SỞ GIÁO DỤC VÀ ĐÀO TẠO HÀ NỘI

GIÁO TRÌNH

Lý thuyết truyền tin

DÙNG TRONG CÁC TRƯỜNG TRUNG HỌC CHUYÊN NGHIỆP

NHÀ XUẤT BẢN HÀ NỘI

SỞ GIÁO DỤC VÀ ĐÀO TẠO HÀ NỘI

TRẦN THỊ NGÂN

GIÁO TRÌNH
CƠ SỞ LÝ THUYẾT TRUYỀN TIN

(Dùng trong các trường THCN)

NHÀ XUẤT BẢN HÀ NỘI - 2007

Lời giới thiệu

Nước ta đang bước vào thời kỳ công nghiệp hóa, hiện đại hóa nhằm đưa Việt Nam trở thành nước công nghiệp văn minh, hiện đại.

Trong sự nghiệp cách mạng to lớn đó, công tác đào tạo nhân lực luôn giữ vai trò quan trọng. Báo cáo Chính trị của Ban Chấp hành Trung ương Đảng Cộng sản Việt Nam tại Đại hội Đảng toàn quốc lần thứ IX đã chỉ rõ: "Phát triển giáo dục và đào tạo là một trong những động lực quan trọng thúc đẩy sự nghiệp *công nghiệp hóa, hiện đại hóa*, là điều kiện để phát triển nguồn lực con người - yếu tố cơ bản để phát triển xã hội, tăng trưởng kinh tế nhanh và bền vững".

Quán triệt chủ trương, Nghị quyết của Đảng và Nhà nước và nhận thức đúng đắn về tầm quan trọng của chương trình, giáo trình đối với việc nâng cao chất lượng đào tạo, theo đề nghị của Sở Giáo dục và Đào tạo Hà Nội, ngày 23/9/2003, Ủy ban nhân dân thành phố Hà Nội đã ra Quyết định số 5620/QĐ-UB cho phép Sở Giáo dục và Đào tạo thực hiện đề án biên soạn chương trình, giáo trình trong các trường Trung học chuyên nghiệp (THCN) Hà Nội. Quyết định này thể hiện sự quan tâm sâu sắc của Thành ủy, UBND thành phố trong việc nâng cao chất lượng đào tạo và phát triển nguồn nhân lực Thủ đô.

Trên cơ sở chương trình khung của Bộ Giáo dục và Đào tạo ban hành và những kinh nghiệm rút ra từ thực tế đào tạo, Sở Giáo dục và Đào tạo đã chỉ đạo các trường THCN tổ chức biên soạn chương trình, giáo trình một cách khoa học, hệ

thống và cập nhật những kiến thức thực tiễn phù hợp với đối tượng học sinh THCN Hà Nội.

Bộ giáo trình này là tài liệu giảng dạy và học tập trong các trường THCN ở Hà Nội, đồng thời là tài liệu tham khảo hữu ích cho các trường có đào tạo các ngành kỹ thuật - nghiệp vụ và đông đảo bạn đọc quan tâm đến vấn đề hướng nghiệp, dạy nghề.

Việc tổ chức biên soạn bộ chương trình, giáo trình này là một trong nhiều hoạt động thiết thực của ngành giáo dục và đào tạo Thủ đô để kỷ niệm "50 năm giải phóng Thủ đô", "50 năm thành lập ngành" và hướng tới kỷ niệm "1000 năm Thăng Long - Hà Nội".

Sở Giáo dục và Đào tạo Hà Nội chân thành cảm ơn Thành ủy, UBND, các sở, ban, ngành của Thành phố, Vụ Giáo dục chuyên nghiệp Bộ Giáo dục và Đào tạo, các nhà khoa học, các chuyên gia đầu ngành, các giảng viên, các nhà quản lý, các nhà doanh nghiệp đã tạo điều kiện giúp đỡ, đóng góp ý kiến, tham gia Hội đồng phản biện, Hội đồng thẩm định và Hội đồng nghiệm thu các chương trình, giáo trình.

Đây là lần đầu tiên Sở Giáo dục và Đào tạo Hà Nội tổ chức biên soạn chương trình, giáo trình. Dù đã hết sức cố gắng nhưng chắc chắn không tránh khỏi thiếu sót, bất cập. Chúng tôi mong nhận được những ý kiến đóng góp của bạn đọc để từng bước hoàn thiện bộ giáo trình trong các lần tái bản sau.

GIÁM ĐỐC SỞ GIÁO DỤC VÀ ĐÀO TẠO

Lời nói đầu

Thông tin là một trong những nhu cầu không thể thiếu đối với con người, là một điều kiện cần cho sự tồn tại và phát triển. Ngành công nghiệp thông tin liên lạc cũng được coi là ngành công nghiệp trí tuệ hoặc công nghiệp của tương lai, là nền tảng để phát triển và tăng cường sức mạnh quốc gia cũng như sự cạnh tranh trong công nghiệp.

Khi khoa học kỹ thuật và xã hội càng phát triển thì thông tin càng thể hiện được vai trò quan trọng. Cùng với lịch sử phát triển của con người, kỹ thuật truyền tin cũng không ngừng phát triển. Sự phát minh ra sóng vô tuyến dùng cho thông tin liên lạc cùng các định lý lấy mẫu, định lý về dung lượng kênh đã làm nền tảng cho thông tin số nhằm nâng cao tốc độ truyền tin và tăng độ tin cậy cho thông tin nhận được. Tiếp theo là công nghệ chinh phục vũ trụ, công nghệ vi mạch và sự thâm nhập lẫn nhau giữa khoa học máy tính và truyền thông đã tạo điều kiện thuận lợi để phát triển các hệ thống thông tin hiện đại, tốc độ cao, dung lượng lớn, độ tin cậy cao, truyền ở cự ly rất xa, địa hình hiểm trở để phục vụ cuộc sống của con người như các hệ thống phát thanh, truyền hình, Internet,...

Tất cả các công nghệ truyền tin hiện đại đều phải dựa trên kiến thức nền tảng của lý thuyết truyền tin. Nói cách khác, cơ sở lý thuyết truyền tin là kiến thức cơ bản không thể thiếu được đối với các ngành Điện tử - Viễn thông và Công nghệ thông tin.

Mục tiêu của giáo trình này là cung cấp cho học sinh chuyên ngành Điện tử - Viễn thông, Công nghệ thông tin trong các trường trung học chuyên nghiệp những kiến thức cơ bản nhất về lý thuyết thông tin, mã hóa, điều chế và mối liên hệ chặt chẽ giữa chúng trong một hệ thống truyền tin. Trên cơ sở đó, học sinh tự tìm hiểu và làm việc trong những ngành chuyên môn có liên quan và có thể tiếp cận với sự phát triển của các hệ thống truyền tin tương lai.

Giáo trình được được biên soạn với dung lượng 45 tiết, được chia thành 5 chương:

Chương 1: ***Nhập môn lý thuyết truyền tin.***

Chương 2: ***Thông tin và lượng tin.***

Chương 3: ***Mã hiệu.***

Chương 4: ***Mã hóa.***

Chương 5: ***Điều chế.***

Giáo trình được trình bày cơ bản, ngắn gọn, dễ hiểu, phù hợp với đối tượng học sinh trung học. Do khuôn khổ có hạn nên một số nội dung chỉ có thể giới thiệu tóm tắt. Vì thế, người dạy và người học cần tham khảo thêm các tài liệu liên quan với ngành học để việc sử dụng giáo trình có hiệu quả hơn.

Do hạn chế về thời gian và kinh nghiệm biên soạn, giáo trình không thể tránh khỏi những thiếu sót nhất định. Tác giả rất mong nhận được ý kiến đóng góp của các bạn đồng nghiệp và độc giả để có thể hoàn thiện hơn nữa nội dung giáo trình này.

Cuối cùng xin chân thành cám ơn tiến sĩ Phạm Thế Quế và thạc sĩ Phạm Ngọc Đĩnh đã đọc và đóng góp nhiều ý kiến quý báu cho chúng tôi trong quá trình biên soạn cuốn giáo trình.

TÁC GIẢ

Chương 1

NHẬP MÔN LÝ THUYẾT TRUYỀN TIN

Mục tiêu

- Nghiên cứu tổng quát hệ thống truyền tin.
- Đánh giá được những vấn đề cơ bản của hệ thống truyền tin.
- Nêu và giải thích được sơ đồ khối chức năng chức năng của hệ thống truyền tin và nêu được chức năng của từng khối trong sơ đồ đó.
- Giải thích được các khái niệm mã hoá và điều chế.

I. TIN TỨC - THÔNG TIN

Vật liệu ban đầu được gia công trong một hệ thống thông tin liên lạc (hệ thống truyền tin) là tin tức. Tin tức (news, nouvelles) là sự phản ảnh của sự vật khách quan đối với sự nhận biết của con người. Tin tức có tính chất là sự "mới mẻ".

Thông tin (information) là sự phản ánh mang tính hướng đích (sự quan tâm của người nhận) của sự vật khách quan đối với sự nhận biết của con người. Hay nói cách khác, thông tin là sự cảm hiểu của con người về thế giới xung quanh thông qua sự tiếp xúc với nó. Ví dụ: Hai người nói chuyện với nhau, cái được trao đổi giữa họ chính là thông tin. Một người đang xem tivi hoặc nghe đài hoặc đọc báo, người đó đang nhận thông tin. Đàm thoại, tham dự diễn đàn, gửi/nhận thư điện tử... chỉ là những ví dụ trong hàng nghìn ví dụ khác nhau về thông tin liên lạc. Thông tin xuất hiện dưới nhiều dạng: âm thanh, hình ảnh, ký hiệu,... Những dạng này chỉ là "vỏ bọc" vật chất chứa thông tin, "vỏ bọc" là phần "xác", thông tin là phần "hồn". Ngữ nghĩa của thông tin chỉ có thể hiểu được khi bên nhận hiểu được cách biểu diễn ngữ nghĩa của bên phát.

Một tin nếu được một ai đó quan tâm thì đó chính là thông tin đối với người đó. Như vậy, càng tiếp xúc với thế giới xung quanh, lượng thông tin mà con người thu nhận được càng nhiều, vì thế họ càng tăng thêm sự hiểu biết và nhận thức tốt hơn về những đối tượng trong đời sống xã hội, trong thiên nhiên, .. giúp cho họ thực hiện hợp lý công việc cần làm để đạt tới mục đích một cách tốt nhất. Khi tiếp nhận được thông tin, con người có thể truyền, lưu trữ, nhân bản hoặc phải xử lý nó để tạo ra những thông tin mới có ích hơn, từ đó có những phản ứng nhất định.

Ví dụ:

- Những đám mây đen đùn lên ở chân trời phía đông chứa đựng thông tin báo hiệu về trận mưa lớn sắp xảy ra.

- Những nốt nhạc trong bản xô-nát Ánh trăng của Beethoven làm cho người nghe cảm thấy được sự tươi mát, êm dịu của đêm trăng.

- Người tài xế chăm chú quan sát người, xe cộ đi lại trên đường, độ tốt xấu của mặt đường, tính năng kỹ thuật cũng như vị trí chiếc xe để quyết định cần tăng tốc độ hay hãm phanh, cần bẻ lái sang trái hay phải,....để đảm bảo an toàn tối đa cho chuyến đi.

Thông tin là một hiện tượng vật lý, nó thường tồn tại và được truyền đi dưới một dạng vật chất nào đó. Về nguyên tắc, bất kỳ cấu trúc vật chất hoặc bất kỳ dòng năng lượng nào cũng có thể mang thông tin. Chúng được gọi là những *vật mang tin* (carrier). Vật mang tin đã chứa thông tin trong nó và là một đại diện của thông tin, sẽ được gọi là *tín hiệu* (signal).

Thông tin là một quá trình *ngẫu nhiên*. Tín hiệu mang tin tức cũng là tín hiệu ngẫu nhiên và mô hình toán học của nó cũng là các quá trình ngẫu nhiên. Vì vậy, lý thuyết truyền tin là *lý thuyết ngẫu nhiên của tin tức*, có nghĩa là nó xét đến tính *bất ngờ* của tin tức đối với nơi nhận tin.

Trước đây, người ta nghiên cứu định lượng hệ thống truyền tin bằng cách tính toán và thực nghiệm sự biến đổi năng lượng mang tin trong các hệ thống đó. Trên quan điểm năng lượng, lý thuyết mạch và tín hiệu đã giải quyết những vấn đề tổng quát về phân tích, tổng hợp mạch và tín hiệu, nhờ đó mà kỹ thuật truyền tin đã có những bước tiến bộ khá dài. Nhưng đồng thời với sự phát triển

mạnh mẽ của mình, trong ngành kỹ thuật truyền tin đã nảy sinh những vấn đề mà lý thuyết xây dựng trên quan điểm năng lượng không giải thích được trọn vẹn như: mối liên hệ cơ bản giữa các hệ thống truyền tin sử dụng những năng lượng khác nhau; vấn đề bảo tồn tin tức trong các hệ thống truyền tin vũ trụ mà ở đó năng lượng tải tin rất nhỏ... Do đó, các lý thuyết xây dựng trên quan điểm năng lượng cần phải được bổ sung bằng những lý thuyết xây dựng trên quan điểm thông tin.

II. LƯỢNG ĐO THÔNG TIN

Chương sau sẽ trình bày về vấn đề lượng đo thông tin (measure of information - còn được dịch là lượng tin) một cách chi tiết hơn. Mục này chỉ nêu một khái niệm ban đầu về lượng tin nhằm vật thể hoá thông tin để có một phương tiện so sánh, định lượng các tin với nhau. Từ đấy sẽ giúp chúng ta dễ nhận thức hơn những chỉ tiêu chất lượng đề ra trong khi xây dựng các phương pháp xử lý thông tin.

Một tin đối với người nhận mang hai nội dung: độ bất ngờ của tin và ý nghĩa của tin. Để so sánh các tin với nhau, chúng ta có thể lấy một hoặc cả hai khía cạnh trên để làm thước đo. Khía cạnh ngữ nghĩa chỉ có ý nghĩa đối với con người. Khía cạnh quan trọng nằm ở chỗ tin thật sự là một cái được chọn từ một **tập các tin** (tập các khả năng) có thể. Số tin trong tập tin càng nhiều thì sẽ mang lại một "lượng tin" càng lớn khi nhận được một tin (giả sử các tin là bình đẳng như nhau về khả năng xuất hiện).

Để truyền tin đạt hiệu quả cao, chúng ta không thể đối đãi các tin như nhau nếu chúng xuất hiện ít nhiều khác nhau.

Xét một tin x có xác suất xuất hiện là $p(x)$, thì chúng ta có thể xem tin này như là một tin trong một tập có $\frac{1}{p(x)}$ tin với các tin có xác suất xuất hiện như nhau. Nếu $p(x)$ càng nhỏ thì $\frac{1}{p(x)}$ càng lớn và vì vậy, "lượng tin" khi nhận được tin này cũng sẽ càng lớn. Vậy "lượng tin" của một tin tỷ lệ thuận với số khả năng của một tin và tỷ lệ nghịch với xác suất xuất hiện của tin đó. Mà xác suất xuất hiện của một tin tỷ lệ nghịch với độ bất ngờ khi nhận được tin đó. Một tin có xác suất xuất hiện càng nhỏ thì độ bất ngờ càng lớn và vì thế, có lượng tin càng lớn.

Vấn đề đặt ra là: hàm f dùng để biểu thị lượng tin phải thỏa mãn những điều kiện gì?

- *Phải phản ánh được tính chất thống kê của tin tức.* Ví dụ có hai nguồn tin K và L với số tin tương ứng là k, l (giả thiết đều đẳng xác suất). Nếu k > l thì độ bất ngờ khi nhận một tin bất kỳ của nguồn K phải lớn hơn độ bất ngờ khi nhận một tin bất kỳ của nguồn L, vậy $f(k) > f(l)$

- *Phải hợp lý khi tính toán:* Giả thiết hai nguồn độc lập K và L với số tin tương ứng là k, l. Cho việc nhận một cặp k_i và l_j bất kỳ đồng thời là một tin của nguồn hỗn hợp KL. Số cặp k_i và l_j mà nguồn này có thể có bằng tích kl. Độ bất ngờ khi nhận được một cặp như vậy phải bằng tổng lượng tin khi nhận được k_i hay l_j. Vì vậy chúng ta phải có:

$$f(kl) = f(k) + f(l)$$

- *Khi nguồn chỉ có một ký hiệu, lượng tin chứa trong ký hiệu duy nhất đó phải bằng không* ($f(1) = 0$). Hay nói cách khác, một tin không cho chúng ta lượng tin nào khi chúng ta biết trước nó hay nó có xác suất bằng 1.

Những điều kiện trên đưa đến việc chọn hàm logarit để làm lượng đo tin là hợp lý vì hàm này thoả mãn được cả ba điều kiện trên.

Định nghĩa: *Lượng đo thông tin của một tin được đo bằng logarit của độ bất ngờ của tin hay nghịch đảo xác suất xuất hiện của tin đó.*

$$I(x) = \log\frac{1}{p(x)} = -\log p(x) \qquad (1\text{-}1)$$

Xét trường hợp nguồn A có m ký hiệu $A = \{ a_1, a_2, ..., a_m \}$ với các xác suất xuất hiện từng ký hiệu tương ứng là $p(a_i)$, $i = 1,2,...m$. Một tin do nguồn A hình thành là một dãy n ký hiệu $x = a_1\, a_2 \ldots a_n$ với $a_i \in A$. Lượng tin chứa trong một tin x như vậy sẽ là :

$$I(x) = \log\frac{1}{p(x)} = -\sum_{i=1}^{n} \log p(a_i) \qquad (1\text{-}2)$$

Trong trường hợp m ký hiệu của nguồn A đẳng xác suất với nhau thì lượng tin chứa trong một ký hiệu là:

$$I(a_i) = \log\frac{1}{p(a_i)} = \log m$$

Lượng tin của tin x lúc đó sẽ bằng n lần lượng tin của một ký hiệu (vì đẳng xác suất).

$$I(x) = n\log m \qquad (1\text{-}3)$$

Đơn vị của lượng tin: Đơn vị lượng tin tuỳ theo cách chọn cơ số của logarit. Hiện nay người ta thường dùng các đơn vị đo sau đây:

- Bit hay đơn vị nhị phân khi cơ số logarit là 2
- Nat hay đơn vị tự nhiên khi cơ số logarit là e
- Hartley hay đơn vị thập phân khi cơ số logorit là 10

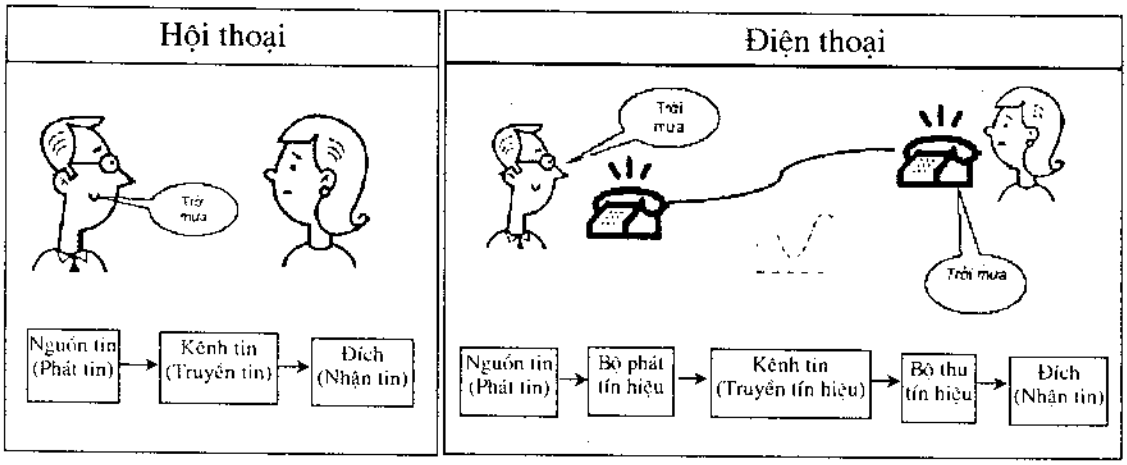
Khi m ký hiệu của nguồn có những xác suất khác nhau và không độc lập thống kê với nhau thì lượng tin riêng từng ký hiệu phụ thuộc vào xác suất xuất hiện $p(a_i)$ của nó và lượng tin của một tin (dãy n ký hiệu của nguồn) không những phụ thuộc vào xác suất xuất hiện của từng ký hiệu mà còn phụ thuộc vào xác suất có điều kiện. Khái niệm này sẽ được đề cập đến một cách cặn kẽ hơn trong chương 2.

III. HỆ THỐNG TRUYỀN TIN

1. Khái niệm và phân loại

Trong cuộc sống, con người luôn có nhu cầu giao lưu trao đổi thông tin với nhau, có nghĩa là có nhu cầu truyền tin (communication). Các dạng trao đổi thông tin có thể như: đàm thoại giữa người với người, đọc sách báo, nghe radio, gửi và nhận thư, nói chuyện qua điện thoại, xem truyền hình, tham dự diễn đàn, truy cập thông tin trên internet,... Nếu không có giao lưu trao đổi thì sẽ không thành tin tức hoặc thông tin.

Ví dụ: Anh A muốn thông báo cho chị B một thông tin là tại một địa điểm nào đó đang có mưa thì sự truyền tin có thể xảy ra như sau:



Hình 1-1: Hai ví dụ về hệ thống truyền tin

Một hệ thống mà thông tin được truyền tải từ nơi phát đến nơi nhận được gọi là hệ thống truyền tin (hoặc hệ thống viễn thông - telecommunication).

Trong các hệ thống truyền tin có sự tham gia của máy tính, thông tin được biểu thị dưới dạng dữ liệu (*data* - còn gọi là số liệu). Mạng truyền và xử lý thông tin dưới dạng dữ liệu được gọi là *mạng truyền số liệu.*

Những hệ thống truyền tin mà con người sử dụng và khai thác có rất nhiều dạng và khi phân loại chúng có thể dựa trên nhiều cơ sở khác nhau. Ví dụ, dựa trên cơ sở năng lượng mang tin, người ta có thể phân hệ thống truyền tin thành các loại:

- Hệ thống điện tín: dùng năng lượng điện một chiều.

- Hệ thống thông tin vô tuyến điện: dùng năng lượng sóng điện từ.

- Hệ thống thông tin quang năng (hệ thống báo hiệu, thông tin hồng ngoại, lazc, cáp quang): dùng năng lượng quang học.

- Hệ thống thông tin dùng sóng âm, siêu âm...: dùng năng lượng cơ học.

Chúng ta cũng có thể phân loại hệ thống truyền tin dựa trên cơ sở biểu hiện bên ngoài của thông tin như:

- Hệ thống truyền thanh, truyền hình.

- Hệ thống truyền số liệu.

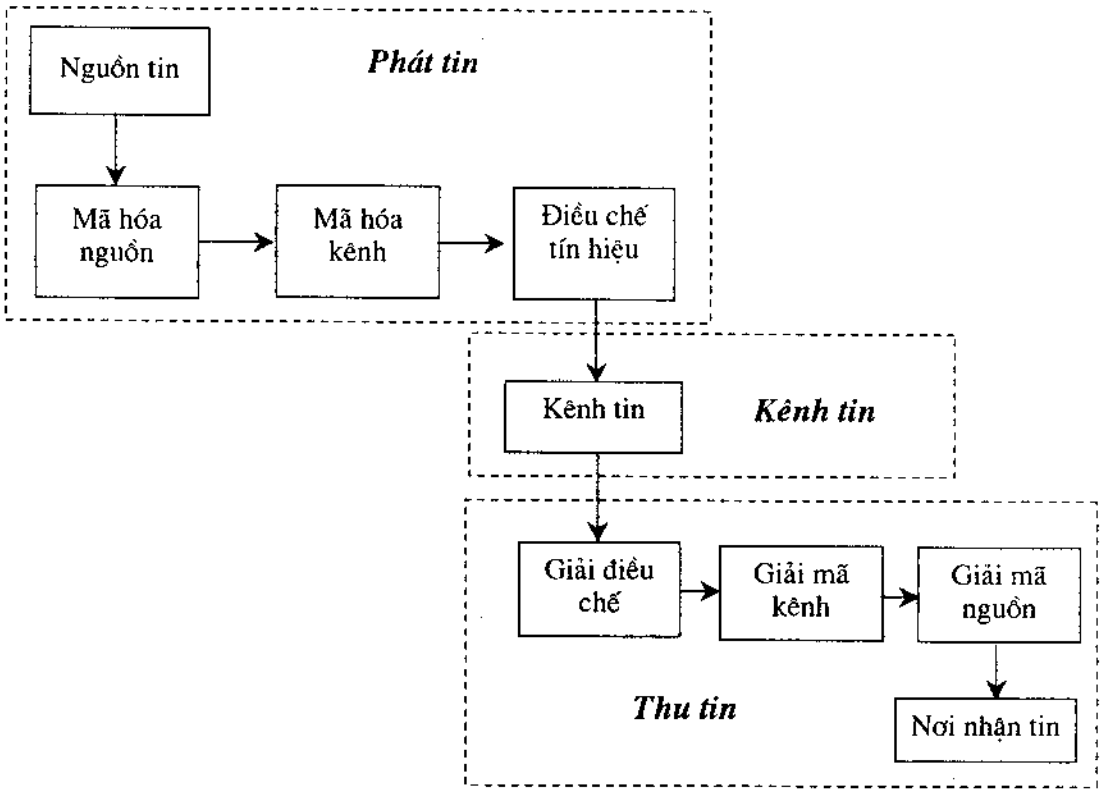
- Hệ thống thông tin thoại...

Kỹ thuật hiện đại ngày nay có thể cho phép truyền tin tức, thông tin dưới các dạng thoại, hình ảnh, số liệu (thoại và phi thoại) trên một hệ thống truyền tin chung.

Những phương pháp phân loại dựa theo nhu cầu kỹ thuật sẽ giúp cho các cán bộ kỹ thuật nhận thức vấn đề một cách cụ thể và khiến cho sự tìm hiểu khai thác các loại hệ thống được dễ dàng nên chúng được ứng dụng rộng rãi. Nhưng ở đây để đảm bảo tính logic của vấn đề được trình bày, chúng ta căn cứ đặc điểm của thông tin đưa vào kênh là rời rạc hay liên tục để phân các hệ thống truyền tin thành hai loại: ***hệ thống truyền tin liên tục*** và ***hệ thống truyền tin rời rạc***

2. Mô hình hệ thống truyền tin

Mô hình tổng quát của một hệ thống truyền tin như sau:



Hình 1-2: Mô hình hệ thống truyền tin

**** Nguồn tin (information source):***

- Là nơi sản sinh ra thông tin.

- Trong quá trình truyền tin, nguồn tin có thể truyền đi một chuỗi các tin (còn gọi là bản tin). Có thể coi nguồn là một tập các tin và khả năng xuất hiện tại mỗi thời điểm của mỗi tin.

- Thông tin có thể thuộc nhiều loại như:

+ Một dãy ký tự như trong điện tín của hệ thống gởi điện tin.

+ Một hàm theo chỉ một biến thời gian $f(t)$, như trong radio và điện thoại.

+ Một vài hàm của một vài biến, như trong trường hợp tivi mầu - ở đây, thông tin bao gồm ba hàm $f(x,y,t)$, $g(x,y,t)$, $h(x,y,t)$ biểu diễn cường độ sáng của ba thành phần cơ bản (xanh lá cây, đỏ, xanh dương).

**** Thiết bị mã hóa***

- Biến đổi các cấu trúc thống kê của nguồn, làm cho các thông số thống kê của nguồn thích ứng với các thông số của kênh như: tốc độ hình thành tin gần với khả năng cho thông qua của kênh, tính chống nhiễu của tin khi truyền qua kênh tăng lên.

- Đối với việc mã hóa thống kê tối ưu : Những tin có xác suất xuất hiện nhiều sẽ được thay thế bởi những từ mã ngắn, và ngược lại, để đảm bảo độ dài trung bình của mã hiệu tối thiểu.

- Đối với mã hóa chống nhiễu: Sử dụng thêm một số bit bổ sung phục vụ cho việc chống nhiễu (phát hiện, sửa lỗi) nhằm tăng độ tin cậy truyền tin.

**** Thiết bị điều chế:***

- Sử dụng sóng mang có tần số phù hợp với môi trường truyền tin, thay đổi các tính chất dữ liệu (tín hiệu điều chế) theo dữ liệu đầu vào (tín hiệu mang tin). Ví dụ: điều chế theo tần số, theo biên độ, theo góc pha...

- Sử dụng tối ưu môi trường truyền tin.

- Tuỳ thuộc vào tạp nhiễu trong kênh mà xây dựng những hệ thống tín hiệu có độ phân biệt với nhau rõ ràng để quá trình giải điều chế dễ dàng phân biệt dù có bị tạp nhiễu làm biến dạng.

**** Kênh tin (channel):***

- Là môi trường truyền tín hiệu từ nguồn tin đến nơi nhận tin.

- Môi trường truyền tin gồm: môi trường định hướng (cáp đồng trục, cáp xoắn, điện thoại,..) hoặc môi trường không định hướng (không khí, tầng điện ly, sóng âm...)

- Trong môi trường truyền tin luôn có các tạp nhiễu (noise) phá huỷ tin tức.

- Trong lý thuyết truyền tin kênh tin; đặc trưng bởi hỗn hợp tín hiệu và tạp nhiễu.

**** Thu tin (sink)***

- Là nơi tiếp nhận thông tin từ kênh tin và cố gắng khôi phục lại thành thông tin ban đầu như ở nguồn tin đã phát đi, bao gồm:

+ Giải điều chế: Xử lý tín hiệu bị biến đổi sau khi truyền như lọc nhiễu, chỉnh méo, lọc các thành phần tín hiệu mang tin.

+ Giải mã kênh: Xây dựng lại thông tin trước khi điều chế, căn cứ vào: thông tin bổ sung, phương thức mã hóa kênh, lọc bỏ các thông tin phục vụ cho việc truyền tin.

+ Giải mã nguồn: Biến đổi thông tin thành dữ liệu cần thiết.

- Nơi nhận tin : Hiển thị thông tin chuyển đến.

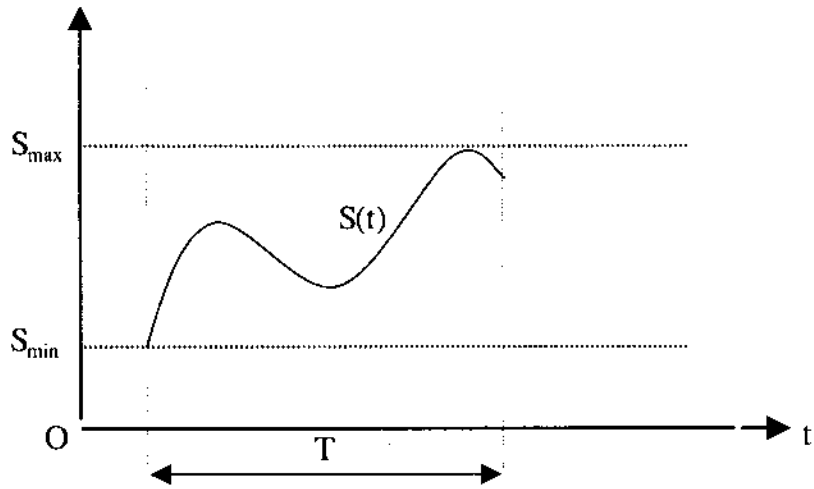
Để tìm hiểu chi tiết hơn; chúng ta đi sâu từng khối chính với chức năng, nhiệm vụ cụ thể của từng khối này.

3. Nguồn tin

3.1. Nguồn tin nguyên thuỷ

Các nguồn tin thường thấy trong tự nhiên được gọi là các *nguồn tin nguyên thuỷ*. Đây là các nguồn tin chưa qua bất kỳ một phép biến đổi nhân tạo nào.

Các tín hiệu âm thanh, hình ảnh được phát ra từ các nguồn tin nguyên thuỷ này thường là các *hàm liên tục theo thời gian và theo mức*, nghĩa là có thể biểu diễn một thông tin nào đó dưới dạng một hàm s(t) tồn tại trong một quãng thời gian t và lấy bất kỳ một giá trị bất kỳ trong phạm vi (S_{min}÷ S_{max})



Hình 1-3: Hàm tin của tín hiệu liên tục

Các nguồn như vậy được gọi là các *nguồn liên tục* (continuous source), các tin được gọi là *tin liên tục* (continuous information) và kênh tin truyền các tín hiệu liên tục gọi là *kênh liên tục* (continuous channel).

Tuy nhiên vẫn còn có những nguồn nguyên thuỷ là rời rạc như: bảng chữ cái của một ngôn ngữ, các tin trong hệ thống điện tín, các lệnh điều khiển trong hệ thống điều khiển. Trong trường hợp này; các nguồn được gọi là *nguồn rời rạc* (discrete source), các tin được gọi là *tin rời rạc* (discrete information) và kênh tin được gọi là *kênh rời rạc* (discrete channel).

Sự phân biệt về bản chất của nguồn rời rạc với nguồn liên tục là số lượng các tin trong nguồn rời rạc là hữu hạn, số lượng các tin trong nguồn liên tục là không đếm được.

Những tin nguyên thuỷ có thể trực tiếp đưa vào kênh để truyền đi nhưng chỉ ở trong phạm vi rất nhỏ. Muốn truyền tin trong phạm vi xa hơn thì phải qua các phép biến đổi nhân tạo (xử lý tín hiệu) như thông qua các mạch điện cảm biến (Sensor) để biến đổi các đại lượng phi điện ra các đại lượng điện. Ví dụ, microphôn biến đổi thanh áp ra dòng điện một chiều biến đổi rồi được khuếch đại, điều chế,.. mới phát vào kênh truyền.

3.2. Rời rạc hoá nguồn tin liên tục

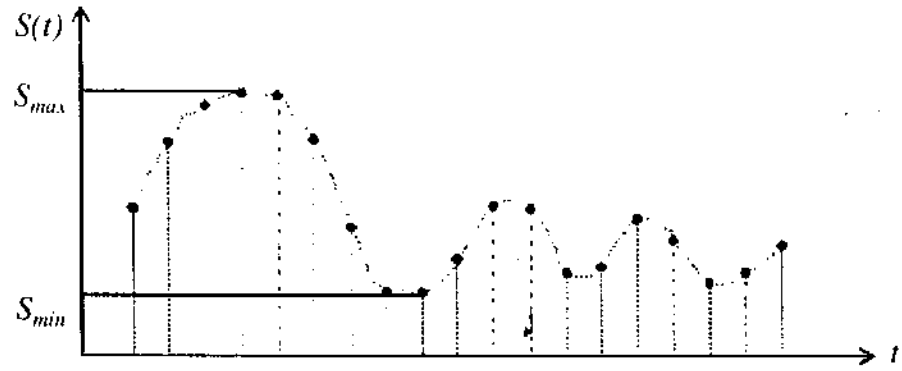
Hệ thống liên tục có nhiều nhược điểm như cồng kềnh, không hiệu quả và chi phí cao. Hệ thống truyền tin rời rạc có nhiều ưu điểm hơn, khắc phục được những nhược điểm của hệ thống liên tục và đặc biệt đang ngày càng được phát triển để hoàn thiện dần những sức mạnh và ưu điểm của nó.

Chính vì vậy trong các hệ thống truyền tin mà bộ thu là những thiết bị xử lý tin tức rời rạc (máy tính) như các hệ thống truyền thống truyền số liệu, hay là các hệ thống thống thông tin chuyển tiếp điều chế mã xung (PCM), nguồn tin có thể là rời rạc hoặc liên tục. Nếu các nguồn tin là liên tục, nhất thiết trước khi đưa vào kênh tin phải thông qua phép biến đổi liên tục thành rời rạc rồi sau đó sẽ áp dụng các phương pháp mã hoá để đáp ứng được các chỉ tiêu kỹ thuật của hệ thống truyền tin cụ thể.

Phép biến đổi nguồn tin liên tục thành rời rạc gồm hai khâu cơ bản: một là khâu rời rạc hóa theo thời gian hay còn gọi là khâu *lấy mẫu*, hai là khâu lượng tử hóa theo mức (viết tắt là *lượng tử hóa)*. Cơ sở lý thuyết của phép biến đổi này gồm các định lý lấy mẫu và luật lượng tử hóa.

3.2.1 Lấy mẫu

Lấy mẫu một hàm tin liên tục, có nghĩa là trích từ hàm đó ra các mẫu tại những những thời điểm nhất định. Nói một cách khác là thay một hàm tin liên tục bằng một hàm rời rạc là những mẫu của hàm trên lấy ra tại những thời điểm gián đoạn. Vấn đề được đặt ra là xét các điều kiện để cho sự thay thế đó là một sự thay thế tương đương. Tương đương ở đây là về mặt ý nghĩa thông tin, nghĩa là hàm thay thế không bị mất mát thông tin so với hàm được thay thế. Các điều kiện này phải thoả mãn định lý lấy mẫu nổi tiếng của Shannon.

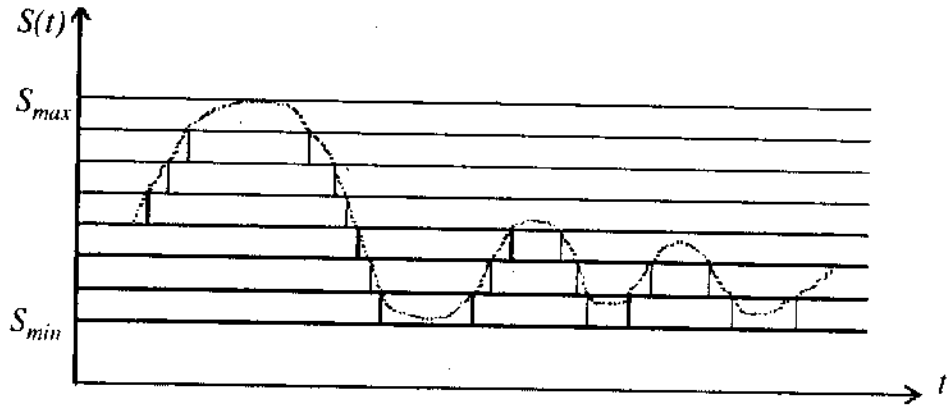


Hình 1-4: Lấy mẫu hàm tin liên tục

3.2.2. Lượng tử hoá

Biên độ của tín hiệu thường là một miền liên tục trong phạm vi ($S_{min} \div S_{max}$). Lượng tử hóa là phân chia miền này thành một số mức nhất định, đánh số các mức từ $S_{min} = S_0, S_1, S_2,...,S_n = S_{max}$. Việc gián đoạn sự biến đổi biên độ của s(t) là cho biên độ lấy mức S_i nhất định khi nó tăng hoặc giảm gần đến mức đó.

Việc lượng tử hoá sẽ biến đổi hàm s(t) ban đầu thành một hàm s'(t) biến đổi theo bậc thang. Sự khác nhau giữa s(t) và s'(t) được gọi là sai số lượng tử. Sai số lượng tử càng nhỏ thì s'(t) càng biểu diễn chính xác s(t) và đồng thời, làm giảm sai nhầm trong quá trình truyền tin.



Hình 1-5: Lượng tử hóa hàm tin liên tục

Một nguồn tin liên tục sau khi được lấy mẫu và lượng tử hoá sẽ trở thành một nguồn rời rạc. Nếu quy ước các mức là những con số thì thay vì gửi đi tín hiệu liên tục là gửi đi những con số tại các thời điểm lấy mẫu - đây chính là quá trình số hóa tin liên tục.

3.3. Mô hình hóa toán học nguồn tin

Để phân tích, nghiên cứu các hệ thống thông tin, người ta thường sử dụng các mô hình toán học hoặc thống kê nguồn, kênh và người sử dụng. Nếu các mô hình của nguồn, kênh và người sử dụng được xây dựng sơ sài thì các bộ thu, bộ phát dù được thiết kế cẩn thận đến đâu cũng không phục vụ hiệu quả quá trình truyền tin. Nhưng nếu các mô hình này được xây dựng quá phức tạp về mặt thống kê hoặc toán học thì cũng không thể thiết kế được các bộ thu, bộ phát thích hợp. Vì vậy, phải có sự phù hợp ở các mặt này.

Có thể xây dựng mô hình toán học cho nguồn tin như sau: Một bản tin xuất phát từ một nguồn tin nào đó đều phản ánh tính chất thống kê của nguồn, bản tin càng dài sự phản ánh càng trung thực. Có thể xem một bản tin cụ thể là thể hiện một quá trình ngẫu nhiên và đứng trên quan điểm toán học, xem nguồn tin là cấu trúc thống kê của quá trình đó. Như vậy, để xác định một nguồn tin, hay nói cách khác để xác định cấu trúc thống kê của một quá trình ngẫu nhiên, chúng ta cần phải biết được các quy luật thống kê của quá trình .

Để nghiên cứu định lượng nguồn tin cũng như hệ thống truyền tin, người ta mô hình hoá toán học nguồn tin bằng bốn quá trình sau:

- Quá trình ngẫu nhiên liên tục: Nguồn tiếng nói, âm nhạc, hình ảnh là tiêu

biểu cho quá trình này. Trong các hệ thống thông tin thoại, truyền thanh, truyền hình với các tín hiệu điều biên, điều tần thông thường chúng ta gặp các nguồn như vậy.

- Quá trình ngẫu nhiên rời rạc: Một quá trình ngẫu nhiên liên tục sau khi lượng tử hoá theo mức sẽ trở thành quá trình này. Một ngôn ngữ, tín hiệu điện tín, các lệnh điều khiển là những nguồn rời rạc thuộc loại này

- Dãy ngẫu nhiên liên tục: Đây là trường hợp nguồn liên tục đã được gián đoạn hoá theo thời gian, như thường gặp trong các hệ thống thông tin điều biên xung (PAM), điều pha xung (PPM), điều tần xung (PFM),... không bị lượng tử hoá.

- Dãy ngẫu nhiên rời rạc: Dãy ngẫu nhiên liên tục được tiếp tục lượng tử hoá theo mức. ví dụ các hệ thống điều biên (pha, tần), xung lượng tử hóa, điều chế xung mã PCM.

4. Kênh tin

Môi trường truyền tin thường rất đa dạng:

- Môi trường không khí, nơi mà tin được truyền dưới dạng âm thanh và tiếng nói, sóng điện từ, ngoài ra cũng có thể bằng lửa hay bằng ánh sáng.

- Môi trường tầng điện ly trong khí quyển, nơi thường xuyên xảy ra sự truyền tin giữa các vệ tinh nhân tạo với các trạm thu phát ở mặt đất...

- Các đường truyền định hướng như dây song hành, cáp đồng trục, cáp quang...

Kênh tin có thể hiểu là một môi trường để truyền lan tín hiệu mang tin đồng thời cũng chịu sự tác động của tạp nhiễu. Tạp nhiễu (thường gọi tắt là nhiễu) là loại tín hiệu người ta không mong muốn. Nó tác động vào tín hiệu mang tin trên kênh truyền làm cho thông tin sai lệch và mất chính xác. Nhiễu do bản thân hệ thống tác động gây nên được gọi là tạp âm và do tác động bên ngoài hệ thống gây nên được gọi là can nhiễu.

Nhiễu rất đa dạng và thường đi kèm với môi trường truyền tin tương ứng. Chẳng hạn: nếu truyền dưới dạng sóng điện từ đi qua các vùng của trái đất có từ trường mạnh, tín hiệu mang tin bị ảnh hưởng ít nhiều bởi từ trường này, thì có thể coi từ trường là một loại nhiễu. Nếu truyền âm thanh trong không khí thì tiếng ồn xung quanh cũng có thể coi là một loại nhiễu.

Như vậy, khi tín hiệu đi qua các môi trường truyền tin, ngoài sự biến đổi về năng lượng, dạng tín hiệu cũng bị thay đổi do tác động của tạp nhiễu tồn tại trong các môi trường vật lý hoặc do các phương thức truyền lan. Sự biến đổi các thông số vật lý của môi trường gây ra sự điều chế tín hiệu không cần thiết. Rõ ràng, tác động của nhiễu lên tín hiệu tiêu biểu cho môi trường truyền lan của tín hiệu. Vậy có thể lấy tạp nhiễu làm đặc tính chung của môi trường truyền lan và lấy sự phân tích, phân loại tạp nhiễu để phân tích và phân loại môi trường. Tuy môi trường truyền lan trong thực tế rất khác nhau, song vẫn có thể quy nạp chúng theo các dạng cơ bản sau:

- Môi trường có tác động nhiễu cộng là chủ yếu.

- Môi trường có tác động của nhiễu nhân là chủ yếu.

- Môi trường gồm cả nhiễu cộng và nhiễu nhân.

Ngoài ra, trong trường hợp sự truyền tin xảy ra giữa hai vật di động so với nhau, tín hiệu sẽ bị điều tần phụ do hiệu ứng Doppler gây nên, chúng ta xếp riêng một loại, gọi là kênh có hiệu ứng Doppler.

Tóm lại, để mô tả kênh, chúng ta dùng một mạng hai cửa và sự quan hệ giữa tín hiệu đầu ra và tín hiệu đầu vào được mô tả như hình 1- 6

Hình 1-6: Mô hình kênh với tín hiệu vào và ra

Với giả thiết mạng hai cửa này có hàm truyền đơn vị (bằng1) trên mọi tần số và trên toàn miền thời gian, chúng ta có:

$$S_r(t) = N_N(t)S_V(t) + N_C(t)$$

Trong đó, $N_N(t)$ ký hiệu nhiễu nhân và $N_C(t)$ ký hiệu cho nhiễu cộng.

- Nhiễu cộng do các nguồn nhiễu công nghiệp và vũ trụ tạo ra, luôn luôn tồn tại trong các môi trường truyền lan của tín hiệu. Dải phổ của nhiễu cộng rất rộng, cho nên với bất kỳ tín hiệu có phổ ở đoạn tần số nào, chúng cũng tạo thành cái nền trùm lên tín hiệu. Nhiễu cộng làm cho tín hiệu đến máy thu bị sai lạc.

- Nhiễu nhân là do phương thức truyền lan của tín hiệu hay là sự thay đổi

thông số vật lý của bộ phận môi trường truyền lan khi tín hiệu đi qua. Trong trường hợp đầu, nhiễu sẽ tác động nhanh lên tín hiệu và tác động chậm trong trường hợp thứ hai vì các biến động của môi trường thường xảy ra với những chu kỳ vài phút đến vài giờ hoặc hơn nữa. Hiện tượng này thường gặp trong khi thu các tín hiệu vô tuyến điện ở dải sóng ngắn bằng nhiều con đường truyền lan khác nhau, tuỳ theo sai trình (dài ngắn khác nhau) của các đường đó thay đổi làm cho tổng cường độ điện trường ở đầu thu biến đổi, gây ra biên độ tín thu khi lớn, khi bé và đôi khi mất hẳn, chúng ta gọi là hiện tượng Phading.

Nhiệm vụ của kênh truyền là chuyển tín hiệu mang tin từ nguồn tin đến nơi nhận tin và đảm bảo tính toàn vẹn của tin tức với thời gian quy định. Vì môi trường truyền lan của tín hiệu rất khác nhau nên vấn đề đặt ra là từ các dạng khác nhau đó, cần tìm ra được những đặc điểm chung để có thể tổng quát hóa kênh.

5. Mã hóa - Giải mã

Mã hoá là một phép biến đổi cấu trúc thống kê của nguồn. Phép biến đổi ấy tương đương trên quan điểm thông tin (nghĩa là phép biến đổi không thay đổi lượng tin) và nhằm cải tiến các chỉ tiêu kỹ thuật của hệ thống truyền tin. Nói một cách khác, lớp tin ở đầu vào của thiết bị mã hoá được thay thế bằng một lớp tin khác ở đầu ra, tương đương và kinh tế hơn, như: tốc độ hình thành tin gần với khả năng cho thông qua của kênh, tính chống nhiễu của tin khi truyền qua kênh cũng tăng lên.

Ví dụ chúng ta có một nguồn tin gồm có 4 tin đẳng xác suất với sơ đồ thống kê như sau:

$$A = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

Lượng tin $I(a_i)$ chứa trong một tin của A bằng:

$$I(a_i) = 1 \times \log_2 4 = 2 \text{ bit}$$

Bằng một phép mã hóa như sau:

$$\left.\begin{array}{l} a_1 \longrightarrow b_1b_1 \\ a_2 \longrightarrow b_1b_2 \\ a_3 \longrightarrow b_2b_1 \\ a_4 \longrightarrow b_2b_2 \end{array}\right\}$$

Chúng ta đổi thành một nguồn tin mới B gồm hai ký hiệu đẳng xác suất:

$$B = \begin{bmatrix} b_1 & b_2 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Lượng tin chứa trong một tin của B cũng vẫn bằng lượng tin chứa trong tin tương ứng của A. Ví dụ tin b_1b_2 của B tương ứng với tin a_2 của A có lượng tin là: $I(b_1b_2) = 2\times\log_2 2 = 2$bit

Ở ví dụ ở trên, giả thiết các ký hiệu của nguồn là đẳng xác suất, nhưng thông thường các ký hiệu trong một nguồn rời rạc, hoặc một nguồn liên tục được rời rạc hoá đều có những xác suất xuất hiện khác nhau. Ví dụ, trong nguồn chữ viết bằng tiếng Việt, xác suất xuất hiện của chữ b lớn hơn nhiều so với xác suất xuất hiện của chữ y. Khi mã hóa để tăng tốc độ hình thành các tin, đối với các ký hiệu của nguồn cũ có xác suất xuất hiện lớn thì phải mã hóa bằng một dãy ít ký hiệu của nguồn mới và ngược lại, đối với nguồn cũ có xác suất xuất hiện nhỏ hơn thì phải mã hóa bằng dãy nhiều ký hiệu hơn. Tóm lại, độ dài của mã hiệu (số ký hiệu hợp thành) lớn hay bé tuỳ theo xác suất xuất hiện các tin của nguồn cũ mà chúng thay thế. Quy luật mã hóa như thế nào để đạt được một tốc độ hình thành tin cực đại, nghĩa là độ dài trung bình của mã hiệu tối thiểu, thuộc phạm vi giải quyết của mã hóa thống kê. Phương pháp mã hóa này chỉ tùy thuộc tính chất thống kê của nguồn.

Trường hợp truyền tin trong kênh có nhiễu, các tín hiệu điện đại diện luồng bit truyền rất dễ bị lỗi do sự thâm nhập điện từ cảm ứng lên các đường dây từ các thiết bị điện gần nó, đặc biệt là các đường dây tồn tại trong một môi trường xuyên nhiễu. Để xác suất thông tin thu được ở đích giống như thông tin đã truyền đạt, cần phải có một vài biện pháp để nơi thu có khả năng biết được thông tin nó thu được có chứa lỗi hay không. Hơn nữa, khi phát hiện được lỗi, cần phải có một cơ cấu thích hợp để thu về bản copy chính xác của thông tin.

Biện pháp đó là truyền thêm những thông tin bổ sung nhằm giúp phía thu không chỉ phát hiện được lỗi mà còn xác định được lỗi nằm ở đâu để khắc phục chúng. Đó chính là phương pháp mã hóa chống nhiễu.

Tóm lại, hai nhiệm vụ lớn mà mã hóa cần phải đạt được là: tăng hiệu suất truyền tin và tăng độ tin cậy.

Giải mã là phép biến đổi ở phía thu để chuyển đổi dữ liệu đã được mã hóa thành thông tin cần thiết.

6. Điều chế - giải điều chế

Trong các hệ thống truyền tin liên tục, các tin hình thành từ nguồn tin liên tục được biến đổi thành các đại lượng điện (điện áp hoặc dòng điện) và chuyển vào kênh truyền như trường hợp điện thoại trong thành phố. Đây là các dao động điện có tần số thấp, biên độ nhỏ nên không thể bức xạ thành sóng điện từ để chuyển đi xa được. Khi muốn chuyển các tín hiệu ấy qua một cự ly lớn thì phải cho qua một phép biến đổi khác là điều chế.

Điều chế có nghĩa là chuyển tin tức trong một dạng năng lượng thích hợp với môi trường truyền lan. Dạng năng lượng được dùng, phải ít bị tổn hao và bị biến dạng do tác động của nhiễu.

Thực chất của phép điều chế là biến đổi một hoặc nhiều thông số của dạng năng lượng đã chọn theo quy luật đại biểu cho tin tức. Ví dụ, sự thông thoại giữa các vùng trong thành phố được thực hiện bằng các đường tải ba, trong đó quy luật tin tức điều khiển sự biến đổi của một thông số (biên độ, tần số) của năng lượng dòng điện xoay chiều tần số thấp (khoảng vài chục KHz). Thông tin ở cự ly xa hơn sẽ được thực hiện bằng các đường thông tin vô tuyến điện như viba chẳng hạn. Lúc này, quy luật tin tức điều khiển một hoặc nhiều thông số của trường điện từ cao tần

Đối với các hệ thống truyền tin rời rạc, quy luật mã hiệu điều khiển một hoặc nhiều thông số của dạng năng lượng được dùng để mang tin trên đường truyền. Ví dụ, trường hợp điện báo thông thường quy luật mã hiệu điều khiển biên độ dòng điện một chiều. Với các dạng năng lượng khác như dòng điện xoay chiều hay sóng điện từ, chúng ta sẽ có các hệ thống truyền tin bằng điện tải ba hoặc thông tin vô tuyến điện điều chế mã.

Nhiệm vụ của phép điều chế là ngoài việc chọn năng lượng thích hợp với

sự truyền lan trong môi trường (sóng điện từ) còn tuỳ theo tính chất của tạp nhiễu trong kênh mà xây dựng một hệ thống tín hiệu có độ phân biệt với nhau rõ ràng để quá trình giải điều chế có thể dễ dàng phân biệt dù có bị tạp nhiễu làm biến dạng đến phần nào.

Trong các hệ thống truyền tin hiện nay, các phương pháp điều chế thường được dùng đối với tin tức liên tục là điều chế biên độ (AM- Amplitude Modulation), điều chế tần số (FM - Frequency Modulation) và điều chế góc pha (PM - Phase Modulation) cao tần. Để tăng tính chống nhiễu dùng các phương pháp điều chế kép, ngoài điều chế cao tần có thêm một điều chế phụ như *điều chế xung* (điều chế các thông số của một dãy xung tuần hoàn, có chu kỳ lặp lại thỏa mãn điều kiện đã nêu ra trong định lý lấy mẫu) như: điều chế góc pha xung (PPM - Pulse Phase Modulation), điều chế độ rộng xung (PWM - Pulse Width Modulation), điều chế tần số xung (PFM - Pulse Frequency Modulation) và điều chế biên độ xung (PAM Pulse Amplitude Modulation). Một phương pháp điều chế phụ thường dùng là điều chế mã xung (PCM- Pulse Code Modulation) và điều chế delta. Khi đã biến tin liên tục thành tin rời rạc, sự điều chế cao tần hoàn toàn giống như trường hợp đối với hệ thống truyền tin rời rạc.

Đối với các tin tức rời rạc, các phương pháp điều chế cao tần cũng giống như đối với các trường hợp tin tức liên tục nhưng làm việc gián đoạn theo thời gian, được gọi là manip hay khóa dịch. Cụ thể là có các phương pháp manip biên độ (ASK - Amplitude Shift Key), manip pha (PSK- Phase Shift Key) và manip tần số (FSK - Frequency Shift Key), điều chế biên độ cầu phương (Quadrature Amplitude Modulation - QAM)

Những năm gần đây, do sự phát triển của lý thuyết tin tức và lý thuyết tín hiệu, bắt đầu áp dụng đến các tín hiệu dải rộng (tín hiệu giả nhiễu, có phổ và hàm tương quan giống tạp âm trắng), một phương pháp điều chế được nghiên cứu và áp dụng trong kỹ thuật thông tin một cách có hiệu quả, đó là phương pháp điều chế nhiễu. Điều chế nhiễu cũng được dùng như một điều chế phụ đối với tin tức rời rạc để tăng cường tính chống nhiễu của tín hiệu.

Giải điều chế là phép biến đổi ngược lại của phép điều chế. Điều khác là đầu vào của thiết bị giải điều chế không phải là tín hiệu đầu ra của thiết bị điều chế mà nó là hỗn hợp của tín hiệu điều chế và tạp nhiễu. Nhiệm vụ của thiết bị

giải điều chế là từ trong hỗn hợp đó lọc ra được những tin tức dưới dạng một hàm điện áp liên tục hoặc là một dãy xung rời rạc giống như tin tức ở đầu vào của thiết bị điều chế với một sai số trong phạm vi cho phép.

Phương pháp giải điều chế, nói cách khác là phép lọc tin, được thực hiện tuỳ theo hỗn hợp tín hiệu nhiễu với các chỉ tiêu tối ưu về sai số (độ chính xác) đề ra cho nó. Chúng ta có các phương pháp lọc tin thông thường như tách sóng biên độ, tách sóng tần số, tách sóng pha, tách sóng đồng bộ, lọc tin liên kết, lọc tin bằng phương pháp tương quan, lọc tối ưu.

IV. NHỮNG VẤN ĐỀ CƠ BẢN CỦA HỆ THỐNG TRUYỀN TIN

Yêu cầu tối đa đối với bất kỳ một hệ thống truyền tin nào chính là việc thực hiện sự truyền tin một cách nhanh chóng và chính xác. Như vậy, các vấn đề cơ bản của hệ thống truyền tin gồm có:

- *Nâng cao hiệu suất truyền tin* : hay là nâng cao tốc độ truyền tin của hệ thống - chính là nâng cao lượng thông tin mà hệ thống có khả năng truyền đi trong một đơn vị thời gian.

- *Nâng cao độ chính xác truyền tin*, nói cách khác là nâng cao khả năng chống nhiễu của hệ thống. Cụ thể là:

+ Truyền tin với sai số nhỏ tuỳ ý.

+ Truyền tin khoảng cách xa, tín hiệu ít bị suy yếu.

+ Nhiễu nhỏ hơn nhiều so với tín hiệu.

Hai vấn đề trên dường như mâu thuẫn với nhau, bởi tốc độ truyền tin tăng lên sẽ làm giảm độ chính xác của tin tức truyền đi. Nhưng trong thực tế, tốc độ hình thành tin của nguồn còn rất thấp so với khả năng cho thông qua của kênh, nghĩa là vẫn có thể cho phép tốc độ truyền tối đa mà không bị sai nhầm. Vì thế, bên cạnh hướng tìm tòi chủ yếu để giải quyết tốc độ hình thành tin của nguồn như thay đổi cấu trúc thống kê của nguồn bằng phương pháp mã hóa thống kê, một hướng tìm tòi khác là tận dụng khả năng cho thông qua của kênh để xây dựng những hệ thống tín hiệu có khả năng chống nhiễu cao nhằm nâng cao độ chính xác của tin tức truyền đi như những mã hiệu chống nhiễu, những tín hiệu phức tạp. Hướng nghiên cứu thứ ba là dựa trên cơ sở tính chất thống kê của tin tức, xây dựng những cấu trúc lọc tin tối ưu sau khi đã bị tạp nhiễu phá hoại. Những khái niệm về lý thuyết thông tin cho biết giới hạn tốc độ truyền tin

trong một kênh tin, nghĩa là khối lượng thông tin lớn nhất mà kênh cho truyền qua với một độ sai nhầm nhỏ tuỳ ý. Những lý thuyết ở phần sau sẽ giải đáp những vấn đề này.

Khi sự truyền tin tiến hành trên những cự ly rất lớn, người ta thường dùng năng lượng mang tin là sóng điện từ. Trong trường hợp công suất máy phát bị hạn chế, năng lượng tín hiệu và tạp nhiễu ở đầu thu sẽ xấp xỉ bằng nhau. Vì thế một vấn đề được đặt ra là xác định cấu trúc của thiết bị thu tín hiệu lý tưởng, nghĩa là có thể phát hiện và tách tín hiệu trong nền tạp âm lớn. Đó là nội dung của lý thuyết chống nhiễu.

Câu hỏi ôn tập

1/ Nêu khái niệm thông tin, tín hiệu.

2/ Nêu khái niệm hệ thống truyền tin và phân loại chúng.

3/ Nêu mô hình tổng quát một hệ thống truyền tin và giải thích chức năng của từng khối trong mô hình đó.

4/ Làm thế nào để rời rạc hóa một nguồn tin liên tục?

5/ Nêu khái niệm và nhiệm vụ của mã hóa.

6/ Nêu khái niệm và nhiệm vụ của điều chế.

7/ Các vấn đề cơ bản của một hệ thống truyền tin là gì?

8/ Nêu khái niệm về độ đo thông tin.

Chương 2

THÔNG TIN VÀ ĐỊNH LƯỢNG THÔNG TIN

Mục tiêu

- Nghiên cứu phương pháp định lượng thông tin trong nguồn tin, kênh tin.
- Phân biệt được các khái niệm lượng tin trung bình, tốc độ lập tin của nguồn, entropi, entropi đồng thời, entropi có điều kiện, thông lượng của kênh và ghi nhớ các công thức xác định chúng.
- Giải thích được ý nghĩa các định lý Shannon
- Vận dụng để giải các bài tập định lượng thông tin trong nguồn tin, kênh tin.

Trong chương 1, chúng ta đã xét một cách tổng quát hệ thống truyền tin và các vấn đề cơ bản cũng như cơ cấu của hệ thống. Chương này sẽ khảo sát kỹ hơn các cơ sở lý thuyết và các phép xử lý tin tức ở khối nguồn tin, cụ thể là sẽ lần lượt giải quyết các vấn đề cơ bản sau:

- Định lượng tin tức của nguồn tin.
- Khái niệm Entropi.
- Khái niệm tốc độ hình thành tin của nguồn.
- Khái niệm thông lượng của kênh.
- Các định lý cơ sở của lý thuyết mã hóa.

Mặt khác, vì thông tin là một quá trình *ngẫu nhiên*, tín hiệu mang tin tức cũng là tín hiệu ngẫu nhiên, cho nên lý thuyết xác suất và quá trình ngẫu nhiên chính là công cụ toán học để nghiên cứu các hệ thống truyền tin. Trước khi giải quyết các vấn đề chính của chương này, mục I sẽ trình bày một số kiến thức cơ bản về xác suất thống kê để rút ra những kết quả cần thiết ứng dụng vào việc nghiên cứu hệ thống truyền tin.

I. XÁC SUẤT

1. Khái niệm về xác suất

**** Không gian mẫu (Sample space):***

Là tập hay không gian tất cả các kết quả có thể có của một thí nghiệm và thường được ký hiệu là S hoặc E. Nếu không gian mẫu là rời rạc thì S có thể biểu diễn bằng $S = \{ s_1, s_2, ..., s_n \}$

**** Sự kiện (Event), sự kiện cơ bản (Elementary event):***

Mỗi phần tử của S (không gian mẫu) được gọi là một sự kiện cơ bản, mỗi tập con của S được gọi là một sự kiện.

Ví dụ: Trong thí nghiệm tung một con xúc sắc thì tập các giá trị có thể xuất hiện là $S = \{1, 2, 3, 4, 5, 6\}$. Ở đây, mỗi giá trị xuất hiện chính là số lượng điểm chấm có trên mặt ngửa của con xúc sắc và chúng được gọi là sự kiện cơ bản. Khi gieo một số con xúc sắc thì một tập con các giá trị của S sẽ xuất hiện và sự xuất hiện này là một sự kiện. Chẳng hạn sự kiện A là việc xuất hiện hai giá trị 2 và 4 thì $A = \{2,4\}$

Sự kiện bù của sự kiện A được ký hiệu là $\bar{A}$, là một tập con gồm các phần tử của S nhưng không thuộc A. Ví dụ, $\bar{A} = \{1,3,5,6\}$.

Hai sự kiện được gọi là loại trừ nhau nếu chúng không chứa một giá trị chung nào. Ví dụ, nếu $B = \{1,3\}$ thì A và B là loại trừ nhau.

Hợp của hai sự kiện là sự kiện chứa tất cả các giá trị có trong hai sự kiện. Phép hợp được ký hiệu là '$\cup$'. Ví dụ: nếu $C = \{1,2,6\}$ thì $D = A \cup C = \{1,2,4,6\}$

Giao của hai sự kiện là sự kiện chứa các giá trị chung trong hai sự kiện. Giao được ký hiệu là '$\cap$'. Ví dụ: $E = \{3,4,6\}$ thì $H = E \cap D = \{4,6\}$

**** Định nghĩa xác suất:***

Thực hiện phép thử n lần. Giả sử sự kiện A xuất hiện m lần. Khi đó m được gọi là tần số của sự kiện A và tỷ số $\frac{m}{n}$ được gọi là tần suất xuất hiện sự kiện A trong loạt phép thử. Cho số phép thử tăng lên vô hạn, tần suất xuất hiện sự kiện A dần về một số xác định gọi là xác suất xuất hiện sự kiện A, ký hiệu là P(A) và được xác định như sau:

$$P(A) = \lim_{n\to\infty} \frac{m}{n} \quad (2\text{-}1)$$

Ví dụ: Một xạ thủ bắn 100 viên đạn vào bia, có 70 viên trúng bia. Khi đó, xác suất mà xạ thủ bắn trúng bia là $\frac{70}{100} = 70\%$.

Ví dụ 2: Nếu tung súc sắc (súc sắc đồng nhất) thì xác suất:

$P(1) = P(2) = ...= P(6) = \frac{1}{6}$, $P(2,5) = \frac{1}{3}$, $P(1,3,5) = \frac{1}{2}$

Rõ ràng với định nghĩa trên thì chúng ta luôn có $0 \leq P(A) \leq 1$

Nếu có một tập các sự kiện loại trừ nhau A_i, i =1,2,.... thuộc tập S, thì xác suất xuất hiện của sự kiện hợp của chúng bằng tổng các xác suất xuất hiện của các sự kiện thành phần:

$$P(\cup A_i) = \sum P(A_i) \qquad (2\text{-}2)$$

**** Sự kiện đồng thời và xác suất đồng thời:***

Sự kiện đồng thời là sự kiện mà hai sự kiện riêng đồng thời xuất hiện.

Cụ thể: nếu thực hiện một phép thử làm xuất hiện sự kiện A_i, i = 1,2,.. n và phép thử thứ hai làm xuất hiện sự kiện B_j, j = 1,2,.. m thì phép thử đồng thời sẽ làm xuất hiện sự kiện đồng thời (A_i,B_j), i =1,2,..., n, j = 1,2,..., m

Ứng với mỗi sự kiện đồng thời này là một khả năng xuất hiện của nó và được gọi là xác suất xuất hiện đồng thời $P(A_i,B_j)$.

Nếu các sự kiện B_j loại trừ nhau thì : $\sum_{j=1}^{m} P(A_i, B_j) = P(A_i)$.

Tương tự, nếu các sự kiện A_i loại trừ nhau thì : $\sum_{i=1}^{n} P(A_i, B_j) = P(B_j)$.

Nếu tất cả các sự kiện A_i và B_j loại trừ nhau thì : $\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i, B_j) = 1$.

**** Xác suất có điều kiện:***

Giả thiết rằng một thực nghiệm đồng thời đã được xác định và sự kiện đồng thời với xác suất P (A,B). Khi đang thực nghiệm, giả thiết sự kiện B đã xuất

hiện và cần xác định xác suất xuất hiện sự kiện A. Xác suất này gọi là xác suất có điều kiện của sự kiện A với điều kiện sự kiện B đã xuất hiện và được định nghĩa :

$$P(A \mid B) = \frac{P(A,B)}{P(B)} \qquad (2\text{-}3)$$

Tương tự, xác suất có điều kiện của sự kiện B với điều kiện sự kiện A đã xuất hiện là:

$$P(B \mid A) = \frac{P(A,B)}{P(A)} \qquad (2\text{-}4)$$

Với điều kiện P(A) > 0 hoặc P(B) > 0 kết hợp cả hai (2-3) và (2-4) ta có:

$$P(A,B) = P(A).P(B \mid A) = P(B).P(A \mid B) \qquad (2\text{-}5)$$

Một quan hệ thường dùng của xác suất có điều kiện là công thức Bayes. Công thức này nói rằng: nếu các sự kiện A_i , i = 1,2,..n là loại trừ nhau và B là một sự kiện xuất hiện đồng thời với các sự kiện A_i và P(B) > 0 thì:

$$P(A_i \mid B) = \frac{P(A_i, B)}{P(B)} = \frac{P(B \mid A_i).P(A_i)}{\sum_{j=1}^{n} P(B \mid A_j).P(A_j)} \qquad (2\text{-}6)$$

Trong hệ thống truyền tín hiệu số, các sự kiện A_i sẽ được coi là các tin có thể được phát, B được coi là tin nhận được khi phía nguồn phát tin A_i và có nhiễu tác động, xác suất $P(A_i \mid B)$ được coi là xác suất để nguồn tin phát tin A_i khi phía thu đã nhận được tin B. Trong công thức Bayes, $P(A_i \mid B)$ được gọi là xác suất hậu nghiệm, còn $P(A_i)$ được gọi là xác suất tiên nghiệm.

**** Tính độc lập thống kê của các sự kiện***

Tính độc lập thống kê của các sự kiện là một khái niệm quan trọng trong lý thuyết xác suất. Để giải thích khái niệm này, chúng ta xem xét hai sự kiện A và B và xác suất có điều kiện của chúng là P(A|B) hoặc P(B|A). Giả thiết việc xuất hiện của sự kiện A không phụ thuộc vào sự xuất hiện của sự kiện B hoặc ngược lại thì ta có:

$$\left.\begin{array}{l} P(A|B) = P(A) \\ P(B|A) = P(B) \end{array}\right\} \Longrightarrow \quad P(A,B) = P(A).P(B) \qquad (2\text{-}7)$$

Lúc này xác suất đồng thời là tích trực tiếp của hai xác suất thành phần. Trong lý thuyết xác suất, người ta nói hai sự kiện là độc lập thống kê khi chúng thỏa mãn quan hệ (2-7)

2. Đại lượng ngẫu nhiên và phân phối xác suất

**** Đại lượng ngẫu nhiên:***

Đại lượng ngẫu nhiên là đại lượng biến đổi biểu thị giá trị kết quả của một phép thử ngẫu nhiên. Ta dùng các chữ cái hoa như X, Y, Z,... để ký hiệu đại lượng ngẫu nhiên.

**** Đại lượng ngẫu nhiên rời rạc:***

Đại lượng ngẫu nhiên được gọi là rời rạc nếu nó chỉ nhận một số hữu hạn hoặc một số vô hạn đếm được các giá trị. Có thể liệt kê các giá trị của đại lượng ngẫu nhiên rời rạc $x_1, x_2, ..., x_n$. Ta ký hiệu đại lượng ngẫu nhiên X nhận giá trị x_n là $X= x_n$ và xác suất để X nhận giá trị x_n là $P(X=x_n)$

Ví dụ : Gọi X là số chấm xuất hiện trên mặt con xúc sắc thì X là một đại lượng ngẫu nhiên. Số học sinh vắng mặt trong một buổi học cũng là một đại lượng ngẫu nhiên rời rạc.

**** Đại lượng ngẫu nhiên liên tục:***

Đại lượng ngẫu nhiên được gọi là liên tục nếu các giá trị có thể của nó lấp đầy một khoảng trên trục số. Ví dụ: nhiệt độ không khí ở mỗi thời điểm nào đó, sai số khi đo lường một đại lượng vật lý.

Hàm mật độ xác suất:

Hàm mật độ xác suất của đại lượng ngẫu nhiên liên tục X là hàm không âm $f(x)$, xác định với mọi $x \in (-\infty, +\infty)$ và thoả mãn:

$$P(X \in B) = \int_B f(x)dx \qquad (2\text{-}8)$$

với mọi tập số thực B

Tính chất của hàm mật độ xác suất:

- $f(x) \geq 0, \forall\ x \in (-\infty, +\infty)$
- $\int_{-\infty}^{+\infty} f(x)dx = 1$

** Hàm phân phối xác suất:*

Hàm phân phối xác suất của đại lượng ngẫu nhiên X, ký hiệu F(x) là hàm được định nghĩa như sau:

$$F(x) = P(X < x) \qquad (2\text{-}9)$$

Nếu X là đại lượng ngẫu nhiên rời rạc nhận các giá trị có thể $x_1, x_2, ..., x_n$ thì:

$$F(x) = \sum_{x_i < x} P(X = x_i) = \sum_{x_i < x} pi \quad (\text{với } p_i = P(X = x_i)) \qquad (2\text{-}10)$$

Nếu X là đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ thì:

$$F(x) = \int_{-\infty}^{x} f(x)dx \qquad (2\text{-}11)$$

Ý nghĩa: Hàm phân phối xác suất $F(x)$ phản ánh mức độ tập trung xác suất về bên trái của điểm x.

Tính chất của hàm phân phối xác suất:

- $0 \le F(x) \le 1$; $\forall x$.
- $F(x)$ là hàm không giảm ($x_1 \le x_2 \rightarrow F(x_1) \le F(x_2)$)
- $F'(x) = f(x)$, $\forall x$.

**** Đại lượng ngẫu nhiên nhiều chiều, hàm phân phối, hàm mật độ xác suất của biến ngẫu nhiên nhiều chiều***

Khi thực hiện đồng thời nhiều phép thử hoặc thực hiện một phép thử phức tạp, chúng ta sẽ gặp một đại lượng ngẫu nhiên nhiều chiều. Đại lượng này nhận một bộ giá trị ngẫu nhiên mà mỗi giá trị có thể là một giá trị đại lượng ngẫu nhiên thành phần. Hàm phân bố xác suất và hàm mật độ phân bố xác suất của đại lượng này được gọi là hàm phân bố xác suất và mật độ phân bố xác suất nhiều chiều.

Giả sử biến ngẫu nhiên X gồm hai đại lượng ngẫu nhiên thành phần X_1, X_2. Hàm phân bố xác suất hai chiều sẽ là:

$$F(x_1, x_2) = P(X_1 \le x_1, X_2 \le x_2) \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} p(u_1, u_2)du_1 du_2 \qquad (2\text{-}12)$$

Hàm mật độ phân bố xác suất hai chiều là:

$$p(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F(x_1, x_2) \qquad (2\text{-}13)$$

Khi lấy tích phân hàm mật độ phân bố xác suất hai chiều theo một đại lượng, ta được hàm mật độ phân bố xác suất của đại lượng kia, có nghĩa là:

$$\int_{-\infty}^{+\infty} p(x_1, x_2) dx_1 = p(x_2) \qquad (2\text{-}14)$$

$$\int_{-\infty}^{+\infty} p(x_1, x_2) dx_2 = p(x_1) \qquad (2\text{-}15)$$

II. LƯỢNG TIN CỦA NGUỒN RỜI RẠC

1. Nguồn tin rời rạc

Định nghĩa: *Nguồn tin rời rạc là nguồn tin tạo ra các tin (biến ngẫu nhiên) dưới dạng rời rạc: $x_1, x_2, ..., x_n$.*

Ký hiệu (x_i) là phần tử nhỏ nhất có chứa thông tin.

Bộ ký hiệu là tập hợp tất cả các ký hiệu có thể, còn được gọi là bảng chữ cái.

$X = \{x_1, x_2, ..., x_n\}$.

Từ là tập hợp hữu hạn các ký hiệu (trong trường hợp đặc biệt, mỗi từ có thể chỉ chứa một ký hiệu).

Bộ từ là tập hợp tất cả các từ mà một bộ ký hiệu có thể tạo ra.

Nguồn rời rạc đặc trưng bởi xác suất $\{X, p(x)\}$, $X = \{ x_1, x_2, ..., x_n\}$

Nguồn rời rạc không nhớ là nguồn rời rạc mà xác suất xuất hiện một ký hiệu không phụ thuộc vào các ký hiệu xuất hiện trước.

$$p(x_n | x_1, x_2 ... x_{n-1}) = p(x_n)$$

Trong đó $x_n \in X$ là một ký hiệu nào đó của bộ ký hiệu X do nguồn tạo ra tại thời điểm n.

Nguồn rời rạc có nhớ là nguồn rời rạc mà xác suất xuất hiện một ký hiệu phụ thuộc vào một hay nhiều ký hiệu đã xuất hiện trước nếu khả năng nhớ của nguồn đủ lớn

$$p(x_n | x_1, x_2 ... x_{n-1}) < p(x_n)$$

Nguồn dừng là nguồn rời rạc mà xác suất xuất hiện các ký hiệu không phụ thuộc vào gốc thời gian mà chỉ phụ thuộc vào vị trí tương quan giữa các ký hiệu, có nghĩa là :

$$p(x_{i_1}) = p(x_{i_1+\tau})$$

Nguồn có tốc độ thông tin điều khiển được là nguồn có thể tạo ra các tin với tốc độ phụ thuộc vào các yếu tố bên ngoài nguồn và không có các ràng buộc nội tại về mặt thời gian trong việc tạo ra các tin. Ví dụ về loại nguồn này là nguồn điện báo. Rõ ràng đối với hệ thống điện báo thì không có các ràng buộc về mặt thời gian đối với các ký hiệu tạo ra được truyền đi.

Nguồn có tốc độ thông tin không điều khiển được là nguồn tạo ra các bản tin với tốc độ cố định, không điều khiển được từ bên ngoài nguồn, tốc độ này là một tính chất nội tại của nguồn. Ví dụ trong trường hợp này là nguồn rời rạc tạo ra khi lấy mẫu một tín hiệu liên tục theo thời gian. Các mẫu được tạo ra liên tiếp nhau, cách nhau một khoảng thời gian cố định phụ thuộc vào các tín hiệu liên tục.

Nguồn Markov giữ vai trò trong trọng trong lĩnh vực truyền thông. Đây là nguồn rời rạc mà xác suất của một ký hiệu chỉ phụ thuộc vào ký hiệu trước đó:

$$p(x_{i_n} | x_{j_{n-1}}, x_{k_{n-2}}, ...) = p(x_{i_n} | x_{j_{n-1}})$$

2. Lượng tin riêng, lượng tin tương hỗ, lượng tin có điều kiện

Lượng tin của mỗi tin $x_i \in X$ là $I(x_i) = - \log p(x_i)$, được gọi là lượng tin riêng của tin x_i

Nếu nguồn X thông qua một phép biến đổi trở thành nguồn Y, ví dụ như thông qua sự truyền lan trong kênh thì phép biến đổi đó có thể không phải là một - một. Trong quá trình truyền lan trong kênh, nhiễu phá hoại làm cho một tin $x_i \in X$ có thể chuyển thành một tin y_j bất kỳ trong nguồn Y ở đầu ra của kênh với những xác suất chuyển đổi khác nhau tuỳ theo tính chất của nhiễu trong kênh.

**** Bài toán thu tin:***

Các tin của nguồn X truyền qua hệ thống biến đổi trở thành đầu ra Y. Cho biết:

- Cấu trúc thống kê của nguồn.

- Cấu trúc thống kê của tạp nhiễu và phép biến đổi (cho bằng xác suất chuyển đổi).

- Với mỗi đầu ra $y_j \in Y$, xác định đầu vào $x_i \in X$ đã sinh ra $y_j \in Y$.

Lời giải:

- Chính xác: Không có

- Xác suất: Xác định đầu vào có khả năng nhất.

- Thông tin: Tách thông tin đầu vào chứa trong đầu ra.

+ Xác định lượng thông tin của mỗi tin x_i chứa trong y_j. Lượng tin này chính là lượng tin tương hỗ = lượng tin ban đầu – lượng tin bị mất đi do nhiễu.

+ Chọn đầu vào là lượng tin tương hỗ lớn nhất.

Xác suất của x_i khi biết y_j là $p(x_i|y_j)$

Lượng tin bị tạp nhiễu phá huỷ không đến được đầu thu chính là lượng tin còn lại của tin x_i sau khi đã nhận được y_j là:

$$I(x_i \mid y_j) = -\log p(x_i \mid y_j) \qquad (2\text{-}16)$$

còn được gọi là lượng tin có điều kiện.

Lượng tin tương hỗ của x_i trong y_j là:

$$I(x_i, y_j) = I(x_i) - I(x_i \mid y_j)$$

$$= \log\frac{p(x_i \mid y_j)}{p(x_i)} = \log\frac{p(y_j \mid x_i)}{\sum_j p(y_j)p(y_j \mid x_i)} \qquad (2\text{-}17)$$

3. Tính chất của lượng tin

Tính chất 1: *Lượng tin riêng của một tin x_i bao giờ cũng lớn hơn lượng tin tương hỗ trong một tin khác y_j.*

Khi x_i và y_j độc lập thống kê với nhau thì lượng tin tương hỗ bằng không.

Nếu từ y_j xác định được x_j thì lượng tin tương hỗ cực đại, lượng tin riêng chính là lượng tin tương hỗ cực đại.

$$I(x_i) = -\log p(x_i) \geq I(x_i, y_j) = \log\frac{p(x_i \mid y_j)}{p(x_i)} = \log\frac{p(y_j \mid x_i)}{p(y_j)}$$

Từ đấy cũng có thể giải thích ý nghĩa lượng tin riêng như là lượng tin tương hỗ cực đại giữa x_i và y_j

Tính chất 2: *Lượng tin riêng là một đại lượng không âm vì $p(x_i) \leq 1$, nhưng lượng tin tương hỗ có thể dương hoặc âm.*

Khi xác suất xuất hiện x_i với điều kiện đã xảy ra y_j lớn hơn xác suất của x_i khi nó xảy ra độc lập với y_j, thì lượng tin tương hỗ dương, ngược lại lượng tin tương hỗ sẽ là một đại lượng âm.

Tính chất 3: *Lượng tin của một cặp (x_iy_j) bằng tổng lượng tin riêng của từng tin trừ đi lượng tin tương hỗ giữa chúng.*

$$I(x_i, y_j) = I(x_i) + I(y_j) - I(x_i, y_j) \qquad (2\text{-}18)$$

Khi chúng độc lập thống kê với nhau thì lượng tin tương hỗ với nhau bằng không, khi đó:

$$I(x_i y_j) = I(x_i) + I(y_j) \qquad (2\text{-}19)$$

4. Lượng tin trung bình

Nguồn tin là một tập hợp các tin. Lượng tin riêng chỉ có ý nghĩa với một tin nào đó nhưng không thể phản ánh được tin tức của nguồn tin. Nói cách khác, $I(x_i)$ mới đánh giá được về mặt tin tức của một tin x_i khi nó đứng riêng rẽ, nhưng không thể dùng để đánh giá về mặt tin tức của một tập hợp trong đó x_i tham gia. Trong thực tế, điều mà ta quan tâm là giá trị tin tức của một tập hợp chứ không phải là của một phần tử nào đó trong tập hợp.

Ví dụ: một nguồn tin chỉ gồm có hai tin $X = \{x_1, x_2\}$ với xác suất xuất hiện là $p(x_1) = 99\%$, $p(x_2) = 1\%$. Khi nhận được một ký hiệu của nguồn, người ta có thể cầm chắc đến 99% là tin x_1 và có thể xem như là một tin đã biết trước. Đứng về quan điểm tin tức mà xét, lượng tin nhận được của nguồn tin này ít có giá trị mặc dù lượng tin riêng của tin x_2 ($= \log 100 \approx 6{,}5$ bit/ký hiệu) khá lớn. Từ đấy, trong thực tế, để đánh giá một tin nhận được của một nguồn đã cho, người ta dùng khái niệm lượng tin trung bình (trị trung bình theo tập hợp):

$$I(X) = -\sum_X p(x) \log p(x) \qquad (2\text{-}20)$$

Lượng tin trung bình là lượng tin tức trung bình chứa trong một ký hiệu bất kỳ của nguồn đã cho.

Đối với ví dụ trên lượng tin trung bình của nguồn là:

$I(X) = - 0,99\log_2 0,99 - 0,01\log_2 0,01 = 0,081$(bit/ký hiệu)

Tương tự lượng tin riêng, lượng tin tương hỗ không mang đầy đủ ý nghĩa thực tế cần thiết, nó chỉ cho biết lượng tin về một ký hiệu đã cho chứa trong một ký hiệu xác định, nghĩa là mới cho biết sự ràng buộc thống kê giữa một cặp (x_i, y_j) nào đó. Nhưng điều quan trọng hơn là cần phải xác định trong thực tiễn mối liên hệ thống kê giữa hai tập X, Y, nghĩa là lượng tin trung bình về một tin bất kỳ của nguồn X chứa trong một tin bất kỳ thuộc nguồn Y hay còn được gọi là lượng tin tương hỗ trung bình:

$$I(X,Y) = \sum_{XY} p(x,y) \log \frac{p(x \mid y)}{p(x)} \qquad (2\text{-}21)$$

Tương tự, *lượng tin riêng trung bình có điều kiện I(X/Y)* là lượng tin trung bình của một tin bất kỳ của X khi đã biết một tin bất kỳ của Y và được xác định như sau:

$$I(X \mid Y) = -\sum_{XY} p(x,y) \log p(x \mid y) \qquad (2\text{-}22)$$

Ta cũng có quan hệ giữa các lượng tin trung bình:

$$I(X,Y) = I(X) - I(X \mid Y) \qquad (2\text{-}23)$$

$$I(X,Y) = I(Y,X) \geq 0 \qquad (2\text{-}24)$$

5. Entropi của nguồn rời rạc

5.1. Khái niệm Entropi

Khi chúng ta nhận được một tin, độ bất ngờ về tin đó được giải thoát (tin đã biết, độ bất ngờ = 0) đồng thời nhận được một lượng tin. Vì vậy, độ bất ngờ bằng lượng tin về số đo nhưng trái ngược nhau về ý nghĩa vật lý.

Độ bất ngờ của tin x_i sẽ là:

$$H(x_i) = -\log p(x_i) = I(x_i) \qquad (2\text{-}25)$$

Độ bất ngờ trung bình của một nguồn tin sẽ là:

$$H(X) = -\sum_{X} p(x) \log p(x) = I(X) \qquad (2\text{-}26)$$

Độ bất ngờ $H(X)$ gọi là entropi của nguồn, được đo bằng lượng tin trung bình của các tin do nguồn phát ra và đây là thông số phản ánh khả năng phát tin (trung bình) của nguồn và là một thông số thống kê cơ bản của nguồn.

5.2. Tính chất của entropi

Tính chất 1: *Entropi là một đại lượng luôn luôn không âm* $H(X) \geq 0$

Tính chất 2: *Entropi sẽ bằng không khi nguồn có một ký hiệu có xác suất xuất hiện bằng một và tất cả các ký hiệu còn lại có xác suất bằng không.*

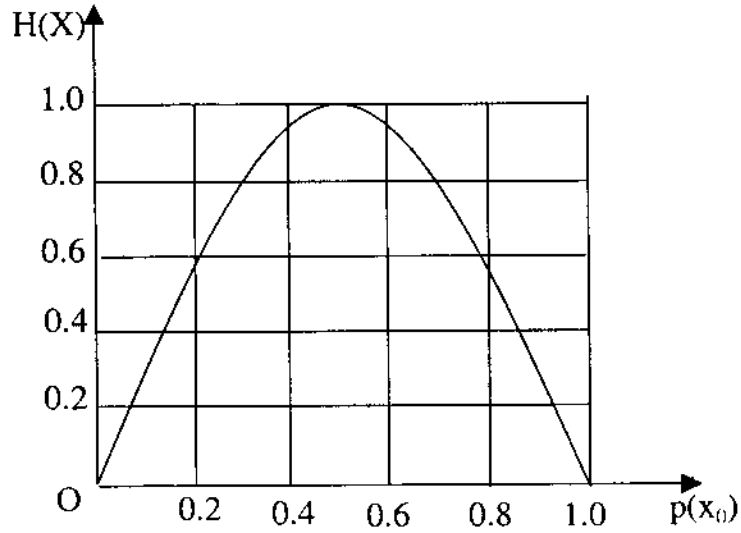
Tính chất 3: *Entropi có giá trị cực đại khi tất cả các ký hiệu có cùng xác suất*

Lấy một ví dụ đơn giản về nguồn có hai tin $X = \{x_0, x_1\}$ với xác suất tương ứng là p_0 và p_1 để minh hoạ điều trên, chúng ta có quy luật phân bố xác suất:

$$p_0 + p_1 = 1 \rightarrow p_1 = 1 - p_0$$

Entropi của nguồn sẽ là:

$$H(X) = - p_0 \log p_0 - p_1 \log p_1 = - p_0 \log p_0 - (1 - p_0) \log(1 - p_0)$$



Hình 2-1: Biểu diễn H(X) theo p_0

Đường biểu diễn cho ta thấy $H(X)_{max}$ khi $p_0 = \frac{1}{2}$ (khi đó $p_1 = 1 - p_0 = \frac{1}{2} = p_0$ và $H(X)_{max} = \log 2$, Nếu dùng cơ số 2 thì $H(X) = 1$ (bit/ký hiệu).

Từ đấy có thể giải thích ý nghĩa của *đơn vị bit là entropi của một nguồn gồm hai ký hiệu đẳng xác suất.*

Tổng quát, nếu nguồn X gồm m ký hiệu, entropi sẽ có giá trị lớn nhất khi các ký hiệu đẳng xác suất:

$$p_1 = p_2 = .. = p_m = \frac{1}{m}$$

lúc đó sẽ có :

$$H(X)_{\max} = -\sum_{i=1}^{m} p_i \log p_i = \log m \qquad (2\text{-}27)$$

Nếu dùng loga cơ số 2, ta sẽ có $H(X)_{max} = \log_2 m$ (bit/ký hiệu)

6. Tốc độ lập tin của nguồn

Thông số thống kê cơ bản thứ nhất của nguồn là entropi, nó tuỳ thuộc vào cấu trúc thống của nguồn. Nhưng tốc độ tạo ra các tin (các ký hiệu) của nguồn nhanh hay chậm còn phụ thuộc vào các tính chất vật lý của nguồn như quán tính, độ phân biệt,... Ví dụ: kết cấu cơ quan phát âm của con người bị hạn chế, mỗi giây chỉ có thể phát được từ 5 đến 7 ký hiệu (âm tiết) trong lời nói thông thường, trong khi đó máy điện báo có thể phát từ 50 đến 70 ký hiệu trong một giây.

Như vậy, thông số thống kê cơ bản thứ hai của nguồn tin là lượng tin mà nguồn tạo ra trong một đơn vị thời gian, còn gọi là tốc độ lập tin của nguồn, ký hiệu là R.

$$R = n_0 \, H(X) \text{ (bit/sec)} \qquad (2\text{-}28)$$

với n_0 là số ký hiệu mà nguồn tạo ra trong một đơn vị thời gian.

- Để có tốc độ lập tin lớn nhất với n_0 (nguồn vật lý) cố định, cần $H(X)_{max}$.

- Để $H(X)_{max}$ thì phải thay đổi cấu trúc thống kê của nguồn bằng các phương pháp ***mã hóa thống kê.***

III. LƯỢNG TIN TRONG KÊNH RỜI RẠC

1. Kênh rời rạc

Nguồn tin và thu tin liên hệ với nhau qua kênh tin. Kênh tin thực hiện một phép biến đổi từ không gian các kí hiệu vào đến không gian kí hiệu ở đầu ra của kênh.

Kênh được gọi là *rời rạc* nếu không gian tín hiệu vào và không gian tín hiệu ra là rời rạc. Kênh được gọi là *liên tục* nếu cả hai không gian ký hiệu vào và ra là liên tục.

Nếu sự truyền tin trong kênh liên tục theo thời gian thì kênh được gọi là *liên tục theo thời gian*. Nếu sự truyền tin chỉ thực hiện ở những thời điểm rời rạc theo thời gian thì kênh được gọi là *rời rạc theo thời gian*.

Nếu sự chuyển đổi ký hiệu vào là x thành ký hiệu ra là y không phụ thuộc vào các chuyển đổi trước đó thì kênh được gọi là *không nhớ*. Nếu sự chuyển đổi đó phụ thuộc vào việc chọn gốc thời gian thì kênh được gọi là *dừng*.

Trong phần này, chúng ta quan tâm chủ yếu đến kênh rời rạc dừng, không nhớ.

2. Entropi đồng thời và Entropi có điều kiện

Trong trường hợp mã hóa hay truyền lan tin trong kênh, ngoài tập tin X của nguồn còn các tập tin Y ở đầu ra của kênh, giữa chúng tồn tại một sự liên hệ thống kê trong một tập tích (XY). Từ đó hình thành một số khái niệm mới là entropi đồng thời và entropi có điều kiện.

- Entropi đồng thời là độ bất định trung bình của tất cả các cặp (x,y)

$$H(XY) = -\sum_{XY} p(x,y) \log\ p(x,y) \qquad (2\text{-}29)$$

- Độ bất định trung bình của một ký hiệu $x_i \in X$ khi biết một ký hiệu $y_j \in Y$ gọi là Entropi có điều kiện.

$$H(X \mid Y) = -\sum_{XY} p(x,y) \log\ p(x \mid y) \qquad (2\text{-}30)$$

Tương tự: $$H(Y \mid X) = -\sum_{XY} p(x,y) \log\ p(y \mid x) \qquad (2\text{-}31)$$

+ $H(X \mid Y)$ còn gọi là 'độ mập mờ' vì nó cho biết sự mập mờ của đầu vào khi đầu ra đã biết.

+ $H(Y \mid X)$ được gọi là sai số trung bình vì nó cho biết độ bất định (sai số) của đầu ra khi đã biết đầu vào.

Tính chất:

$$H(XY) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y) \qquad (2\text{-}32)$$

$$H(Y) \geq H(Y \mid X) \qquad (2\text{-}33)$$

$$H(X) \geq H(X \mid Y) \qquad (2\text{-}34)$$

Nghĩa là độ bất định trung bình của một tập tin bất kỳ bao giờ cũng lớn hơn độ bất định trung bình của tập tin khi đã biết một tập tin bất kỳ khác có liên hệ thống kê với nó. Dấu bằng ở các bất đẳng thức trên sẽ xảy ra trong trường hợp hai tập tin X và Y độc lập thống kê với nhau.

- Sự liên hệ giữa lượng tin tương hỗ trung bình và entropi được biểu diễn như sau:

$$I(X,Y) = H(X) - H(X \mid Y) \qquad (2\text{-}35)$$

$$I(X,Y) = H(Y) - H(Y \mid X) \qquad (2\text{-}36)$$

$$I(X,Y) = H(X) + H(Y) - H(XY) \qquad (2\text{-}37)$$

Các biểu thức này có một ý nghĩa cụ thể khi dùng chúng để mô tả sự truyền tin trong một kênh có nhiễu. Lượng tin trung bình nhận được về tin phát bằng tổng các độ bất định trung bình về tin phát và tin thu được xét một cách độc lập với nhau trừ cho độ bất định trung bình về sự phát đồng thời của chúng. Nếu giữa tin thu và tin phát không có liên quan gì đến nhau nghĩa là tập phát X và tập thu Y độc lập thống kê với nhau, thì lượng tin trung bình nhận được về tin phát sẽ bằng không vì:

$$H(XY) = H(X) + H(Y) \text{ và } I(X,Y) = 0$$

Nếu trên kênh không có nhiễu, giữa đầu vào và đầu ra có một quan hệ một – một, sai số trung bình cũng như độ mập mờ bằng không, khi đó :

$$H(XY) = H(X) = H(Y)$$

Từ đấy có thể giải thích lượng tin tương hỗ như một số đo mỗi liên hệ thống kê giữa X và Y.

3. Thông lượng kênh rời rạc

3.1. Khái niệm thông lượng của kênh

- Thông lượng của kênh là lượng tin tối đa mà kênh cho đi qua trong một đơn vị thời gian mà không gây sai nhầm. Ký hiệu là C. Đơn vị bit/sec.

- Là tốc độ lập tin tối đa ở đầu ra của kênh.

- Tốc độ lập tin của nguồn thường nhỏ hơn nhiều so với thông lượng của kênh.

$$R << C \qquad (2\text{-}38)$$

- Tận dụng thông lượng của kênh như sau:

+ Tối đa tốc độ lập tin của nguồn cho phù hợp với kênh: mã hóa thống kê để có tốc độ lập tin cực đại, gần với thông lượng của kênh (đồng bộ kênh - nguồn) - cơ sở lý thuyết là định luật Shannon cho kênh không nhiễu.

+ Sử dụng phần còn lại của thông lượng kênh để chống nhiễu (mã chống nhiễu) - cơ sở lý thuyết là định luật Shannon cho kênh có nhiễu.

3.2. Thông lượng của kênh rời rạc không nhiễu

Khi kênh không có nhiễu thì thông tin do nguồn thiết lập truyền sẽ không có sai nhầm, do đó thông lượng kênh khi đó bằng tốc độ lập tin cực đại của nguồn:

$$C = R_{max} = n_0 \, H(X)_{max} \qquad (2\text{-}39)$$

Để tối ưu hệ thống cần cực đại entropi của nguồn. Hai câu hỏi đặt ra là liệu có tồn tại một phương pháp mã hóa làm tăng entropi của nguồn đạt đến mức cực đại? và giới hạn của tốc độ truyền tin khi đó là bao nhiêu?

Điều này được Shannon phát biểu trong một định lý cơ bản của lý thuyết tin tức với nội dung như sau:

Giả sử nguồn có entropi H(bit/ký hiệu) và kênh có thông lượng C(bit/sec), có thể mã hoá tin tức ở đầu ra của nguồn làm cho sự truyền tin trong kênh không nhiễu theo một tốc độ trung bình $\frac{C}{H} - \varepsilon$ *(ký hiệu/sec) với* ε *là lượng bé tùy ý và không thể truyền nhanh hơn* $\frac{C}{H}$ *(ký hiệu /sec).*

Tốc độ lập tin tối đa tiệm cận và có thể bằng thông lượng của kênh. Phép mã hóa tương ứng gọi là phép mã hóa thống kê tối ưu. Có thể nói, định luật trên chính là cơ sở lý thuyết của phương pháp mã hóa thống kê tối ưu. Phép mã hóa thống kê tối ưu không sử dụng hết thông lượng của kênh.

Khi tốc độ lập tin của nguồn chưa đạt cực đại, còn khả năng để tối ưu nguồn thì khả năng này được đo bằng độ dư của nguồn.

Độ dư của nguồn được định nghĩa như sau:

$$R_s = H(X)_{max} - H(X) \qquad (2\text{-}40)$$

Độ dư tương đối của nguồn được định nghĩa như sau:

$$r_s = \frac{H(X)_{max} - H(X)}{H(X)_{max}} = 1 - \frac{H(X)}{H(X)_{max}} \qquad (2\text{-}41)$$

Để có thể xây dựng mã chống nhiễu, điều kiện đầu tiên là phải có độ dư.

3.3. Thông lượng của kênh rời rạc có nhiễu

Xét kênh không nhớ, nghĩa là những tin nhận được sau không phụ thuộc vào những tin nhận được trước, nói cách khác chúng độc lập thống kê với nhau. Độ chính xác của tin truyền đi trong kênh chỉ còn bị ảnh hưởng của nhiễu làm giảm đi mà thôi. Ở đây có thể xảy ra hai trường hợp:

- Nếu các tin nhận được sai lệch với các tin gửi đi nhưng vẫn phân biệt được các tin đầu vào thì sự truyền tin như vậy vẫn đảm bảo chính xác.

- Nếu các tin đầu vào bị lẫn nhau ở đầu ra (nhiều tin đầu vào cho một tin đầu ra), trường hợp này làm giảm độ chính xác truyền tin và xuất hiện sai số truyền tin. Chúng ta sẽ khảo sát trường hợp này.

Xét hệ thống gồm đầu vào $X = \{x_i\}$, $i = 1....m$ và đầu ra của kênh nhận được các tin tương ứng hợp lại thành tập $Y = \{y_j\}$, $j = 1...n$. Do kênh có nhiễu, nên phép biến đổi giữa X và Y được biểu diễn bằng ma trận xác suất chuyển đổi $p(y_j \mid x_i)$.

Trong trường hợp này lượng tin tương hỗ trung bình giữa X và Y được xác định bằng:

$$I(X,Y) = H(X) - H(X \mid Y)$$
$$I(X,Y) = H(Y) - H(Y \mid X)$$

Tốc độ lập tin ở đầu ra của kênh trong trường hợp kênh có nhiễu bằng:

$$R = n_0 I(X,Y) = n_0 (H(X) - H(X \mid Y)) \text{ (bit/sec)} \qquad (2\text{-}42)$$

n_0 I(X|Y) là lượng tin bị nhiễu phá hủy trong một đơn vị thời gian.

Khi các thông số của kênh đã được xác định, muốn nâng cao tốc độ lập tin ở đầu ra của kênh nhất thiết phải tăng entropi bằng phương pháp mã hóa. Thông lượng kênh lúc này chính là tốc độ lập tin tối đa ở đầu ra của kênh

$$C = R_{max} = n_0 I(X,Y) = n_0 (H(X) - H(X \mid Y)) \qquad (2\text{-}43)$$

Nếu xem giải thông của kênh $\Delta f = n_0$, thì thông lượng của kênh có nhiễu là:

$$C = \Delta f(H(X) - H(X \mid Y))_{\max} \qquad (2\text{-}44)$$

Vấn đề đặt ra cho sự truyền tin trong kênh có nhiễu là bằng cách nào có thể truyền tin chính xác và mức độ chính xác là bao nhiêu?. Trả lời cho vấn đề này, Shannon đã phát biểu trong một định lý cơ bản thứ hai của lý thuyết tin tức. Nội dung định lý như sau:

Kênh rời rạc có thông lượng C(bps), tốc độ lập tin của nguồn R.

- Nếu R < C: Phần dư của nguồn được dùng để bổ sung các thông tin chống nhiễu. Cần truyền lượng tin lớn hơn so với thông tin cần truyền.

- Nếu R > C: Phần thông tin không được truyền đi sẽ trở thành sai số (tối thiểu). Tồn tại cách mã hóa để có sai số bé hơn R-C+ ε (ε là lượng bé tùy ý). Không tồn tại mã hiệu đảm bảo độ sai nhầm của sự truyền tin bé hơn R-C.

IV. ENTROPI CỦA NGUỒN VÀ THÔNG LƯỢNG CỦA KÊNH LIÊN TỤC

1. Nguồn liên tục

- Nguồn liên tục là một quá trình ngẫu nhiên liên tục. Để có thể nghiên cứu, xem xét nguồn liên tục cần rời rạc hóa nguồn liên tục.

- Với điều kiện nguồn có phổ hữu hạn, có thời gian tồn tại hữu hạn, nguồn liên tục có thể được lấy mẫu với tần số $2\Delta f$ tại các thời điểm $\{t_i\}$, i = 1..n.

- Mỗi thể hiện của nguồn liên tục là một hàm x(t) theo thời gian, được đặc trưng bởi n giá trị tức thời $\{x_i\}$, i = 1..n.

- Tính chất thống kê của nguồn được đặc trưng bởi phân bố xác suất đồng thời (nhiều chiều): $p(x_1, x_2, ..., x_n)$.

- Trong thực tiễn truyền tin, thường gặp các quá trình ngẫu nhiên dừng, tính chất thống kê của các quá trình này không phụ thuộc vào gốc thời gian.

$$p(x_{t_i}) = p(x_{t_i+\tau})$$

2. Entropi nguồn liên tục

Entropi là đại lượng đo độ bất định trung bình của một giá trị bất kỳ mà mẫu $x(t_i)$ có thể lấy được. Về số đo, entropi cũng là lượng tin trung trình của một trị bất kỳ thuộc $x(t_i)$.

Entropi của nguồn liên tục dừng (các mẫu $x(t_i)$ độc lập thống kê với nhau và cùng một quy luật phân bố xác suất $p(x)$)

$$H(X) = -\int p(x)\log p(x)dx \qquad (2\text{-}45)$$

Entropi của nguồn liên tục không dừng:

$$H(X) = -\int_{-\infty}^{+\infty}.....\int_{-\infty}^{+\infty} p(x_{t_1}, x_{t_2}, ...x_{t_n})\log p(x_{t_1}, x_{t_2}, ...x_{t_n})dx_1 dx_2 ...dx_n \qquad (2\text{-}46)$$

Tốc độ lập tin của nguồn liên tục sẽ là :

$$R = n_0\ H(X) = 2\Delta f\ H(X) \text{ (bit/sec)} \qquad (2\text{-}47)$$

Khi sự truyền tin xảy ra trong kênh có nhiễu, ứng với một thể hiện x(t) ở đầu vào kênh có một thể hiện y(t) ở đầu ra đã bị sai lệch ít nhiều. Tại một thời điểm bất kỳ một trị nào đó $x(t_i)$ qua kênh sẽ chuyển đổi thành một trị nào đó $y(t_i)$ theo quy luật phân bố có điều kiện p(x|y).

Quy luật phân bố p(x|y) cho phép xác định độ bất định trung bình về một trị x nào đó khi nhận được một trị y, nghĩa là xác định được entropi có điều kiện của nguồn liên tục.

$$H(X \mid Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(x \mid y)dxdy \qquad (2\text{-}48)$$

$$H(Y \mid X) = -\int_{-\infty}^{+\infty+\infty}\int p(x,y)\log p(y \mid x)dxdy \qquad (2\text{-}49)$$

p(x,y) là quy luật phân bố đồng thời của x và y. Quy luật này dùng để tính entropi đồng thời:

$$H(XY) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(x,y)dxdy \qquad (2\text{-}50)$$

Lượng tin tương hỗ trung bình giữa một tin bất kỳ của tập nhận Y và một tin bất kỳ của tập gửi X được xác định theo biểu thức:

$$I(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{p(x \mid y)}{p(x)}dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{p(y \mid x)}{p(y)}dxdy$$

Quan hệ giữa các entropi với lượng tin tương hỗ như sau:

$$H(XY) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y) \quad (2\text{-}51)$$

$$I(X,Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) \quad (2\text{-}52)$$

3. Thông lượng của kênh liên tục

Giả thiết tín hiệu đầu vào của kênh là x(t), tín hiệu ở đầu ra của kênh là y(t) tương ứng là hỗn hợp của x(t) và nhiễu cộng n(t):

$$y(t) = x(t) + n(t)$$

Giả thiết $y(t) \in Y$, $x(t) \in X$ và $n(t) \in N$ với X là tập nguồn, Y là tập nhận, N là tập nhiễu và H(X), H(Y), H(N) là entropi các tập tương ứng. Giữa nguồn tin X và nguồn nhiễu N không có liên hệ thống kê nào, ta có thể viết entropi đồng thời của X và Y như sau:

$$H(XY) = H(X, X+N) = H(X,N) = H(X) + H(N) \quad (2\text{-}53)$$

Mặt khác entropi đồng thời có thể viết lại như sau:

$$H(XY) = H(X) + H(Y \mid X)$$

So sánh hai công thức cho thấy:

$$H(Y \mid X) = H(N) \quad (2\text{-}54)$$

Vậy $H(Y \mid X)$ là entropi của nguồn nhiễu.

Tốc độ lập tin ở đầu ra của kênh được xác định theo:

$$R = n_0(H(Y) - H(Y \mid X)) = 2\Delta f(H(Y) - H(N)) \quad (2\text{-}55)$$

Thông lượng của kênh bằng tốc độ lập tin cực đại ở đầu ra:

$$C = R_{max}$$

Tốc độ lập tin cực đại khi H(Y) cực đại. Các tin của Y gồm hai thành phần từ X và N. Phụ thuộc vào các hạn chế của X,N có thể xác định qui luật phân bố của X để thông lượng kênh đạt cực đại.

Bài tập

1. Nguồn tin X là một đèn LED 7 thanh. Hãy xác định H(X).

2. Tính entropi trong trường hợp tung đồng xu làm nguồn tin.

3. Nguồn tin X có 4 lớp tin được ký hiệu là x_1, x_2, x_3, x_4 có xác suất xuất hiện tương ứng là 0,5; 0,25; 0,125; 0,125. Hãy tính lượng tin riêng tương ứng với từng lớp tin và entropi của nguồn.

4. Hai biến ngẫu nhiên X và Y nhị phân có phân bố đồng thời

p(X=Y=0) = p(X=0,Y=1) = P(X=Y=1) = 1/3

Tính H(X), H(Y), H(X|Y), H(Y|X) và H(XY)

5. Gọi X là một biến ngẫu nhiên với hàm mật độ phân bố xác suất $p_X(x)$ và đặt Y = aX+b là một biến đổi tuyến tính của X, a và b là hai hằng số. Tính H(Y) theo H(X).

6. Xác định entropi của biến ngẫu nhiên X có phân bố đồng đều

$$p(x)=\begin{cases} a\text{-}1 & \text{với } 0\le x\le a \\ 0 & \text{với các giá trị khác} \end{cases}$$

Trong các trường hợp:

a/ a =1 c/ a= 1/4

b/ a = 4 d/ a = 1/8

7. Đặt Y = g(X), với g là hàm xác định. Chứng tỏ rằng H(Y) ≤ H(X). Khi nào đẳng thức xảy ra.

8. Tìm entropi của các biến ngẫu nhiên liên tục sau đây:

X là biến ngẫu nhiên mũ với tham số λ > 0

$$f_x(x)=\begin{cases} \lambda^{-1} e^{-x/\lambda} & \text{với } x > 0 \\ 0 & \text{với } x \le 0 \end{cases}$$

Chương 3

MÃ HIỆU

Mục tiêu

- Phân biệt được các khái niệm mã hóa, mã hiệu, giải thích được các thông số cơ bản của mã hiệu.
- Nêu được điều kiện thiết lập mã.
- Biết được các phương pháp biểu diễn mã.
- Vận dụng các kiến thức xác định các thông số cơ bản của mã hiệu và biểu diễn được bộ mã.

I. MÃ HIỆU VÀ THÔNG SỐ CƠ BẢN CỦA MÃ HIỆU

1. Khái niệm mã hiệu

Trong các hệ thống truyền tin, các nguồn tin nguyên thủy thường phải qua các phép biến đổi để trở thành các nguồn tin trung gian phù hợp với các quá trình xử lý, với môi trường truyền... nhằm nâng cao hiệu suất truyền tin hoặc nâng cao tính chống nhiễu của hệ thống. Tại đầu thu tín hiệu, thông qua những phép biến đổi ngược lại, biến đổi nguồn tin trung gian trở thành nguồn tin có dạng ban đầu. Quá trình mã hóa biến đổi nguồn tin thành các nguồn tin trung gian và mã hiệu dùng để biểu diễn các nguồn tin trung gian này.

Mã hiệu được định nghĩa là một tập hữu hạn các ký hiệu (thường là các chữ số) và phép ánh xạ các tin/bản tin của nguồn tin thành các dãy ký hiệu tương ứng. Tập các ký hiệu và phép ánh xạ này phải đáp ứng các yêu cầu của từng hệ thống truyền tin đặt ra.

Mã hóa (encoding) là phép biến đổi từ nguồn tin thành mã hiệu hay mã hóa là phép biến đổi nguồn tin này thành nguồn tin khác có tính thống kê (entropi, chiều dài các tin, độ chính xác) theo yêu cầu.

Sự mã hóa thông tin cho phép chúng ta ký hiệu hóa thông tin hay sử dụng các ký hiệu quy ước để biểu diễn bản tin ở dạng phù hợp.

Chính nhờ mã hoá, chúng ta có thể nhìn thấy hay hiển thị được thông tin có bản chất là các khái niệm (sự hiểu biết của con người). Đối với một hệ thống truyền tin, việc mã hóa cho phép tăng tính hữu hiệu và độ tin cậy của hệ thống truyền tin, nghĩa là tăng tốc độ truyền tin và khả năng chống nhiễu của hệ thống.

**** Các khái niệm liên quan đến mã hiệu:***

- Mã hiệu gồm một tập hữu hạn các ký hiệu có phân bố xác suất nào đó gọi là *dấu mã* hay *ký hiệu mã*.

- Tập hợp một số dấu mã gọi là *tổ hợp mã*.

- Trong tập hợp tất cả các tổ hợp mã, một tập hợp các tổ hợp mã được xây dựng theo một quy luật nào đó gọi là tổ hợp mã hợp lệ

- Trong quá trình mã hóa, một tin nguyên thủy được ánh xạ vào một tổ hợp mã. Một tổ hợp mã như vậy gọi là *từ mã*. Những tổ hợp mã khác gọi là tổ hợp cấm.

- Nguồn tin rời rạc gồm nhiều tin tạo thành các bản tin. Các nguồn tin trong thực tế có số lượng các tin rất lớn. Ngược lại các mã hiệu thường có số lượng các ký hiệu tương đối nhỏ. Do đó một tin của nguồn ban đầu thường được mã hóa thành một chuỗi các ký hiệu mã (từ mã).

- Một dãy từ mã bất kỳ tạo thành một *từ thông tin*.

- Quá trình ngược lại của quá trình mã hóa được gọi là *giải mã*.

2. Các thông số cơ bản của mã hiệu

2.1. Cơ số mã

Tập các ký hiệu mã dùng để biểu diễn các tin được gọi là bảng ký hiệu mã, còn số các ký hiệu mã khác nhau trong bảng mã đó được gọi là ***cơ số mã***. Cơ số mã thường được ký hiệu là m.

Ví dụ, điện báo manip tần số FSK hoặc manip pha FPK chỉ dùng hai tần số hoặc hai góc pha ngược nhau 180^0, cơ số của mã bằng hai (m = 2). Điện báo Morse dùng ba loại ký hiệu: gạch, chấm, nghỉ có độ dài khác nhau thì cơ số của mã bằng ba (m = 3). Khi mã có cơ số hai gọi là mã nhị phân, đây là loại

mã được dùng rộng rãi nhất. Mã có cơ số bằng ba, bốn được gọi là mã tam phân, tứ phân,...Trong thực tế máy tính và các thiết bị điện tử thường dùng mã nhị phân, nhưng mã nhị phân thường rất dài nên trong máy tính còn dùng hệ mã Hecxa (mã có cơ số bằng 16).

2.2. Độ dài từ mã

Tổ hợp các ký hiệu mã dùng để mã hoá một tin được gọi là độ dài từ mã. Độ dài từ mã thường được ký hiệu là n. Có thể nói mã hoá là một phép biến đổi một - một giữa một tin của nguồn với một từ mã của bộ mã. Trong một số trường hợp nhất định, người ta không thay thế mỗi tin của nguồn bằng một từ mã mà một khối tin của nguồn mới được mã hóa bằng một từ mã thì từ mã sẽ là một tổ hợp mã dùng để mã hóa một khối tin của nguồn. Lúc này chúng ta có khái niệm mã khối.

2.3. Độ dài trung bình từ mã

Nếu một bộ mã mà các từ mã có độ dài như nhau gọi là mã đồng đều, ngược lại gọi là mã không đồng đều. Đối với mã không đồng đều, chúng ta còn có thêm một tham số cơ bản nữa đó là *độ dài trung bình của từ mã*.

$$\bar{n} = \sum_{i=1}^{L} n_i p(x_i) \qquad (3\text{-}1)$$

trong đó:

$p(x_i)$ là xác suất xuất hiện tin x_i của nguồn X sẽ được mã hóa

n_i là độ dài từ mã tương ứng với tin x_i

L là tổng số từ mã ứng với tổng số tin của X.

2.4. Tổng số từ mã

Tất cả các từ mã để mã hóa các tin của nguồn làm thành bộ mã hóa nguồn. *Tổng số từ mã* bằng số tin được mã hóa. Trong trường hợp mã khối thì tổng số từ mã bằng tổng số khối tin sẽ được mã hóa. Trong trường hợp mã đều, nếu tất cả các tổ hợp mã gồm n ký hiệu với m trị khác nhau đều được dùng làm từ mã, chúng ta sẽ có:

$$L = m^n \qquad (3\text{-}2)$$

Khi này bộ mã được gọi là mã đầy

Nếu $L < m^n$, bộ mã được gọi là mã vơi.

Căn cứ sự phát triển của lý thuyết và kỹ thuật mã hiện nay, có thể có hai xu hướng nghiên cứu về mã:

Một là, nghiên cứu cấu trúc các loại mã đồng đều, chủ yếu là loại mã với có tính chất phát hiện và sửa sai.

Hai là, nghiên cứu các loại mã không đồng đều để xây dựng những loại mã có độ dài trung bình ngắn nhất, gọi là mã thống kê tối ưu hoặc là mã kinh tế (về quan điểm thời gian truyền tin). Hiện nay cũng có những công trình nghiên cứu phương pháp xây dựng các loại mã không đồng đều có tính chất chống nhiễu.

2.5. Độ đo từ mã (trọng số từ mã)

Để thuận tiện cho việc sử dụng mã hiệu, mỗi từ mã được gán cho một **độ đo** còn gọi là trọng số hoặc giá trị từ mã.

$$b = \sum_{k=0}^{n-1} a_k W_k \qquad (3\text{-}3)$$

- Chỉ số k: số thứ tự của mỗi ký hiệu mã trong từ mã, được đánh số 0 từ phải qua trái.

- Trọng số W_k: hệ số nhân của từng vị trí ký hiệu ($W_k = m^k$).

- Mỗi ký hiệu mã được gán cho một giá trị gọi là *giá trị riêng* a_k.

Ví dụ: trọng số của từ mã 1011 là: $b = 1\times 2^0 + 0\times 2^1 + 1\times 2^2 + 1\times 2^3 = 13$.

II. ĐIỀU KIỆN PHÂN TÁCH CỦA MÃ HIỆU

1. Điều kiện để mã phân tách được

Các tin tức mã hoá khi truyền đi được trình bày dưới dạng một dãy ký hiệu liên tiếp nhau. Muốn giải mã, hay muốn khôi phục các tin từ dãy ký hiệu mã nhận được, điều kiện trước tiên là trong dãy ký hiệu liên tiếp đó phải có quy luật đảm bảo sự phân tách một cách duy nhất các từ mã. Nói một cách khác chúng phải có quy luật để phân biệt được các ký hiệu đầu của mỗi từ mã. Nếu điều kiện này không được thỏa mãn thì từ một chuỗi ký hiệu mã nhận được, ta có thể có nhiều cách để tách nó ra thành các từ mã hoặc không thể tách nó được. Điều này làm cho bản tin nhận được sau giải mã có thể không giống với bản tin được phát. Chúng ta có thể tìm hiểu kỹ hơn qua hai ví dụ sau đây:

Ví dụ 1: Xét bộ mã $X_1 = \{0, 10, 11\}$ mã hóa cho nguồn A = {a, b, c} theo quy luật:

$$a \rightarrow 0$$
$$b \rightarrow 10$$
$$c \rightarrow 11$$

Giả sử bên phát phát đi bản tin x = abaac, lúc đó chuỗi tương ứng từ mã được phát đi là y = 0100011

Vấn đề là bên nhận sau khi nhận được chuỗi từ mã y làm sao có thể nhận biết được bản tin tương ứng mà bên phát đã phát đi.

Cần chú ý rằng giữa từ mã và tin được mã hóa có quan hệ một – một thì việc giải mã ở phía nhận tin sẽ bao gồm việc tách đúng từ mã nhận được (tách mã) và chuyển ngược từ mã thành tin tương ứng. Việc chuyển từ mã thành tin sẽ được thực hiện dễ dàng nhờ một sơ đồ giải mã xác định. Việc tách mã là một thuật toán kiểm tra tính đúng đắn của một số tiêu chuẩn được gọi là điều kiện phân tách của mã hiệu. Việc kiểm tra này sẽ bắt đầu từ ký hiệu mã đầu tiên của chuỗi cho đến khi có thể cắt được một từ mã thì nó sẽ cắt từ mã và lại coi ký hiệu tiếp theo là ký hiệu đầu tiên của chuỗi và kiểm tra tiếp.

Chẳng hạn, với chuỗi ký hiệu mã nhận được như trên thì bên nhận chỉ có một khả năng để tách mã hợp lý là:

$$0 \mid 10 \mid 0 \mid 0 \mid 11$$

và xác định được bản tin đã được gửi đi là abaac.

Ví dụ 2: Xét bộ mã $X_2 = \{0, 10, 01\}$ mã hóa cho nguồn A = {a, b, c} theo quy luật :

$$a \rightarrow 0$$
$$b \rightarrow 10$$
$$c \rightarrow 01$$

Giả sử bên nhận nhận được chuỗi ký hiệu y = 01010 và thực hiện quá trình tách mã. Ở đây chúng ta thấy bên nhận có thể thực hiện được ba khả năng tách mã hợp lý sau:

$$0 \mid 10 \mid 10$$
$$01 \mid 0 \mid 10$$
$$01 \mid 01 \mid 0$$

Vì vậy bên nhận sẽ không biết được chính xác bên phát đã phát đi bản tin nào trong 3 bản tin sau: abb hay cab hay ca. Một bộ mã như vậy không phù hợp cho việc tách mã và được gọi là mã không phân tách được (uniquely undecodable code).

Vì vậy điều kiện để một bộ mã là phân tách được (uniquely decodable code) là bất kỳ dãy các từ mã nào của bộ mã cũng không được trùng với một dãy các từ mã khác của cùng bộ mã.

Ví dụ 3: Xét bộ mã X_3 = {010, 0101, 10100} mã hóa cho nguồn A = {a, b, c} theo quy luật :

$$a \rightarrow 010$$

$$b \rightarrow 0101$$

$$c \rightarrow 10100$$

Giả sử bên nhận nhận được một chuỗi ký hiệu là 01010100101 và thực hiện quá trình tách mã. Ở đây ta thấy chỉ có một cách tách mã duy nhất là:

$$0101 \mid 010 \mid 0101$$

nhưng việc tách mã trở nên khó khăn hơn so với bộ mã X_1. Chẳng hạn, lúc gặp chuỗi 010 thì chưa chắc chắn đó là một từ mã vì có thể là phần đầu của từ mã 0101, điều này phụ thuộc vào ký hiệu đi ngay sau chuỗi 010. Nếu ký hiệu đi ngay sau chuỗi này là 0 thì có thể khẳng định được 010 là từ mã và 0 là phần đi đầu của một từ mã sau đó. Còn nếu ký hiệu ngay sau là 1 thì không thể khẳng định vì có hai khả năng: hoặc 010 là một từ mã và 1 là phần đi đầu của từ mã khác hoặc 0101 là một từ mã.

Nguyên nhân của điều này là do trong bộ mã có một từ mã này là *tiếp đầu ngữ* (phần đầu - prefix) của một từ mã khác. Và đó cũng chính là nguyên nhân và bản chất của việc một dãy ký hiệu có thể tách thành hai dãy từ mã khác nhau.

Mã có tính prefix là bộ mã không có bất kỳ từ mã nào là prefix của một từ mã khác trong cùng bộ mã. Mã prefix chỉ có duy nhất một cách tách thành các từ mã thành phần nên đây là bộ mã phân tách được.

2. Bảng thử mã và độ chậm giải mã

Dựa vào tính *tiếp đầu ngữ* trên, để nhận biết một bộ mã (dĩ nhiên không phải mã prefix) có phân tách được hay không, người ta thường dùng một công

cụ gọi là *bảng thử mã*. Bản chất của bảng thử mã là phân tích những từ mã dài thành những từ mã ngắn dần.

Cách thực hiện bảng thử mã như sau :

- Đem các từ mã xếp thành một cột, theo thứ tự chiều dài của từ mã từ nhỏ đến lớn, đánh dấu cột là cột 1.

- Trong cột này, đối chiếu các từ mã ngắn với từ mã dài hơn, nếu từ mã ngắn là tiếp đầu ngữ của từ mã dài thì ghi tiếp vĩ ngữ (phần còn lại) vào cột tiếp theo và đánh dấu là cột 2.

- Tiếp tục đối chiếu các tổ hợp mã trong cột 1 và cột 2 với nhau, nếu chuỗi nào trong cột này là tiếp đầu ngữ của chuỗi trong cột kia, tiếp vĩ ngữ sẽ được ghi vào cột 3.

- Tiếp tục theo khuôn mẫu này nếu đang xét ở cột thứ j thì đối chiếu các chuỗi trong cột này với cột 1. Nếu có chuỗi nào trong cột này là tiếp đầu ngữ trong cột kia thì tiếp vĩ ngữ sẽ được ghi vào cột j +1. Thực hiện cho đến khi không thể điền thêm được nữa hoặc cột mới thêm vào trùng với một cột trước đó hoặc có một chuỗi trong cột mới trùng với một từ mã.

Ta có định lý sau:

Điều kiện cần và đủ để mã có tính phân tách là không có một tổ hợp mã nào trong các cột $j \geq 2$, trùng với một từ mã trong cột 1.

Thực vậy, vì từ cột 2 là những tiếp vĩ ngữ (chuỗi cuối) của từ mã dài mà tiếp đầu ngữ của nó đã trùng với từ mã khác hoặc phần cuối của từ mã khác. Nếu một chuỗi cuối trùng với một từ mã khác thì chắc chắn cơ lớn hơn một cách phân tách chuỗi mã nhận được thành các từ mã. Hay nói cách khác nếu định lý trên không thoả mãn thì sẽ tồn tại một cách ghép các từ mã thành chuỗi mã, mà trong nó tạo ra một đoạn ký hiệu mã trùng với các từ mã được ghép thành chuỗi.

Độ chậm giải mã là số ký hiệu cần phải nhận được đủ để có thể nhận dạng được từ mã. Đối với mã có thể phân tách, độ chậm giải mã thường là hữu hạn, nhưng cũng có trường hợp độ chậm giải mã là vô hạn.

Ví dụ 1: Lập bảng thử mã phân tách với bộ mã 00,01,100,1010,1011

Cột 1	Cột 2
00	
01	
100	
1010	
1011	

Bảng thử cho thấy cột 2 trống rỗng, nghĩa là bộ mã phân tách được. Rõ ràng trong trường hợp này không có từ mã nào trùng với phần đầu của từ mã khác nên khi nhận đủ số ký hiệu của một từ mã thì không có từ mã nào khác nhận các ký hiệu mã này làm phần đầu nữa nên đây chính là thời điểm cắt mã. Độ chậm giải mã của mã này bằng độ dài từ mã.

Ví dụ 2 :

Xét tính phân tách của bộ mã 10,100,01,011.

Bảng thử mã phân tách của bộ mã này như sau :

Cột 1	Cột 2	Cột 3	Cột 4	Cột 5	.....
10	0	1	0	1	
100	1	11	00	11	
01		0	1	0	
011		00	11	00	

Trong các cột từ cột 2, 3,..của bảng thử này không có tổ hợp mã nào trùng với từ mã trong cột 1, nhưng có thể điền các cột j đến vô hạn mà không gặp cột trống. Như vậy độ chậm giải mã là vô hạn, tuy vậy mã vẫn có thể phân tách được.

Bảng thử mã phân tách cho phép đánh giá độ chậm giải mã. Nếu j là số thứ cột rỗng, thì độ chậm giải mã T_{ch} được tính theo :

$$\left[\frac{j-1}{2}\right]n_{\min} \le T_{ch} \le \left[\frac{j}{2}\right]n_{\max} \qquad (3\text{-}4)$$

$n_{\min}$ và $n_{\max}$ là độ dài từ mã ngắn nhất và dài nhất, [x] là ký hiệu chỉ phần nguyên của x.

3. Bất đẳng thức Kraft

Điều kiện tồn tại một mã thỏa mãn tính prefix được xác định bằng bất đẳng thức **Kraft** như sau :

Nếu một dãy số nguyên $n_1, n_2,..n_L$ thoả mãn điều kiện:

$$\sum_{k=1}^{L} 2^{-n_k} \leq 1 \qquad (3\text{-}4)$$

thì sẽ tồn tại một mã có tính prefix, cơ số m, dãy số nguyên trên là độ dài các từ mã của bộ mã.

Độ chậm giải mã của mã prefix bằng độ dài của từ mã dài nhất.

III. PHƯƠNG PHÁP BIỂU DIỄN MÃ

Phương pháp biểu diễn mã là cách trình bày toàn bộ mã hiệu với các tin được mã hóa. Các dạng biểu diễn mã thường được dùng là các bảng mã, đồ hình kết cấu và các hàm cấu trúc mã.

1. Các bảng mã

Các bảng mã thường được dùng rộng rãi để mô tả một mã hiệu (bộ mã) là bảng đối chiếu và mặt tọa độ mã.

1.1. Bảng đối chiếu mã

Là cách trình bày đơn giản nhất: liệt kê trong một bảng các tin của nguồn và chỉ rõ những từ mã tương ứng với nó. Bảng đối chiếu mã có ưu điểm là cho thấy cụ thể và tức thời bản tin và từ mã của nó.

Ví dụ: Nguồn tin X = { $x_1, x_2, x_3, x_4, x_5,$} các tin của nó được mã hoá như sau:

Tin	x_1	x_2	x_3	x_4	x_5
Từ mã	00	01	100	1010	1011

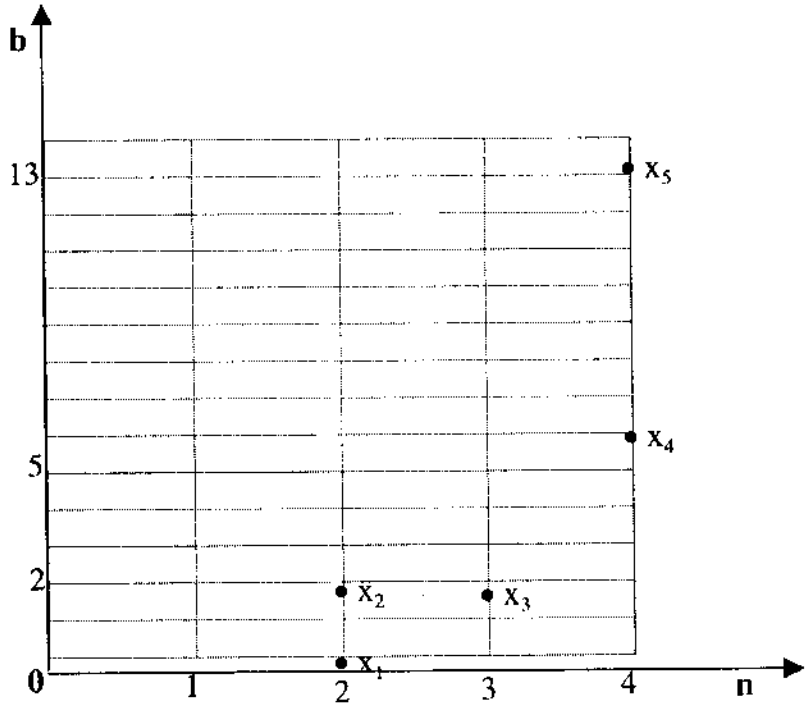
Trong bảng đối chiếu mã, số thứ tự của từ mã trong bộ mã là số thứ tự người ta liệt kê từ mã.

Cách biểu diễn này tuy rõ ràng nhưng không thích hợp với các bộ mã lớn và cồng kềnh.

1.2. Mặt toạ độ mã

Mặt toạ độ mã là một cách biểu diễn mã dựa vào hai thông số: độ dài từ mã n và trọng số b của từ mã để lập một mặt phẳng toạ độ, trên đó mỗi từ mã được biểu diễn bằng một điểm .

Bộ mã trình bày trong bảng đối chiếu mã ở trên được biểu diễn trong mặt tọa độ mã ở hình 3-1 như sau:



Hình 3-1: Biểu diễn mã bằng mặt toạ độ mã

Mỗi từ mã sẽ hoàn toàn xác định khi ta xác định được cặp (n,b) của nó. Như vậy mỗi từ mã được biểu diễn bởi một cặp tọa độ (n,b) duy nhất .

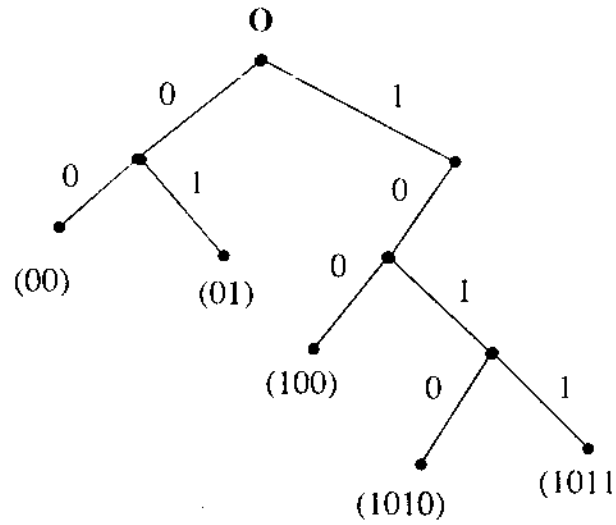
2. Đồ hình mã

Các phương pháp đồ hình cho phép trình bày bộ mã một cách gọn hơn các bảng mã, đồng thời cho thấy rõ hơn các tính chất của mã hiệu như tính phân tách được hay tính prefix... các phương pháp đồ hình thường dùng gồm có: cây mã và đồ hình kết cấu.

2.1. Cây mã

Cây mã là một đồ hình gồm các nút và các nhánh. Gốc của cây gọi là nút gốc (mức 0). Từ nút gốc phân đi m nhánh hoặc ít hơn, mỗi nhánh đại diện cho một trị của ký hiệu. Mỗi nhánh kết thúc tại một nút ở cấp cao hơn nút xuất phát. Tiếp tục từ mỗi nút ở cấp i sẽ phân ra m nhánh hoặc ít hơn. Mỗi nhánh đại diện cho một trị của ký hiệu có tại vị trí i + 1 (tính từ bên trái) của những từ mã mà i ký hiệu bên trái nó được biểu diễn bằng một đường liên tục các nhánh tính từ nút gốc đến nút cấp thứ i được phân nhánh này. Mỗi nhánh phân ra từ nút cấp i này sẽ kết thúc ở một nút cấp i +1.

Mỗi nút cuối (còn gọi là nút lá) là nút kết thúc của nhánh sẽ đại diện cho một từ mã, thứ tự các trị ký hiệu mã là thứ tự các trị trên các nhánh đi từ nút gốc đến nút cuối qua các nút trung gian.



Hình 3-2: Ví dụ về cây mã

Rõ ràng có thể có những nút cuối mà không có nhánh nào đi ra từ nó, và cũng có thể có những nút cuối của từ mã này là nút trung gian của từ mã khác.

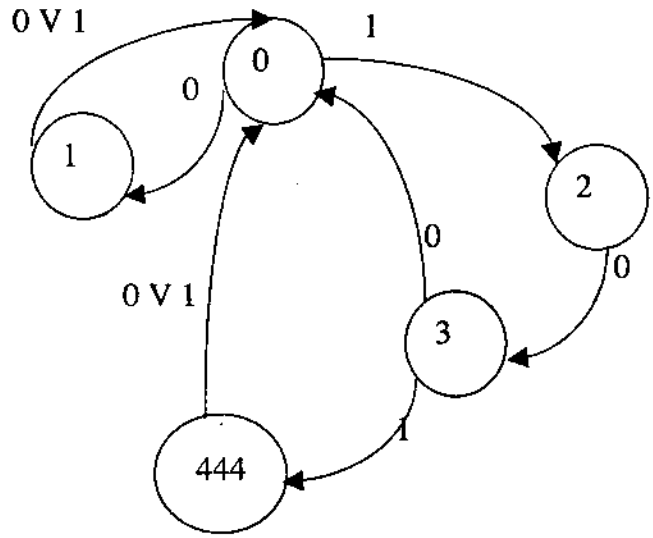
Mã hiệu có nút cuối trùng với nút trung gian của từ mã khác sẽ có đặc điểm là từ mã ngắn hơn là phần đầu từ mã dài hơn và nó không cho phép phân tách một chuỗi mã bất kỳ thành một dãy duy nhất các từ mã.

Nhìn cây mã, chúng ta có thể nhận biết mã đã cho thuộc loại mã đều (khi

các nút lá cùng bậc) hay không đều, mã đầy (tất cả các nút trung gian bậc trước các nút lá đều có k nhánh) hay mã vơi.

2.2. Đồ hình kết cấu

Đồ hình kết cấu gồm có những nút và nhánh có hướng, đây là cách biểu diễn cây mã rút gọn.



Hình 3-3: Ví dụ về đồ hình kết cấu

Một từ mã được biểu diễn bởi một vòng kín xuất phát từ nút gốc theo các nhánh có hướng qua các nút trung gian và trở về kết thúc tại nút gốc. Trị của các nhánh được ghi ở đầu xuất phát của nhánh và ở phía bên trái. Dấu V là dấu hoặc (hay là) có nghĩa là một nhánh có thể đại diện cho trị bên trái hoặc trị bên phải của dấu V; các nút được đánh số theo thứ tự xa dần nút gốc (số ghi trong vòng tròn của nút). Thứ tự trị các ký hiệu lấy theo thứ tự các nhánh trên đường đi.

Đồ hình kết cấu không chỉ được dùng để mô phỏng bản thân mã mà còn được dùng để xét cách vận hành của thiết bị mã hoá và giải mã.

3. Hàm cấu trúc mã

Hàm cấu trúc mã thể hiện một đặc tính quan trọng của mã là sự phân bố các từ mã theo độ dài, ký hiệu bằng $G(n_i)$.

Ví dụ: Với bộ mã 00, 01, 100, 1010, 1011; ta có G(n) dưới dạng sau:

$$G(n_i) = \begin{cases} 2, \text{ khi } n_i = 2 \\ 1, \text{ khi } n_i = 3 \\ 0, \text{ khi } n_i = 3 \end{cases}$$

Từ hàm cấu trúc có thể phân biệt được mã đều hoặc không đều. Trường hợp hàm cấu trúc bằng không với tất cả các từ mã trừ tại một điểm, chúng ta có loại mã đều. Nếu hàm cấu trúc có hai hoặc nhiều điểm khác không, chúng ta có loại mã không đều. Cũng từ hàm cấu trúc có thể xác minh mã thỏa mãn điều kiện phân tách được hay không.

Kết luận:

Trên đây là những phương pháp biểu diễn tổng quát có thể áp dụng cho bất cứ loại mã nào, nhưng cũng còn những phương pháp biểu diễn tổng quát có tính chất bao trùm hơn mà chúng ta có thể phân làm hai phương pháp chủ yếu: phương pháp biểu diễn hình học và phương pháp biểu diễn đại số. Với phương pháp biểu diễn hình học, các từ mã có độ dài n được biểu diễn bằng một véctơ hay một điểm. Bộ mã là một hệ điểm trong không gian đó. Với phương pháp biểu diễn đại số, bộ mã được xem như một cấu trúc đại số nhất định. Các loại mã được phân loại và nghiên cứu theo các cấu trúc đại số.

IV. MÃ HỆ THỐNG

1. Mã hệ thống có tính prefix

Mã hệ thống là một loại mã mà mỗi từ mã của nó được xây dựng bằng cách liên kết một số từ mã của một bộ mã gốc. Vì bộ mã gốc có tính phân tách nên bộ mã hệ thống cũng có tính phân tách.

Hiện nay, người ta thường sử dụng một số từ mã của bộ mã gốc làm các tổ hợp tạo thành phần đầu của từ mã hệ thống và gọi nó là tổ hợp sơ đẳng. Các tổ hợp còn lại của bộ mã gốc được sử dụng làm các tổ hợp kết thúc của từ mã hệ thống và gọi là các tổ hợp cuối. Từ mã của mã hệ thống được tạo ra bằng cách nối các tổ hợp sơ đẳng với nhau và nối thêm một tổ hợp cuối.

Việc giải mã đối với mã hệ thống phải thông qua hai bước. Bước một là tách các chuỗi ký hiệu mã nhận được thành chuỗi các tổ hợp sơ đẳng và các tổ hợp cuối. Bước hai là tìm các tổ hợp cuối và xác định điểm kết thúc từ mã tại đây.

Nếu bộ mã gốc có tính prefix thì mã hệ thống cũng có tính prefix. Mã hệ thống có tính prefix được xây dựng từ một bộ mã gốc có tính prefix nào đó bằng cách lấy một số từ mã của mã prefix gốc làm tổ hợp sơ đẳng và các từ mã còn lại làm tổ hợp cuối. Ghép các tổ hợp sơ đẳng với nhau và nối một trong các tổ hợp cuối thành một từ mã mới gọi là mã hệ thống có tính prefix .

Ví dụ: Lấy bộ mã frefix 1, 00, 010, 011 làm gốc, trong đó các tổ hợp 1, 00, 010 là tổ hợp sơ đẳng, 011 làm tổ hợp cuối. Các tổ hợp được hình thành sau đều có thể là từ mã của mã hệ thống:

1011, 11011, 00011, 100011, 010011, 11010011 ,...

Khi giải mã phải qua hai bước.

Bước một, là từ dãy ký hiệu nhận được phân tách thành dãy các tổ hợp sơ đẳng và tổ hợp cuối. Bước hai, phân tách thành dãy các tổ hợp mã của hệ thống.

Với bộ mã trên, khi nhận được tin dưới dạng dãy ký hiệu mã:

1101101001110110001101001l

Bước 1: tách thành dãy các tổ hợp sơ đẳng và tổ hợp cuối:

1-1-011-010-1-011-00-011-010-011

Bước 2: phân tách thành dãy các tổ hợp mã của hệ thống

11011-0101011-00011-010011

2. Mã có dấu phân tách

Các loại mã hệ thống có tính prefix đều có tính phân tách được, nhưng các thiết bị tách từ mã đối với các loại mã đó đều phức tạp. Để có thể sử dụng những thiết bị đơn giản để phân tách từ mã và cho phép phân tách mã trong điều kiện có nhiễu, cần phải xây dựng những loại mã có dấu phân tách. Những lý luận về mã hệ thống giúp chúng ta nhận thấy rằng nếu mỗi từ mã có một tổ hợp cuối và tổ hợp cuối của mọi từ mã là giống nhau thì mỗi khi tìm thấy tổ hợp cuối này chúng ta tìm thấy điểm kết thúc của một từ mã. Như vậy, người ta có thể sử dụng tổ hợp cuối này để phân tách mã và nó được gọi là dấu phân tách. Mã dùng dấu phân tách để cắt các từ mã được gọi là mã có dấu phân tách.

Tóm lại, mã có dấu phân tách là một loại mã mà mỗi từ mã được giới hạn bởi một dấu hiệu đặc biệt gọi là dấu phân tách. Dấu phân tách là một dấu hiệu

có thể là một tổ hợp các ký hiệu mã mà cũng có thể không phải là ký hiệu mã. Ví dụ, một bài viết là một cách mã hóa thông tin trao đổi giữa con người với nhau. Nếu coi bài viết này được mã hóa bởi các ký hiệu mã là chữ cái thì dấu cách (là dấu phân tách) không phải là ký hiệu mã. Nhưng nếu chúng ta coi bài viết được mã hóa bằng bộ chữ mở rộng (gồm cả dấu) thì dấu cách là một ký hiệu mã. Thường trong mã hóa, dấu phân tách là một tổ hợp đặc biệt mà ta không gặp nó trong từ mã hoặc chuỗi liên tục các từ mã. Một trường hợp thường gặp trong mã hóa, mã có dấu phân tách là một trường hợp riêng của mã hệ thống có tính prefix .

Trường hợp mã có tính prefix, khi phân tách một từ mã, ta phải biết được ít nhất tất cả các ký hiệu của từ mã cần phân tách. Việc phân tách một từ mã là việc nhận dạng dãy ký hiệu của một từ mã cụ thể. Việc này phải được thực hiện với mọi từ mã của bộ mã.

Trường hợp mã có dấu phân tách, để tách một từ mã đã cho, chúng ta chỉ cần nhận dạng dấu phân tách. Nếu dấu phân tách là một dấu hiệu đặc biệt mà nhiễu trong đường truyền, khi không lớn, khó làm nó chuyển thành một ký hiệu mã thì mã này cho phép chống nhiễu. Nếu dấu phân tách là tổ hợp ký hiệu mã của dấu phân tách thì việc nhận dạng này được thực hiện bằng cách trễ đoạn mã dài bằng dấu phân tách bằng một thanh ghi dịch và so sánh đoạn mã này với dấu phân tách. Nếu đầu ra của mạch so sánh cho biết sự giống nhau thì chúng ta nhận được một dấu phân tách và ta biết đã kết thúc một từ mã. Cấu trúc phân tách từ mã nêu trên được gọi là bộ lọc tuyến tính điều chỉnh theo dấu phân tách.

Với mã có dấu phân tách là một trường hợp riêng của mã hệ thống được gọi là mã hệ thống có dấu phân tách thì từ mã của nó gồm phần đầu (là các từ mã của bộ mã gốc có tính prefix) và phần cuối (là một từ mã của bộ mã gốc) là dấu phân tách. Dấu phân tách này sẽ không trùng với bất kỳ một đoạn mã nào trong từ mã cũng như do xếp các từ mã tạo thành.

Một ví dụ về mã hệ thống có dấu phân tách là bộ mã được xây dựng từ bộ mã gốc 1, 01,001, với tổ hợp cuối hay dấu phân tách là 001 và các tổ hợp sơ đẳng là 1,01, chúng ta sẽ có các từ mã:

01111001, 111101001, 010101001,...

Nếu nhận được chuỗi ký hiệu mã:

01111001010101001111101001011111001

ta dễ dàng tách ra được các từ mã:

01111001- 010101001-111101001-01111001

bằng bộ lọc tuyến tính điều chỉnh theo dấu phân tách 001 gồm thanh ghi dịch 3 bit mà các ký hiệu mã của chuỗi mã nhận được cho đi qua nó. Đầu ra mỗi bít thanh ghi dịch được đưa đến đầu vào một mạch đồng dấu. Ba đầu vào còn lại của mỗi mạch sẽ theo thứ tự là 0 0 1. Đầu ra 3 mạch đồng dấu sẽ qua một mạch ngưỡng mà chỉ khi cả 3 đầu vào của mạch ngưỡng là 1 thì đầu ra của nó mới bằng 1. Khi đầu ra của mạch ngưỡng là 1, chúng ta có dấu phân tách trên thanh ghi dịch hay nói cách khác, chúng ta xác định được điểm kết thúc của từ mã.

Câu hỏi ôn tập

1/ Nêu ý nghĩa của mã hóa

2/ Khái niệm mã hiệu và các thông số cơ bản của mã hiệu

3/ Nêu điều kiện để mã phân tách được

4/ Khái niệm về mã có tính prefix, mã hệ thống, mã hệ thống có tính prefix và mã có dấu phân tách

5/ So sánh các phương pháp biểu diễn mã?

6/ Cho bộ mã X = {00,10,110, 1110, 11110, 11111}

- Biểu diễn bằng bảng đối chiếu
- Biểu diễn bằng mặt phẳng toạ độ
- Biểu diễn bằng cây nhị phân
- Biểu diễn bằng đồ hình kết cấu

7/ Hãy lập bảng thử mã cho những bộ mã sau. Cho biết mã có phân tách được hay không?, nếu được thì độ chậm giải mã (trong trường hợp xấu nhất, là bao nhiêu?

- X_1 = { 00,01,100,1010,1011}
- X_2 = { 00,01,101,1010}
- X_3 = { 00,01,100,1010,1011}
- X_4 = { 00,01,110,111,1110}
- X_5 = { 00,01,110,111,0111}

Chương 4

MÃ HÓA

Mục tiêu

- Nghiên cứu các phương pháp mã hóa nguồn, mã hóa kênh
- Giải thích được khái niệm mã hóa thống kê tối ưu, nêu được các bước mã hóa và giải mã của các phương pháp mã hóa thống kê tối ưu.
- Giải thích được khái niệm mã hóa chống nhiễu, nêu được các bước mã hóa và giải mã của các phương pháp mã hóa chống nhiễu.
- Vận dụng kiến thức để giải các bài toán mã hóa thống kê tối ưu, mã hóa chống nhiễu.

I. MỞ ĐẦU

Hệ thống truyền tin được sử dụng để truyền thông tin từ nguồn tin đến nơi nhận tin. Nguồn tin có thể có nhiều dạng, trong hệ thống radio, nguồn tin là nguồn âm thanh (tiếng nói hay âm nhạc); trong hệ thống truyền hình, nguồn tin là nguồn video, nó tạo ra các hình ảnh chuyển động. Những nguồn như thế được gọi là nguồn tương tự. Ngược lại, máy tính tạo ra và sử dụng các tín hiệu rời rạc (các số nhị phân hay các chữ cái trong bảng mã ASCII) được gọi là nguồn rời rạc.

Hệ thống truyền tin rời rạc, khi truyền các tín hiệu liên tục thường phải thông qua một số phép biến đổi rời rạc (rời rạc hóa, lượng tử hóa) để đổi thành tín hiệu số (thường là nhị phân) rồi mã hóa. Ở đầu thu tín hiệu phải thông qua một số phép biến đổi ngược lại như giải mã, liên tục hóa, để phục hồi tin tức.

Sự mã hóa tin tức nhằm mục đích tăng tính hiệu quả và độ tin cậy của hệ thống truyền tin, nghĩa là tăng tốc độ truyền tin và khả năng chống nhiễu của tín hiệu khi truyền qua kênh. Thông thường tốc độ lập tin của nguồn thường còn rất xa mới đạt được thông lượng của kênh. Để tăng tốc độ lập tin, người ta

dùng phép mã hóa để thay đổi tính chất thống kê, làm giảm độ dư của thông tin không cần thiết của nguồn nhờ đó tiếp cận với thông lượng của kênh. Các phương pháp mã loại này gọi là mã hóa nguồn.

Mặt khác, các tin mang các mã hiệu khi truyền trong kênh thường bị nhiễu phá hoại. Vì vậy, mã hiệu phải được xây dựng theo những quy tắc nhất định nhằm đảm bảo cho phía thu phát hiện được các sai nhầm đồng thời sửa được chúng. Loại mã hóa này được gọi là mã hóa kênh còn được gọi là mã chống nhiễu. Mã hóa kênh nhằm cải tiến kỹ thuật truyền tin, cho phép tín hiệu phát đi có khả năng chống lại ảnh hưởng của nhiễu. Mã hóa kênh làm giảm lỗi bit hoặc giảm tỷ số năng lượng bit trên mật độ nhiễu tại đầu thu.

II. MÃ HÓA NGUỒN

1. Một số khái niệm chung

Mã hóa nguồn là phép biến đổi đầu tiên cho nguồn nguyên thủy, đầu vào của phép biến đổi này có thể là nguồn tin rời rạc hoặc nguồn tin nguyên thủy. Trong cả hai trường hợp mục đích chính của phép mã hóa nguồn là biểu diễn thông tin với tài nguyên tối thiểu.

Các vấn đề cần nghiên cứu đối với mã hóa nguồn là: mã hóa nguồn liên tục, mã hóa nguồn rời rạc và nén dữ liệu.

1.1. Nguồn rời rạc

- Nguồn rời rạc tạo ra các tin rời rạc thể hiện là chuỗi các ký hiệu ngẫu nhiên.

- Đối với mã hóa nguồn rời rạc, vấn đề cơ bản là thay đổi bảng chữ cái và phân bố xác suất để giảm bớt lượng ký hiệu cần dùng. Như vậy cần quan tâm:

+ Entropi của nguồn trước khi mã hóa

+ Entropi của nguồn sau khi mã hóa

+ Hiệu quả của phép mã hóa

+ Giới hạn của phép mã hóa

- Đối với nguồn rời rạc không nhớ: sử dụng từ mã có độ dài cố định hoặc từ mã có độ dài thay đổi.

- Đối với nguồn dừng rời rạc: thường sử dụng thuật toán mã hoá nguồn Lempel-Ziv.

1.2. Nguồn liên tục

- Nguồn liên tục tạo ra tín hiệu liên tục thể hiện một quá trình ngẫu nhiên liên tục. Trong các hệ thống truyền thông, nguồn liên tục thường được biến đổi thành nguồn rời rạc, xử lý và truyền rồi ở đầu nhận lại biến đổi thành nguồn liên tục.

- Rời rạc hóa nguồn liên tục:

+ Lấy mẫu nguồn tương tự: biến đổi nguồn tương tự thành một chuỗi các giá trị ngẫu nhiên liên tục tại các thời điểm lấy mẫu.

+ Lượng tử hóa nguồn tương tự: mã hóa các giá trị liên tục bằng nguồn rời rạc.

- Tại đích, nguồn rời rạc được tổng hợp thành nguồn tương tự, cụ thể là:

+ Tái tạo lại các giá trị liên tục của chuỗi giá trị ban đầu từ các ký hiệu của nguồn rời rạc.

+ Kết nối các giá trị liên tục thành một tín hiệu ngẫu nhiên đầu ra.

+ Do quá trình lượng tử, đầu ra sai khác so với đầu vào gọi là sai số lượng tử.

- Người ta thường sử dụng các kỹ thuật: mã hóa miền thời gian, mã hóa miền tần số, mã hóa mô hình nguồn để mã hóa nguồn liên tục.

Trong mục này chúng ta chỉ khảo sát phương pháp mã hóa nguồn rời rạc.

2. Mã hoá nguồn rời rạc không nhớ

2.1. Mã hóa nguồn rời rạc với từ mã có độ dài cố định

Nguyên tắc: mã hóa một ký hiệu nguồn bằng một chuỗi ký hiệu mã có độ dài n.

Để đảm bảo phép mã hóa là một - một thì một ký hiệu nguồn ứng với một chuỗi ký hiệu nhị nhân, số lượng chuỗi ký hiệu nhị phân phải lớn hơn số ký hiệu nguồn.

$$2^n \geq L \text{ hay } n \geq \log_2 L \qquad (4\text{-}1)$$

Nếu L là luỹ thừa của 2 thì giá trị nhỏ nhất của R là $\log_2 L$. Nếu L không phải là luỹ thừa của 2 thì giá trị của $n = [\log_2 L]+1$. (Ở đây [x] chỉ số nguyên lớn nhất nhỏ hơn x).

Hiệu quả của phép mã hóa được xác định bằng:

$$K = \frac{H(X)}{n} . \qquad (4\text{-}2)$$

2.1.1. Với nguồn rời rạc đẳng xác suất

- Hiệu quả của mã hóa đạt giá trị cực đại (bằng 1) khi L là luỹ thừa của 2

- Nếu nguồn tin ban đầu đẳng xác suất nhưng L không phải là luỹ thừa của 2 thì n sai khác H(X) tối đa là 1 bit/ký hiệu. Khi $\log_2 L >> 1$ thì phương pháp mã hóa này có hiệu quả cao. Ngược lại, khi $\log_2 L < 1$ thì hiệu quả của phương pháp mã hóa với độ dài cố định có thể nâng lên bằng cách mã hóa từng khối gồm J ký hiệu nguồn. Để mã hoá này duy nhất, ta cần L^J từ mã, nếu gọi N là số ký hiệu mã được sử dụng để mã hóa thì giá trị nguyên nhỏ nhất có thể chấp nhận của N là $N = [J\log_2 L] + 1$. Khi đó, số ký hiệu mã trung bình ứng với một ký hiệu nguồn là $n = \frac{N}{J}$, độ không tối ưu của mã giảm xuống xấp xỉ $\frac{1}{J}$ so với việc mã hóa riêng từng ký hiệu nguồn. Khi J đủ lớn, hiệu quả của phương pháp được đo bằng tỷ số $\frac{JH(X)}{N}$ rất gần tới 1.

- Các phương pháp trên không bị sai số, mỗi chuỗi ký hiệu nguồn luôn ứng với một từ mã duy nhất.

2.1.2. Với nguồn rời rạc không đẳng xác suất

- Trong trường hợp nguồn không đẳng xác suất, để có thể tiếp cận với hiệu quả tối đa (bằng 1) bằng cách giảm n, cần chấp nhận một sai số nào đó. Ví dụ chỉ một phần trong LJ khối ký hiệu được mã hóa một cách duy nhất còn LJ – (2N –1) khối J ký hiệu được mã hóa thành một từ mã duy nhất. Phương pháp mã hóa này gây ra việc giải mã sai đối với các khối có xác suất xuất hiện thấp, gọi P_e là xác suất giải mã sai.

- Dựa trên phương pháp mã hóa này Shannon đã chứng minh định lý mã hóa nguồn như sau:

" Gọi X là một nguồn rời rạc không nhớ có entropi hữu hạn H(X). Các khối J ký hiệu của nguồn được mã hóa thành các từ mã nhị phân có độ dài N. Với mọi $\varepsilon > 0$, xác suất giải mã khối sai P_e có thể nhỏ tuỳ ý nếu:

$$n = \frac{N}{J} \geq H(X) + \varepsilon$$

Ngược lại, nếu : $n = \frac{N}{J} \le H(X) + \varepsilon$

thì P_e dần tới 1 khi J tiến tới vô hạn."

- Từ định lý này ta thấy số lượng nhị phân trung bình để mã hóa cho một nguồn rời rạc không nhớ với xác suất giải mã sai nhỏ tuỳ ý bị chặn dưới bởi entropi H(X). Ngược lại nếu n < H(X) thì xác suất giải sai tiến tới 100% khi J tăng tới vô hạn.

2.2. Mã hóa nguồn rời rạc với từ mã có độ dài thay đổi

Mã hoá nguồn sử dụng với từ mã có độ dài thay đổi thường sử dụng phương pháp mã hóa thống kê tối ưu.

2.2.1. Khái niệm mã hoá thống kê tối ưu

Mục tiêu của mã hoá thống kê tối ưu là mã hóa tin rời rạc với số lượng các ký hiệu mã nhị phân tối thiểu. Do đó mã hóa thống kê tối ưu thuộc loại mã nén dữ liệu.

Phương pháp mã hóa này khai thác đặc tính: không phải tất cả các ký hiệu trong một frame truyền có cùng tần suất xuất hiện, ví dụ trong cùng một chuỗi ký tự, sẽ có một số ký tự xuất hiện nhiều hơn các ký tự khác. Thay vì dùng một số bit nhất định mã hoá một ký tự, chúng ta dùng các từ mã có độ dài nhỏ để mã hoá các ký tự có xác suất xuất hiện cao và ngược lại. Như vậy, số ký hiệu cần thiết để mã hoá cho chuỗi ký tự sẽ nhỏ hơn và tính kinh tế cũng cao hơn. Ví dụ: ký tự a, b xuất hiện nhiều nên dùng 1 bit để mã hoá, ký tự z xuất hiện ít có thể dùng 10 bit để mã hóa.

Tiêu chuẩn của mã thống kê tối ưu đạt đến là *độ dài trung bình từ mã là nhỏ nhất*. Một nguồn tin liên tục sau khi được rời rạc hóa, hoặc một nguồn tin rời rạc hóa đều được mô tả bằng cấu trúc thống kê của chúng. Trường hợp các lớp tin trong nguồn xuất hiện độc lập thống kê với nhau, sự mô tả nguồn tin được đơn giản hóa đi rất nhiều, lúc đó chỉ cần nêu quy luật phân bố xác suất xuất hiện của các ký hiệu trong bộ mã là đủ. Thông thường quy luật phân bố xác suất xuất hiện của các ký hiệu không đều, lúc đó nhiệm vụ của phép mã hoá nguồn rời rạc là mã hóa sao cho độ dài trung bình của từ mã là nhỏ nhất. Cụ thể là dùng những từ mã ngắn cho các ký hiệu mã có xác suất xuất hiện lớn và ngược lại mã hóa các ký hiệu ít xuất hiện bằng những từ mã dài.

Mã hóa thống kê tối ưu có ưu điểm là hệ số nén tương đối cao, phương pháp thực hiện tương đối đơn giản, đòi hỏi ít bộ nhớ, có thể xây dựng trên các mảng nhỏ hơn 64KB. Nhược điểm của nó là phải chứa cả bảng mã vào tệp tin nén thì phía nhận mới có thể giải mã được, do đó hiệu suất nén chỉ cao khi thực hiện các tệp tin lớn.

2.2.2. Định lý về giới hạn trên và dưới của chiều dài trung bình từ mã

Định lý này phát biểu như sau:

Gọi $X = \{x_i\}$, các xác suất xuất hiện tương ứng $p(x_i)$ là nguồn rời rạc không nhớ với entropi hữu hạn $H(X)$. Có thể xây dựng một mã hiệu nhị phân có tính prefix và có độ dài trung bình từ mã $\bar{n}$ thoả mãn điều kiện bất đẳng thức:

$$H(X) \le \bar{n} < H(X) + 1 \qquad (4\text{-}3)$$

Mã thống kê tối ưu là bộ mã có độ dài trung bình từ mã thoả mãn (4-3) .

2.2.3. Mã thống kê tối ưu Shannon - Fano

Độc lập với nhau nhưng Shannon và Fano đã xây dựng phương pháp lập mã thống kê tối ưu trên cùng một cơ sở: độ dài từ mã tỷ lệ nghịch với xác suất xuất hiện.

a. Các bước để lập mã

Bước 1: Sắp xếp các tin x_i của nguồn X theo thứ tự xác suất giảm dần.

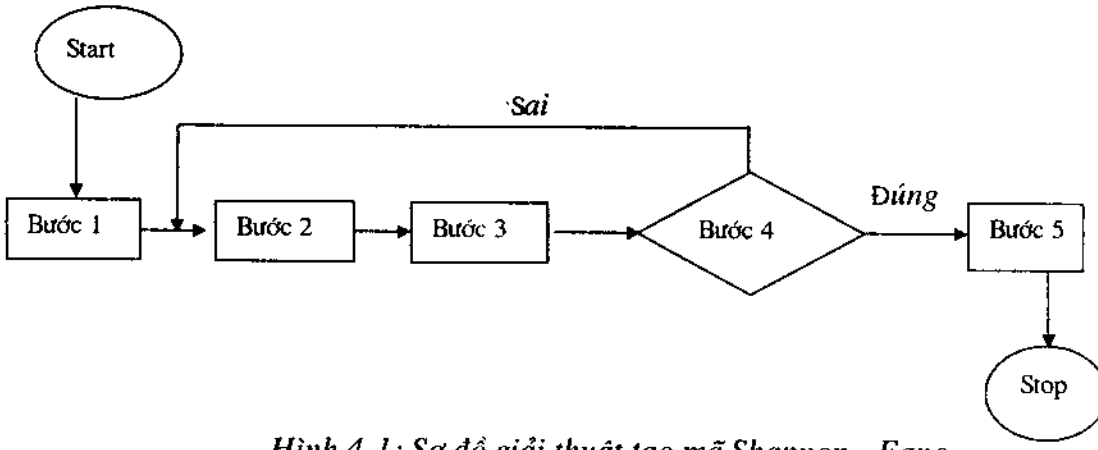
Bước 2: Chia thành hai nhóm tin sao cho tổng xác suất mỗi nhóm gần bằng nhau.

Bước 3: Mỗi nhóm gán cho một ký hiệu mã (0,1).

Bước 4: Lặp lại bước 2 và 3 cho các nhóm con cho đến khi không thể tiếp tục được nữa.

Bước 5: Đọc từ mã tương ứng với mỗi tin là tổ hợp tất cả các ký hiệu mã của các nhóm mà lớp tin đó phụ thuộc, lấy từ nhóm lớn đến nhóm nhỏ (tính từ trái sang phải).

b. Sơ đồ giải thuật tạo mã Shannon - Fano



Hình 4-1: Sơ đồ giải thuật tạo mã Shannon - Fano

c. Các thông số tối ưu

Thử bất đẳng thức Kraft ($\sum_{k=1}^{L} 2^{-n_k} \leq 1$), dấu bằng được xảy ra.

Hiệu quả của mã (hiệu suất của mã hay còn gọi là hệ số tối ưu tương đối)

$$K = \frac{H(X)}{\overline{n}} = \frac{-\sum_{i=1}^{L} p(x_i) \log_2 p(x_i)}{\sum_{i=1}^{L} p(x_i) n_i}$$

Ví dụ 1:

Cho nguồn tin X = { x_i} i = 1...8, có các lớp tin và xác suất xuất hiện tương ứng p(x_i) = {0,25; 0,125; 0,0625; 0,0625; 0,25; 0,125; 0,0625; 0,0625}. Hãy lập mã thống kê tối ưu Shannon - Fano cho nguồn tin trên và tính hiệu suất mã.

Đầu tiên, ta sắp xếp nguồn tin giảm dần theo xác suất xuất hiện, ghi vào bảng theo mẫu:

x_i	$p(x_i)$	Số lần chia nhóm				Từ mã	n_i
		1	2	3	4		
x_1	0,25	**0**	**0**			00	2
x_5	0,25		**1**			01	2
x_2	0,125	**1**	**0**	**0**		100	3
x_6	0,125			**1**		101	3
x_3	0,0625		**1**	**0**	**0**	1100	4
x_4	0,0625				**1**	1101	4
x_7	0,0625			**1**	**0**	1110	4
x_8	0,0625				1	1111	4

$$H(X) = \sum_{i=1}^{L} p(x_i) \log p(x_i) = 2{,}75$$

$$\bar{n} = \sum_{i=1}^{L} p(x_i) n_i = 2{,}75$$

$$K = \frac{H(X)}{\bar{n}} = 100\%$$

d. Nhận xét

Mã Shannon - Fano là bộ mã có tính prefix.

Mã Shannon - Fano có từ mã tương ứng với lớp tin có xác suất xuất hiện lớn sẽ ngắn, những từ mã ứng với lớp tin có xác suất xuất hiện ít sẽ dài.

Mã Shannon - Fano không duy nhất do phương pháp lấy tổng xác suất xấp xỉ nên trong nhiều trường hợp có nhiều hơn một cách chia. Ứng với mỗi cách chia có thể sẽ cho ra các bộ mã có chiều dài trung bình khác nhau.

Ví dụ 2:

Mã hóa nguồn X= { x_i}, i = 1,2..,8 với các xác suất lần lượt là:

$p(x_i)$ = {0,23; 0,2; 0,14; 0,12; 0,1; 0,09; 0,06; 0,06}

x_i	$p(x_i)$					Từ mã
		1	2	3	4	
x_1	0,23	0	0			00
x_2	0,2		1			01
x_3	0,14	1	0	0		100
x_4	0,12			1		101
x_5	0,1		1	0	0	1100
x_6	0,09				1	1101
x_7	0,06			1	0	1110
x_8	0,06				1	1111

$\bar{n} = 2,88$

x_i	$p(x_i)$					Từ mã
		1	2	3	4	
x_1	0,23	0	0			00
x_2	0,2		1	0		010
x_3	0,14			1		011
x_4	0,12	1	0	0		100
x_5	0,1			1		101
x_6	0,09		1	0		110
x_7	0,06			1	0	1110
x_8	0,06				1	1111

$\bar{n} = 2,89$

2.2.4. Mã thống kê tối ưu Huffman

Năm 1952 Huffman đã đưa ra một thuật toán mã hóa dựa trên xác suất xuất hiện của các ký hiệu.

a. Các bước lập mã

Bước 1: Sắp xếp các tin x_i của nguồn X theo thứ tự xác suất giảm dần.

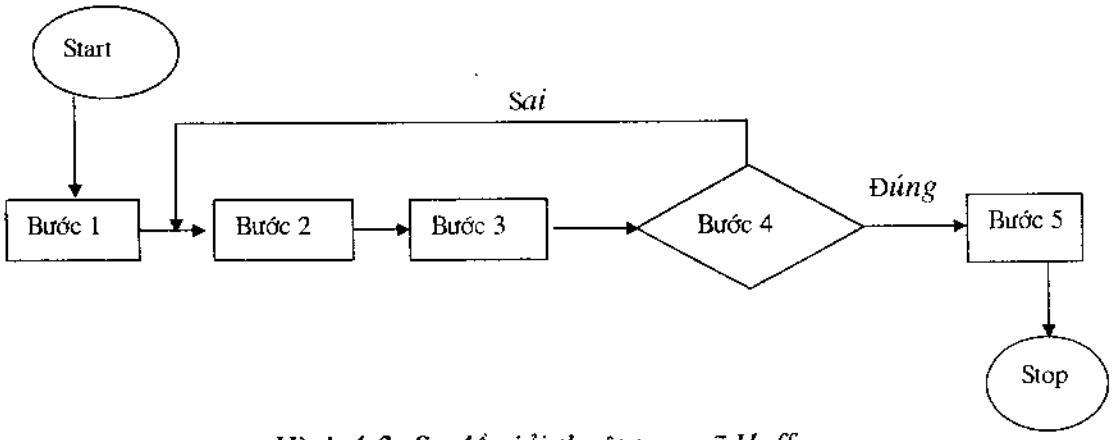
Bước 2: Chọn hai lớp tin có xác suất nhỏ nhất gán cho mỗi lớp tin một ký hiệu mã (0,1).

Bước 3: Thay thế hai lớp tin nhỏ nhất này bằng một tin mới có xác suất bằng tổng hai xác suất.

Bước 4: Coi như có nguồn tin mới, quay lại làm từ bước 1 cho đến khi tổng hai xác suất bằng 1 thì dừng lại.

Bước 5: Đọc từ mã tương ứng với mỗi lớp tin là tổ hợp các ký hiệu mã gán cho các nhóm mà lớp tin đó phụ thuộc vào, lấy từ nút gốc đến nút cuối.

b. Sơ đồ giải thuật tạo mã



Hình 4-2: Sơ đồ giải thuật tạo mã Huffman

c. Các thông số tối ưu.

Thử bất đẳng thức Kraft ($\sum_{k=1}^{L} 2^{-n_k} \leq 1$), dấu bằng được xảy ra.

Hiệu quả của mã (hiệu suất của mã hay còn gọi là hệ số tối ưu tương đối)

$$K = \frac{H(X)}{\bar{n}} = \frac{-\sum_{i=1}^{L} p(x_i)\log_2 p(x_i)}{\sum_{i=1}^{L} p(x_i)n_i}$$

Ví dụ 3:

Cho nguồn tin X = { x_i} i = 1, 2,..8, có các lớp tin và xác suất xuất hiện tương ứng p (x_i) = { 0,36; 0,14; 0,13; 0,12; 0,1; 0,09; 0,04; 0,02}

Hãy lập mã thống kê tối ưu Huffman cho nguồn tin trên và tính hiệu suất bộ mã

Đầu tiên ta sắp xếp nguồn tin giảm dần theo xác suất xuất hiện, ghi vào bảng theo mẫu:

x_i	$p(x_i)$	Số lần chia nhóm							Từ mã	n_i
		1	2	3	4	5	6	7		
x_1	0,36						**0**		00	2
x_2	0,14				**0**		**1**	**0**	010	3
x_3	0,13				**1**				011	3
x_4	0,12			**0**		**0**		**1**	100	3
x_5	0,1			**1**					101	3
x_6	0,09		**0**			**1**			110	3
x_7	0,04	**0**	**1**						1110	4
x_8	0,02	**1**							1111	4

$$H(X) = -\sum_{i=1}^{L} p(x_i)\log_2 p(x_i) = 2,63$$

$$\bar{n} = \sum_{i=1}^{L} n_i p(x_i) = 2,7$$

$$K = \frac{H(X)}{\bar{n}} \approx 97,4\%$$

d. Nhận xét

- Mã Huffman là bộ mã có tính prefix.

- Mã Huffman có từ mã tương ứng với lớp tin có xác suất xuất hiện lớn sẽ ngắn, những từ mã ứng với lớp tin có xác suất xuất hiện ít sẽ dài.

- Mã Huffman là mã duy nhất .

- Phương pháp mã hoá tối ưu Huffman sẽ trở nên nặng nề khi số tin nguồn quá lớn. Trong trường hợp này người ta dùng một biện pháp phụ để giảm nhẹ việc mã hoá. Trước tiên liệt kê các tin nguồn theo thứ tự xác suất giảm dần, sau đó ghép thành từng nhóm những tin có xác suất gần bằng nhau. Dùng mã đều để mã hoá các tin trong cùng một nhóm. Sau đó, xem các nhóm tin như một khối tin và dùng phương pháp mã hoá Huffman để mã hoá các khối tin. Từ mã cuối cùng tương ứng mỗi tin nguồn gồm 2 phần: mã Huffman và mã đều.

Kết luận:

Mã hóa thống kê tối ưu có các đặc điểm:

- Mã thống kê tối ưu là mã có độ dài từ mã của các lớp tin tỷ lệ nghịch với xác suất xuất hiện của chúng.

- Đây là bộ mã của phép mã hóa tối ưu cho nguồn vì kết quả mã hoá là một bộ mã có chiều dài trung bình là nhỏ nhất trong tất cả các phép mã hóa cho nguồn.

- Các ký hiệu khác nhau của bộ mã phải đồng xác suất, có như vậy lượng tin mỗi ký hiệu mới đạt trị số cực đại.

- Xác suất xuất hiện ký hiệu trong từ mã không phụ thuộc vào sự có mặt của các ký hiệu ra trước.

- Thử bất đẳng thức Kraft ($\sum_{k=1}^{L} 2^{-n_k} \leq 1$), dấu bằng được xảy ra vì vậy đây là bộ mã có tính prefix.

3. Mã hoá nguồn dừng rời rạc

Các thuật toán mã hóa thống kê tối ưu phải biết xác suất xuất hiện của tất cả các tin. Tuy nhiên, trong thực tế, tính chất thống kê của nguồn thường không biết trước và mỗi chúng ta thường ước lượng các giá trị xác suất của nguồn rời rạc bằng cách quan sát một chuỗi dài các ký hiệu.

Ngược lại, thuật toán mã hóa nguồn Lampel-Ziv lại độc lập với tính chất thống kê của nguồn do đó rất thích hợp để mã hóa nguồn dừng rời rạc. Trong thuật toán này, một dãy các ký hiệu tin của nguồn rời rạc được chia thành các khối có độ dài thay đổi gọi là các câu. Một câu mới - là một khối các ký hiệu của nguồn và là một câu đã thêm vào một ký hiệu cuối cùng. Các câu được liệt kê trong một từ điển kèm với vị trí xuất hiện. Để mã hoá câu mới, ta chỉ ra vị trí trong từ điển và chèn thêm ký hiệu mới vào cuối.

Ví dụ xét dãy ký hiệu nhị phân sau:

101011010010011101010000110011101011000011011

Chia dãy ký hiệu trên thành các câu như sau:

1, 0, 10, 11, 01, 00, 100, 111, 010, 1000, 011, 001, 110, 101, 10001, 1011

Ta thấy rằng mỗi câu trong dãy là ghép của câu cũ và một ký hiệu mới. Để

mã hóa các câu, ta xây dựng một từ điển như hình (4-3). Các vị trí của từ điển liên tiếp nhau, bắt đầu bằng 1 và tăng dần (trong trường hợp này là 16). Các từ mã được xác định bởi vị trí của một câu có trước trong từ điển và ký hiệu mới được thêm vào trong từ mã. Khởi đầu, vị trí 0000 dùng để mã hóa cho một câu chưa xuất hiện trong từ điển.

Vị trí trong từ điển	Nội dung	Từ mã
0001	1	00001
0010	0	00000
0011	10	00010
0100	11	00011
0101	01	00101
0110	00	00100
0111	100	00110
1000	111	01001
1001	010	01010
1010	1000	01110
1011	011	01011
1100	001	01101
1101	110	01000
1110	101	00111
1111	10001	10101
	1011	11101

Hình 4-3: Từ điển dùng trong thuật toán Lempel - Ziv

Để giải mã cần xây dựng lại từ điển ở phía thu giống như ở phía phát và sau đó giải mã lần lượt các từ mã nhận được.

Ở ví dụ trên, mã hóa 44 ký hiệu nhị phân của nguồn thành 16 từ mã, mỗi từ mã có độ dài 5 bit, như vậy là không thực hiện việc nén số liệu do chuỗi ký hiệu được quan sát quá ngắn. Nếu chuỗi ký hiệu được quan sát thêm thì thuật toán sẽ hiệu quả hơn và có nén số liệu ở đầu ra của nguồn.

Độ lớn của từ điển chỉ phụ thuộc vào bộ nhớ dùng trong lưu trữ. Thuật toán Lempel - Ziv được dùng rộng rãi trong việc nén số liệu các tệp trong máy tính. Các tiện ích như compress và uncompress trong hệ điều hành DOS, UNIX xuất phát từ thuật toán này.

III. MÃ HOÁ KÊNH

1. Khái niệm

Thông tin truyền qua kênh thường bị nhiễu phá hoại gây ra các lỗi nên cần phải có các phương pháp mã hóa có khả năng phát hiện lỗi/ sửa lỗi.

Định lý Shannon mã hóa kênh có nhiễu ở chương II đã chỉ ra rằng: nếu thông lượng kênh lớn hơn tốc độ lập tin của nguồn thì có thể truyền tin với sai số nhỏ tuỳ ý. Định lý này là cơ sở lý thuyết của các loại mã chống nhiễu.

Mã hóa kênh là biện pháp dùng các mã hiệu cho phép phát hiện lỗi hoặc có khả năng tự sửa lỗi để mã hóa tín hiệu truyền qua kênh. Mã hóa kênh nhằm cải tiến kỹ thuật truyền tin cho phép tín hiệu phát đi có khả năng chống lại ảnh hưởng của nhiễu. Mã hóa kênh làm giảm lỗi bit hoặc giảm tỷ số năng lượng bit trên mật độ nhiễu yêu cầu tại đầu thu.

Trong mục này chúng ta sẽ khảo sát kỹ bốn loại mã chống nhiễu được dùng rộng rãi nhất trong truyền số liệu đó là: Mã kiểm tra chẵn lẻ (Patity), mã tuyến tính, mã vòng kiểm tra (CRC), mã Hamming.

2. Nguyên tắc phát hiện lỗi và sửa lỗi

Việc phát hiện lỗi và sửa lỗi phụ thuộc vào tính chất thống kê của kênh và nhiễu. Có hai loại lỗi:

- Lỗi độc lập thống kê: các lỗi xuất hiện riêng lẻ, không liên quan lẫn nhau.
- Lỗi chùm: lỗi liên quan chặt chẽ với nhau, thường xuất hiện cùng một lúc.

2.1. Nguyên tắc phát hiện lỗi của mã hiệu

Khả năng phát hiện sai của một mã hiệu dựa trên một ý tưởng rất đơn giản:

nếu mã hiệu là tập hợp những từ mã có độ dài n thì số các từ mã (tổ hợp mã mang tin) được chọn phải nhỏ hơn tổng số các tổ hợp n ký hiệu mã. Số các tổ hợp không làm từ mã được gọi là tổ hợp cấm. Khi chịu tác động của nhiễu, một từ mã chuyển đổi sai thành một từ mã khác thì không thể phát hiện được lỗi sai, nếu chuyển đổi thành tổ hợp cấm thì sẽ phát hiện được đã thu sai, do vậy khả năng phát hiện sai được tăng cường nếu số tổ hợp cấm tăng lên.

Một bộ mã cho phép phát hiện sai phải là bộ mã được xây dựng sao cho mọi từ mã bị sai trên đường truyền phải được chuyển thành tổ hợp cấm.

2.2. Nguyên tắc sửa lỗi của mã hiệu:

Cơ chế sửa sai của mã hiệu đồng thời cũng là nguyên lý giải mã phải dựa vào tính thống kê của kênh để đảm bảo sai nhầm tối thiểu. Nói một cách khác khi nhận được một tổ hợp cấm, thiết bị sửa sai có nhiệm vụ quy về từ mã phát đi với xác suất sai tối thiểu. Muốn vậy cần phải dựa vào tính chất nhiễu trong kênh để phân nhóm các tổ hợp cấm, mỗi nhóm tương ứng với một từ mã mà chúng có khả năng bị chuyển đổi sang nhiều nhất.

Một bộ mã cho phép sửa được sai là một bộ mã được xây dựng sao cho mỗi từ mã khi bị sai sẽ chuyển thành những tổ hợp cấm và là những tổ hợp cấm của chỉ riêng nó.

Ví dụ mã hiệu gồm 4 từ mã 0001, 0101, 1000, 1100 khi bị sai cụm gồm 2 ký hiệu kế cận (véc tơ sai 0011,0110,1100) sẽ chuyển thành 12 tổ hợp cấm theo cụm. Cơ chế sửa chữa sai dựa theo bảng giải mã sau:

Từ mã	0001	0101	1000	1100
Tổ hợp cấm	0010	0110	1011	1111
	0111	0011	1110	1010
	1101	1001	0100	0000

Khi nhận được các tổ hợp cấm 0010 hoặc 1011 sẽ giải mã về 0001,1000

3. Mã Parity

3.1. Khái niệm

Phương pháp thông dụng nhất được dùng để phát hiện các lỗi bit trong

truyền dữ liệu không đồng bộ và đồng bộ hướng ký tự là phương pháp mã Parity. Phương pháp này được xây dựng dựa trên nhận xét sau: nếu số ký tự 1 của mỗi từ mã luôn là chẵn thì khi truyền nếu bị sai lẻ số vị trí thì tổng số ký tự 1 sẽ là lẻ và ngược lại. Do đó có hai phương pháp kiểm tra đó là Parity chẵn và Parity lẻ, người ta thường dùng kiểm tra chẵn cho truyền không đồng bộ và kiểm tra lẻ cho truyền đồng bộ.

Xét phương pháp tạo mã và giải mã với Parity chẵn. Thuật toán tạo mã đơn giản nhất là thêm vào cuối mỗi tổ hợp mã mang tin một bit Parity sao cho nếu tổng các ký hiệu 1 của tổ hợp mã mang tin là lẻ (thực hiện bằng cách cộng modul 2 các bit 1 với nhau) thì bit Parity thêm vào có giá trị 1, nếu tổng các ký hiệu 1 là chẵn thì giá trị thêm vào có giá trị 0.

Khi giải mã, mạch kiểm tra sẽ xác định xem số bit 1 trên mỗi từ mã nhận được có đúng là chẵn không. Nếu không nó sẽ phát hiện được các lỗi đơn bít (số lượng bit lỗi là một số lẻ) và yêu cầu truyền lại. Nếu truyền đúng, nó sẽ loại bỏ bít Parity để lấy lại thông tin gốc.

3.2. Kiểm tra Parity theo ký tự

Một ví dụ điển hình của việc sử dụng mã parity đó chính là bộ mã ASCII. Trong bộ mã này, mỗi ký tự có 7 bit và một bit kiểm tra, với kiểm tra chẵn, giá trị của bit kiểm tra là 0 nếu số lượng các bit có giá trị 1 trong 7 bit là chẵn và có giá trị 1 trong trường hợp ngược lại. Còn kiểm tra lẻ thì ngược lại. Thông thường người ta sử dụng kiểm tra chẵn và bit kiểm tra gọi là P. Giá trị kiểm tra đó cho phép ở đầu thu phát hiện những sai sót đơn giản

Thí dụ

Kí tự	Mã ASCII	Từ mã phát đi	
A	1000001	10000010	bit P
E	1010001	10100011	

3.3. Phương pháp kiểm tra theo ma trận

Khi truyền đi một khối thông tin, mỗi ký tự được truyền đi sẽ được kiểm tra tính chẵn lẻ theo chiều ngang, đồng thời cả khối thông tin này cũng được kiểm tra tính chẵn lẻ theo chiều dọc. Như vậy cứ sau một số byte nhất định thì một byte kiểm tra chẵn lẻ cũng được gửi đi. byte chẵn lẻ này được tạo ra bằng cách kiểm ta tính chẵn lẻ của khối ký tự theo cột. Dựa vào các bit kiểm tra

ngang và dọc ta xác định được toạ độ của bit sai và sửa được bit sai này. Một Frame coi như một khối ký tự sắp xếp có 2 chiều, mỗi ký tự có bit kiểm tra chẵn lẻ P. Nếu ta sắp xếp các bit của ký tự đúng vị trí tương ứng từ trên xuống thì ta có một khối các ký tự:

	bit 1	bit 2		bit n	bit Parity
Ký tự 1	b_{11}	b_{21}		b_{n1}	R_1
Ký tự 2	b_{12}	b_{22}		b_{n2}	R_2
Ký tự n	b_{1m}	b_{2m}		b_{nm}	R_m
Bit kiểm tra cột	C_1	C_2		C_n	C_{n+1}

Hình 4-3: Kiểm tra khối BCC

Tính theo chiều ngang, giá trị bit chẵn lẻ P của dòng thứ i sẽ là :

$R_j = b_{1j} \oplus b_{2j} \oplus b_{3j} \oplus ..... \oplus b_{nj}$

ở đây $\oplus$ là ký hiệu của phép cộng modul 2.

Với R_j: bit kiểm tra thứ tự thứ j

b_{ij} : bit thứ i của ký tự thứ j

n : số lương bit trong một ký tự

Nếu tính theo chiều dọc, ta có:

$C_i = b_{i1} \oplus b_{i2} \oplus b_{i3} ..... \oplus b_{im}$

Với C_i: bit kiểm tra cột thứ i

m: số lượng ký tự trong một Frame

Tuy nhiên phương pháp này cũng không hoàn toàn hiệu quả. Giả sử bit thứ nhất và bit thứ ba của ký tự thứ nhất bị sai, kiểm tra hàng không thấy sai, nhưng kiểm tra chẵn lẻ của cột sẽ phát hiện bit thứ nhất và thứ ba bị sai, ta biết sự truyền bị sai nhưng không biết sai ở vị trí nào. Bây giờ, lại giả thiết rằng bit

thứ nhất và bit thứ ba của ký tự thứ 5 cũng bị sai đồng thời với bit thứ nhất và bit thứ ba của ký tự thứ nhất, lúc đó ta không phát hiện được cột bị sai, kết quả thu được bị sai nhưng ta không phát hiện được.

4. Mã tuyến tính

Gọi là mã khối tuyến tính vì bắt nguồn từ chỗ các tin của nguồn được phân thành từng khối tách bạch, và sự mã hóa hay giải mã được tiến hành theo từng khối, mỗi khối được mã hóa bằng một từ mã riêng rẽ. Các từ mã có độ dài bằng nhau cho nên mã này thuộc loại mã đồng đều. Đứng trên quan điểm toán học mà xét người ta gọi mã này là mã tuyến tính vì chúng thuộc phạm vi đại số tuyến tính cho nên cách biểu diễn thuận tiện và gọn, nhất là phương pháp biểu diễn ma trận.

4.1. Định nghĩa

Ta có véctơ thông tin i của k bit

$$i = (i_1, i_2, ..., i_k) \qquad i_j \in \{0,1\}$$

Tạo ra từ mã có độ dài n

$$c = (i_1\, i_1, i_2, ..., i_k, C_{k+1}, ..., C_n) \qquad C_j \in \{0,1\}$$

Thông qua cách chuyển đổi tuyến tính

$$C = i.G \qquad (4\text{-}4)$$

$$\text{với } G = [I_k.P] \qquad (4\text{-}5)$$

Trong đó: I_k là ma trận nhận dạng (đơn vị) có kích thước k×k,

P là ma trận k dòng và (n-k) cột.

Tập 2^k từ mã c được gọi là mã tuyến tính. Ma trận G gọi là ma trận tạo mã (ma trận sinh).

Chú ý rằng tổng của hai từ mã bất kỳ c' và c'' cũng là một từ mã

$$c' + c'' = (i' + i'', ..., c_n' + c_n'') \qquad (4\text{-}6)$$

$$c', c'' \text{ và } c'+c'' \in C$$

Theo các công thức trên bit kiểm tra thứ j của C_{k+j} bằng tích vô hướng của véctơ j với véctơ dòng thứ j của P, như vậy, có thể nói rằng C là một từ mã khi và chỉ khi:

$$C.H = 0 \qquad (4\text{-}7)$$

$$H = [P^T.I_{n-k}] \qquad (4\text{-}8)$$

P^T là ma trận chuyển vị của ma trận T và H được gọi là ma trận kiểm tra chẵn lẻ của mã.

4.2. Tính chất

4.2.1. Hội chứng (Syndrome)

Xét một mã tuyến tính C(n,k) với ma trận sinh G và ma trận kiểm tra H. Gọi $c = (i_1\ i_1, i_2, ..., i_k, C_{k+1}, ..., C_n)$ là từ mã được truyền đi qua kênh có đặc tính nhiễu. Gọi x là véctơ có độ dài n nhận được từ kênh truyền:

$$x = (x_1, x_2,...,x_n)$$

x có thể khác c do có nhiễu và:

$$x = c + e \qquad (4\text{-}9)$$

với $e = (e_1, e_2, ...,e_n)$ gọi là véc tơ sai số.

Dĩ nhiên là phía thu không biết được c và e khi nhận được r. Bộ giải mã phải xác định r có chứa lỗi truyền hay không. Nếu lúc đó lỗi được phát hiện thì bộ giải mã sẽ thực hiện định vị lỗi và sửa lỗi hoặc yêu cầu truyền lại.

Khi nhận được r bộ giải mã sẽ tính bộ (n-k) thành phần.

$$S(x) = x.H^T \qquad (4\text{-}10)$$

S(x) được gọi là hội chứng (Syndrome) của x.

Ta có thể thiết lập tính chất sau:

Tính chất 1: *Một véc tơ x là một từ mã khi và chỉ khi nó hội chứng với 0.*

S(x) = 0 khi và chỉ khi x là từ mã, S(x) # 0 nghĩa là x không là từ mã. Có thể có vài véctơ lỗi không bị phát hiện, điều đó chứng tỏ x là có lỗi nhưng $S(x) = x.H^T = 0$ (xảy ra khi véctơ lỗi e giống với một từ mã khác 0) và x là tổng hai từ mã và cũng là một từ mã do đó không phát hiện được sai.

4.2.2. Khoảng cách tối thiểu giữa những từ mã

Khoảng cách tối thiểu giữa những từ mã là một thông số phụ thuộc vào dung lượng đúng (correction) đúng của mã. Các tính chất sau cho phép ta nhận biết được khoảng cách tối thiểu đó.

Tính chất 2: *Khoảng cách tối thiểu giữa những từ mã của một mã tuyến tính bằng trọng lượng Hamming tối thiểu của nó.*

Ở đây trọng lượng Hamming của một từ mã được định nghĩa là số thành phần khác 0 của từ mã. Trọng lượng Hamming tối thiểu của bộ mã là trọng lượng Hamming nhỏ nhất trong tập trọng lượng Hamming của các từ mã trong bộ mã.

Tính chất 3: *Trong một mã tuyến tính, khoảng cách Hamming tối thiểu là m tồn tại ít nhất một tập m dòng của ma trận H mà tổng bằng 0. Nói cách khác, mỗi tập con của ít nhất m dòng của H có tổng bằng 0.*

Từ đó, nếu c là một từ mã, trọng lượng Hamming tối thiểu là m thì $c.H^T = 0$, m dòng của H cũng có tổng bằng 0.

4.2.3. Sửa và tìm sai, bảng tiêu chuẩn và dung lượng sửa sai của mã

Giả thiết c là từ mã truyền trên kênh trực tiếp.

Khi nhận được x, nhiệm vụ của bộ giải mã là đánh giá e với khả năng chính xác lớn nhất và trả lại bộ phận thu véc tơ c = x + e.

Để xác định được e, bộ giải mã sắp xếp véctơ nhận được x và ma trận H cho phép, ta tính ra hội chứng S(x) của x.

Nếu S(x) = 0, x là từ mã và cũng là véc tơ sai số. Thông thường khả năng sai P trên các bit là nhỏ, sai số là độc lập và kênh truyền đối xứng nhị phân và vì vậy bộ giải mã quyết định e = 0 và hoàn trả bộ thu giá trị x.

Nếu S(x) # 0, thật sự có lỗi. Dựa vào chiến lược chấp nhận người ta có thể cảnh báo hoặc cho truyền lại hoặc có thể sửa sai trực tiếp. Trong trường hợp sửa sai trực tiếp người ta chọn cho véctơ sai một hội chứng S(x) với trọng lượng bé nhất.

Dung lượng sửa sai của một mã có thể quyết định bởi tính chất sau:

Tính chất 4: *Trên một kênh nhị phân đối xứng và sai số độc lập, một mã tuyến tính C(n,k) cho phép phát hiện và sửa sai với trọng lượng ≤ e khi và chỉ khi trọng lượng Hamming của tất cả các từ mã là tốt nhất với 1+.*

5. Mã vòng

5.1. Định nghĩa

Một mã tuyến tính C (n,k) (với n là độ dài từ mã, k là số ký hiệu mã mang tin, r = n-k là số ký tự điều khiển kiểm soát lỗi) nếu sự hoán vị vòng của một từ mã cũng là một từ mã khác. Nghĩa là dịch vòng (sang phải hay trái) của một từ mã thì kết quả cũng là một từ mã. Ở đây quy ước dịch phải

$c = (c_{n-1}, c_{n-2}, ..., c_0) \in C \Leftrightarrow (c_{n-2}, c_{n-3}, ..., c_0, c_{n-1}) \in C$

Ta có thể viết từ mã dưới dạng đa thức:

$$c(x) = c_{n-1}x^{n-1} + c_{n-2}x^{n-2} + ...+c_1x+c_0 \qquad (4\text{-}11)$$

Ví dụ: Bảng sau đây trình bày một mã vòng C (7,4)

Tin	Từ mã	Đa thức	Tin	Từ mã	Đa thức
0000	0000000	0	0001	0001101	$x^3 + x^4 + x^6$
1000	1101000	$1 + x^2 + x^3$	1001	1100101	$1+ x + x^4 + x^6$
0100	0110100	$x + x^2 + x^4$	0101	0111001	$x+ x^2 + x^3 + x^6$
1100	1011100	$1+x^2 + x^3 + x^4$	1101	1010001	$1 + x^2 + x^6$
0010	0011010	$x^2 + x^3 + x^5$	0011	0010111	$x^2 + x^4 + x^5 + x^6$
1010	1110010	$1+ x + x^2 + x^5$	1011	1111111	$1+ x+ x^2 +x^3 + x^4 + x^5 + x^6$
0110	0101110	$x+ x^3 + x^4 + x^5$	0111	0100011	$x + x^5 + x^6$
1110	1000110	$1+ x^4 + x^5$	1111	1001011	$1+ x^3 + x^5 + x^6$

5.2. Đa thức sinh, ma trận sinh và ma trận kiểm tra của mã vòng

Một đặc điểm quan trọng của mã vòng là những từ mã đều chia chẵn cho g(x) gọi là đa thức sinh (đa thức tạo mã). Sở dĩ gọi như vậy là vì từ đa thức này có thể tạo ra các từ mã của mã vòng xuất phát từ những đa thức mang tin đơn giản.

Một khối dữ liệu k phần tử $(u_0, u_1,...,u_{k-1})$ có thể xem như một đa thức thông tin bậc k.

$$u(x) = u_{k-1}x^{k-1} + u_{k-2}x^{k-2} +...+ u_1x + u_0 \qquad (4\text{-}12)$$

Bậc của mã vòng C(n,k) là n thì bậc của đa thức sinh g(x) là r = n-k

$$g(x) = g_r x^r + g_{r-1}x^{r-1} +.....+g_1x + g_0 \qquad (4\text{-}13)$$

Nếu đa thức sinh được biểu diễn dưới dạng tổ hợp $g_o, g_1, .., g_r$ thì mã vòng có thể biểu diễn bằng một ma trận sinh có dạng:

$$G = \begin{vmatrix} g_0 & g_1 & g_2 \ldots\ldots g_r \ldots\ldots 0 & 0 \ldots\ldots & 0 \\ 0 & g_0 & g_1 \ldots\ldots \ldots & g_r\ 0 \ldots\ldots & 0 \\ 0 & 0 & g_0 \ldots\ldots \ldots\ldots \ldots & g_r \ldots\ldots & 0 \\ 0 & 0 & 0 \ldots\ldots g_0 \ldots\ldots & \ldots\ldots \ldots & \ldots g_r \end{vmatrix} \quad (4\text{-}14)$$

Hàng thứ nhất của ma trận sinh đại diện tổ hợp của đa thức sinh cộng k-1 số 0, như vậy số cột của ma trận bằng n = r+k. Các hàng còn lại được thiết lập bằng cách lặp lại hàng ngay trên nó và dịch đi một vị trí. Ma trận có k hàng đại diện cho k tổ hợp mã, những tổ hợp còn lại (2^k - k - 1) sẽ được xác định bằng cách lấy những tổ hợp tuyến tính bất kỳ của những hàng trong số k hàng của ma trận.

Bằng những phép biến đổi tuyến tính ví dụ như tổng các tổ hợp khác nhau của các hàng của ma trận sinh để có thể chuyển ma trận dưới dạng chuẩn tắc:

$$G^*_{n,k} = \begin{vmatrix} p_{11} & p_{12} \ldots\ldots\ldots p_{1r} & 1\ 0 \ldots\ldots\ldots\ldots 0 \\ p_{21} & p_{22} \ldots\ldots\ldots p_{2r} & 0\ 1 \ldots\ldots\ldots\ldots 0 \\ \ldots\ldots\ldots\ldots\ldots & & \\ p_{k1} & p_{k2} \ldots\ldots\ldots p_{kr} & 0\ 0 \ldots\ldots\ldots\ldots 1 \\ \underbrace{\qquad\qquad\qquad\qquad}_{r} & & \underbrace{\qquad\qquad\qquad\qquad}_{k} \end{vmatrix} \quad (4\text{-}15)$$

Từ cách biểu diễn ma trận sinh cho thấy một hàng bất kỳ của ma trận đều là bản thân đa thức sinh thêm một vài số không cho nên một từ mã bất kỳ là tổ hợp tuyến tính của một số hàng ma trận đều mang tính chất chia chẵn cho đa thức sinh. Cách biểu diễn của một ma trận sinh dưới dạng chuẩn tắc cho thấy: một từ mã (một hàng của ma trận hoặc tổng một số hàng của ma trận) gồm hai phần: một phần có r ký hiệu thử p_{ij} (i = 1,2,..k; j = 1,2..r) và phần còn lại có k ký hiệu mang tin. Để xác định đa thức sinh, chúng ta dựa trên tính chất:

Đa thức sinh g(x) của mã vòng C(n,k) là một thừa số của x^n+1

Cũng từ những tính chất trên ta có thể xây dựng ma trận sinh dạng chuẩn tắc bằng cách chia x^{n-1} cho đa thức sinh g(x), các số dư trong phép chia sẽ thành ma trận các số dư dạng $A_{(r,k)}$ (có r cột, k hàng), hợp với ma trận $\overline{I_k}$ có

kích thước k × k (là ma trận của cả các phần tử trên đường chéo chính từ phải sang trái bằng 0, còn tất cả các phần tử còn lại đều bằng 0).

$$\overline{I_K} = \begin{vmatrix} 0\ 0\ \ldots\ldots\ldots\ldots 1 \\ 0\ 0\ldots\ldots\ldots\ldots 1\ \ 0 \\ \ldots\ldots\ldots\ldots \\ 1\ \ 0\ldots\ldots\ldots\ldots 0\ \ 0 \end{vmatrix} \qquad (4\text{-}16)$$

Tạo thành ma trận sinh:

$$G^*_{n,k} = \left| \overline{I_k}\ A_{(r,k)} \right| \qquad (4\text{-}17)$$

Phương pháp thực hiện sửa sai có thể thực hiện dựa trên ma trận phát hiện và sửa sai M được xây dựng như sau:

$$M = \begin{vmatrix} \overline{I}_n & \overline{I}_{n-k} \\ & A_{(r,k)} \end{vmatrix} \qquad (4\text{-}18)$$

Ma trận $\overline{I}_n$ đại diện tất cả các véctơ sai của từ mã.

5.3. Mã hóa mã vòng

Mã hóa mã vòng C (n,k) gồm 3 bước:

Bước 1: Nhân đa thức thông tin u (x) với x^{n-k} .

Bước 2: Chia $u(x).\ x^{n-k}$ cho đa thức sinh g(x) ta được phần dư b(x), đây là các bit kiểm tra còn gọi là các bít CRC.

Bước 3: Hình thành từ mã $c(x) = b(x) + u(x).\ x^{n-k}$.

Ví dụ 1:

Mã hoá bản tin u(x) = 1010 bằng bộ mã vòng C(7,4).

Trước tiên chúng ta xác định đa thức sinh g(x) và ma trận phát hiện và sửa sai M của bộ mã vòng này:

Do $x^7 + 1 = (x+1)(x^3 + x^2+1)(x^3 + x+1)$ đa thức này có hai thừa số có bậc n - k = 3, mỗi thừa số sẽ là một đa thức sinh và tạo ra mã vòng C(7,4). Giả sử ta thống nhất chọn đa thức sinh $g(x) = x^3 + x+1$.

Ma trận dư A(r,k) là các số dư của phép chia x^{n-1} cho g(x).

$$A(r,k) = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{vmatrix}$$

Ma trận sửa sai của mã vòng C(7,4) với đa thức sinh $g(x) = x^3 + x + 1$ là:

$$M = \left|\begin{array}{ccccccc:ccc} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{array}\right| \quad (4\text{-}19)$$

Ta tiến hành mã hoá như sau:

Bước 1: Nhân đa thức thông tin với x^{n-k}.

$$u(x) = x^3 + x$$

$$u(x).\ x^{n-k} = x^3(x^3 + x) = x^6 + x^4$$

Bước 2: Chia $u(x).\ x^{n-k}$ cho đa thức sinh g(x)

$$\begin{array}{l|l} x^6 + x^4 & x^3 + x + 1 \\ \hline x^6 + x^4 + x^3 & x^3 + 1 \\ \hline x^3 & \\ \hline x^3 + x + 1 & \\ x + 1 & \end{array}$$

Vậy $b(x) = x + 1 \Leftrightarrow (011)$ đây là các bít kiểm soát lỗi.

Bước 3: Hình thành từ mã $c(x) = b(x) + u(x).\ x^{n-k}$

$$c(x) = x^6 + x^4 + x + 1 \Leftrightarrow 1010011$$

Điều này chứng tỏ từ mã được tạo thành từ k bit tin cần mã hóa ghép với r = n-k bít kiểm soát lỗi.

Mạch tạo CRC:

Để tạo mã CRC có thể dùng phần mềm để tính CRC cho từng gói dữ liệu, hoặc tính toán sẵn lưu vào bảng giá trị CRC cho 256 byte sau đó khi tính CRC cho từng byte thì tra bảng. Tuy nhiên trong thực tế để nhanh và giảm thời gian hoạt động của bộ vi xử lý người ta thường dùng phần cứng để tạo CRC và kiểm tra. Mạch điện sẽ bao gồm các bộ ghi dịch và các bộ cộng modul 2, số lượng cột của bộ ghi dịch phụ thuộc vào giá trị C đã chọn cho g (x) và cũng là số bít cho mã CRC.

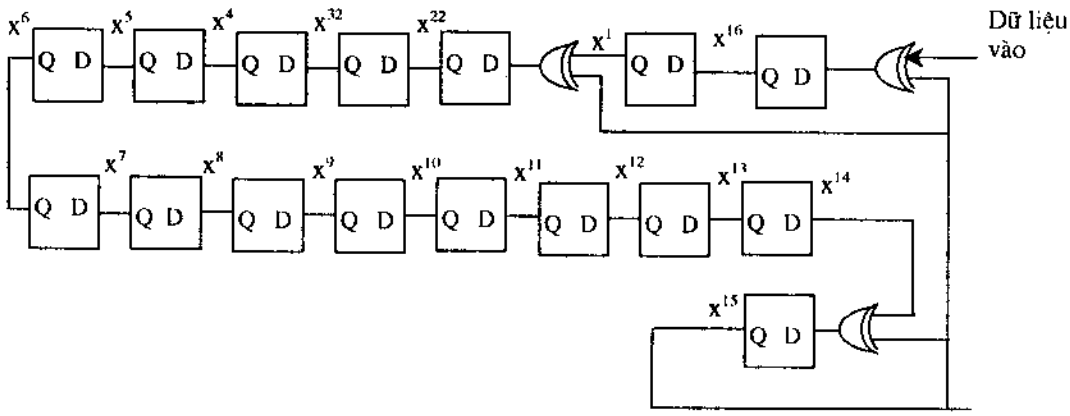
Người ta có thể tạo mã CRC dài 12 bit, 16 bit, 32 bit... Vấn đề là chọn g(x) sao cho phù hợp. Hiện nay có 3 đa thức sinh được coi là chuẩn quốc tế đó là:

CRC- 16: $g(x) = x^{16} + x^{15} + x^2 + 1$

CRC- CCITT: $g(x) = x^{16} + x^{12} + x^2 + 1$

CRC-12 $g(x) = x^{12} + x^{11} + x^3 + x^2 + x + 1$

Mạch tạo CRC-16 cho 16 bit như hình vẽ sau:



Hình 4-4: Mạch tạo CRC với $g(x) = x^{16} + x^{15} + x^2 + 1$
(không hiển thị CLK +RST)

5. 4. Giải mã mã vòng

Bước 1: Đem tổ hợp mã nhận được chia cho đa thức sinh. Nếu phép chia chẵn chứng tỏ tổ hợp mã nhận được chính là từ mã đã phát đi. Nếu phép chia có dư chứng tỏ tổ hợp mã nhận được có ký hiệu sai.

Bước 2: Lấy số dư này đối chiếu với ma trận phát hiện và sửa sai để biết vectơ sai.

Bước 3: Để sửa sai đem tổ hợp mã nhận được cộng (modul 2) với véc tơ sai để tìm từ mã đã phát đi.

Bước 4: Loại bỏ các r bít kiểm soát lỗi đã ghép vào cuối từ mã phát đi để thu được tin ban đầu.

Ví dụ: Với bộ mã vòng C(7,4) sử dụng đa thức sinh $g(x) = x^3 + x+1$. Giả sử nhận được tổ hợp mã là 1000011. Hãy tìm tin tương ứng đã phát.

Việc giải mã được tiến hành như sau:

Bước 1: Lấy tổ hợp mã nhận được chia cho đa thức sinh g(x).

$$p(x) = 1000011 = x^6 + x + 1$$

$x^6 + x + 1$	$x^3 + x+1$
$x^6 + x^4 + x^3$	$x^3 + 1$
$x^4 + x^3 + x + 1$	
$x^4 + x^2 + x$	
$x^3 + x^2 + 1$	
$x^3 + x + 1$	
$x^2 + x$	

Vậy số dư = $x^2 + x \Leftrightarrow (110)$ chứng tỏ tổ hợp mã nhận được có ký hiệu sai.

Bước 2: Lấy số dư đối chiếu với ma trận phát hiện và sửa sai M của bộ mã C(7,4) sử dụng đa thức sinh $g(x) = x^3 + x+1$ ở (4-19) ta được véctơ sai

$$e = 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0$$

Bước 3: Đem tổ hợp mã nhận được cộng modul 2 với véctơ sai để tìm từ mã đã phát.

e	=	0	0	1	0	0	0	0
p (x)	=	1	0	0	0	0	1	1
c (x)	=	1	0	1	0	0	1	1

Bước 4: Loại bỏ 3 bít kiểm soát lỗi ở cuối từ mã để được tin đã phát

u(x) = 1010

6. Mã chống nhiễu Hamming

6.1. Khái niệm mã Hamming

Mã Hamming là một loại mã tuyến tính, mã này được R.W. Hamming đưa ra và được dùng trong hệ thống thông tin có sử dụng FEC. Mã này có khả năng sửa sai một lỗi. Mã có đặc điểm là sơ đồ tạo mã và giải mã đơn giản.

Các từ mã của mã Hamming có độ dài n = k + r. Số bít kiểm tra r và số bít mang tin k của mã phải thoả mãn hệ thức

$$k \le 2^r - 1 - r. \qquad (4\text{-}21)$$

Khi số bít mang tin tăng thì số bít kiểm tra cũng tăng nhưng tốc độ tăng của số bít mang tăng nhanh hơn nhiều so với tốc độ tăng của số bít kiểm tra cho nên khi số bít mang tin lớn tính kinh tế của mã càng cao vì vậy mã Hamming được sử dụng rộng rãi trong thực tế.

Việc tạo và giải mã Hamming dựa vào ma trận kiểm tra H(r,n). Ma trận này có tính chất quan trọng sau:

Ma trận gồm r hàng và n cột, với các cột là các số nhị phân r bit có giá trị tăng từ 1 đến n(giá trị của cột h_i bằng i).

Nếu một véc tơ v = (a_1, a_2,..,a_n) là một từ mã của mã Hamming thì với bất kỳ hàng nào của H ta cũng có:

$$\sum_{j=1}^{n} a_j H_{ij} = 0 \ (*) \qquad (4\text{-}22)$$

Ví dụ : Bộ mã Hamming C(7,4) có 2^4 =16 từ mã, mỗi từ mã có độ dài 7 bít, có 4 bít mang tin và 6 bít kiểm tra. Ma trận kiểm tra của mã là H (3,7)

$$H(3,7) = \begin{vmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{vmatrix} \qquad (4\text{-}23)$$

Các từ mã của bộ mã là c(x) = (0000000, 1101001, 0101010, 1000011, 1001100, 0100101, 1100110, 0001111, 1110000, 0011001, 1011010, 0110011, 0111100, 1010101, 0010110, 1111111) từ mã nào cũng thoả mãn điều kiện (*) chẳng hạn từ mã 1100110 luôn có:

$$\begin{cases} 1\otimes 0 \oplus 1\otimes 0 \oplus 0\otimes 0 \oplus 0\otimes 1 \oplus 1\otimes 1 \oplus 1\otimes 1 \oplus 0\otimes 1 = 0 \\ 1\otimes 0 \oplus 1\otimes 1 \oplus 0\otimes 1 \oplus 0\otimes 0 \oplus 1\otimes 0 \oplus 1\otimes 1 \oplus 0\otimes 1 = 0 \\ 1\otimes 1 \oplus 1\otimes 0 \oplus 0\otimes 1 \oplus 0\otimes 0 \oplus 1\otimes 1 \oplus 1\otimes 0 \oplus 0\otimes 1 = 0 \end{cases}$$

Các phép nhân(⊗) và cộng(⊕) ở đây là nhân và cộng modul 2.

6.2. Tạo mã

Để tạo mã trước hết cần xác định ma trận kiểm tra của bộ mã sau đó xác định vị trí và giá trị các bit kiểm tra trong mỗi từ mã sau đó tiến hành xác định bit từ mã. Vị trí các bit kiểm tra thêm vào chính là các bit có thứ tự thứ 2^{i-1} trong từ mã với $0 \le i \le r-1$. Giá trị của các bít kiểm tra dựa vào các phương trình.

$$\sum_{j=1}^{n} a_j H_{ij} = 0 \ (*)$$

Ví dụ: mã Hamming C(7,4) có ma trận kiểm tra:

$$H(3,7) = \begin{vmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{vmatrix}$$

Các bit kiểm tra ở các vị trí $2^0 = 1$, $2^1 = 2$, $2^2 = 4$

Một từ mã của bộ mã trên có dạng (x_1, x_2, u_1, x_3, u_2, u_3, u_4), trong đó u_i là các bit mang tin, x_i là các bit kiểm tra sai thêm vào.Giá trị các bít kiểm tra tính dựa vào phương trình (*). Chẳng hạn 1001 được mã hoá thành $x_1x_21x_3001$, trong đó:

$$x_3 \oplus 1 = 0 \rightarrow x_3 = 1$$
$$x_2 \oplus 1 \oplus 1 = 0 \rightarrow x_2 = 0$$
$$x_1 \oplus 1 \oplus 1 = 0 \rightarrow x_1 = 0$$

Vậy từ mã để mã hoá 1001 là 0011001

6.3. Giải mã

Khi nhận được một từ mã, ta lần lượt lấy từ mã nhân tương ứng với các véc tơ hàng của H. Nếu kết quả s là 000 thì không có sai khi truyền, ngược lại thì

có sai, so sánh giá trị s với từng cột của H nếu trùng với cột j nào thì sai ở vị trí thứ j. Do cách thiết lập ma trận kiểm tra H của mã Hamming khá đặc biệt nên chỉ cần đổi s ra giá trị thập phân tương ứng thì giá trị thập phân này chính là vị trí bit sai.

Ví dụ: khi nhận được dãy bit 0011011 tiến hành giải mã bằng cách nhân tương ứng với 3 hàng của ma trận H ở (4-23).

$$1 \oplus 1 \oplus 1 = 1$$
$$1 \oplus 1 \oplus 1 = 1$$
$$1 \oplus 1 \quad = 0$$

110 ứng với hàng thứ 6 của ma trận kiểm tra (hoặc đổi ra số thập phân cũng là 6), vậy bit bị sai là bit thứ 6, sửa sai bit thứ 6 từ 1→0. Từ mã đúng là 0011001, các bit mang tin là 1001.

Bài tập

Bài 1: Cho nguồn tin rời rạc X_8 có các lớp tin và xác suất xuất hiện tương ứng với chúng trong bảng sau:

Lớp tin	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$p(x_i)$	0,2	0,2	0,15	0,15	0,1	0,1	0,05	0,05

a/ Hãy lập mã Shannon - Fano cho nguồn tin trên.

b/ Tính các thông số tối ưu của bộ mã

c/ Biểu diễn bộ mã bằng đồ hình kết cấu

Bài 2: Cho nguồn tin rời rạc X_8 có các lớp tin và xác suất xuất hiện tương ứng với chúng trong bảng sau:

Lớp tin	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$p(x_i)$	0,25	0,2	0,12	0,13	0,1	0,1	0,05	0,05

a/ Hãy lập mã Huffman cho nguồn tin trên.

b/ Tính các thông số tối ưu của bộ mã.

c/ Thử tính prefix của bộ mã khi nhận được một chuỗi bit.

Bài 3: Cho nguồn tin rời rạc X_7 có các lớp tin và xác suất xuất hiện tương ứng với chúng trong bảng sau:

Lớp tin	x_1	x_2	x_3	x_4	x_5	x_6	x_7
$p(x_i)$	0,01	0,34	0,23	0,1	0,19	0,07	0,06

a/ Hãy lập mã Huffman cho nguồn tin trên.

b/ Tính các thông số tối ưu của bộ mã.

c/ Biểu diễn bộ mã bằng cây mã.

Bài 4: Hãy mã hoá chống nhiễu bằng mã vòng C(7,4) sử dụng đa thức sinh g(x)= 1101 các tin sau:

a/ $u_1 = 1011$

b/ $u_2 = 1000$

c/ $u_3 = 1000$

Bài 5: Hãy giải mã bằng mã vòng C(7,4) sử dụng đa thức sinh g(x) = 1011 các tổ hợp ký hiệu nhận được sau:

a/ $p_1 = 1000010$

b/ $p_2 = 1001100$

Bài 6: Với bộ mã Hamming C(7,4). Hãy mã hoá các tin

a/ $u_1 = 1000$

b/ $u_2 = 0100$

c/ $u_3 = 0110$

Bài 7: Với bộ mã Hamming C(7,4). Hãy giải mã thành tin tương ứng khi nhận được các tổ hợp mã sau:

a/ $p_1 = 11111100$

b/ $p_2 = 1011011$

Bài 8: Cho nguồn tin rời rạc X_8 có các lớp tin và xác suất xuất hiện tương ứng với chúng trong bảng sau:

Lớp tin	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$p(x_i)$	0,2	0,18	0,15	0,14	0,12	0,09	0,08	0,04

1. Nếu mã hóa nguồn tin trên bằng phương pháp Shannon- Fano thì tin x_5 được mã hóa thành :

a. 011 b. 010 c. 101 d. 110 e. 111

2. Nếu mã hóa nguồn tin trên bằng phương pháp Huffman thì tin x_7 được mã hóa thành:

a.1001 b. 1101 c. 1011 d. 1111 e. 1000

Bài 9: Cho mã vòng C(8,5) có đa thức sinh $g(x) = 1 + x + x^2 + x^3$

1. Tin u_1 = 10100 sẽ được mã hóa thành từ mã

   a. 01100110 b. 00110011 c. 100110011 d. 11101110 e.11001100

2. Tổ hợp nào sau đây không phải là một từ mã của mã vòng trên

   a. 10100100 b. 10110100 c. 01011010 d. 10001000 e.10100101

Bài 10: Đa thức nào sau đây có thể làm đa thức sinh cho mã vòng C(9,6)

a/ $x^3 + 1$

b/ $x^3 + x^2 + 1$

c/ $x^3 + x + 1$

d/ $x^3 + x^2 + x + 1$

e/ Cả bốn câu trên đều sai

Chương 5

ĐIỀU CHẾ TÍN HIỆU

Mục tiêu

- Giải thích được khái niệm và phân loại được các kỹ thuật điều chế.
- Nêu được nguyên lý các kỹ thuật điều chế thông dụng.
- Nhận biết các quá trình điều chế và giải điều chế thông dụng.

I. KHÁI NIỆM VỀ ĐIỀU CHẾ

1. Định nghĩa

Phần lớn các tín hiệu mang thông tin như hình ảnh, tiếng nói, ... đều là các tín hiệu tần số thấp và có phổ giới hạn từ f_m đến f_M với $f_m < f_M$.

Ví dụ:

- Tín hiệu tiếng nói có phổ từ 300 Hz đến 3400 Hz
- Tín hiệu truyền hình có phổ từ 25 Hz đến 6,5 MHz

Do đó ta không thể truyền trực tiếp các tín hiệu này đi xa bằng sóng điện từ vì chúng không thể bức xạ trực tiếp được bằng Ăngten. Trong khi đó các tín hiệu radio tần số cao lại có khả năng bức xạ tốt. Vì vậy cần phải dịch chuyển phổ của tín hiệu lên miền tần số cao. Quá trình này gọi là điều chế (Modulation).

Ngoài ra để có thể tận dụng mạng điện thoại công cộng (Public Switched Telephone Network-PSTN) cho việc truyền số liệu giữa các máy tính ở những khoảng cách xa (ví dụ như liên lạc trong mạng Internet) thì cần phải chuyển đổi các tín hiệu điện đi ra từ các máy tính cho phù hợp với đường truyền PSTN. Bởi vì đường truyền PSTN được thiết kế cho thông tin tiếng nói có băng thông trong khoảng 300 - 3400Hz, nếu các tín hiệu dạng số tần số cao được truyền

trực tiếp trên đường dây thoại thì chúng sẽ bị suy giảm và biến dạng, khi đến đầu thu sẽ không nhận ra được. Quá trình chuyển đổi này cũng được gọi là điều chế.

Quá trình số hóa tín hiệu tương tự, quá trình chuyển từ tín hiệu điện sang tín hiệu quang... đều là các quá trình điều chế, chúng có nhiệm vụ là chuyển tin tức trong một dạng năng lượng thích hợp với môi trường truyền lan, trong đó dạng năng lượng được dùng ít bị tổn hao, biến dạng do tác động của nhiễu và có độ phân biệt rõ ràng để quá trình giải điều chế có thể dễ dàng nhận dạng dù có bị tạp nhiễu làm biến dạng phần nào.

Về mặt toán học để mô tả quá trình điều chế, ta sử dụng hai hàm:

a/ Hàm tin s(t)

b/ Hàm tải tin u(t)

Định nghĩa: *Điều chế là quá trình cho hàm tin s(t) tác động lên tải tin u(t) sao cho nhận được một tín hiệu có phổ nằm ở miền tần số phù hợp với kênh truyền.*

Sau điều chế ta nhận được tín hiệu x(t), x(t) được gọi là tín hiệu điều chế, s(t) được gọi là tín hiệu bị điều chế.

2. Phân loại

Tuỳ theo tính chất của tải tin và tín hiệu mang tin ta có các loại điều chế chính như sau:

Điều chế cao tần các tín hiệu liên tục:

Tín hiệu mang tin s(t) là tín hiệu tương tự, tần số thấp. Tải tin u(t) là một tín hiệu điều hòa tần số cao:

$u(t) = A_0\cos(\omega_0 + \varphi_0) = A_0\cos(2\pi f_0 + \varphi_0)$, với f_0 đủ lớn (5-1)

Ta có thể chia điều chế cao tần làm 3 loại chính:

a/ Cho s(t) tác động vào biên độ của tải tin: ***điều biên*** (AM - Amplitude Modulation): $A(t) = A_0 + \Delta A.s(t)$ (5-2)

b/ Cho s(t) tác động vào tần số của tải tin: ***điều tần*** (FM- Frequency Modulation) : $\omega(t) = \omega_0 + \Delta\omega.s(t)$ (5-3)

c/ Cho s(t) tác động vào pha của tải tin: ***điều pha*** (PM- Phase Modulation:

$$\varphi(t) = \varphi_0 + \Delta\varphi \, .s(t) \qquad (5\text{-}4)$$

Điều tần và điều pha được gọi là điều chế góc vì bản chất của chúng đều làm thay đổi góc pha của tải tin.

Điều chế xung:

Tín hiệu mang tin s(t) là tín hiệu tương tự. Tải tin u(t) là một dãy xung tuần hoàn:

$$u(t) = \sum_{k=-\infty}^{\infty} A \prod_{t}(t - kT_s - t_0) \qquad (5\text{-}5)$$

Cho s(t) tác động vào các tham số của tải tin, ta có các loại điều chế xung sau:

a/ Cho s(t) tác động vào biên độ của tải tin: ***điều biên xung*** (PAM - Pulse Amplitude Modulation)

b/ Cho s(t) tác động vào tần số của tải tin: ***điều tần xung*** (PFM - Pulse Frequency Modulation).

c/ Cho s(t) tác động vào pha của tải tin: ***điều pha xung*** (PPM - Pulse Phase Modulation

d/ Cho s(t) tác động vào độ rộng của tải tin: ***điều rộng xung*** (PWM - Pulse Width Modulation).

e/ Khi lượng tử hóa và mã hóa tín hiệu điều biên xung, ta nhận được tín hiệu ***điều xung mã*** (PCM - Pulse Code Modulation).

Các phương pháp khoá dịch

Tín hiệu mang tin s(t) là dãy bít nhị phân, tải tin u(t) là một dao động điều hoà:

$$u(t) = A_0\cos(\omega_0 + \varphi_0)$$

Cho s(t) tác động vào các tham số biên độ, tần số, pha của tải tin, ta có các phương pháp điều chế khóa dịch sau:

a/ Phương pháp điều chế khoá dịch biên độ (ASK- Amplitude Shift Key)

b/ Phương pháp khóa dịch tần số (FSK - Frequency Shift Key)

c/ Phương pháp khóa dịch pha (PSK - Phase Shift Key)

d/ Kết hợp giữa ASK và PSK ta có điều chế biên độ cầu phương *(Quadrature Amplitude Modulation - QAM)*

II. ĐIỀU CHẾ CAO TẦN CÁC TÍN HIỆU LIÊN TỤC

1. Tín hiệu điều chế góc

1.1. Tín hiệu điều chế góc

Xét tín hiệu điều pha

Độ lệch pha tức thời tỷ lệ với s(t):

$$\varphi(t) = \varphi_0 + \Delta\varphi \, .s(t) \qquad (5\text{-}6)$$

Biểu thức của tín hiệu điều pha là:

$$x_{dp}(t) = A_0 \cos\theta(t) = A_0\cos[\, \omega_0 t + \Delta\varphi.s(t)+\varphi_0] \qquad (5\text{-}7)$$

Xét tín hiệu điều tần

Tần số tức thời của tín hiệu tỷ lệ với s(t):

$$\omega(t) = \omega_0 + \Delta\omega \, .s(t) \qquad (5\text{-}8)$$

Biểu thức của tín hiệu điều tần là:

$$x_{dt}(t) = A_0 \cos\theta(t) = A_0\cos[\int\omega(t)dt] = A_0\cos[\omega_0 t + \Delta\omega\int s(t)dt] \qquad (5\text{-}9)$$

Xét trường hợp riêng s(t) = cosΩ t, biểu thức của tín hiệu điều pha, điều tần trở thành:

$$x_{dp}(t) = A_0\cos[\, \omega_0 t + \Delta\varphi. \cos\Omega t + \varphi_0] \qquad (5\text{-}10)$$

$$x_{dt}(t) = A_0 \cos\left[\omega_0 t + \frac{\Delta\omega}{\Omega}\sin\Omega t + \varphi_0\right] \qquad (5\text{-}11)$$

Nếu đổi gốc thời gian thích hợp, ta có biểu thức chung của tín hiệu điều biên góc như sau:

$$x(t) = A_0 \cos(\omega_0 t + \beta.\sin\Omega t) \qquad (5\text{-}12)$$

trong đó β là chỉ tiêu điều biên góc, là độ lệch cực đại của pha so với giá trị trung bình.

Đối với tín hiệu điều pha, ta có:

$$\beta_{dp} = \Delta\varphi \qquad (5\text{-}13)$$

Đối với tín hiệu điều tần, ta có:

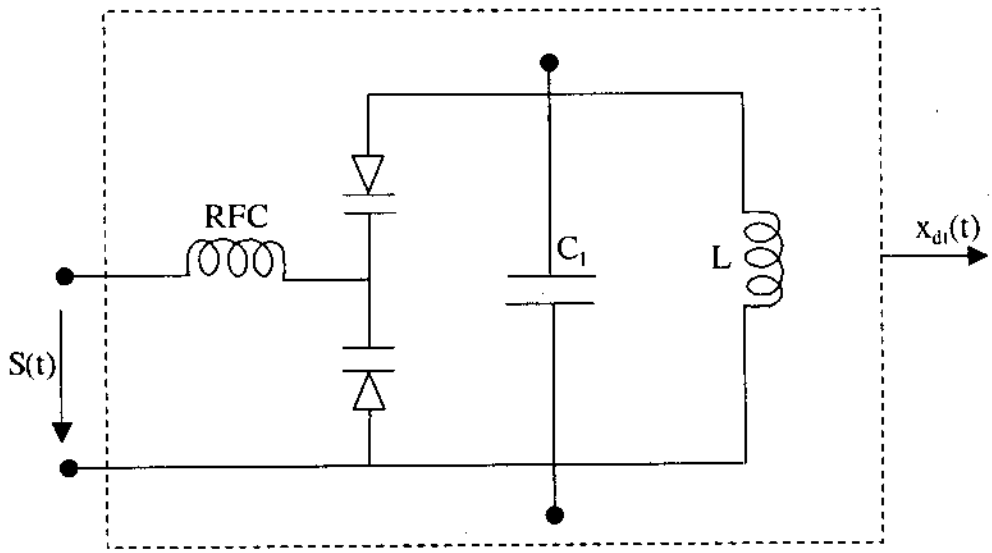
$$\beta dt = \frac{\Delta\omega}{\Omega} \qquad (5\text{-}14)$$

1.2. Thực hiện điều chế và giải điều chế góc

1.2.1. Điều tần

Ta cần phải làm thay đổi tần số tức thời của một bộ tạo dao động hình sin theo hàm tin s(t) bằng cách sử dụng một bộ điều khiển bằng điện áp (VCO):

Để thực hiện bộ tạo dao động này, người ta dùng bộ tạo dao động LC dùng điốt biến dung. Phần thụ động của mạch tạo dao động này được thể hiện trong Hình 5-1



Hình 5-1: Phần thụ động của mạch tạo dao động LC

Trong mạch điện Hình 5-1, ta có:

$$\begin{aligned} C(t) &= C_v(t) + C_1 = [\, C_{v0} - \Delta C_s(t)] + C_1 \\ &= (C_{v0} + C_1) - \Delta C_s(t) = C_0 - \Delta C_s(t) \end{aligned} \qquad (5\text{-}15)$$

Tần số dao động tức thời của mạch tạo dao động LC này là:

$$\omega(t) = \frac{1}{\sqrt{LC\,(t)}} = \frac{1}{\sqrt{L(C_0 - \Delta C.s(t)}} = \omega 0\left[1 - \frac{\Delta C}{c_0} s(t)\right]^{-\frac{1}{2}}$$

$$\omega(t) \approx \omega_0\left[1+\frac{1}{2}\frac{\Delta C}{C_0}s(t)\right] = \omega_0 + \frac{1}{2}\frac{\Delta C}{C0}\omega_0 s(t) = \omega_0 + \Delta\omega s(t) \qquad (5\text{-}16)$$

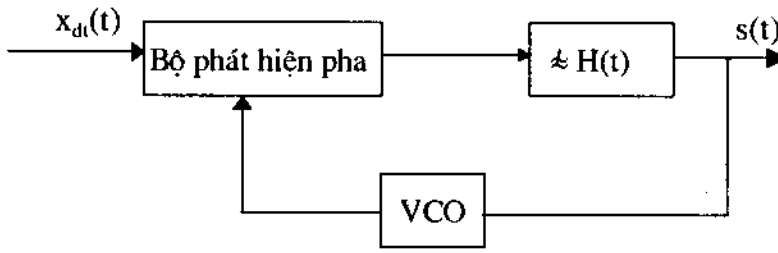
$$(do\ \frac{\Delta C}{C0} << 1)$$

Vậy ta nhận được một tín hiệu điều tần.

1.2.2. Giải điều tần

Có hai phương pháp giải điều tần đó là: giải điều tần bằng mạch tách hình bao và giải điều tần dùng vòng khoá pha.

Vòng khóa pha (PLL: Phase Locked Loop) là một hệ thống kín gồm 3 thành phần chính (Hình 5-2) gồm: một bộ phát hiện pha, một bộ lọc thông thấp và một bộ tạo dao động điều khiển bằng điện áp (VCO).

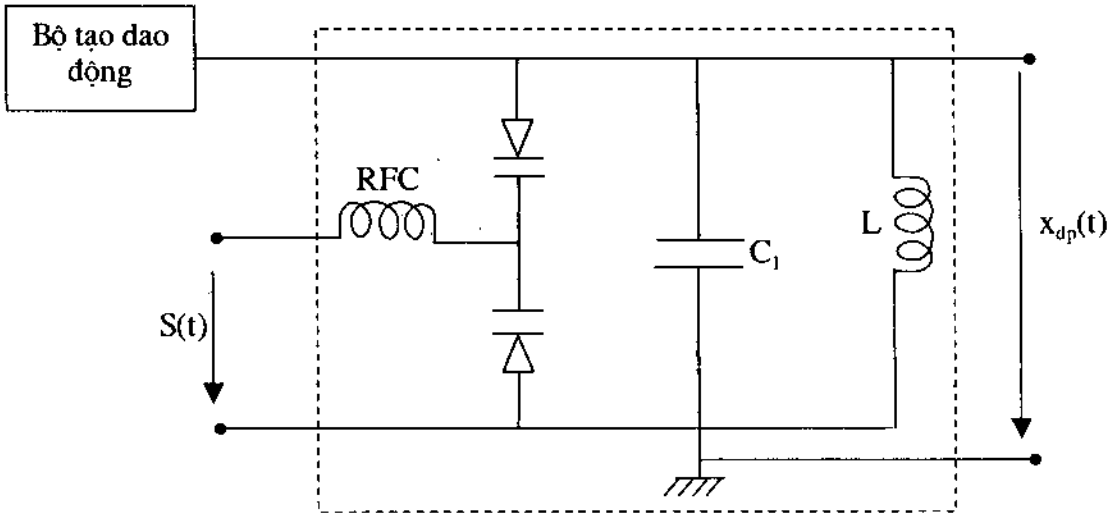


Hình 5-2: Giải điều tần dùng vòng khóa dịch pha

Khi cho một tín hiệu điều tần qua một mạch vòng khoá pha, khi mạch khóa thì tín hiệu ra của bộ lọc thông thấp tỷ lệ với tín hiệu mang tin s(t).

1.2.3. Điều pha

Cho tín hiệu điều hòa tần số ω_0 qua một mạch cộng hưởng có dịch pha thay đổi được theo s(t) như trên Hình 5-3, ta sẽ nhận được tín hiệu điều pha của s(t). Các thông số của mạch tạo dao động này được xác định tương tự như (5-15), (5-16)



Hình 5-3: Phần thụ động của mạch tạo dao động LC

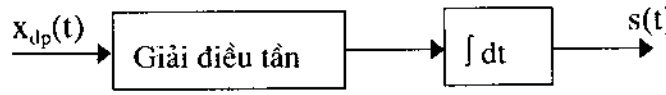
Khi C(t) thay đổi theo s(t) thì tần số cộng hưởng của mạch cũng sẽ thay đổi:

$$\omega_{ch}(t) \approx \omega_0 + \frac{1}{2}\frac{\Delta C}{C_0}\omega_0 s(t) = \omega_0 + \Delta\omega.s(t) \qquad (5\text{-}17)$$

Do đó pha của điện áp ra của mạch cộng hưởng LC sẽ thay đổi theo s(t) do đặc tuyến pha của mạch cộng hưởng xung quanh ω_0 xấp xỉ tuyến tính, vì lệch cộng hưởng nhỏ nên ta nhận được ở đầu ra của mạch tín hiệu điều pha theo s(t).

1.2.4. Giải điều pha

Ta có thể khôi phục lại được s(t) từ tín hiệu điều pha bằng cách đưa tín hiệu điều pha qua một bộ giải điều tần như trên rồi lấy tích phân tín hiệu nhận được như Hình 5 - 4:



Hình 5-4: Giải điều pha

Mạch tích phân thường được thực hiện xấp xỉ bằng bộ lọc thông thấp tần số cắt f_c có hàm truyền:

$$H(f) = \frac{1}{1 + j\frac{f}{f_c}} \qquad (5\text{-}18)$$

2. Điều biên

Tín hiệu mang tin s(t) là thực, biên độ được chuẩn hóa sao cho $|s(t)| \leq 1$ và có phổ bị giới hạn trong khoảng $0 < f_m \leq |f| \leq f_M$. Tín hiệu này có thể coi là tín hiệu dải hẹp có bề rộng phổ :

$$B = f_M - f_m \approx f_M \qquad (5\text{-}19)$$

F_0 là tần số của tải tin (hay sóng mang) u(t), ta giả thiết:

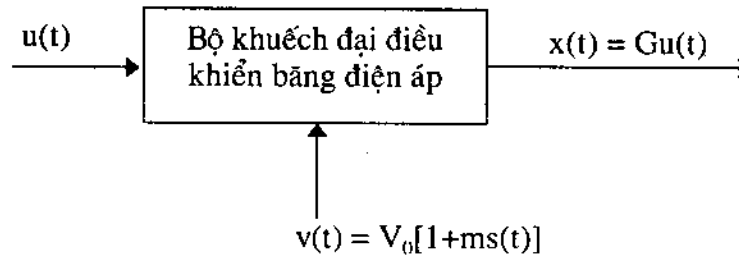
$$\omega_M = 2\pi f_M << \omega_0 = 2\pi F_0 \qquad (5\text{-}20)$$

Tải tin u(t) là một tín hiệu điều hòa tần số F_0:

$$u(t) = A_0\cos(\omega_0 t + \varphi_0) = A_0\cos(2\pi F_0 t + \varphi_0). \qquad (5\text{-}21)$$

Nguyên lý của điều biên là biến đổi biên độ không đổi A_0 của tải tin thành biên độ A(t) phụ thuộc vào s(t).

Thực hiện điều biên



Hình 5-5: Thực hiện điều biên

Ta sử dụng một bộ khuyếch đại điều khiển bằng điện áp G (Hình 5-5)

$$G = \frac{Gv(t)}{V_0} \qquad (5\text{-}22)$$

Khi ta đưa tải tin $u(t) = U\cos(\omega_0 t + \varphi_0)$ vào đầu vào của bộ khuếch đại có điện áp điều khiển là:

$$v(t) = V_0[\ 1+ms(t)] \qquad (5\text{-}23)$$

thì ta nhận được ở đầu ra của bộ khuếch đại:

$$x(t) = Gu(t) = G_0[1+ms(t)]U\cos(\omega_0 t + \varphi_0)$$

$$= A_0[1+ms(t)]\cos(\omega_0 t + \varphi_0)$$

$$x(t) = A(t)\cos(\omega_0 t + \varphi_0) \qquad (5\text{-}24)$$

Tín hiệu x(t) ở đầu ra nhận được chính là tín hiệu điều biên của s(t).

Thực hiện giải điều biên:

Có hai phương pháp thông dụng để giải điều biên, đó là: tách hình bao và giải điều biên đồng bộ.

Xét cụ thể phương pháp tách hình bao như sau:

Một bộ tách hình bao lý tưởng là một mạch điện sinh ra một tín hiệu ở đầu ra tỷ lệ thuận với hình bao thực của tín hiệu vào.

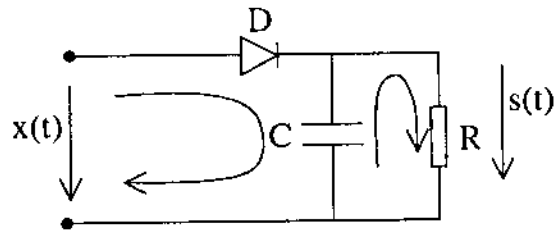
Tín hiệu vào là tín hiệu điều biên x(t):

$$x(t) = A(t)\cos(\omega_0 t + \varphi_0)$$

Tín hiệu ra y(t) của bộ tách hình bao lý tưởng là:

$$y(t) = KA(t) \qquad (5\text{-}25)$$

Mạch tách hình bao (Hình 5-6) là một điốt D (giả thiết là lý tưởng) nối với một điện trở R và tụ điện C. Điốt D mở đối với phần dương của tín hiệu vào. Khi điện áp tăng, D mở tụ C được nạp tới giá trị đỉnh của x(t). Khi điện áp đầu vào giảm đi, D khóa, tụ C phóng điện qua R cho đến khi điện áp vào tăng tới giá trị làm cho D lại mở.



Hình 5-6: Mạch điện tách hình bao

Hằng số phóng:

$$\tau = RC \qquad (5\text{-}26)$$

Khi chọn τ thích hợp:

$$\frac{1}{F_0} << \tau << \frac{1}{F_M} \qquad (5\text{-}27)$$

Trong thực tế, ta chọn:

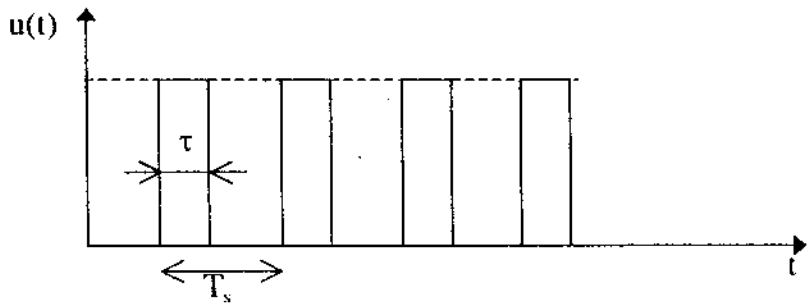
$$\tau = \frac{1}{\sqrt{F_0 f_M}} \qquad (5\text{-}28)$$

Thì tín hiệu ra y(t) xấp xỉ là hình bao của tín hiệu vào s(t) với độ méo không đáng kể. Trái lại, nếu chọn τ quá nhỏ hoặc τ quá lớn thì tín hiệu ra của mạch điện tách hình bao sẽ bị méo lớn.

III. ĐIỀU CHẾ XUNG

Thoạt đầu, để truyền tiếng nói trong các mạng viễn thông, người ta chỉ dựa trên kỹ thuật truyền tương tự, nhưng trong những khoảng cách xa, độ rõ của tiếng nói có thể kém. Do bị suy hao và bị nhiễu rất có thể sẽ khó hiểu hoặc không nhận ra người kia đang nói gì. Do vậy các tín hiệu tương tự được khuếch đại tại những khoảng cách nhất định, tuy nhiên độ nhiễu cũng được khuếch đại và mỗi tầng khuếch đại sau khi nối mạch đều làm tích tụ nhiễu. Để cải thiện các chất lượng truyền dẫn người ta sử dụng các phương pháp điều chế tín hiệu xung (còn gọi là số hóa) để chuyển đổi từ tín hiệu tiếng nói (liên tục) sang tín hiệu dạng số và ở đầu nhận cần phải giải điều chế một lần nữa thành âm thanh. Ở đây có sự khác nhau quan trọng giữa các đặc tính của hai phương pháp truyền dẫn này là trong các hệ thống chuyển mạch số, thông tin được tái tạo và được phát đi hoàn toàn không bị nhiễu. Việc khảo sát các phương pháp điều chế ra tín hiệu số nhằm xác định phương pháp tốt nhất cho mọi loại môi trường truyền là điều hết sức quan trọng. Chúng ta sẽ cụ thể kỹ thuật số hoá như sau:

Tải tin u(t) là một dãy xung vuông tuần hoàn chu kỳ T_s cho ở công thức (5-5) như trên Hình 5-7.



Hình 5-7: Tải tin của tín hiệu điều biên xung

Tải tin u(t) có bốn tham số:

- Biên độ A

- Tần số $F_s = \frac{1}{T_s}$

- Thời điểm xuất hiện của xung thứ k: $t_k = t_0 + kT_s$ (5-29)

- Độ rộng xung: τ

Cho hàm tin s(t) tác động vào từng tham số trên, ta nhận được bốn loại điều biên xung tương ứng: điều biên xung (PAM), điều pha xung (PPM), điều tần xung (PFM), điều rộng xung (PWM)

Tín hiệu điều biên xung sẽ có dạng tổng quát:

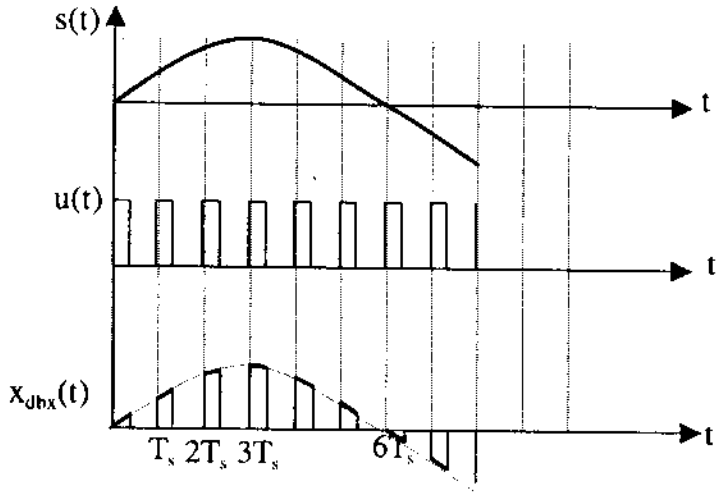
$$x(t) = \sum_{k=\infty}^{\infty} A_k \prod_{\tau_k} (t - kT_{s_k} - t_{o_k}) \quad (5\text{-}30)$$

Các tham số biên độ A_k, độ rộng τ_k, chu kỳ Ts_k, thời điểm xuất hiện xung t_{0k} phụ thuộc vào t, tương ứng với điều biên xung (PAM), điều rộng xung (PWM), điều tần xung (PFM), điều pha xung (PPM).

1. Điều biên xung (PAM)

Tín hiệu điều biên xung là tín hiệu có biên độ mang tin (Hình 5-8) tuân theo quy luật sau:

$$A_k = s(kT_s) \qquad (5\text{-}31)$$

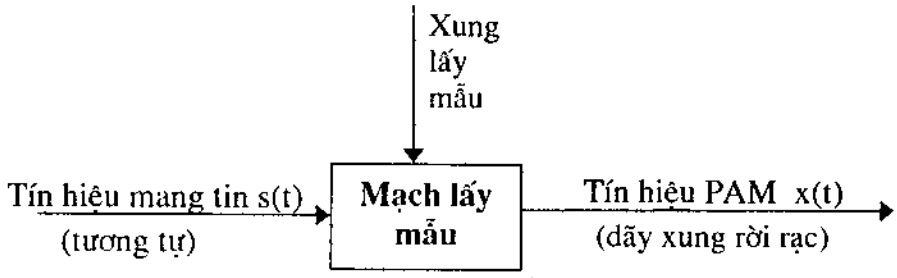


Hình 5 - 8: Tín hiệu điều biên xung

Thực hiện điều biên xung là lấy mẫu tín hiệu s(t) dựa trên định luật lấy mẫu của Shannon, thay thế nó bằng một hàm rời rạc là những mẫu của hàm trên lấy ra tại những thời điểm gián đoạn.

Nếu tần số lấy mẫu ≥ 2 lần tần số cao nhất của tín hiệu(f_{max}) thì phép lấy mẫu bảo toàn thông tin của tín hiệu. Ví dụ tiếng nói có tần số lớn nhất là 4300Hz thì cần lấy mẫu với tần số nhỏ nhất là 8600HZ.

Hình 5-9 trình bày lược đồ tổng quát sự lấy mẫu.



Hình 5-9: Điều biên xung

Tín hiệu mang tin s(t) được đưa đến đầu vào của mạch lấy mẫu. Xung lấy mẫu u(t) là một dãy xung vuông tuần hoàn chu kỳ T_s. Đầu ra là tín hiệu điều biên xung x(t) là một dãy xung mà biên độ mỗi xung bằng với biên độ của tín hiệu tương tự gốc tại thời điểm lấy mẫu. Các xung này được tạo ra có tần số f_s đảm bảo định luật lấy mẫu. Do đó, ta có thể khôi phục (quá trình giải điều biên) lại chính xác tín hiệu s(t) từ tín hiệu điều biên xung nếu tín hiệu s(t) được lấy mẫu ở tần số f_s

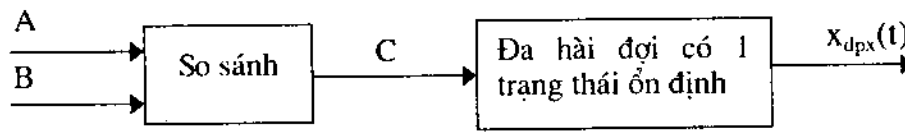
Điều biên xung(PAM) vẫn chưa phải là quá trình số hoá vì dạng điều chế này lấy xung có bề rộng cố định nhưng có chiều cao thay đổi theo tín hiệu điều chế. Điều chế biên độ xung rất nhạy cảm với nhiễu nhiệt và nhiễu xuyên âm, nhiễu sẽ tác động lên đỉnh xung và hình thành tạp âm trên kênh khi giải điều chế.

2. Điều pha xung (PPM)

Tín hiệu điều pha xung là tín hiệu có pha (hay thời điểm xuất hiện của xung) mang tin tuân theo quy luật sau:

$$t_{0_k} = t_0 + \Delta T.s(kT_s) \qquad (5\text{-}32)$$

Ta có thể thực hiện điều pha xung bằng mạch điện như trên hình 5-10.



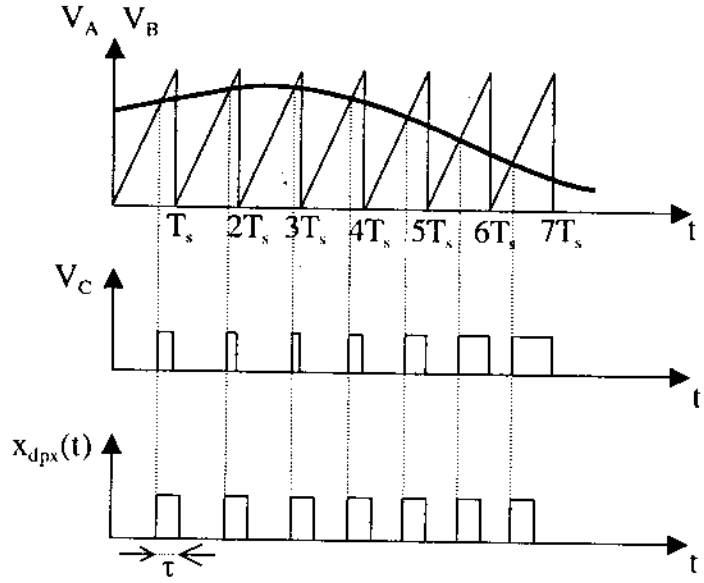
Hình 5-10: Mạch điều pha xung

Đầu vào A là tín hiệu V_A, đó là xung răng cưa tuần hoàn chu kỳ T_s.

Đầu vào B là tín hiệu :

$$V_B = V[1+\alpha.s(t)] \qquad (5\text{-}33)$$

Dạng tín hiệu trong mạch được trình bày ở hình 5-11



Hình 5 - 11: Tín hiệu điều pha xung

Tín hiệu đầu vào ở A được so sánh với tín hiệu đầu vào ở B, cho thấy một dãy xung vuông có thời điểm xuất hiện xung mang tin nhưng có độ rộng xung khác V_C. Nếu ta cho tín hiệu ở C đi qua một đa hài đợi một trạng thái ổn định (độ rộng xung τ) thì sẽ nhận được một tín hiệu điều pha xung

Pha của xung được xác định một cách rõ ràng như sau:

Xét trong một chu kỳ:

$$V_A = V_0.\Delta t \qquad (5\text{-}34)$$

mà $V_A = V_B$ tại thời điểm mạch so sánh chuyển trạng thái, do đó:

$$V_0.\Delta t = V[1 + \alpha.s(t)] \qquad (5\text{-}35)$$

$$\rightarrow t_{0_k} = \Delta t_k = \frac{V}{V_0} + \frac{\alpha V}{V_0} s(KT_s) \qquad (5\text{-}36)$$

$$\rightarrow t_k = \left(kT_s + \frac{V}{V_0}\right) + \frac{\alpha V}{V_0} s(kT_s) \qquad (5\text{-}37)$$

Đây là tín hiệu điều pha xung.

3. Điều rộng xung (PWM)

Tín hiệu điều rộng xung là tín hiệu có độ rộng xung mang tin tuân theo quy luật:

$$\tau_k = \tau + \Delta\tau.s(kT_s) \qquad (5\text{-}38)$$

Trong dạng điều chế này, giữ biên độ xung cố định, nhưng bề rộng xung thay đổi. Cạnh của xung dịch theo tín hiệu mang tin làm bề rộng xung thay đổi theo. Khi nhiễu làm thay đổi biên độ xung sẽ không ảnh hưởng đến tín hiệu gốc vì biên độ xung không mang thông tin nào. Tuy vậy, nhiễu cũng làm xáo trộn cạnh của xung bởi hiện tượng méo dạng khi truyền.

Có 3 cách điều rộng xung:

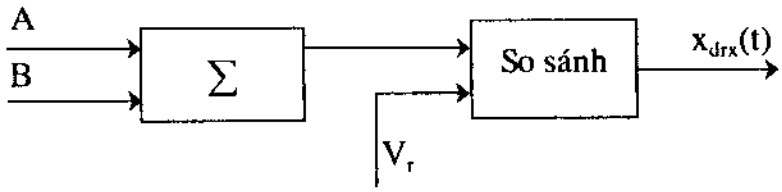
Giữ nguyên sườn trước, thay đổi sườn sau

Giữ nguyên sườn sau, thay đổi sườn trước

Giữ nguyên điểm giữa, thay đổi độ rộng

Ta xét cách thứ nhất là giữ nguyên sườn trước, thay đổi sườn sau.

Mạch điện thực hiện điều rộng xung như Hình 5-12



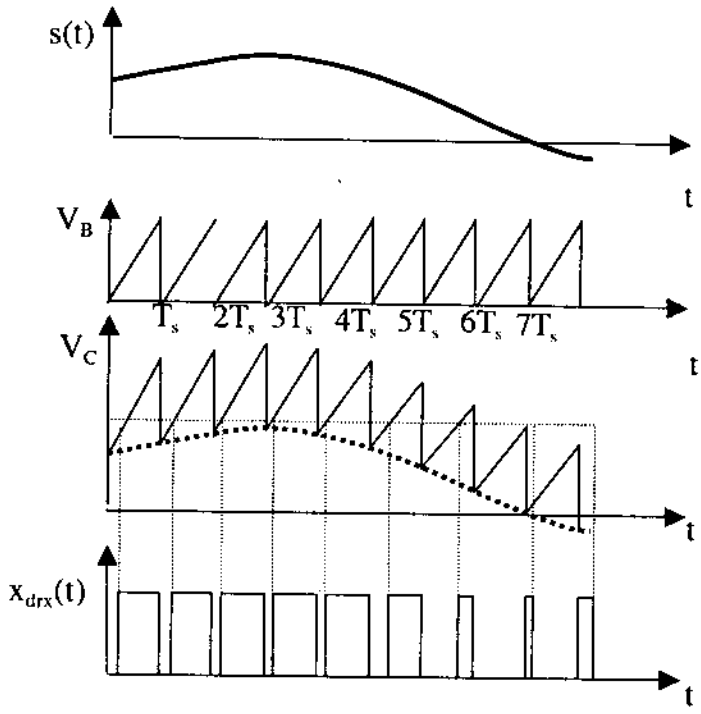
Hình 5-12: Mạch điều rộng xung

Đầu vào A là tín hiệu $V_A = s(t)$

Đầu vào B (V_B) là tín hiệu răng cưa tuần hoàn chu kỳ T_s.

V_r là điện áp tham chiếu

Dạng tín hiệu trong mạch điều rộng xung như hình 5-13



Hình 5-13: Dạng tín hiệu trong mạch điều rộng xung

Tín hiệu ở điểm C là tổng hai tín hiệu V_A, V_B, nó được so sánh với tín hiệu V_r, cho ta ở đầu ra một dãy xung vuông có độ rộng xung mang tin.

Độ rộng xung được xác định cụ thể như sau:

Tại thời điểm mạch so sánh chuyển trạng thái, ta có :

$$V_A + V_B = V_C = V_r \qquad (5\text{-}39)$$

$$\rightarrow V_0\Delta t + s(t) = V_r \qquad (5\text{-}40)$$

$$\rightarrow \Delta t = \frac{V_r}{V_0} = \frac{1}{V_0}s(t) \qquad (5\text{-}41)$$

$$\tau_k = T_s - \Delta t_k = \left(T_s - \frac{V_r}{V_0}\right) + \frac{1}{V_0}s(kT_s) \qquad (5\text{-}42)$$

Đây là tín hiệu điều pha xung.

Đối với hai cách còn lại, mạch điện vẫn như trên tuy nhiên chỉ thay đổi tín

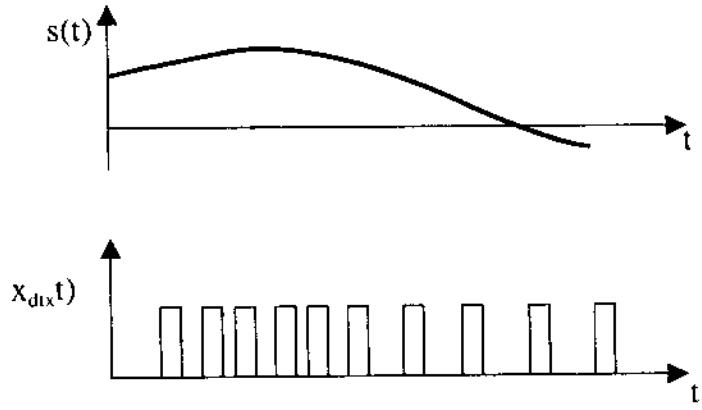
hiệu V_B là tín hiệu xung răng cưa tuần hoàn chu kỳ T dốc sườn sau hoặc dốc cả hai một cách tương ứng.

4. Điều tần xung (PFM)

Tín hiệu điều tần xung là tín hiệu có tần số mang tin.

$$\frac{1}{T_{s_k}} = \frac{1}{T_s} + \Delta F.s(kT_s) \qquad (5\text{-}43)$$

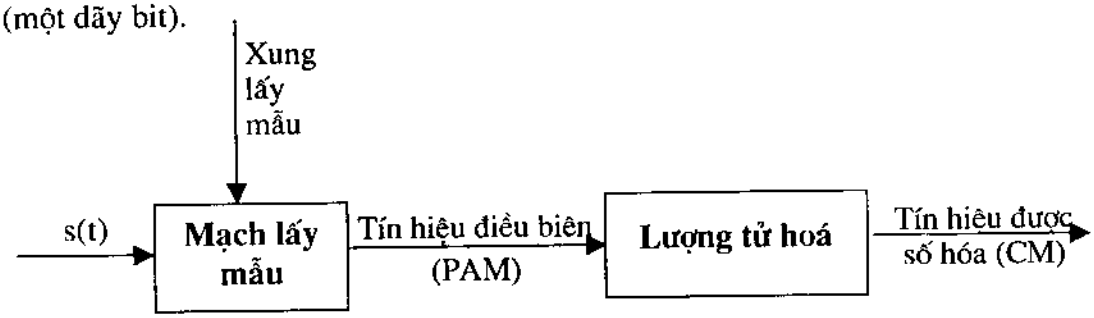
Phương pháp điều chế này được xem như tương đương với điều rộng xung. Dạng tín hiệu điều tần xung được thể hiện trong Hình 5-14



Hình 5-14: Tín hiệu điều tần xung

5. Điều xung mã (PCM)

Điều xung mã là quá trình biến đổi tương tự - số, trong đó thông tin chứa trong các mẫu tín hiệu của tín hiệu tương tự được biểu diễn dưới dạng nhị phân (một dãy bit).



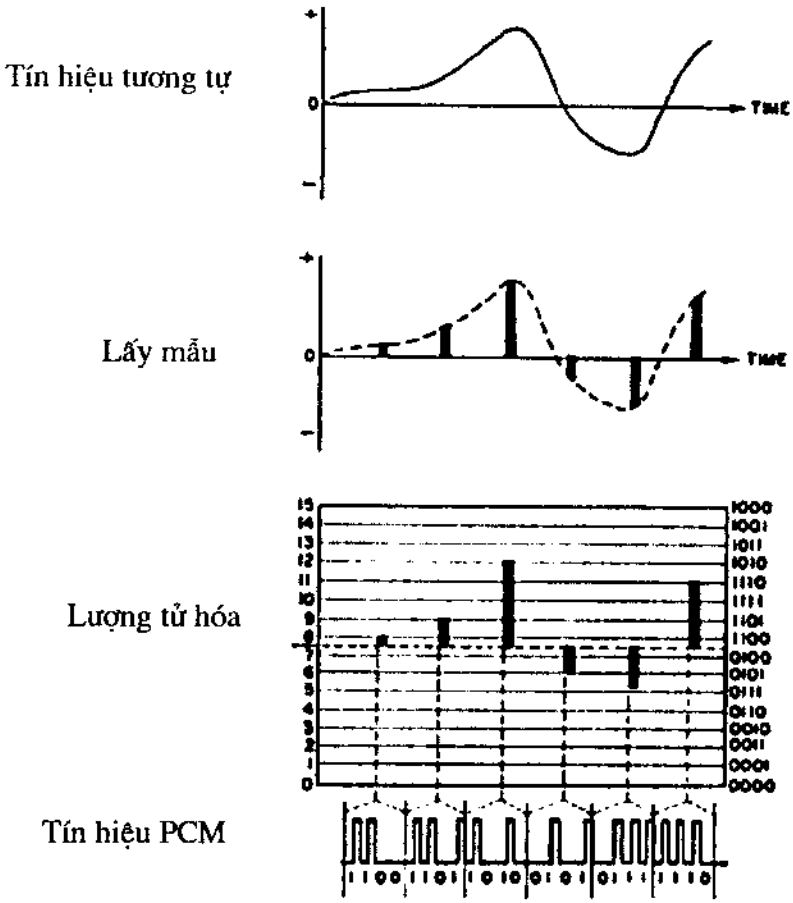
Hình 5-15: Lược đồ hóa PCM

Bước 1: lấy mẫu tín hiệu s(t) (điều biên xung PAM), với chu kỳ lấy mẫu T_s, ta nhận được các mẫu s_k.

Bước 2: Lượng tử hóa s_k với bước lượng tử hóa Δ , làm tròn các giá trị biên độ các mẫu tín hiệu tới số nguyên gần nhất.

$$n_k = \begin{cases} \left\lfloor \frac{s_k}{\Delta} + 0{,}5 \right\rfloor, khi s_k \geq 0 \\ \left\lceil \frac{s_k}{\Delta} - 0{,}5 \right\rceil, khi s_k < 0 \end{cases}$$

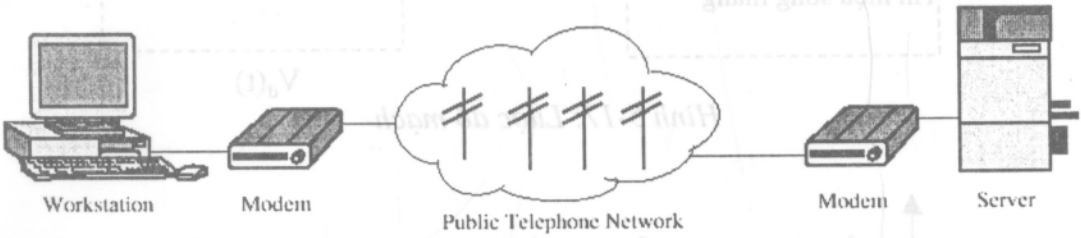
Bước 3: Mã hóa giá trị nguyên của các mẫu n_k dưới dạng nhị phân (dùng mã bù 2 hoặc mã gray), giả thiết mã hóa một mẫu b bit (b được chọn sao cho là tối thiểu), ta nhận được một tín hiệu số là một dãy bit.



Hình 5-16: Dạng tín hiệu PCM

IV. ĐIỀU CHẾ KHÓA DỊCH

Khi muốn truyền số liệu giữa các máy tính trên các PSTN (mạng điện thoại công cộng) có sẵn, cần chuyển đổi các tín hiệu đi ra từ các máy tính sang dạng phù hợp với PSTN. Đường truyền PSTN được thiết kế cho thông tin tiếng nói có băng thông trong khoảng 300-3400 Hz. Chính vì lý do này không thể đặt hai



mức điện áp từ DTE trực tiếp lên đường dây. Vì vậy, dữ liệu nhị phân cần được chuyển đổi sang dạng tương thích với tín hiệu điện thoại tại DTE phát và chuyển đổi ngược lại thành dạng nhị phân tại đầu thu.

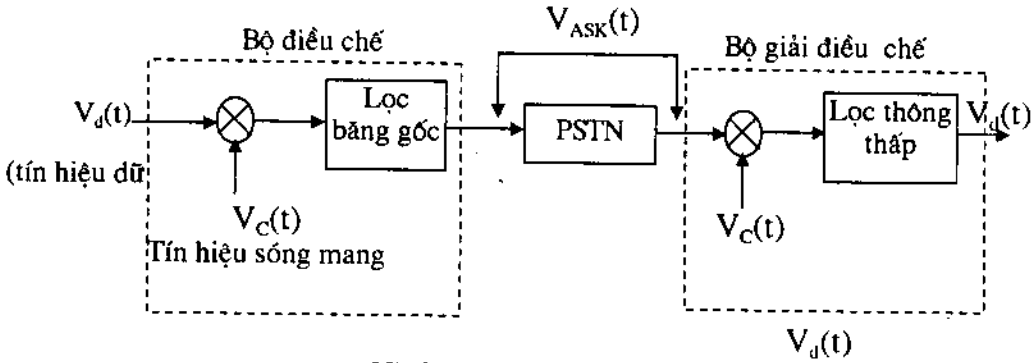
Đối với truyền hình cáp CATV cũng vậy, tín hiệu số từ các trung tâm truyền hình cáp sẽ được chuyển đổi thành tín hiệu tương tự tần số vô tuyến để ghép kênh trên các cáp đồng trục và đến máy thu hình của các thuê bao.

Mạch điện thực hiện hoạt động chuyển đổi từ tín hiệu băng gốc số sang tín hiệu liên tục (hình sin) băng tần thích hợp gọi là bộ điều chế (Modulator), và mạch điện thực hiện hoạt động chuyển đổi ngược lại được gọi là bộ giải điều chế (demodulator). Vì mỗi liên kết đều cần cả hai đầu mạch, chúng kết hợp lại thành một thiết bị chung gọi là Modem.

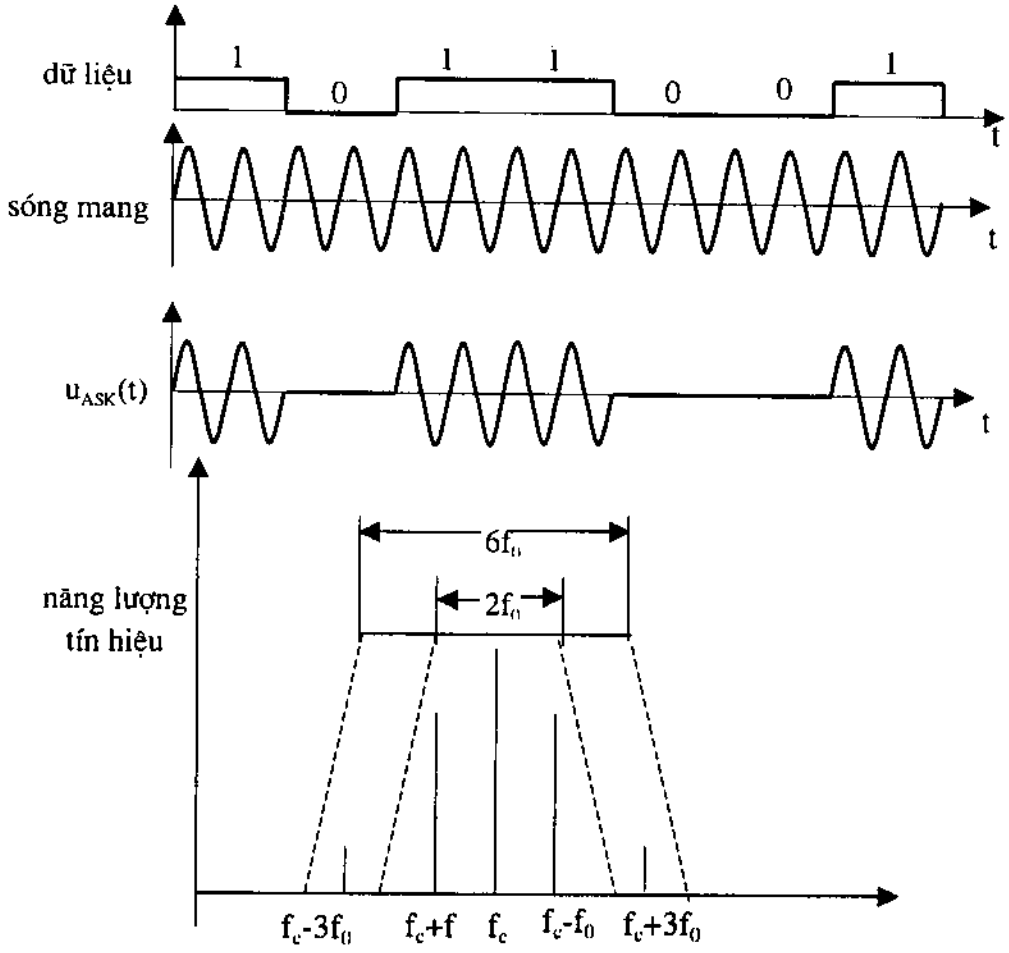
Để thực hiện điều chế, ta phải thay đổi biên độ, tần số và pha của tín hiệu sóng mang hình sin theo sự biến đổi của dữ liệu nhị phân cần truyền. Dữ liệu nhị phân được truyền chỉ yêu cầu hai mức tín hiệu, sự chuyển mạch tín hiệu giữa hai mức mang ý nghĩa khóa (key) vì vậy ba loại điều chế trên được gọi là điều chế khóa dịch biên độ (ASK), khoá dịch tần số (FSK) và khoá dịch pha (PSK). Ngoài ra hiện nay người ta còn kết hợp ASK với PSK gọi là điều chế AQM.

1. Điều chế khóa dịch biên độ *(Amplitude Shift Keying - ASK)*

Nguyên lý hoạt động cơ bản của ASK được minh hoạ trong hình 5-17 và dạng sóng đơn giản ở hình 5- 18.



Hình 5-17: Lược đồ mạch



f_0 = thành phần tần số cơ bản = 1/2 tốc độ bít

Hình 5-18: Dạng sóng và băng thông

Nguyên lý của ASK là biên độ của sóng âm đơn tần được chuyển đổi giữa hai mức với tốc độ được xác định bởi tốc độ bit của tín hiệu nhị phân được truyền. Sóng âm tần hay còn gọi là sóng mang có tần số thuộc băng thông PSTN. Kích thước băng thông yêu cầu được xác định bởi tốc độ của tín hiệu, tốc độ càng cao thì kích thước băng thông yêu cầu càng lớn. Trong thực tế các phương pháp điều chế khác nhau đòi hỏi độ rộng băng thông khác nhau, nên việc đánh giá mức băng thông yêu cầu ứng với mỗi phương pháp là rất cần thiết.

Về mặt toán học, hoạt động điều chế ASK, PSK, FSK tương đương với việc nhân tín hiệu sóng mang với tín hiệu dữ liệu nhị phân. Sóng mang có tần số riêng ω_c và biên độ không đổi biểu diễn dưới dạng biểu thức điện áp

$$u_c(t) = \cos\omega_c t; \quad \omega_c : \text{radian/second} \qquad (5\text{-}44)$$

Một tín hiệu số mang tin tuần hoàn đơn cực với biên độ không đổi và tần số cơ bản ω_0 được khai triển theo chuỗi fourier như sau:

$$s_d(t) = 1/2 + 2/\pi\,(\cos\omega_0 t - 1/3\cos 3\omega_0 t + 1/5\cos 5\omega_0 t - \ldots) \qquad (5\text{-}45)$$

Suy ra ASK có thể biểu diễn :

$$u_{ASK}(t) = u_c(t) \times s_d(t) \qquad (5\text{-}46)$$

Do đó:

$$u_{ASK}(t) = 1/2.\cos\omega_c t + 2/\pi\,.(\cos\omega_0 t.\cos\omega_c t - 1/3.\cos\omega_c t.\cos 3\omega_0 t + \ldots.)$$

Suy ra :

$$u_{ASK} = 1/2\cos\omega_c t + 1/\pi\,[\cos(\omega_c - \omega_0)t + \cos((\omega_c + \omega_0)t - 1/3\cos(\omega_c - 3\omega_0)t$$
$$-1/3\cos(\omega_c + 3\omega_0)t + \ldots.] \qquad (5\text{-}47)$$

Chúng ta thấy tín hiệu ASK tương đương với tín hiệu số liệu nguồn chuyển dịch sang dạng tần số cuả sóng mang, bên cạnh đó còn có hai thành phần tần số cơ bản $((\omega_c - \omega_0)$ và $((\omega_c + \omega_0)$ và hai thành phần hài bậc cao $((\omega_c - 3\omega_0)$ và $((\omega_c + 3\omega_0)$. Tất cả hình thành nên biên tần cơ bản, do đó băng thông của ASK được trình bày trên hình 5-18

Băng thông càng lớn thì tín hiệu thu càng trung thực. Tuy nhiên, hiệu quả hoạt động có thể đạt được nếu băng thông của kênh đủ chỗ để cho thành phần tần số cơ bản của dòng dữ liệu 101010... vì băng thông yêu cầu trong các dòng dữ liệu có thứ tự khác đều nhỏ hơn. Gọi f_0 là tần số cơ bản của tín hiệu dữ liệu

theo tuần tự 101010... f_0 bằng nửa tốc độ bit (bps), suy ra băng thông tối thiểu của ASK bằng với tốc độ bit tính sang đơn vị Hz, $2f_0$ hoặc để thu được thành phần hài bậc 3 thì phải gấp ba lần tốc độ bit tính sang Hz, 6 f_0.

Từ hình 5-18 ta có thể thấy rằng đối với ASK tín hiệu sóng mang hiện diện trong tín hiệu thu ngay cả khi không có tín hiệu thông tin f_0, $3f_0$,... trong đó. Từ công thức Nyquist, suy ra với tín hiệu nhị phân thì tốc độ số liệu tối đa đạt được bằng hai lần băng thông. Vì thế, giả sử:

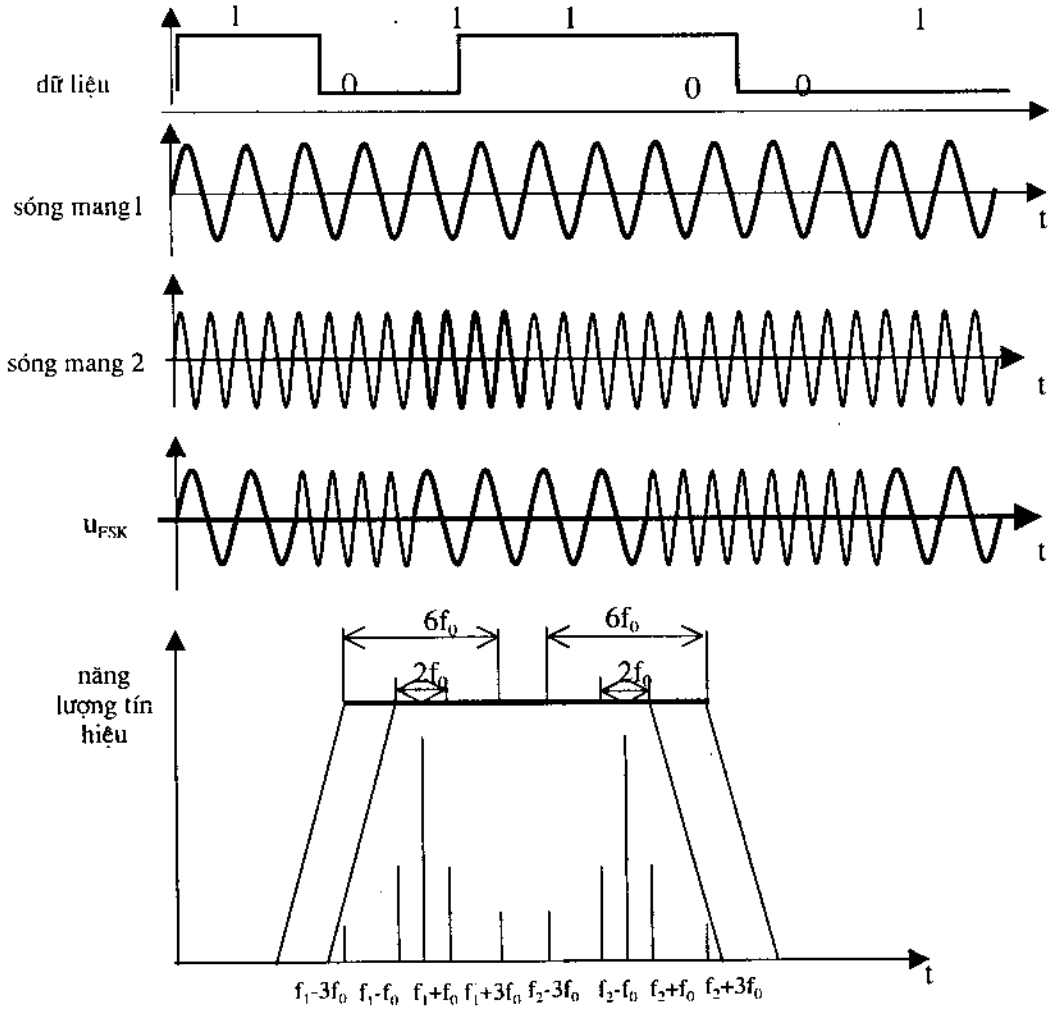
Việc khôi phục lại tín hiệu số liệu được thực hiện bởi mạch giải điều chế. Tại đây tín hiệu thu lại được nhân một lần nữa với sóng mang cùng dạng. Từ đó sinh ra hai phiên bản của tín hiệu thu: thành phần bao quanh tần số $2f_c(f_c + f_c)$ và thành phần bao quanh zero($f_c - f_c$). Cả hai phiên bản đều chứa thông tin trong hai biên tần, nhưng thành phần sau cùng được chọn bằng cách cho đi qua bộ lọc thông thấp. Bộ lọc thông này được thiết kế để chỉ cho các tần số từ 0 đến f_0 hoặc nếu hài bậc 3 nhận được thì từ 0 đến $3f_0$. Do vậy ở đầu ra của bộ lọc chính là phiên bản của tín hiệu dữ liệu truyền bị giới hạn băng tần.

Mặc dù ASK là phương pháp thực hiện đơn giản, dễ bị ảnh hưởng bởi nhiễu, khó đồng bộ và thường dùng trong cáp quang. Gần đây tất cả các hệ thống truyền dẫn và chuyển mạch đều được thực hiện bởi kỹ thuật số và ASK có cơ hội tham gia và thường được dùng kết hợp với PSK để thiết kế các Modem tốc cao.

2. Phương pháp điều chế khoá dịch tần số (*Frequency Shift Keying* - FSK)

FSK là phương pháp điều chế được dùng trong tất cả các Modem tốc độ thấp thế hệ đầu. Nguyên lý hoạt động cơ bản được minh họa trong hình 5-19.

Để tránh vấn đề thay đổi biên độ, với FSK dùng hai tín hiệu sóng mang phân biệt tần số nhưng có cùng biên độ và cố định, một cho bít nhị phân '0' , một cho bit nhị phân '1'.



Hình 5-19: Tín hiệu điều chế FSK và băng thông

Nguyên lý của FSK là hai giá trị nhị phân được biểu diễn bởi hai tín hiệu tần số khác nhau. Hoạt động điều chế tương đương với sự tổng hợp các ngõ ra của hai bộ điều chế ASK riêng biệt: một thực hiện trên sóng mang thứ nhất dùng phần gốc của tín hiệu dữ liệu (mức cao) và một thực hiện sóng mang thứ hai dùng phần bù của tín hiệu dữ liệu (mức 0).

Về mặt toán học, có thể biểu diễn tín hiệu FSK như sau:

Hai sóng mang có tần số riêng ω_1 và ω_2, biên độ không đổi biểu diễn dưới dạng biểu thức điện áp.

$u_1(t) = \cos\omega_1 t$; ω_1 : radian/second

$u_2(t) = \cos\omega_2 t$; ω_2 : radian/second

Một tín hiệu số mang tin tuần hoàn đơn cực với biên độ không đổi và tần số cơ bản ω_0 được khai triển theo chuỗi fourier như sau:

$$s_d(t) = 1/2 + 2/\pi\ (\cos\omega_0 t - 1/3\ \cos 3\omega_0 t + 1/5\ \cos 5\omega_0 t - ...)$$

$$u_{FSK}(t) = \text{Cos}\ \omega_1 t.\ s_d(t) + \text{Cos}\ \omega_2 t.\ s'_d(t) \qquad (5\text{-}48)$$

Trong đó $s'_d(t)$ là phần bù của tín hiệu dữ liệu gốc $s_d(t)$.

Do đó:

$$V_{FSK}(t) = 1/2\cos\omega_1 t + 1/\pi\ [\cos(\omega_1 - \omega_0)t + \cos((\omega_1 + \omega_0)t - 1/3\cos(\omega_1 - 3\omega_0)t - 1/3\cos(\omega_1 + 3\omega_0)t + ...] + 1/2\cos\omega_2 t + 1/\pi\ [\cos(\omega_2 - \omega_0)t + \cos((\omega_2 + \omega_0)t - 1/3\cos(\omega_2 - 3\omega_0)t - 1/3\cos(\omega_2 + 3\omega_0)t + ...] \qquad (5\text{-}47)$$

Có thể suy ra băng thông của FSK một cách đơn giản là tổng hợp hai sóng mang điều chế ASK riêng biệt ở tần số góc ω_1 và ω_2.

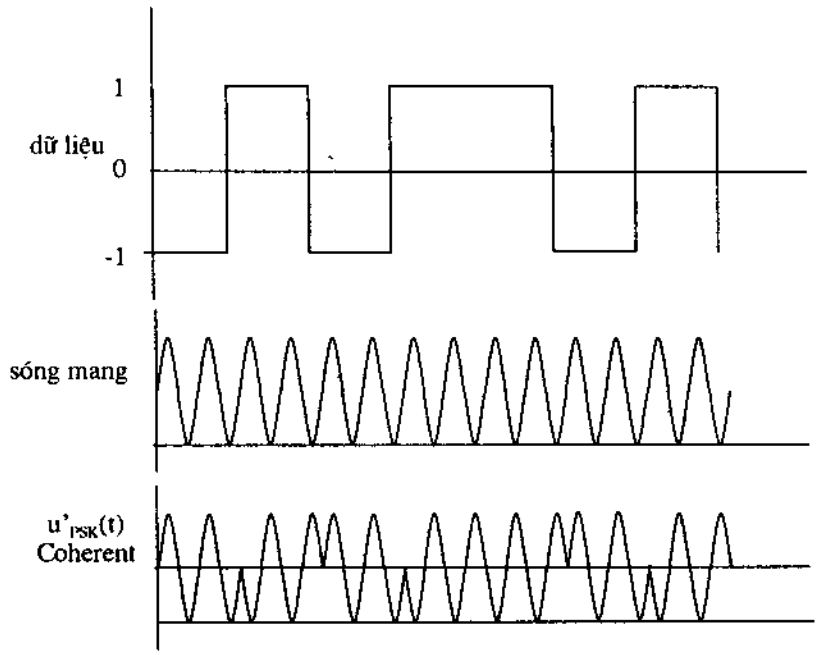
Như đã đề cập ở phần trước, tín hiệu có tần số cao nhất được tạo ra từ dãy bit 1011001... với FSK vì tín hiệu nhị phân 0 và 1 điều chế lên các sóng mang riêng rẽ, nên băng thông tối thiểu cho mỗi sóng mang bằng một nửa tốc độ bit, nghĩa là thành phần tần số cơ bản lớn nhất cho mỗi sóng mang f_0 bằng một nửa so với ASK. Do đó nếu giả sử chỉ thu được thành phần tần số cơ bản thì băng thông tổng cộng bằng $4f_0$ cộng với khoảng dịch tần f_s. Tuy nhiên vì f_0 bằng một nửa so với ASK nên băng thông tổng bằng với băng thông tổng cho ASK cộng với khoảng dịch tần. Tương tự nếu cặp hài bậc 3 được thu thì băng thông bằng $6f_0$ cộng với khoảng dịch tần.

Kỹ thuật FSK có tỷ suất lỗi thấp hơn ASK, thường dùng trong truyền số liệu qua đường điện thoại (tần số thấp), hoặc trong mạng không dây (tần số cao).

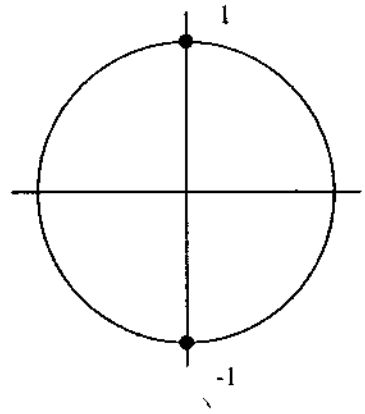
3. Phương pháp điều chế khoá dịch pha (*Phase Shift Keying* - PSK)

Nguyên lý của PSK là tần số và biên độ của sóng mang được giữ không đổi trong khi pha của nó được dịch theo mỗi bit của dòng dữ liệu truyền.

Xét loại PSK thứ nhất dùng hai tín hiệu sóng mang cố định đại diện cho bit 0 và 1, hai sóng mang khác pha nhau 180^0. Vì tín hiệu này chỉ là nghịch đảo của tín hiệu kia nên loại này được gọi là phase coherent PSK (PSK phối hợp pha) còn được gọi là 2-PSK.



Hình 5-20 a: Tín hiệu 2-PSK

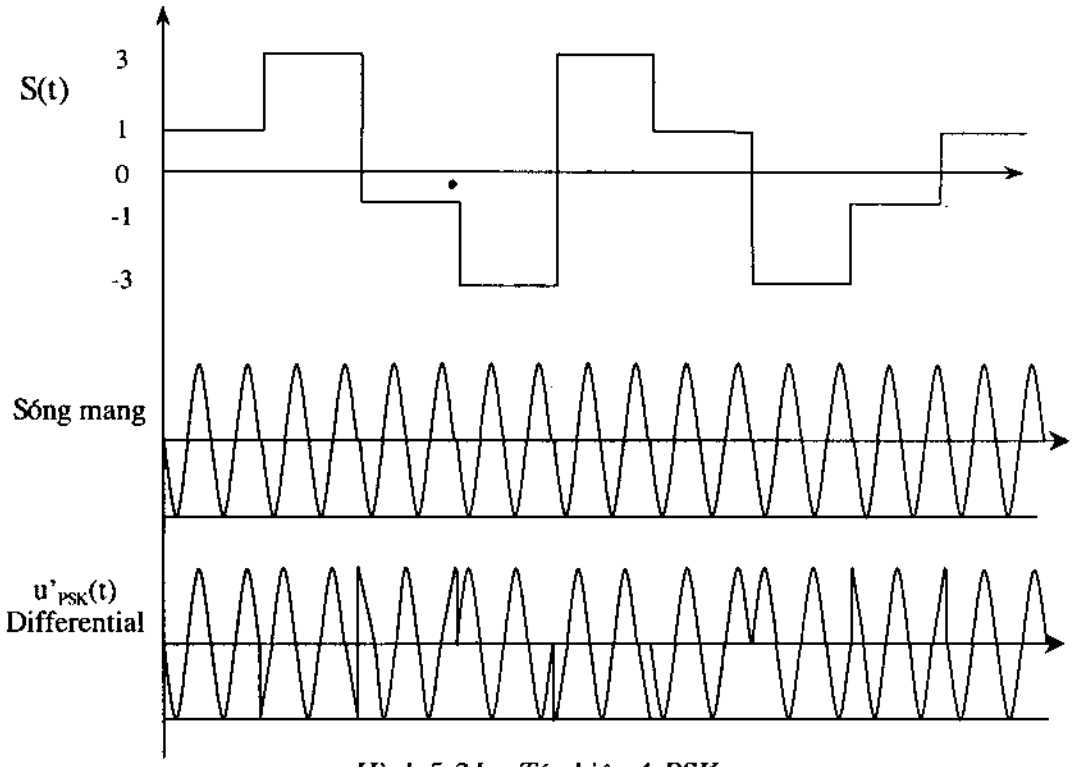


Hình 5-20b: Sơ đồ pha 2-PSK

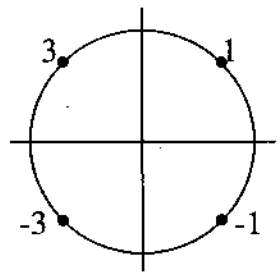
Ở đây, tín hiệu băng gốc s(t) là xung NRZ lưỡng cực nhận n=2 giá trị như ở hình 5-20a, nên dạng sóng đã điều chế có dạng giống như ASK đảo pha, một symbol của tín hiệu điều chế mang thông tin của 1 bit.

Điều bất tiện của loại này là tại máy thu đòi hỏi phải có sóng mang tham chiếu để so pha với tín hiệu thu, do đó cần thực hiện đồng bộ pha giữa máy thu và máy phát. Kết quả là dẫn đến mạch giải điều chế phức tạp hơn.

Loại PSK thứ hai ta xét tới là PSK vi phân (diferential PSK) còn được gọi là 4-PSK. Tín hiệu băng gốc là một xung NRZ lưỡng cực nhận n = 4 giá trị, thì một dạng sóng đã điều chế như hình 5-21 sau:



Hình 5-21a: Tín hiệu 4-PSK



Hình 5-21b : Sơ đồ pha 4-PSK

Với loại này sự dịch pha xảy ra tại mỗi bit không cần quan tâm tới chuỗi bit 0 hay bit 1 đang được truyền. Một sự dịch pha 90^0 tương ứng với tín hiệu hiện hành chỉ định 0 là bit kế tiếp trong khi sự dịch pha 270^0 chỉ bit 1. Như vậy mạch giải điều chế chỉ cần xác định độ lớn của sự dịch pha thay vì phải xác định giá trị tuyệt đối của từng pha. Ở mạch điều chế chỉ khi nào thay đổi trạng thái của dữ liệu mới đổi pha của sóng mang.

Hình 5-21b biểu diễn sơ đồ pha tín hiệu 4-PSK. Giống như 2-PSK, nếu tín hiệu băng gốc không bị giới hạn băng thông, thì tín hiệu 4-PSK cũng dịch chuyển tức thời trên vòng tròn đơn vị.

Về mặt toán học; ta có thể xác định băng thông PSK. Ở đây chúng ta trình bày tín hiệu số nhị phân dưới dạng lưỡng cực vì mức tín hiệu âm sẽ là kết quả đổi pha 180^0 của sóng mang.

Sóng mang có tần số riêng ω_c và biên độ không đổi biểu diễn dưới dạng biểu thức điện áp:

$u_c(t) = \cos\omega_c t$; ω_c : radian/second

Một tín hiệu số mang tin tuần hoàn đơn cực với biên độ tổng hợp và tần số cơ bản ω_0 được khai triển toán học theo chuỗi fourier như sau:

$$s_d(t) = 4/\pi\ [\cos\omega_0 t - 1/3.\cos 3\omega_0 t + 1/5.\cos 5\omega_0 t - ...] \qquad (5\text{-}49)$$

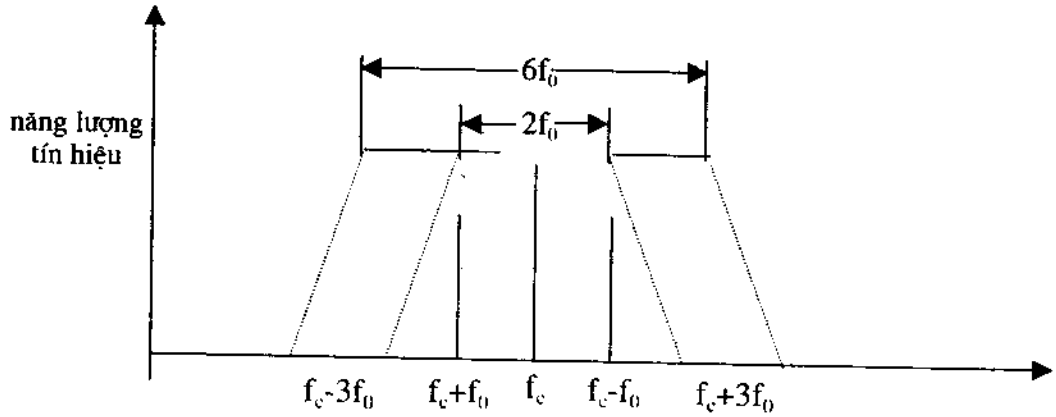
Suy ra :

$$u_{PSK} = 4/\pi.[\cos\omega_c t\ \cos\omega_0 t - 1/3.\cos\omega_c t.\cos 3\omega_0 t + 1/5.\cos\omega_c t\ \cos 5\omega_0 t - ...]$$

$$= 2/\pi.[\cos(\omega_c-\omega_0)t + \cos(\omega_c+\omega_0)t - 1/3.\cos(\omega_c-3\omega_0) -$$

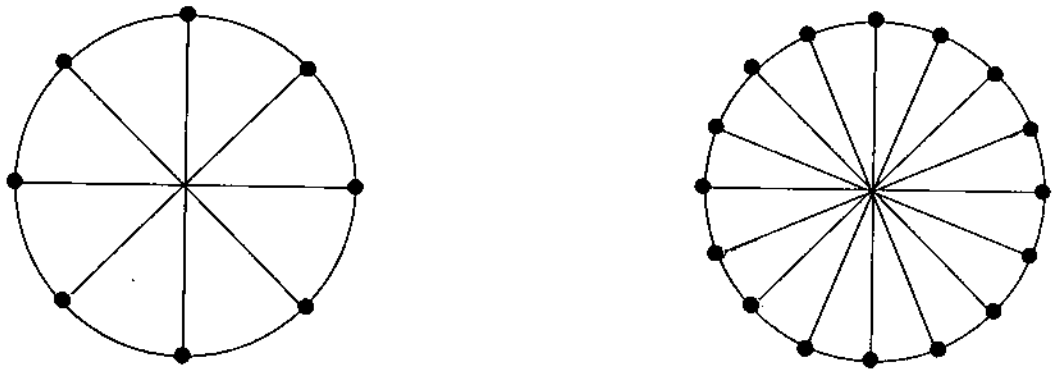
$$1/3.\cos(\omega_c+3\omega_0)t+...] \qquad (5\text{-}50)$$

Từ công thức trên ta thấy phổ tần số PSK giống như ASK chỉ khác là không có thành phần sóng mang. Do đó, băng thông của tín hiệu PSK sẽ được thể hiện như ở hình 5-22. Giả sử chỉ thành phần tần số cơ bản của dãy 101100... được nhận thì băng thông tối thiểu bằng $2f_0$, bằng giá trị tốc độ bit. Tuy nhiên, do vắng mặt thành phần sóng mang nên có nhiều năng lượng trong biên tần đơn (data), điều đó giúp PSK kháng nhiễu tốt hơn ASK hay FSK. Giới hạn băng của tín hiệu từ f_c đến f_c+f_0, có nghĩa là băng thông bằng f_0 và đạt được tốc độ Nyquist. Hầu hết năng lượng nhận được thuộc về tín hiệu mang thông tin, f_c+f_0 (do không có f_c).



Hình 5-22: Băng thông PSK

Bên cạch 2-PSK và 4-PSK, còn có PSK nhiều pha. Hình 5-23 đưa ra các sơ đồ pha đối với 8-PSK và 16-PSK.



(a) 8PSK

(b) 16PSK

Hình 5-23: PSK nhiều pha

Ngày nay các thiết bị truyền dẫn và chuyển mạch kỹ thuật số đã và đang được dùng trong mạng PSTN hiện đại. Kết quả ứng dụng đó, tạo điều kiện đạt được tốc độ bit vượt xa tốc độ có được theo các phương pháp điều chế cơ bản kể trên nhờ sử dụng các phương pháp điều chế phức tạp hơn. Trong các phương

pháp nhằm gia tăng tốc độ, tồn tại hai khuynh hướng: hoặc nhiều mức tín hiệu hoặc trộn lẫn các phương pháp điều chế cơ bản, đặc biệt là ASK và PSK. Loại điều chế kết hợp giữa ASK và PSK gọi là biên độ cầu phương QAM.

4. Điều chế biên độ cầu phương (Quadrature Amplitude Modulation-QAM)

4.1. Biểu diễn tín hiệu cầu phương

Trước khi miêu tả điều chế biên độ cầu phương, chúng ta hãy xét biểu diễn cầu phương của các tín hiệu. Một tín hiệu hình sin $\cos(\omega_0 t + \theta)$ có pha xác định được trình bày bằng định lý cộng của hàm lượng giác như sau:

$$\cos(\omega_0 t + \theta) = \cos\theta.\cos\omega_0 t - \sin\theta.\sin\omega_0 t$$

Trong biểu thức này, $\cos\omega_0 t$ và $\sin\omega_0 t$ là các tín hiệu hàm sin có hiệu số pha là 90^0 và cắt nhau vuông góc trong biểu đồ pha. $\cos\theta$ và tương ứng $\sin\theta$ là các hệ số của chúng có thể biểu diễn tất cả các điểm tín hiệu của sóng điều chế nhiều mức bằng cách chọn thích hợp các hệ số này.

4.2. Biểu diễn điều pha QAM

Bây giờ ta xét biểu diễn cầu phương của 4-PSK. 4-PSK có thể được coi là trường hợp riêng của điều chế QAM trong đó sự thay đổi về biên độ của sóng mang bằng 0. Sóng 4-PSK E(t) có thể được biểu diễn là tổng của hai tín hiệu hình sin vuông góc với nhau như sau:

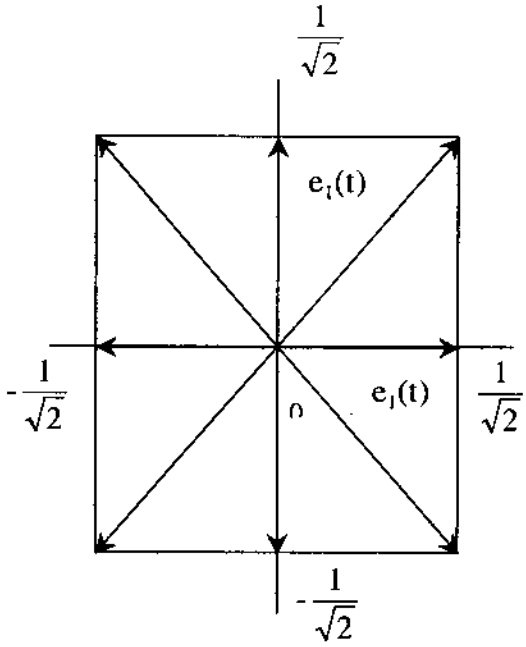
$$E(t) = e_1(t) + e_2(t) = \{S_1(t).\cos\omega_0 t + S_2(t).\sin\omega_0 t\}/\sqrt{2}$$

Bảng dưới đây cho thấy hai tín hiệu băng gốc $S_1(t)$, $S_2(t)$ và các tín hiệu tổng tạo thành từ chúng.

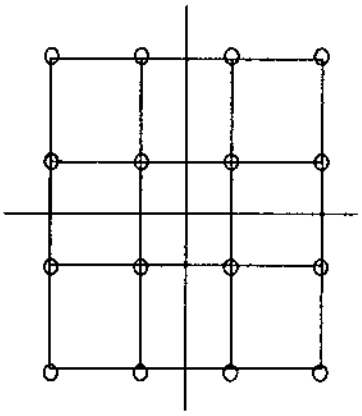
$S_1(t)$	$S_2(t)$	$e_1(t)$	$e_2(t)$	Tín hiệu tổng hợp E(t)
1	-1	$\frac{1}{\sqrt{2}}\cos\omega_c t$	$-\frac{1}{\sqrt{2}}\sin\omega_c t$	$\cos(\omega_c t + \frac{\pi}{4})$
-1	-1	$-\frac{1}{\sqrt{2}}\cos\omega_c t$	$-\frac{1}{\sqrt{2}}\sin\omega_c t$	$\cos(\omega_c t + \frac{3\pi}{4})$
-1	1	$-\frac{1}{\sqrt{2}}\cos\omega_c t$	$-\frac{1}{\sqrt{2}}\sin\omega_c t$	$\cos(\omega_c t + \frac{5\pi}{4})$
1	1	$\frac{1}{\sqrt{2}}\cos\omega_c t$	$-\frac{1}{\sqrt{2}}\sin\omega_c t$	$\cos(\omega_c t + \frac{7\pi}{4})$

Bảng 5-24: Bảng biểu diễn trực giao 4-PSK

Những tín hiệu này được mô tả trong hình 5.24a:



Hình 5.24a: Sóng 4-PSK bao gồm các tín hiệu trực giao



Hình 5.24b: Biểu đồ không gian tín hiệu 16-QAM

Nếu $e_1(t)$ và $e_2(t)$ là các sóng điều biên có hai giá trị thì có thể kết hợp chúng để tạo thành các sóng điều chế có 4 điểm tín hiệu. Vì quỹ tích véctơ của các tín hiệu bị giới hạn băng thông không dịch chuyển trên đường tròn đơn vị nên các sóng thu được khác với tín hiệu 4-PSK chuẩn. Tuy nhiên vì chúng có cùng một biên độ cố định tại thời điểm tách sóng nên có thể coi chúng như là PSK.

Tín hiệu nhận được bằng cách kết hợp hai sóng điều biên (AM) vuông góc với nhau được gọi là sóng điều chế biên độ cầu phương (QAM). Sóng QAM có hai ưu điểm, nó có thể biểu diễn được từ hai tín hiệu điều chế biên độ cơ sở và có thể chọn một điểm bất kỳ trên biểu đồ không gian tín hiệu như là một điểm tín hiệu.

4.3. QAM nhiều trạng thái

QAM cho phép sắp xếp ngẫu nhiên các điểm tín hiệu cũng như dễ dàng thực hiện điều chế và giải điều chế tín hiệu nhờ tính chất cầu phương của các tín hiệu. Ngoài ra, cách sắp xếp hình chữ nhật thường được sử dụng khi thuộc tính C/N đòi hỏi khá cao. Hình 5-24b cho thấy cách sắp xếp các điểm tín hiệu đối với điều chế biên độ cầu phương 16 mức (16 QAM). QAM nhiều trạng thái có thể được tạo thành bởi hai tín hiệu điều chế biên độ trực giao có n mức, vì vậy có 2^n điểm tín hiệu. Khi n=2 thì QAM giống hệt cách sắp xếp tín hiệu của 4-PSK. Khi n=4 thì điều chế là 16-QAM, khi n=8 hoặc 16 thì điều chế tương ứng là 64-QAM hoặc 256-QAM.

Câu hỏi ôn tập

1/ Định nghĩa và nêu nhiệm vụ của điều chế

2/ Phân loại các kỹ thuật điều chế

3/ Nêu bản chất của các kỹ thuật điều chế cao tần

4/ Nêu bản chất của các kỹ thuật điều chế xung

5/ Nêu bản chất của các kỹ thuật điều chế khóa dịch

6/ Nguyên lý điều chế FSK ?

7/ Nguyên lý điều chế PSK ?

8/ Nguyên lý điều chế ASK ?

9/ Nguyên lý điều chế QAM ?

CÁC THUẬT NGỮ VÀ TỪ VIẾT TẮT TIẾNG ANH

Amplitude Modulation - AM	: Điều chế biên độ
Amplitude Shift Key- ASK	: Điều chế khóa dịch biên độ
Carrier	: Vật mang tin
Channel	: Kênh tin
Communication	: Truyền tin
Continuous channel	: Kênh liên tục
Continuous information	: Tin liên tục
Continuous source	: Nguồn liên tục
Data	: Số liệu, dữ liệu
DeModualation	: Giải điều chế
Diferential PSK	: PSK vi phân
Discrete channel	: Kênh rời rạc
Discrete information	: Tin rời rạc
Discrete source	: Nguồn rời rạc
Elementary event	: Sự kiện cơ bản
Encoding	: Mã hóa
Event	: Sự kiện
Frequency Modulation - FM	: điều chế tần số
Frequency Shift Key- FSK	: điều chế khóa dịch tần số
Information source	: Nguồn tin
Information	: Thông tin
Measure of information	: lượng đo thông tin, lượng tin
Modulation	: Điều chế
News, Nouvelles	: Tin tức
Noise	: Tạp nhiễu
Phase coherent PSK	: PSK phối hợp pha

Phase Locked Loop - PLL	: Vòng khóa pha
Phase Modulation - PM	: điều chế góc pha
Phase Shift Key - PSK	: điều chế khóa dịch pha
Prefix	: tiếp đầu ngữ, phần đầu
Public Switched Telephone Network-PSTN	: mạng điện thoại công cộng
Pulse Amplitude Modulation - PAM	: điều chế biên độ xung
Pulse Code Modulation - PCM	: điều chế mã xung
Pulse Frequency Modulation - PFM	: điều chế tần số xung
Pulse Phase Modulation - PPM	: điều chế góc pha xung
Pulse Width Modulation - PWM	: điều chế độ rộng xung
Quadrature Amplitude Modulation - QAM	: điều chế biên độ cầu phương
Sample space	: Không gian mẫu
Sensor	: Cảm biến
Signal	: Tín hiệu
Sink	: Thu tin
Telecommunication	: Hệ thống viễn thông
Uniquely decodable code	: Mã phân tách được
Uniquely undecodable code	: Mã không phân tách được

TÀI LIỆU THAM KHẢO

1. ***Cơ sở lý thuyết truyền tin.*** Tác giả Bùi Minh Tiêu. Nhà xuất bản Đại học, 1985

2. ***Lý thuyết mã.*** Tác giả Nguyễn Thúy Vân. Nhà xuất bản Khoa học kỹ thuật, 1999

3. ***Cơ sở lý thuyết truyền tin.*** Tác giả Hà Quốc Trung. Giáo trình điện tử - Khoa Điện tử viễn thông - Đại học Bách khoa Hà Nội

4. ***Cơ sở lý thuyết truyền tin (2 tập).*** Tác giả Đặng Văn Chuyết và Nguyễn Tuấn Anh. Nhà xuất bản Giáo dục, 2000

5. ***Kỹ thuật truyền số liệu.*** Tác giả Nguyễn Hồng Sơn và Hoàng Đức Hải. Nhà xuất bản Lao động - Xã hội , 2002

6. ***Lý thuyết truyền tin*** . Tác giả Trần Trung Dũng và Nguyễn Thuý Anh. Nhà xuất bản Khoa học và kỹ thuật, 2001

7. ***Giáo trình Xác suất thống kê*** . Tác giả Tống Đình Quỳ. Nhà xuất bản Giáo dục, 1999.

8. ***Digital Communication.*** John G. Proakis, McGraw - Hill International Editions, 1995

9. ***Data and Computer Communication.*** Tác giả Wiliam Stallings. Nhà xuất bản Maxwell MacMillan International, 1991

10. ***Analog and Digital Communication.*** W. David Gregg, John Whiley & Sons, 1977

MỤC LỤC

NHÀ XUẤT BẢN HÀ NỘI
SỐ 4 - TỐNG DUY TÂN, QUẬN HOÀN KIẾM, HÀ NỘI
ĐT: (04) 8252916 - FAX: (04) 9289143

GIÁO TRÌNH

CƠ SỞ LÝ THUYẾT TRUYỀN TIN

NHÀ XUẤT BẢN HÀ NỘI - 2007

Chịu trách nhiệm xuất bản:
NGUYỄN KHẮC OÁNH
Biên tập:
PHẠM QUỐC TUẤN
Bìa:
TRẦN QUANG
Kỹ thuật vi tính:
NGUYỄN HẰNG
Sửa bản in
PHẠM QUỐC TUẤN

In 550 cuốn, khổ 17x24cm, tại Công ty Cổ phần in Khoa học Kỹ thuật. Quyết định xuất bản: 160 - 2007/CXB/444GT - 27/HN, số: 313/CXB ngày 02/3/2007. In xong và nộp lưu chiểu quý III/2007.

BỘ GIÁO TRÌNH XUẤT BẢN NĂM 2007
KHỐI TRƯỜNG TRUNG HỌC CÔNG NGHIỆP

1. THỰC TẬP QUA BAN HÀN
2. THỰC TẬP QUA BAN NGUỘI
3. THỰC TẬP QUA BAN MÁY
4. AN TOÀN LAO ĐỘNG CHUYÊN NGÀNH SCKTTB
5. AN TOÀN LAO ĐỘNG CHUYÊN NGÀNH ĐIỆN
6. VẬT LIỆU ĐIỆN
7. ĐO LƯỜNG ĐIỆN
8. KỸ THUẬT ĐIỆN
9. ĐIỆN TỬ CÔNG SUẤT
10. MÁY CÔNG CỤ CẮT GỌT
11. ĐỒ GÁ
12. CÔNG NGHỆ CHẾ TẠO MÁY
13. TỔ CHỨC SẢN XUẤT
14. MÁY VÀ LẬP TRÌNH CNC
15. CẮT GỌT KIM LOẠI
16. SỬA CHỮA MÁY CÔNG CỤ
17. MÁY ĐIỆN
18. TRUYỀN ĐỘNG ĐIỆN
19. KHÍ CỤ ĐIỆN - TRANG BỊ ĐIỆN
20. CUNG CẤP ĐIỆN
21. KỸ THUẬT ĐIỀU KHIỂN LOGÍC VÀ ỨNG DỤNG
22. ĐỒ ÁN CÔNG NGHỆ CTM
23. THỰC HÀNH CẮT GỌT KIM LOẠI
24. THỰC HÀNH SỬA CHỮA THIẾT BỊ
25. THÍ NGHIỆM KỸ THUẬT ĐIỆN
26. THÍ NGHIỆM MÁY ĐIỆN
27. THỰC TẬP ĐIỆN CƠ BẢN
28. TIẾNG ANH CHUYÊN NGÀNH SCKTTB
29. TIẾNG ANH CHUYÊN NGÀNH ĐIỆN
30. QUẢN TRỊ DOANH NGHIỆP
31. HƯỚNG DẪN ĐỒ ÁN TRANG BỊ ĐIỆN
32. ĐỒ ÁN CUNG CẤP ĐIỆN
33. CƠ SỞ THIẾT KẾ MÁY
34. ĐỒ ÁN CƠ SỞ THIẾT KẾ MÁY (ĐỒ ÁN CHI TIẾT MÁY)
35. CẤU TRÚC DỮ LIỆU VÀ GIẢI THUẬT
36. LÝ THUYẾT TRUYỀN TIN
37. CƠ SỞ KỸ THUẬT TRUYỀN SỐ LIỆU
38. ASSEMBLY
39. THỰC TẬP CHUYÊN NGÀNH ĐIỆN
40. THỰC HÀNH PLC
41. FOXPRO

Giá: 18.000 đ