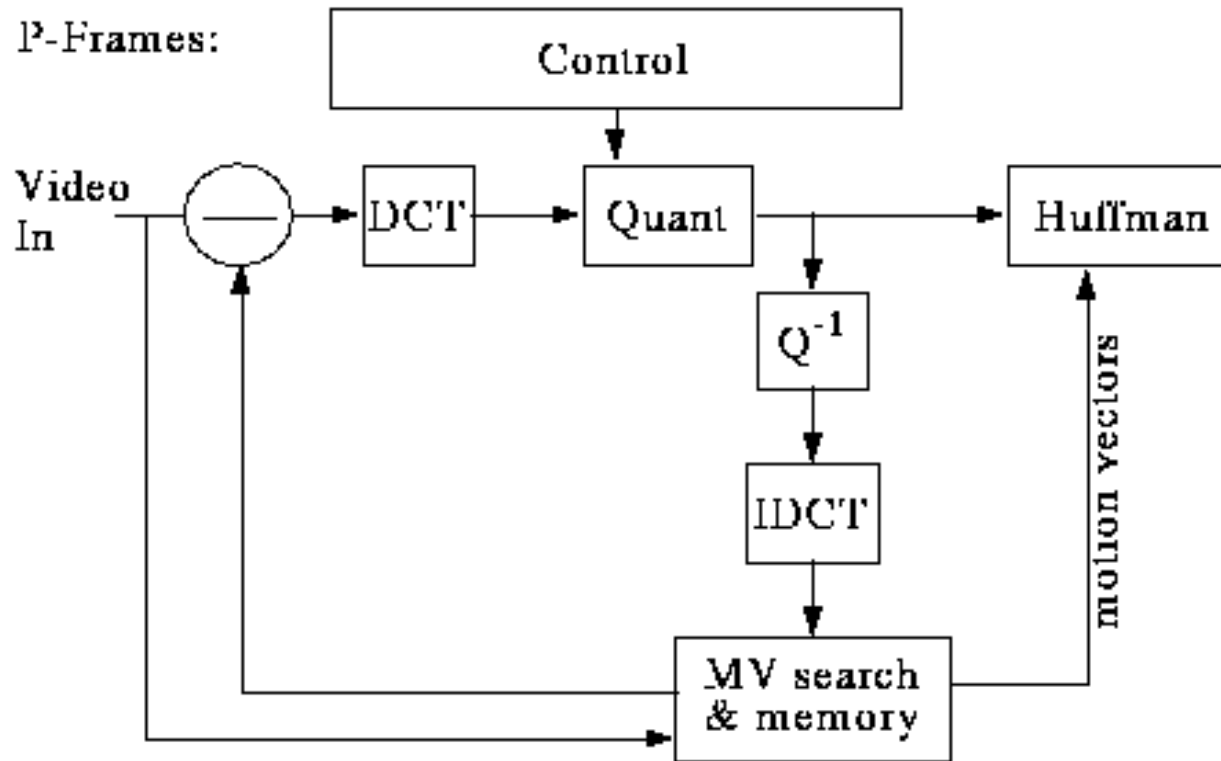
- Mã hoá liên khung



Hình 3-20 Mã hoá liên khung H.261

# H.261

- Sơ đồ mã hoá liên khung



Hình 3-21 Sơ đồ mã hoá khung P

# H.263



- Khái niệm
  - Tiêu chuẩn cải tiến H.261 cho video tốc độ thấp, có thể truyền trên mạng điện thoại công cộng PSTN, được công nhận năm 1996.
  - Giống như H.261, mã hoá DCT cho các MB trong I Frame và DCT sai biệt dự đoán trong P Frame.
  - Tốc độ tối thiểu
- Ưu điểm:
  - Chính xác sai biệt dự đoán với  $\frac{1}{2}$  pixel.
  - Không hạn chế vector chuyển động.
  - Mã hoá số học theo cú pháp.
  - Dự đoán thuận lợi với các khung P.
  - Ngoài CIF, QCIF, H.263 còn hỗ trợ SQCIF, 4 CIF và 16CIF với độ phân giải tín hiệu chói tuần tự là 128x96, 704x576, 1408x1152. Độ phân giải tín hiệu sắc bằng  $\frac{1}{4}$  tín hiệu chói.



# MPEG-1



- Khái niệm

- MPEG-1 là một chuẩn được phát triển và thừa nhận năm 1992 để lưu trữ video dạng CIF và kết hợp với audio khoảng 1.5Mbps trên nhiều môi trường lưu trữ số khác nhau như CD-ROM, DAT, Winchester, đĩa quang với ứng dụng chính là các hệ thống đa môi trường trực tuyến.

- Thuật toán MPEG-1 tương tự như H.261 nhưng với một vài đặc tính bổ sung. Chất lượng của MPEG-1 nén và giải nén tín hiệu CIF khoảng 1.2Mbps (tốc độ video) tương đương với chất lượng video tương tự VHS.

- Đặc tính

- Là tiêu chuẩn tổng quát cú pháp, hỗ trợ ước lượng chuyển động, dự đoán bù chuyển động, DCT, lượng tử và VLC.

- Không định nghĩa các thuật toán xác định cụ thể mà thiết kế bộ mã hoá linh hoạt.

- Khả năng phục vụ các ảnh khác nhau, hoạt động trên các thiết bị, tốc độ khác nhau.

- Truy cập ngẫu nhiên dựa vào các điểm truy cập độc lập (khung I).

- Tìm nhanh xem như quét dòng bit mã hoá, chỉ hiển thị các khung được chọn.

- Trễ mã hoá và giải mã hợp lý (1s), gây ấn tượng tốt cho truy cập video đơn

# MPEG-1



- **Dạng tín hiệu ngõ vào**

- MPEG-1 chỉ xét tín hiệu video lũy tiến. Để đạt được tốc độ bit 1.5Mbps, video ngõ vào thường được chuyển đổi trước khi đưa vào dạng ngõ vào chuẩn MPEG SIF (Standard Input Format).

- Không gian màu ( $Y$ ,  $C_r$ ,  $C_b$ ) đã được thừa nhận theo khuyến nghị CCIR 601. Trong SIF MPEG, kênh chỉ là 352 pixel x 240 dòng và 30 khung/s.

- Số pixel lớn nhất trên một dòng: 720.
    - Số dòng lớn nhất trên một ảnh: 576.
    - Số ảnh trong một giây: 30.
    - Số macroblock trên một ảnh: 396.
    - Số macroblock trên một giây: 9900.
    - Tốc độ bit tối đa: 1.86Mbps.
    - Kích thước bộ đệm giải mã tối đa: 376.832bits.

*Chú ý:* Ràng buộc trên không có nghĩa là tín hiệu video được nén với tốc độ tối đa.

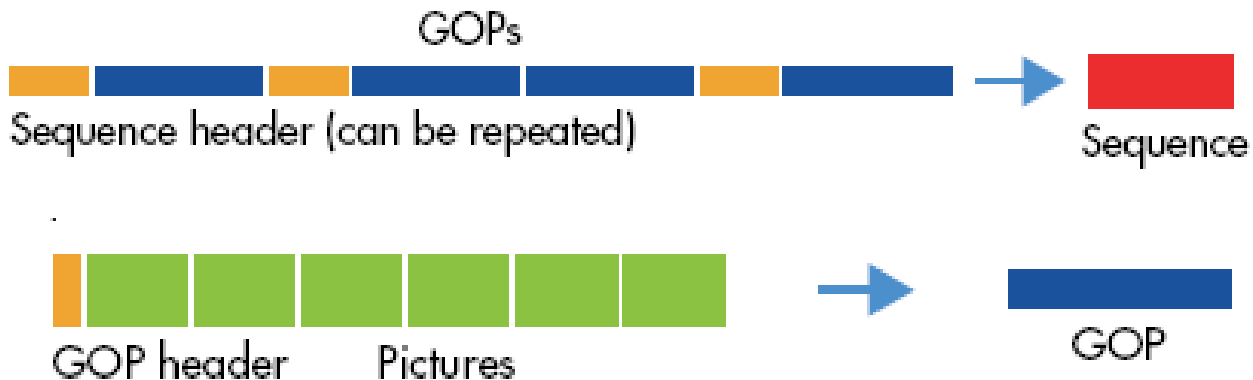
# MPEG-1

## • Cấu trúc dữ liệu và các kiểu nén

• Cấu trúc dữ liệu gồm 6 lớp, cho phép bộ giải mã hiểu được những tín hiệu chưa xác định.

1) Các chuỗi được định dạng bởi một vài nhóm ảnh GOP.

2) Nhóm ảnh tạo nên ảnh. Nhóm ảnh được xác định bởi hai thông số  $m$  và  $n$ . Thông số  $m$  xác định số khung hình P và khung hình B xuất hiện giữa hai khung hình I gần nhau nhất. Số  $n$  xác định số khung hình B giữa hai khung P.

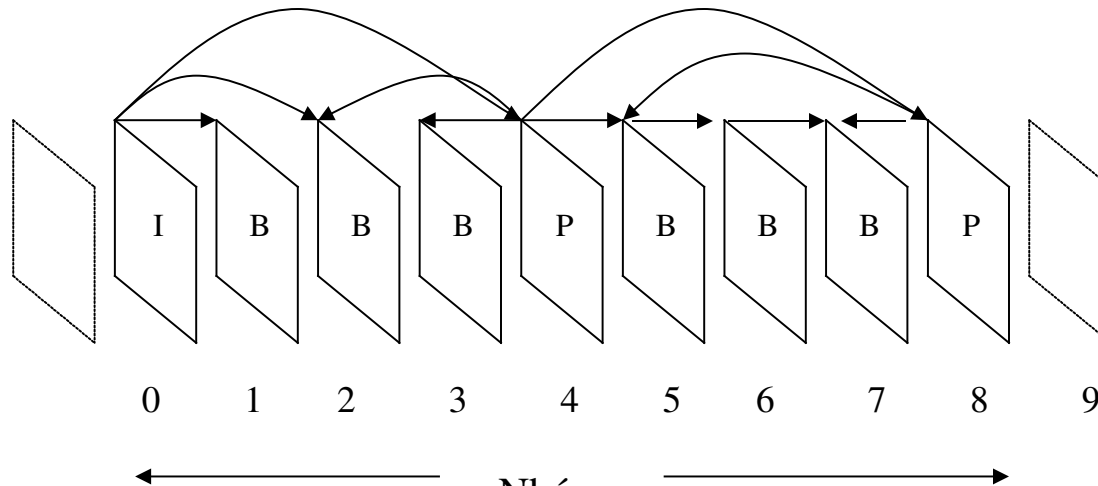


# MPEG-1

• Ví dụ:

• Một GOP như hình vẽ được kết hợp bởi 9 ảnh. Chú ý rằng khung đầu của mỗi GOP luôn là một ảnh I. Trong MPEG, trật tự trong các ảnh không cần thiết giống nhau theo thứ tự liên tiếp. Nghĩa là thứ tự truyền ảnh và thứ tự ảnh hiện lên màn hình là không giống nhau. Ảnh trong hình vẽ có thể được mã hoá trong một trật tự như sau: 0, 4, 1, 2, 3, 8, 5, 6, 7 hay 0, 1, 4, 3, 8, 5, 6, 7

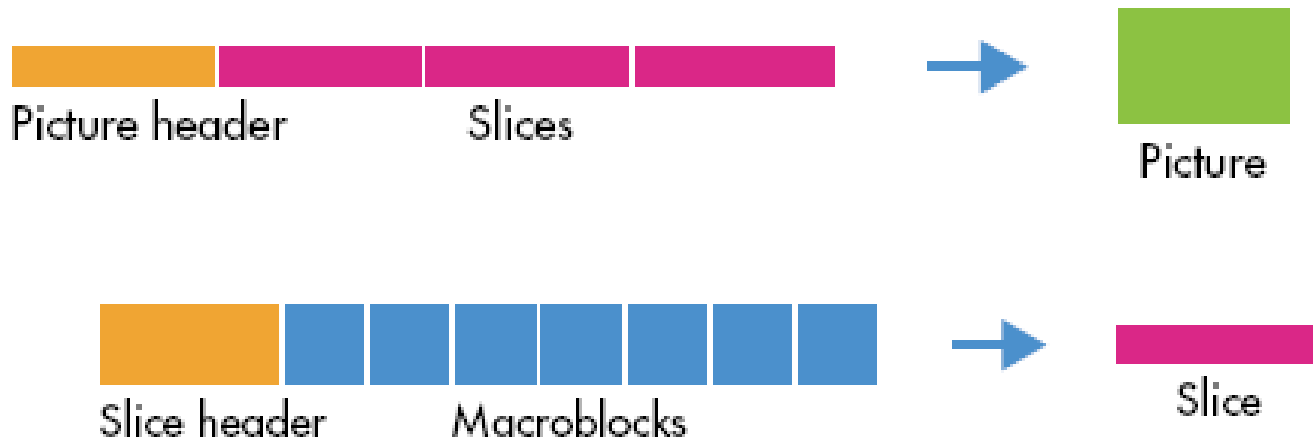
Khi dự đoán với các ảnh P và B nên dựa trên các ảnh đã được truyền



Hình 3-22 Nhóm ảnh trong MPEG-1

# MPEG-1

- 3) Ảnh bao gồm các phần (slice).
- 4) Các phần (slice) được tạo bởi các macroblock. Chúng được đưa ra với mục đích khôi phục lỗi.



# MPEG-1

Có 4 dạng ảnh tương ứng với các kiểu nén, đó là ảnh I, ảnh P, ảnh B và ảnh D.

- Ảnh I là ảnh được mã hoá DCT trong khung sử dụng thuật toán giống như JPEG. Cho phép các điểm truy cập ngẫu nhiên đến chuỗi ảnh.
- Có hai dạng ảnh được mã hoá liên khung là ảnh P và ảnh B. Các ảnh này mã hoá DCT bù chuyển động sai biệt dự đoán. Dự đoán tối là được sử dụng trong kiểu ảnh P, là các ảnh được mã hoá liên quan đến ảnh I và P trước đó. Dự đoán trong ảnh B có thể là tối hoặc lùi hoặc liên hệ hai chiều tới các ảnh I hoặc P khác.
- Các ảnh D chỉ chứa các thành phần DC trong mỗi khối với mục đích cho các tốc độ bit rất thấp. Số khung I, P và B trong một GOP được áp dụng một cách phụ thuộc, ví dụ tùy thuộc vào thời gian truy cập và các tốc độ bit yêu cầu.

# MPEG-1

5) Sự kết hợp các macroblock MB là tương tự như H.261. Một vài tham số nén có thể được mang trong MB cơ sở. Các dạng MB như bảng bên.

I: Nội suy(hai hướng);

B: Lùi

A: Thích Ứng;

F: Tối

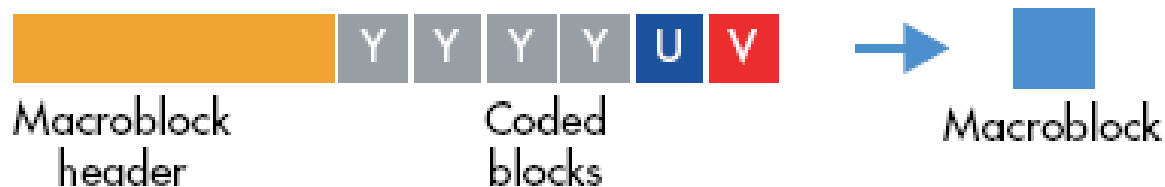
D: DC

<b>Ảnh I</b>	<b>Ảnh P</b>	<b>Ảnh B</b>
Intra	Intra	Intra
Intra-A	Intra-A	Intra-A
	Inter-D	Inter-F
	Inter-DA	Inter-FD
	Inter-F	Inter-FDA
	Inter-FD	Inter-B
	Inter-FDA	Inter-BD
	Skipped	Inter-BDA
		Inter-I
		Inter-ID
		Inter-IDA
		Skipped

# MPEG-1

6) Các khối là các mảng 8x8 pixels. Chúng là một đơn vị DCT nhỏ nhất.

Các tiêu đề được định nghĩa cho các chuỗi, GOP, ảnh, phần và các MB để xác định độc nhất dữ liệu sau nó.





# MPEG-1



- Kiểu nén trong khung

- Các giá trị cường độ của điểm ảnh được mã hoá DCT theo cách tương tự như JPEG và kiểu intra của H.261.

- Việc nén đạt được bằng sự kết hợp của lượng tử hoá và mã hoá chiều dài thay đổi của các hệ số 0.

- MPEG cho phép lượng tử hoá không gian thích ứng bằng việc đưa ra tỷ lệ lượng tử MQUANT trong cú pháp.

- Có hai kiểu MB trong các ảnh I:

- MB “Intra” được mã hoá với ma trận lượng tử hoá hiện thời.

- MB “Intra A” thì ma trận lượng tử hoá được tỷ lệ bởi hệ số MQUANT được truyền trong header.

Chú ý rằng MQUANT có thể thay đổi trong MB cơ bản để điều khiển tốc độ bit hoặc với mục đích lượng tử.

MPEG đã được chứng minh rằng kiểu nén Intra cung cấp 30% nén tốt hơn so với JPEG về tính lượng tử thích ứng.

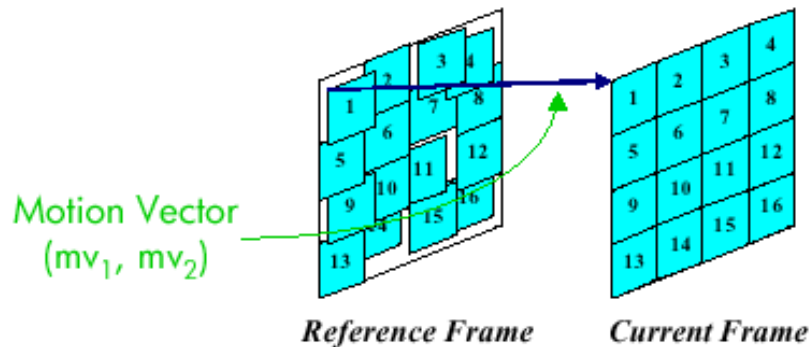
- Mã hoá các hệ số DCT sau lượng tử giống như JPEG.

# MPEG-1

- Kiểu nén liên khung

- Trong các kiểu nén liên khung, sự dự đoán thời gian được tiến hành và kết quả sai biệt dự đoán được mã hoá DCT. Có hai dạng kiểu dự đoán thời gian được cho phép trong MPEG-1, đó là dự đoán tới (khung P) và dự đoán hai hướng (khung B).

- Khung P (dự đoán bù chuyển động)



Hình 3-23 Dự đoán bù chuyển động tới

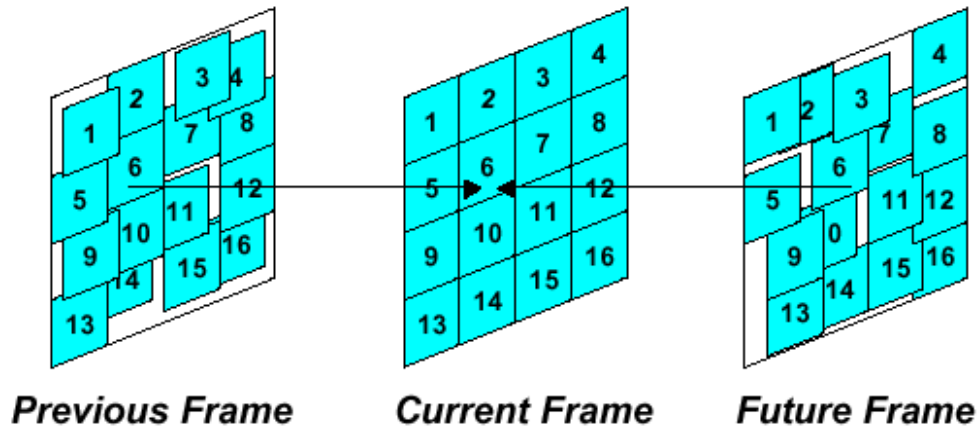
- Sử dụng các MB thích hợp nhất trong khung trước (tham chiếu) để dự đoán bù chuyển động trong khung hiện hành. Thích hợp nhất dựa vào trung bình bình phương tối thiểu hoặc trung bình tuyệt đối tối thiểu.

- Sử dụng thuật toán ước lượng chuyển động nửa pixel với hai bước thời và

# MPEG-1

- **Khung B (dự đoán hai hướng, nội suy)**

- Là đặc trưng của MPEG.



*Hình 3-24 Dự đoán hai hướng*

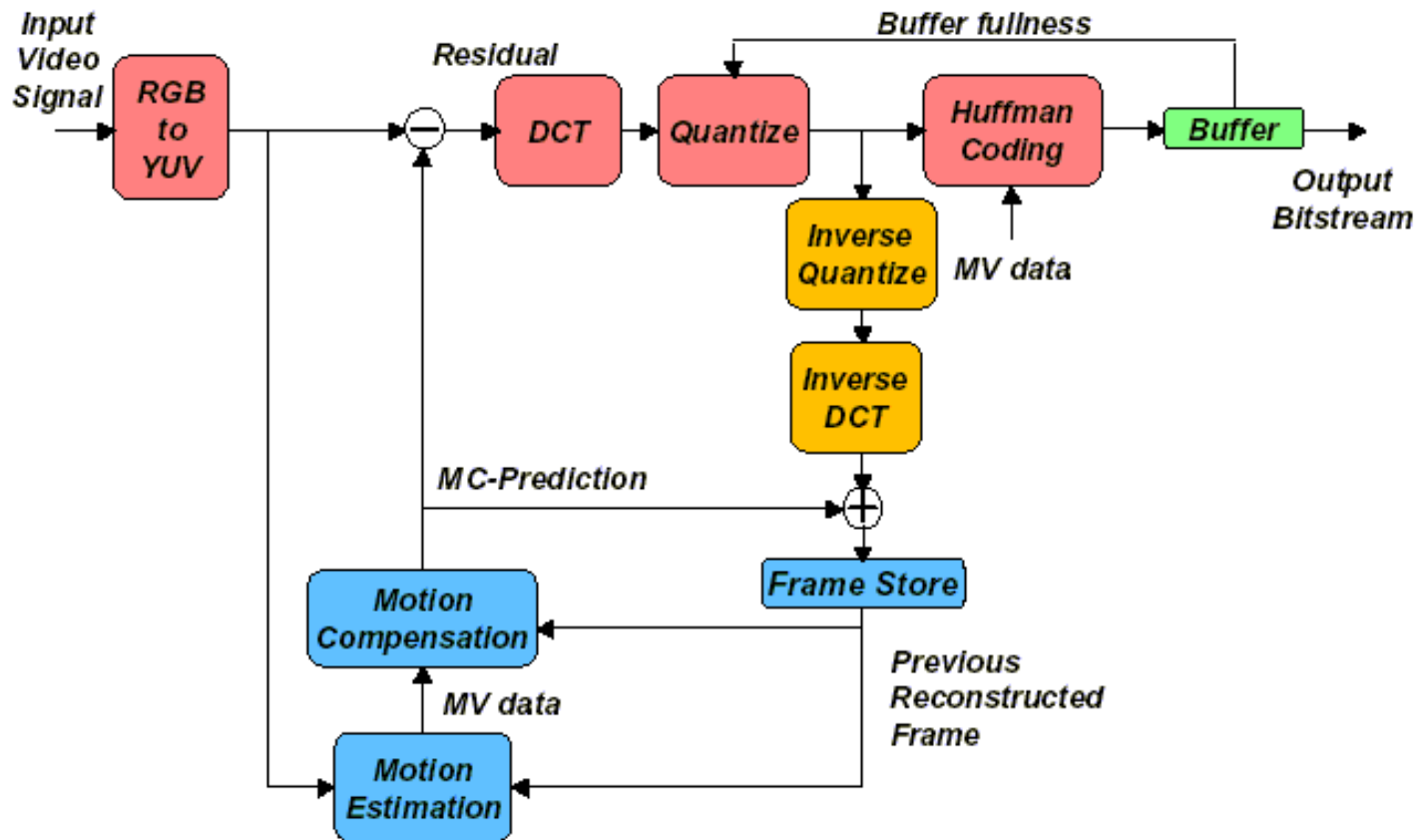
- Dự đoán MB cho khung hiện hành dựa vào: Khung trước; khung sau và trung bình MB của khung trước và khung sau.

- Ưu điểm: Hiệu quả trong việc nén nếu dùng khung B để dự đoán chu khung sau, trung bình MC qua hai khung cho SNR tốt hơn.

- Nhược: Cần ít nhất hai khung tham chiếu I, P; nếu sử dụng khung B quá nhiều thì khoảng cách giữa các khung tham chiếu I, P tăng → trễ mã hoá lớn.

# MPEG-1

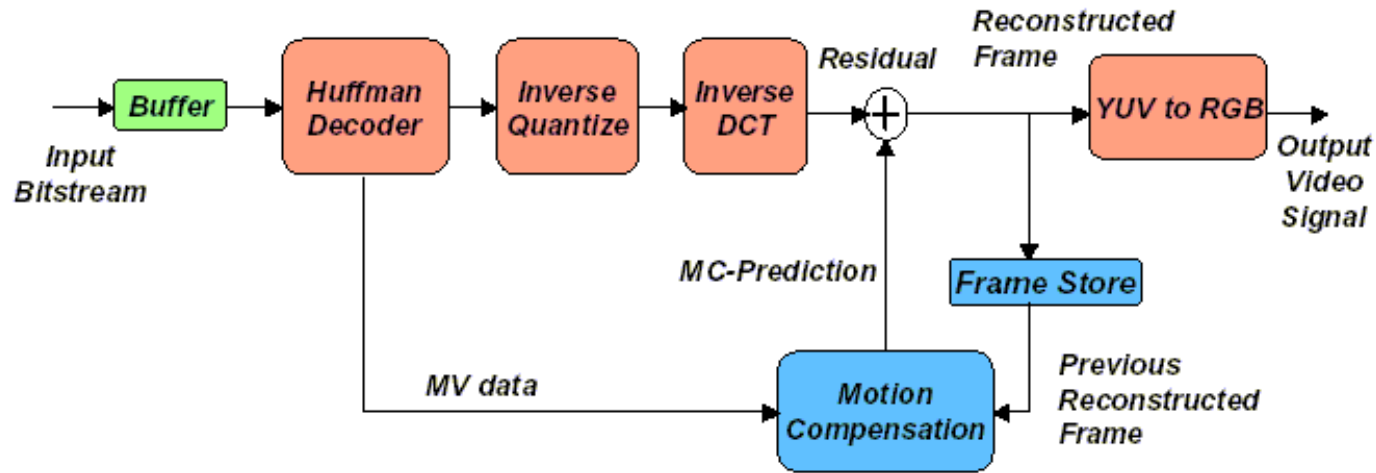
- Mã hoá MPEG-1



Hình 3-25 Mã hoá MPEG-1

# MPEG-1

- Giải mã MPEG-1



Hình 3-26 Giải mã MPEG-1

Tổng quan, bộ mã hoá MPEG-1 tiến hành các bước sau:

- Quyết định nhãn của các khung I, P và B trong một GOP.
- Đánh giá chuyển động cho mỗi MB trong các ảnh P, B.
- Xác định kiểu nén MTYPE cho mỗi MB.
- Đặt tỷ số lượng tử MQUANT nếu lượng tử thích ứng được

chọn.

Bộ giải mã MPEG-1 hoạt động theo hướng ngược lại.

# MPEG-2

- **Khái niệm**

- Chất lượng MPEG-1 ở tốc độ 1.2Mbps cho thấy không thể chấp nhận được với các ứng dụng giải trí.

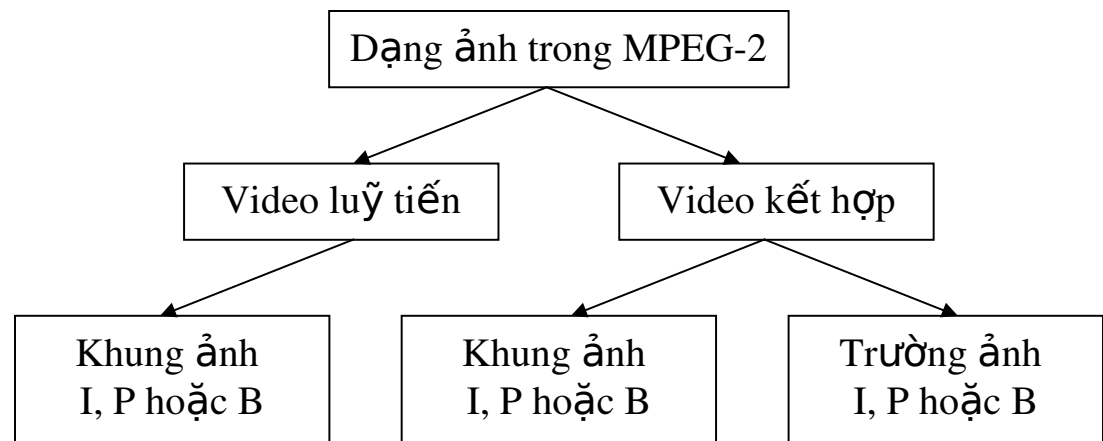
- MPEG-2 là một phiên bản mở rộng của MPEG-1 để cung cấp phạm vi rộng của các ứng dụng với các tốc độ bit và các độ phân giải khác nhau (2-20Mbps).

- **Đặc trưng của MPEG-2:**

- Cho phép ngõ vào kết hợp (interlace), độ phân giải cao, lấy mẫu phụ thay đổi của các kênh sắc.

- Cung cấp dòng bit co giãn.

- Cung cấp các tùy chọn lượng tử hoá và mã hoá.



Hình 3-27 Dạng ảnh trong MPEG-2

# MPEG-2



- Các hình thái của MPEG-2

Hình thái đơn giản: Số bước nén thấp nhất, chỉ cho phép mã hoá các ảnh loại I hoặc P, không cho phép mã hoá các ảnh loại B. Tốc độ tối đa của hình thái này là 15Mbps.

Hình thái chính: Cho phép sử dụng tất cả các loại ảnh nhưng không tạo ra các mức bất kỳ. Chất lượng ảnh tốt hơn hình thái đơn giản nhưng tốc độ bit không đổi.

Hình thái phân cấp theo SNR: Tính phân cấp theo SNR có nghĩa là có sự thoả hiệp về chất lượng hình ảnh với tỷ số tín hiệu trên tạp âm. Hỗ trợ cho các tốc độ 4 và 15Mbps.

Hình thái phân cấp theo không gian: Tính phân cấp theo không gian là có sự thoả hiệp về độ phân giải. Hỗ trợ tốc độ bit 60Mbps.

Hình thái cao: Bao gồm các công cụ của các hình thái trước cộng với khả năng mã hoá các tín hiệu màu khác nhau cùng một lúc. Là hệ thống hoàn hảo được thiết kế cho toàn bộ các ứng dụng mà không bị giới hạn bởi tốc độ cao.

# MPEG-4



- Phiên bản 1: 8/1998.
- Phiên bản 2: 12/1999.
- Phát triển cho tốc độ bit rất thấp: 4.8 đến 64kbps, tuy nhiên, sau đó đã nâng tốc độ lên đáng kể.
  - Video: 5kbps đến 10Mbps.
  - Audio: 2kbps đến 64 kbps.
- Hỗ trợ các phát triển riêng cho các đối tượng khác nhau tạo khả năng thích nghi, linh động để cải thiện chất lượng của các dịch vụ như digital television, **animation** graphics, World Wide Web và các mở rộng của chúng.



# MPEG-4



- Đây là chuẩn nén video giàu tham vọng nhất. Các chuyên gia muốn huy động các yếu tố như sự tổng hợp thoại và hình, hình học bất quy tắc, trực quan máy tính và trí thông minh nhân tạo (AI) để tái tạo lại các hình ảnh.

MPEG-4 tạo ra một bước ngoặt mới cho truyền thông video.

- MPEG-4 là chuẩn thuật toán đồ họa và video dựa trên hai chuẩn MPEG-1, MPEG-2 và công nghệ QuickTime của Apple. Nhờ có kích thước nhỏ hơn và tốc độ truyền tải thấp hơn, MPEG-4 có thể truyền qua một băng thông hẹp hơn, có thể trộn video với text, đồ họa, các lớp 2D và 3D động,... Nó đã được tổ chức ISO công nhận vào tháng 10-1998.

MPEG-4 được coi là một cuộc cách mạng mới trong media số. Nó là chuẩn multimedia toàn cầu thế hệ kế tiếp. Nó được thiết kế để truyền tải video với chất lượng DVD (MPEG-2) qua mạng. Mới nhất là HD-DVD và Blue-ray.

# MPEG-7

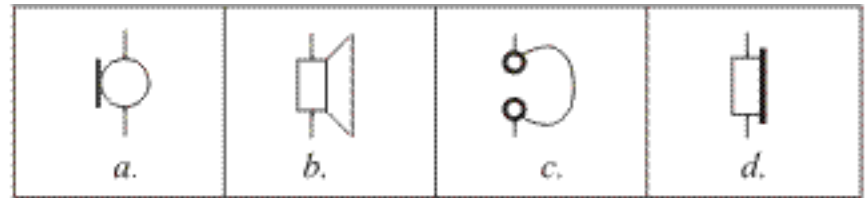


- Được thừa nhận thành tiêu chuẩn quốc tế vào tháng 7/2001.
- Đề xuất cho việc tìm kiếm, chọn lựa, quản lý và phân tích thông tin đa phương tiện.
- Được gọi là “Giao tiếp mô tả nội dung đa môi trường”.
- Không tập trung vào một ứng dụng cụ thể, không phải là một tiêu chuẩn mã hóa thực sự cho audio, video hay multimedia.
- Bao gồm:
  - Tập các sơ đồ mô tả và các bộ mô tả.
  - Ngôn ngữ xác định cho các sơ đồ này gọi là **Description Definition Language (DDL)**
  - Sơ đồ mã hóa mô tả.

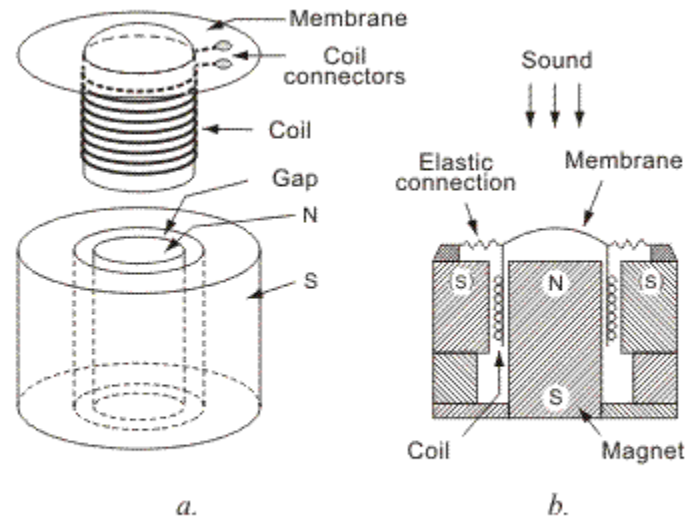
# Một số vấn đề khác về audio-video

# Microphone, speaker và headphone

- Microphone chuyển tín hiệu âm thanh thành tín hiệu điện từ.
- Bao gồm các loại:
- Carbon, dynamic, crystal, capacitive

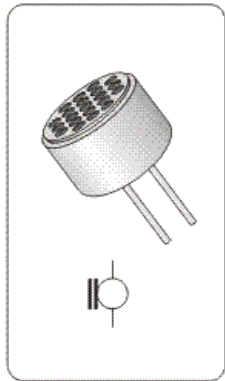


8.1. Symbols: a - microphone, b - speaker, c, d - headphones

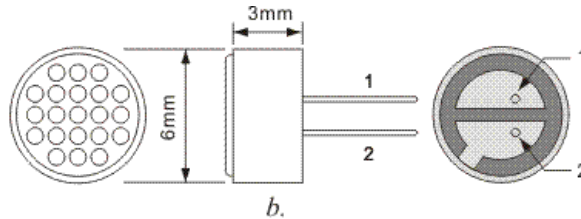


Sl. 8.2. Dynamic microphone: a - internal structure, b - cross section

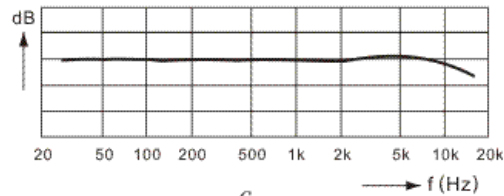
# Microphone, speaker và headphone



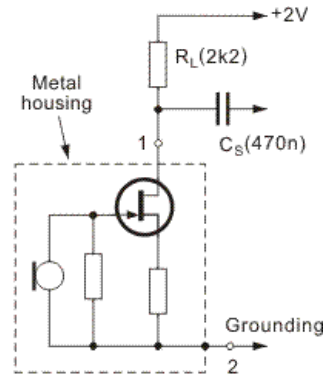
a.



b.



c.

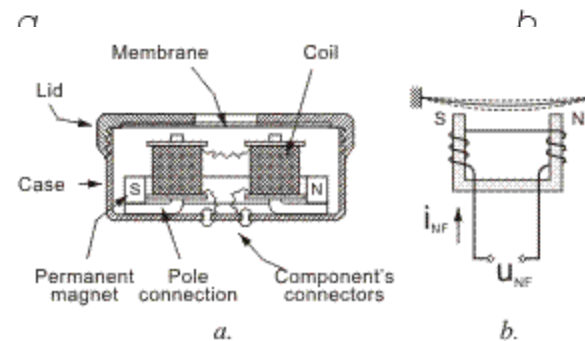
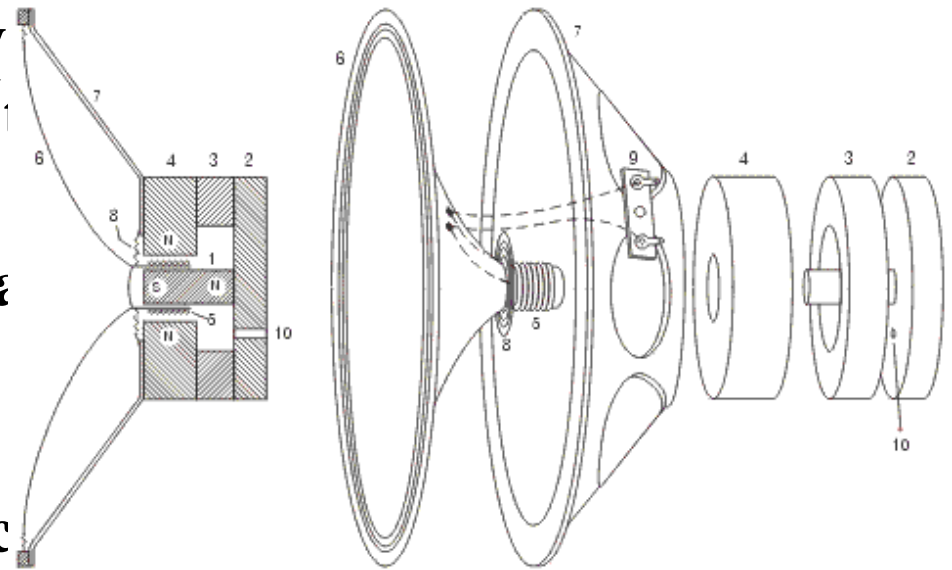


d.

8.3. Electret microphone: a - picture and schematic symbol, b - size and connectors, c - frequency characteristics, d - connection

# Microphone, speaker và headphone

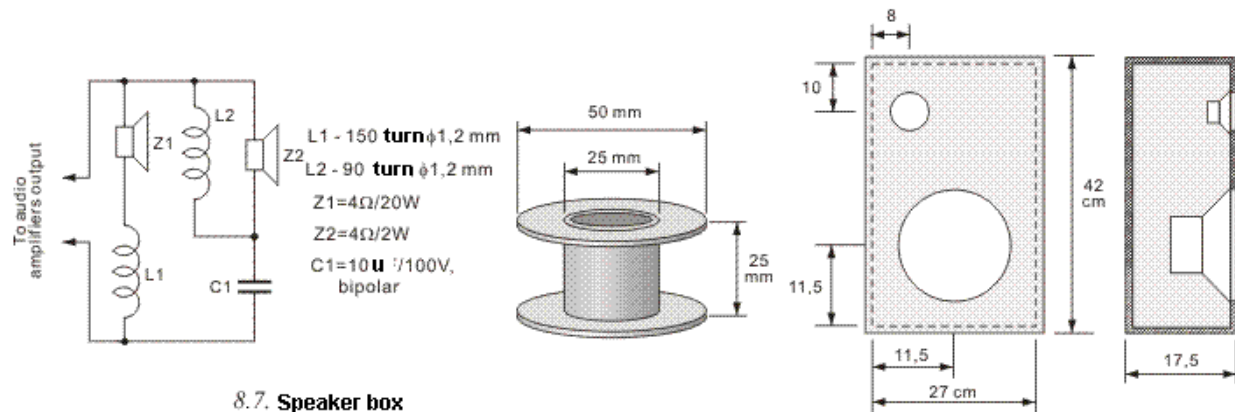
- Loa có kích thước thay đổi từ rất nhỏ đến rất lớn.
- Loa bao gồm 2 loại: loa thạch anh và loa điện động.
- Headphone là loa có kích thước và công suất nhỏ. Nhưng thường là thạch anh hoặc điện từ.



8.5. Elektromagnetic headphone:

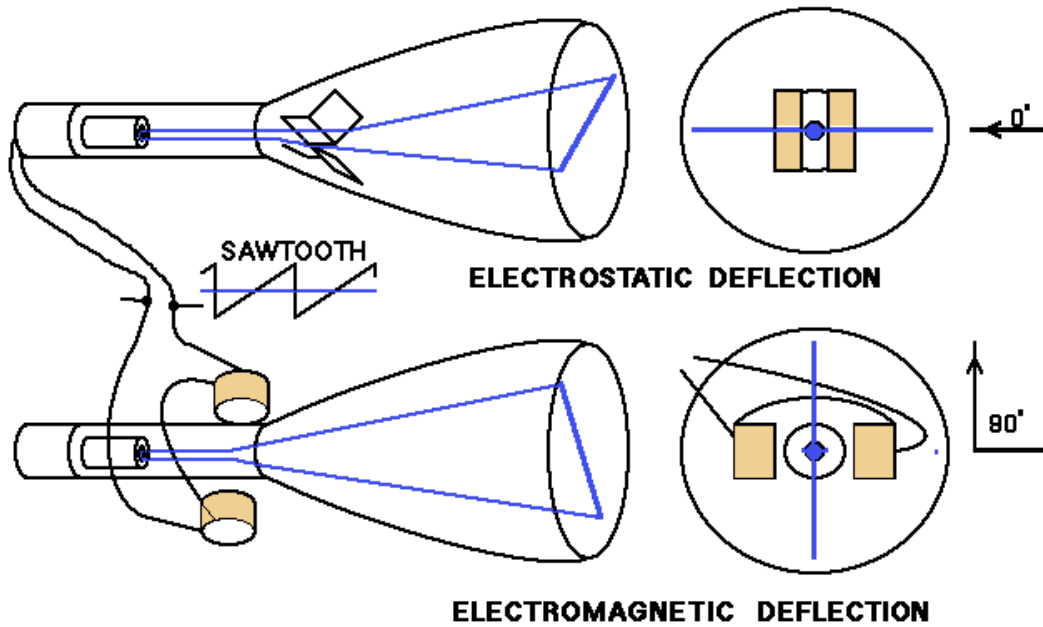
# Microphone, speaker và headphone

- Loa thùng bao gồm nhiều loa bên trong.
- Mỗi loa đáp ứng với một dải tần khác nhau.
- Sử dụng các bộ lọc để chia các dải cho các loa này.
- Các thùng loa lớn thường sử dụng hiệu ứng phản xạ âm thanh để tăng cường âm trầm qua lỗ thoát



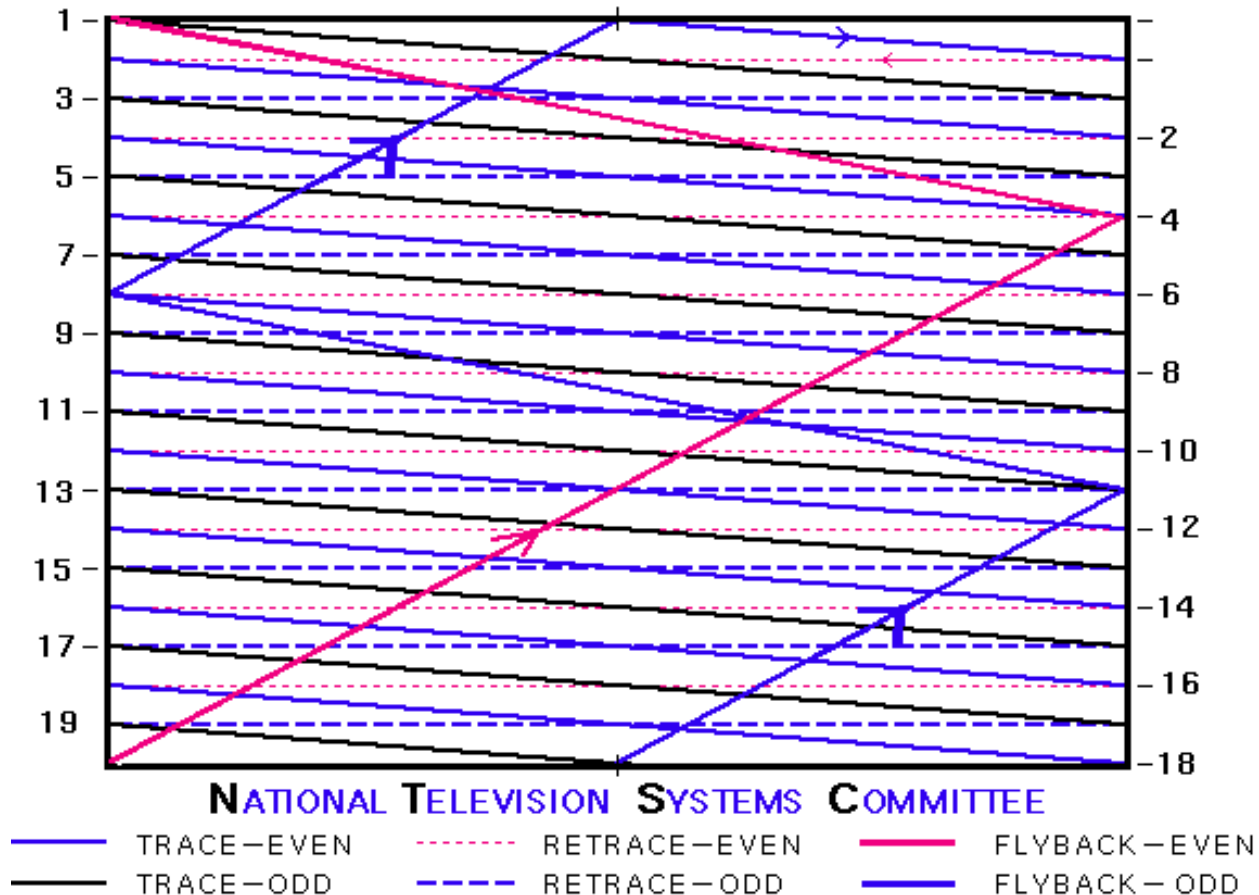
8.7. Speaker box

# CRT (Cathode Ray Tube)

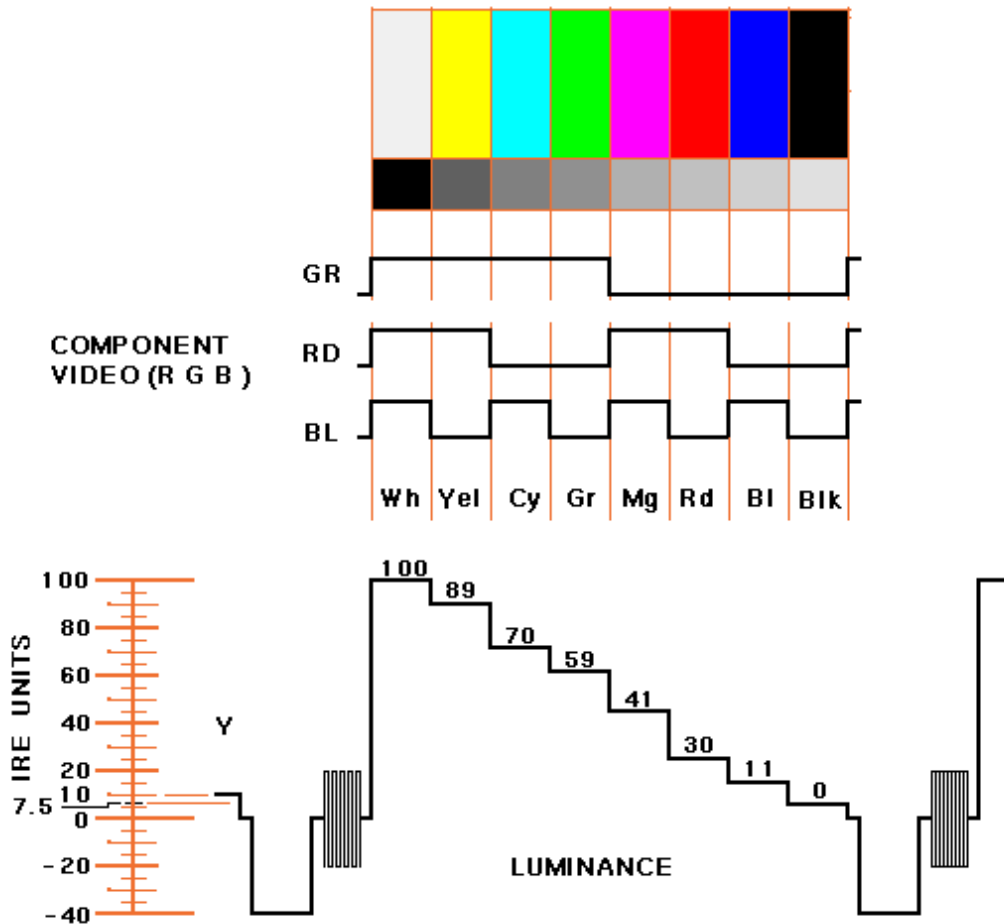




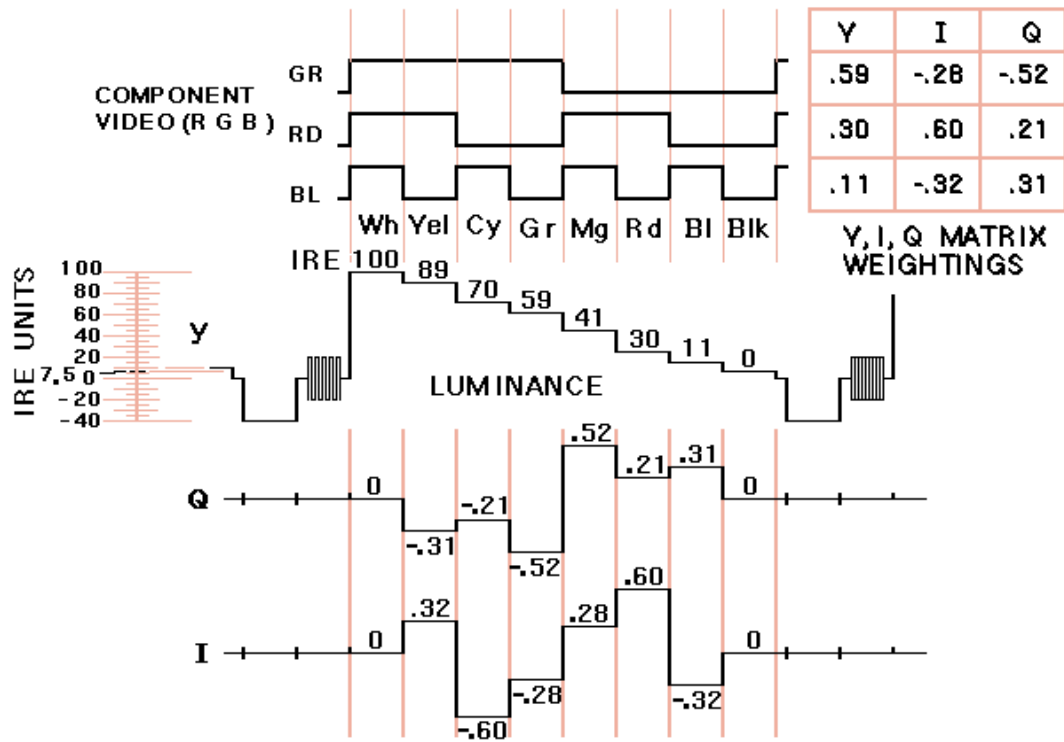
# Kỹ thuật quét dòng trong CRT



# Giải mã màu



# Phát sọc màu



# Vector scope

