



GIỚI THIỆU CHUNG VỀ LÝ THUYẾT THÔNG TIN

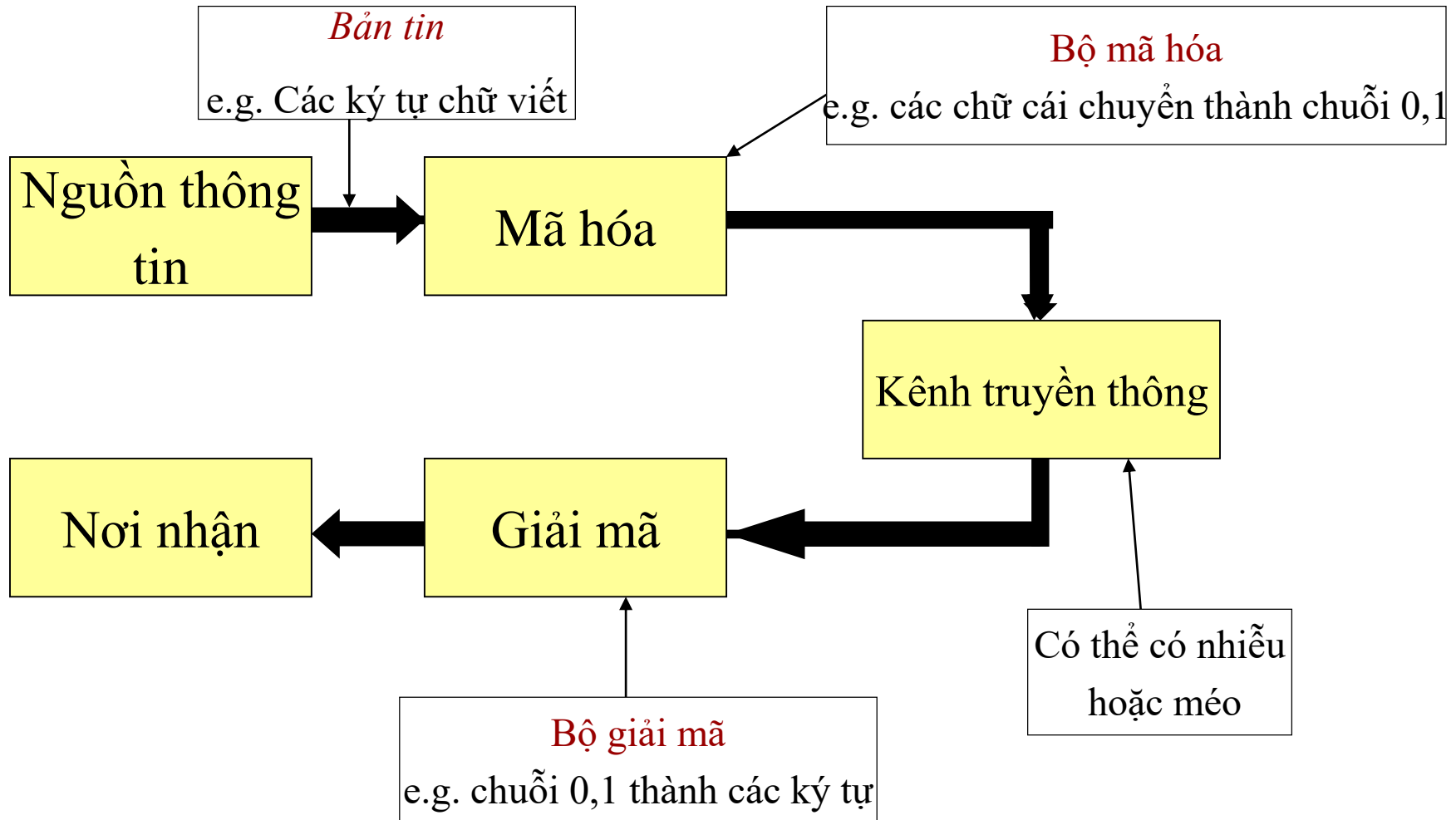


CHA ĐẸ CỦA TRUYỀN THÔNG SỐ

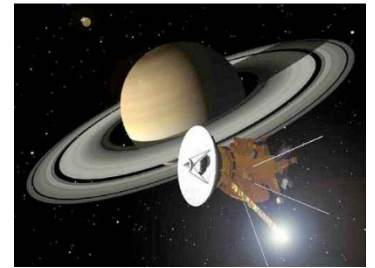


Nguồn gốc của truyền thông kỹ thuật số hiện đại xuất phát từ bài viết mang tính đột phá “Lý thuyết toán học về truyền thông” của Claude Elwood Shannon vào năm 1948.

MÔ HÌNH HỆ THỐNG TRUYỀN THÔNG SỐ



CÁC KÊNH TRUYỀN THÔNG



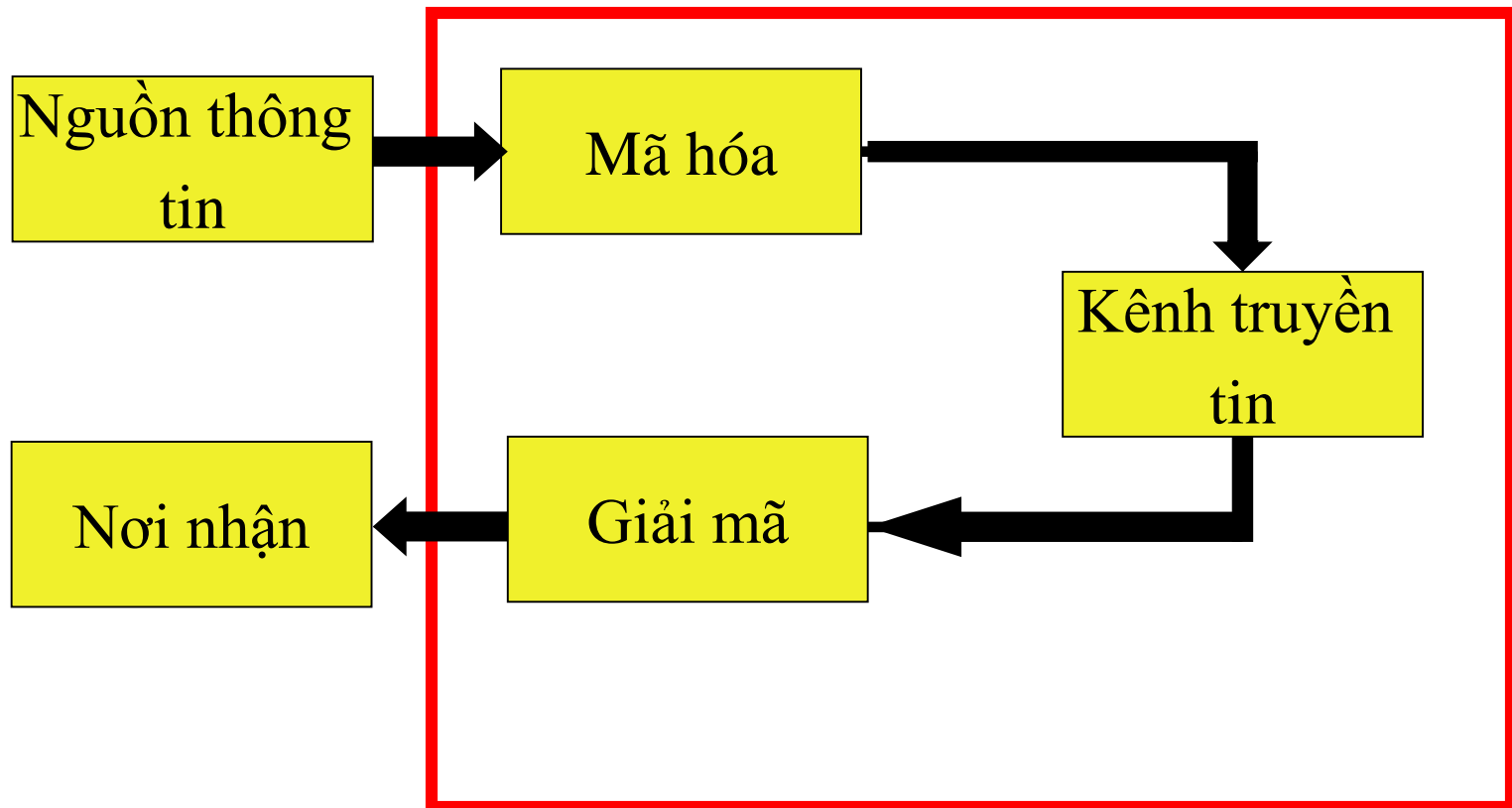
ĐỊNH NGHĨA CỦA SHANNON VỀ TRUYỀN THÔNG



“Bài toán cơ bản của truyền thông là việc tái tạo một bản tin ở điểm này tại một điểm khác một cách chính xác hoặc xấp xỉ”

SHANNON MUỐN...

Shannon muốn tìm ra cách truyền dữ liệu một cách “tin cậy” qua kênh với tốc độ “lớn nhất” có thể.



VÀ KHI NGHĨ VỀ BÀI TOÁN NÀY...

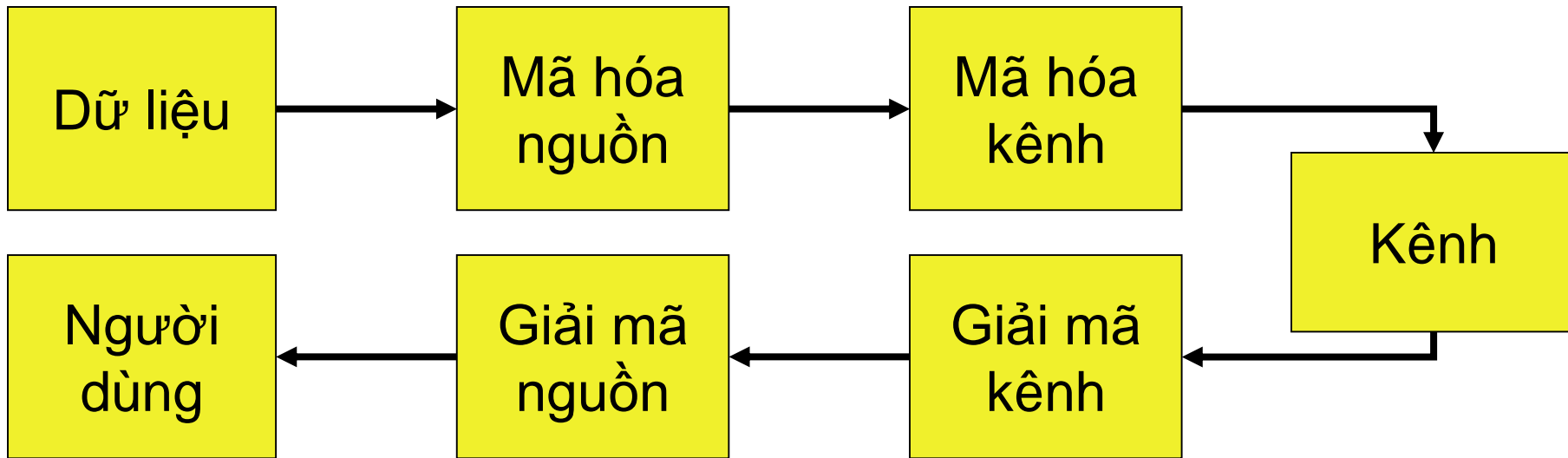


Ông đã tìm ra giải pháp và giới thiệu trong bài báo năm 1948.

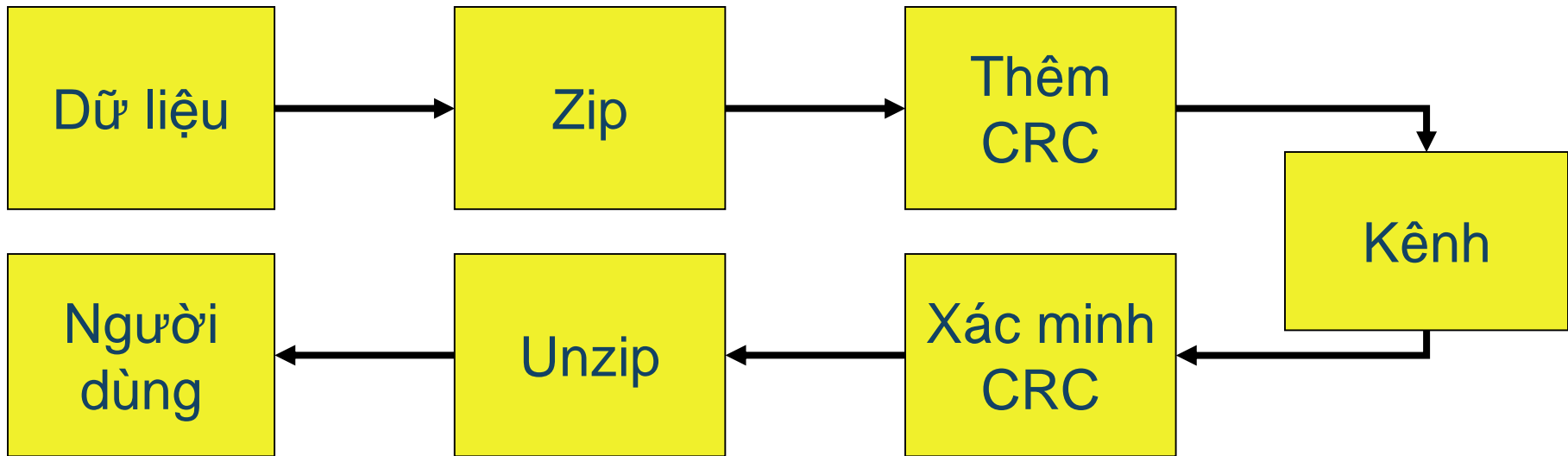


Trong bài báo năm 1948, ông xây dựng một lý thuyết phong phú về vấn đề truyền thông tin cậy, bây giờ được gọi là "**Lý thuyết thông tin**" hoặc "**Lý thuyết Shannon**" để tôn vinh ông.

CÁCH NHÌN CỦA SHANNON



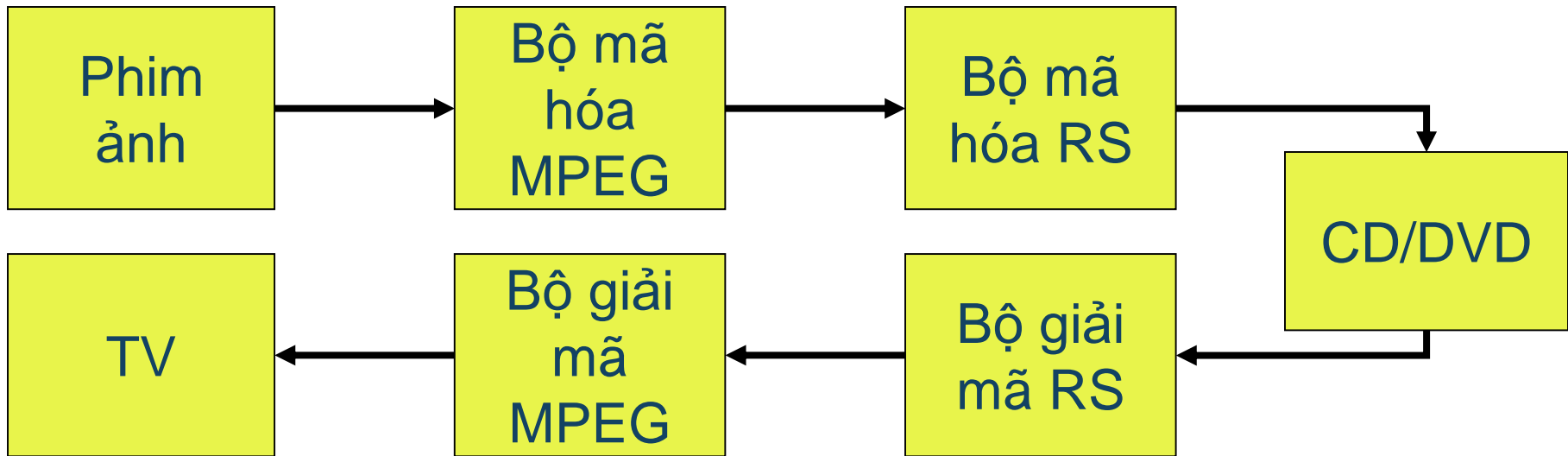
VÍ DỤ: Ổ ĐĨA LƯU TRỮ



THEO THUẬT NGỮ LÝ THUYẾT THÔNG TIN

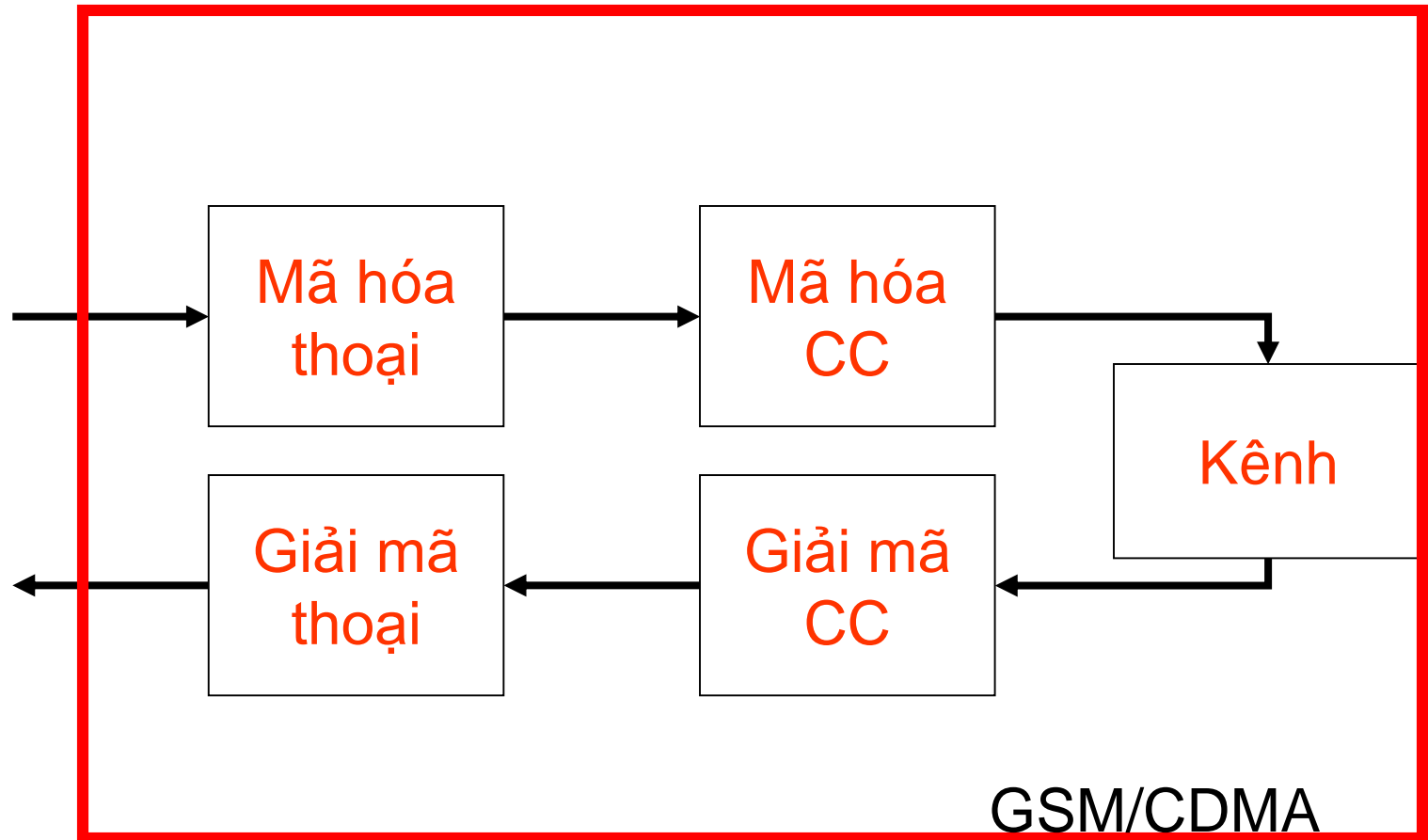
Zip	=	Mã hóa nguồn	Nén dữ liệu
Unzip	=	Giải mã nguồn	Giải nén dữ liệu
Thêm CRC	=	Mã hóa kênh	Phát hiện lỗi
Xác minh CRC	=	Giải mã kênh	Hiệu chỉnh lỗi

VÍ DỤ: VCD VÀ DVD



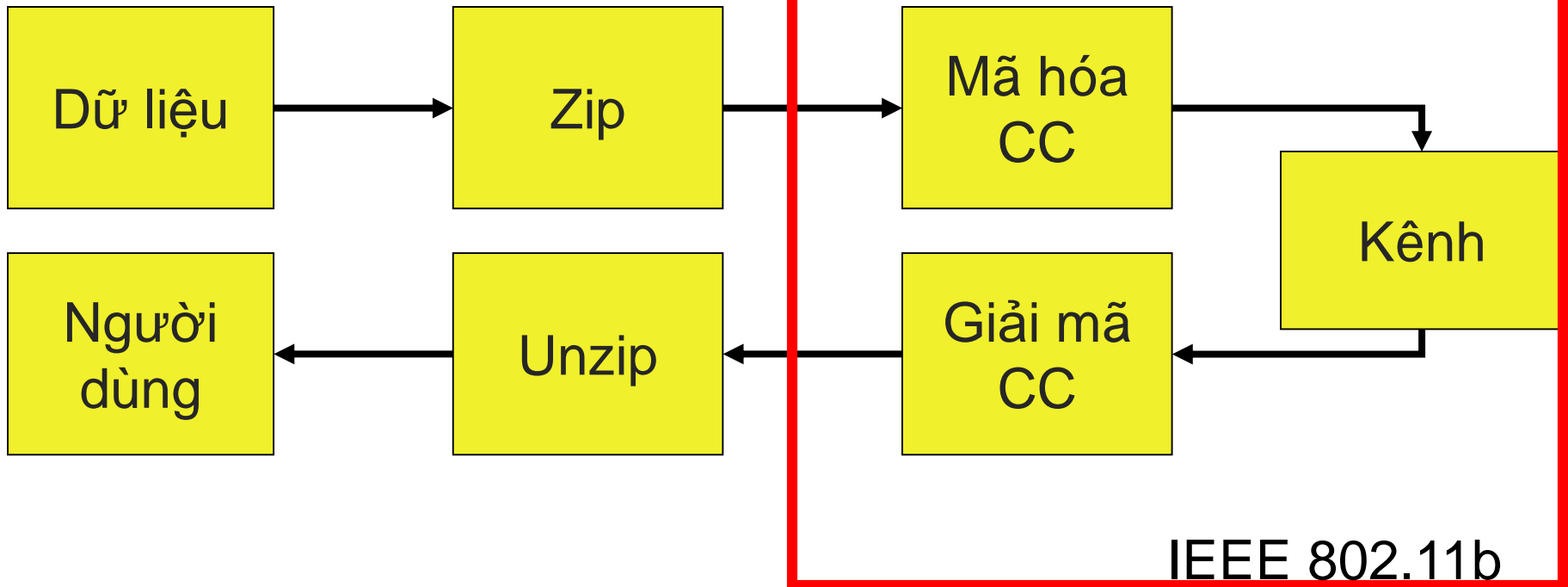
RS: Mã Reed-Solomon

VÍ DỤ: ĐIỆN THOẠI DI ĐỘNG



CC (Convolutional Code): Mã chập

VÍ DỤ: WLAN IEEE 802.11B



CC: Mã chập

LÝ THUYẾT SHANNON

Lý thuyết Shannon năm 1948 bao gồm:

1. Đo lường thông tin
2. Lý thuyết mã hóa nguồn
3. Lý thuyết mã hóa kênh

ĐO LƯỜNG THÔNG TIN

Câu hỏi đầu tiên của Shannon là:

“Làm thế nào để đo thông tin theo đơn vị bit?”



= ? bits



= ? bits

TẤT CẢ CÁC SỰ KIỆN ĐỀU CÓ XÁC SUẤT

Sử dụng lý thuyết xác suất, Shannon đã chứng minh rằng chỉ có một cách để đo thông tin bằng đơn vị bit:

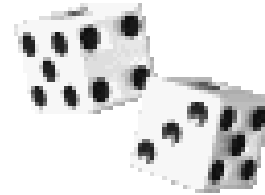
$$H(X) = -\sum_x p(x) \log_2 p(x)$$

$H(X)$ được gọi là hàm **entropy**

VÍ DỤ

Tung một viên xúc xắc:

- Đầu ra có thể là: 1,2,3,4,5,6
- Mỗi đầu ra có thể xảy ra với xác suất $1/6$
- Thông tin nhận được mỗi lần tung xúc xắc là:



$$\begin{aligned} H &= -\sum_{i=1}^6 p(i) \log_2 p(i) = -\sum_{i=1}^6 p(i) \log_2 p(i) \\ &= -\sum_{i=1}^6 \frac{1}{6} \log_2 \frac{1}{6} = \log_2 6 = 2.585 \text{ bits} \end{aligned}$$



WAIT!
KHÔNG HỢP LÝ!

Số 2.585-bits không phải là số nguyên!!

Ý nghĩa của nó là gì??

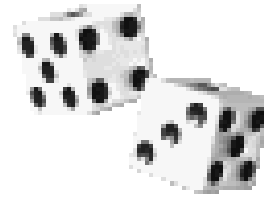
ĐỊNH LÝ MÃ HÓA NGUỒN THỨ NHẤT CỦA SHANNON



Shannon đã chứng minh rằng:

“Để lưu giữ một cách tin cậy lượng thông tin được tạo ra bởi nguồn ngẫu nhiên X nào đó, bạn cần tối thiểu, về mặt trung bình, $H(X)$ bits cho mỗi đầu ra.”

Ý NGHĨA:



Nếu tung viên xúc xắc 1,000,000 lần và ghi lại giá trị cho mỗi lần thử:

1,3,4,6,2,5,2,4,5,2,4,5,6,1,....

Về nguyên tắc, cần 3 bit để lưu giữ cho mỗi kết quả đầu ra vì 3 bit có thể biểu diễn được 8 trường hợp. Vì vậy, chúng ta cần **3,000,000** bits để lưu giữ toàn bộ lượng thông tin này.

Sử dụng cách biểu diễn ASCII, máy tính cần **8 bits=1 byte** để lưu giữ cho mỗi kết quả đầu ra.

File chứa kết quả sẽ có dung lượng **8,000,000** bits

NHƯNG SHANNON NÓI RẰNG...

Bạn chỉ cần 2.585 bits để lưu trữ cho một đầu ra.




Vì vậy, dung lượng file có thể được nén lại với kích thước:

$$2.585 \times 1,000,000 = 2,585,000 \text{ bits}$$

Tỷ lệ nén tối ưu đạt được là:

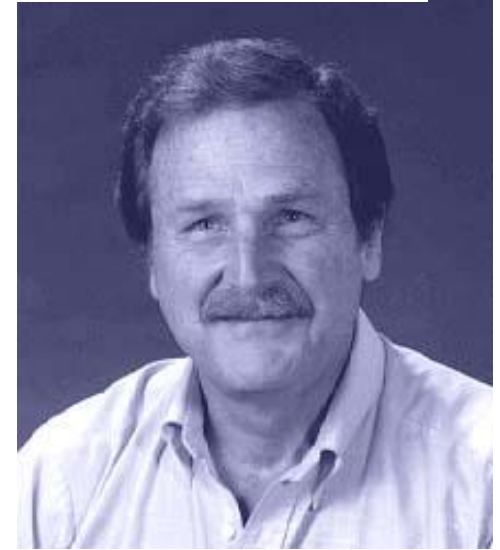
$$\frac{8,000,000}{2,581,000} = 3.09 \quad \frac{2,585,000}{8,000,000} = 0.3231 = 32.31\%$$

HÃY LÀM THỬ MỘT SỐ BÀI KIỂM TRA

	Kích thước file	Tỷ lệ nén
Không nén	8,000,000 bits	100%
Shannon 	2,585,000 bits	32.31%
Winzip 	2,930,736 bits	36.63%
WinRAR 	2,859,336 bits	35.74%

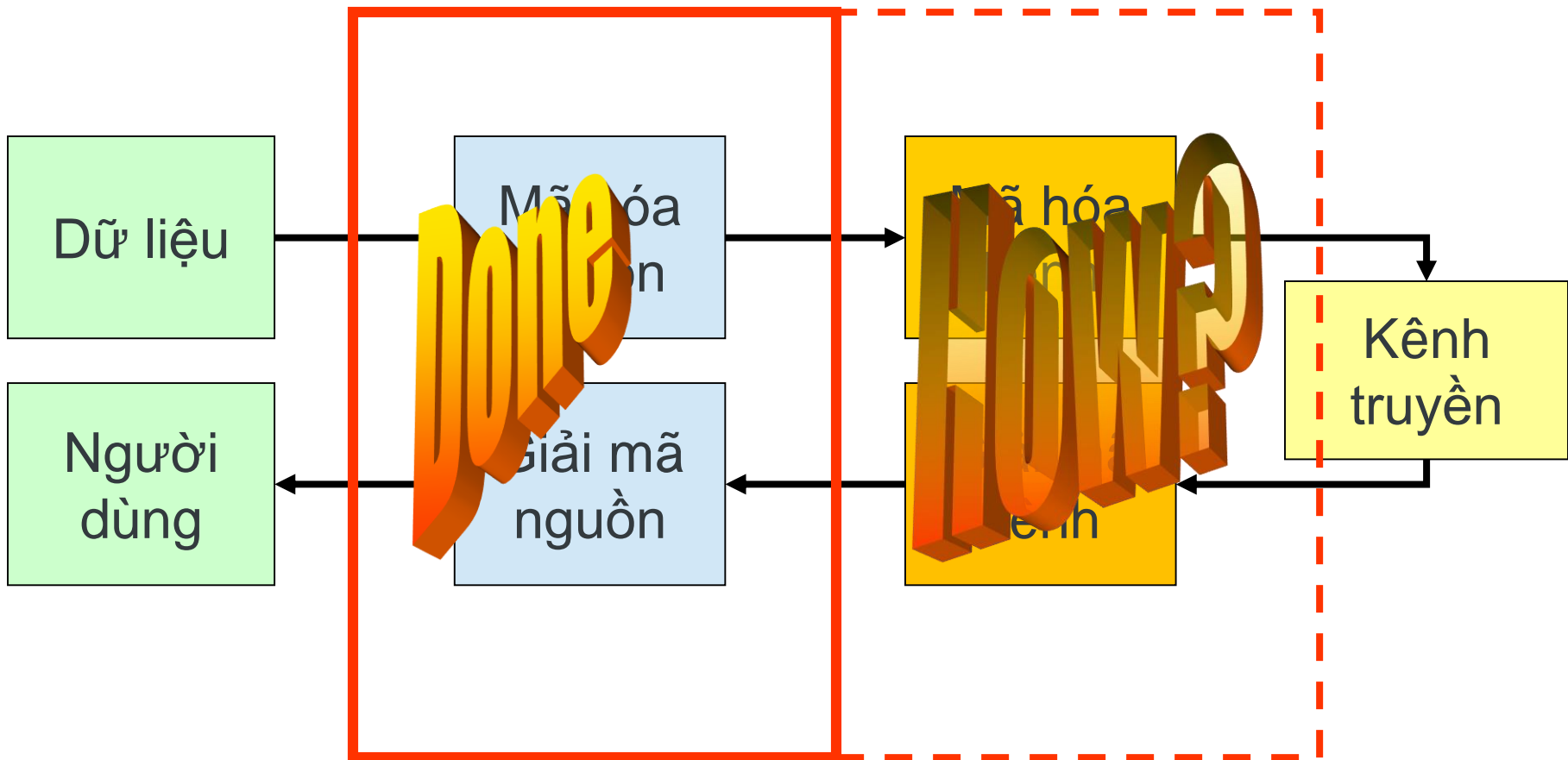
DAVID HUFFMAN

Sau đó vào năm 1952, **David Huffman**, khi là một sinh viên sau đại học MIT, trình bày một phương pháp hệ thống để đạt được tỷ lệ nén tối ưu được đưa ra bởi Shannon. Do đó, kỹ thuật mã hóa được gọi là "**mã Huffman**" để vinh danh thành tích của ông. Mã Huffman được sử dụng trong hầu hết các ứng dụng liên quan đến việc nén và truyền dữ liệu số, chẳng hạn như máy fax, modem, mạng máy tính và truyền hình độ nét cao (HDTV).



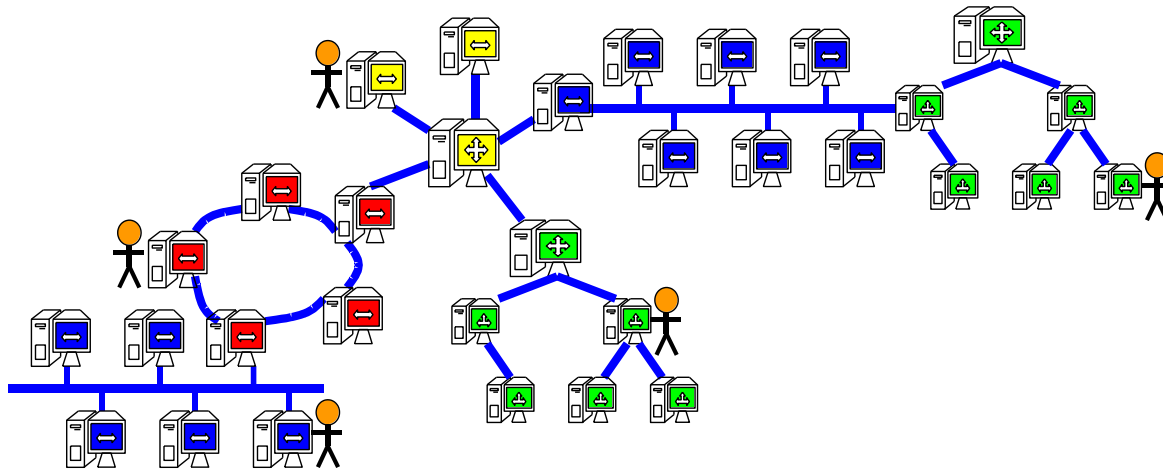
(1925-1999)

CHO ĐẾN NAY



TRƯỜNG HỢP ĐƠN GIẢN NHẤT: MẠNG MÁY TÍNH

Truyền thông qua mạng máy tính, ví dụ Internet

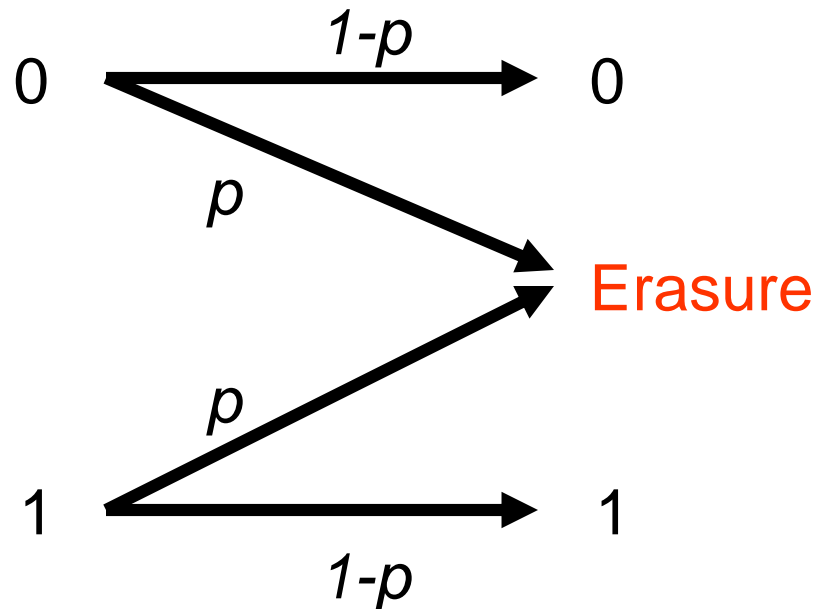


Nhược điểm lớn nhất của kênh ở đây là:

Mất gói

KÊNH NHỊ PHÂN CÓ XÓA

Các sai lỗi như “mất gói” có thể được xem xét như kênh có xóa (**Erasures**). Các dữ liệu bị xóa nghĩa là chúng bị mất trong quá trình truyền.



p tỉ lệ tổn thất gói trong mạng này

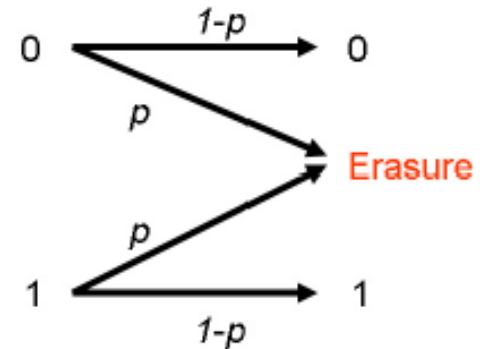
- Khi ký tự nhị phân bị xóa, nó không thể khôi phục lại được....

Ví dụ:

- Alice gửi **0,1,0,1,0,0** tới Bob
- Nhưng chất lượng mạng kém nên Bob nhận được **0,?,0,?,0,0**

- Vì vậy, Bob yêu cầu Alice gửi lại.
- Lần này Bob chỉ nhận được **0,?,?,1,0,0**
- **Và Bob vẫn không thể giải mã được!**

- Vậy Alice có thể làm gì?
- Sẽ ra sao nếu Alice gửi:
0000,1111,0000,1111,0000,0000
Lặp lại các ký tự truyền 4 lần!



GIẢI PHÁP NÀY CÓ ĐỦ TỐT?

Giờ đây Alice gửi: 0000,1111,0000,1111,0000,0000

Vẫn có trường hợp Bob không thể đọc được bản tin Alice gửi, ví dụ:

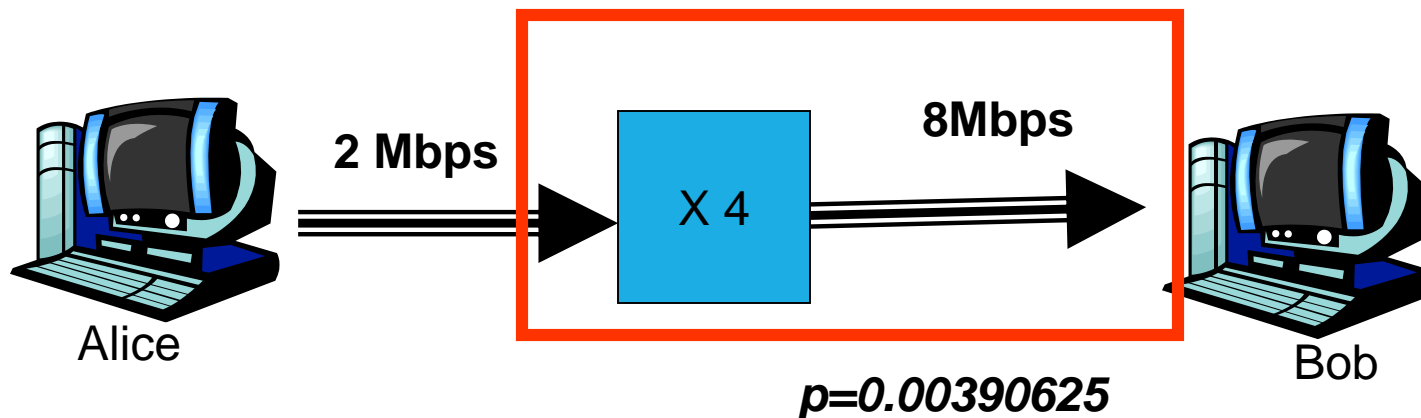
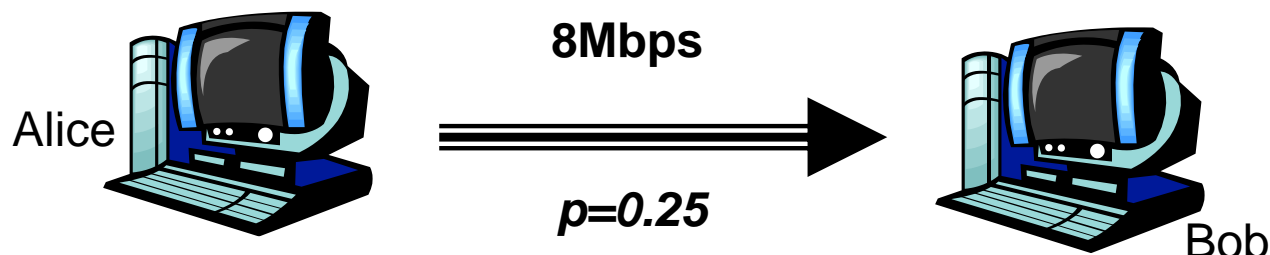
????,1111,0000,1111,0000,0000

tất cả 4 ký tự đều bị xóa.

Tuy nhiên điều này chỉ xảy ra với xác suất p^4

Vì vậy nếu mạng gốc có tỉ lệ mất gói là $p=0.25$, bằng cách lặp lại 4 lần mỗi ký tự, hệ thống mới sẽ có tỉ lệ mất gói là $p^4=0.00390625$

Nhưng nếu tốc độ dữ liệu trong mạng gốc là 8Mb/s thì với việc lặp lại, Alice chỉ có thể truyền với tốc độ 2 Mb/s



THÁCH THỨC MỚI:



Liệu mã lặp lại có phải là giải pháp tốt nhất Alice có thể làm?



• • • • •

ĐỊNH LÝ MÃ HÓA KÊNH CỦA SHANNON



Shannon cho rằng:

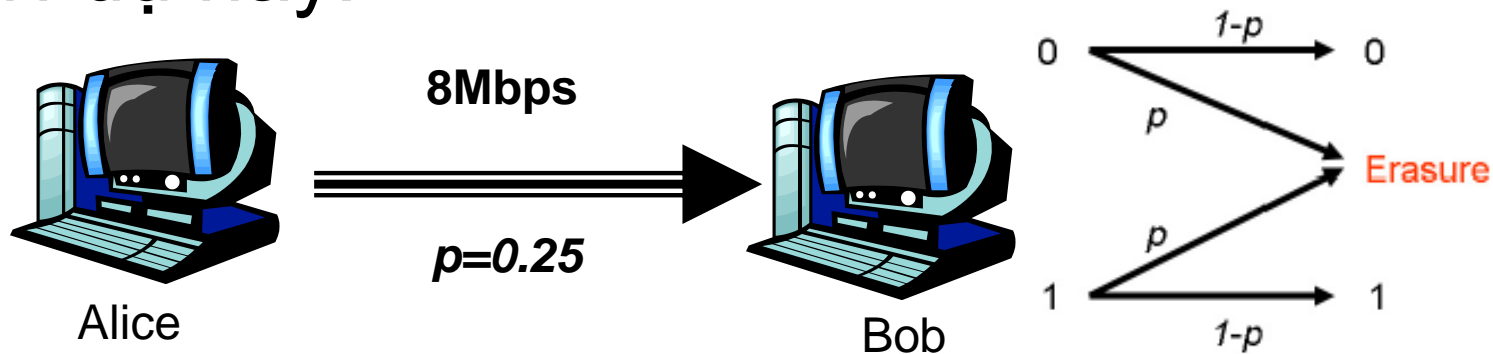
“Hãy cho tôi một kênh và tôi có thể tính một đại lượng gọi là dung lượng kênh C cho kênh đó. Khi đó để đảm bảo việc truyền thông tin cậy thì tốc độ dữ liệu của bạn phải nhỏ hơn C ”



NỘI DUNG ĐỊNH LÝ CỦA
SHANNON LÀ GÌ?

SHANNON CHO RẰNG:

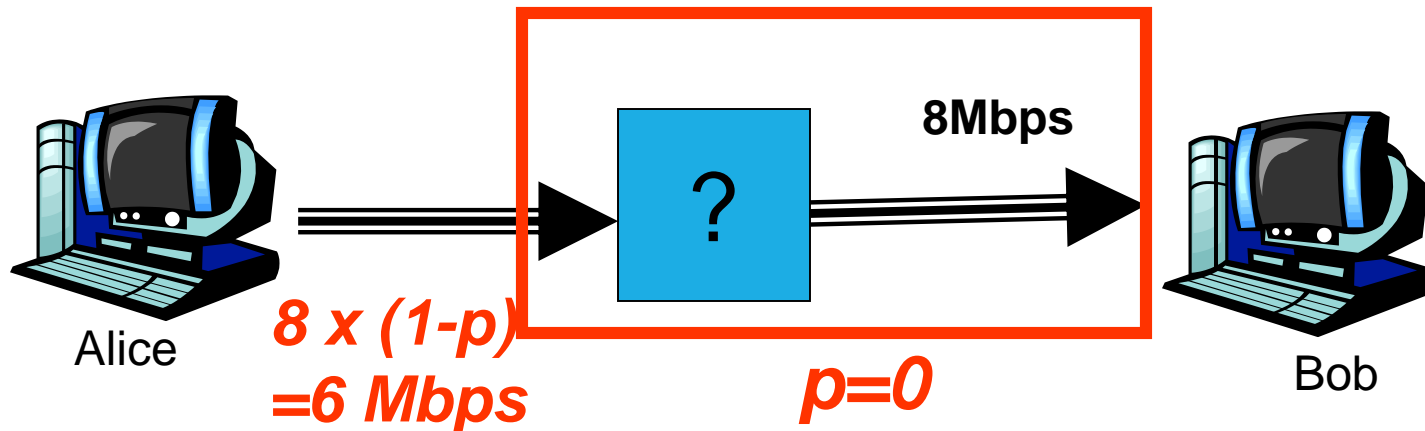
Trong ví dụ này:



Tính toán dung lượng kênh

$$C=1-p=0.75$$

Và sẽ tồn tại một phương pháp mã hóa sao cho:



THẬT KHÔNG MAY...



Shannon không biết chính xác phải làm như thế nào?

CÁC KẾT QUẢ TRONG HƠN 50 NĂM

- Hamming codes
- Convolutional codes,
- Concatenated codes,
- Low density parity check (LDPC) codes
- Reed-Muller codes
- Reed-Solomon codes,
- BCH codes,
- Finite Geometry codes,
- Cyclic codes,
- Golay codes,
- Goppa codes
- Algebraic Geometry codes,
- Turbo codes
- Zig-Zag codes,
- Accumulate codes and Product-accumulate codes,
- ...

Các kết quả này đã rất gần với định lý mà Shannon đã phát biểu cách đây 50 năm! 😊

HIỆN NAY...

Định lý mã hóa nguồn đã được ứng dụng trong:

JPEG
2000

Nén ảnh

MPEG

Nén
Audio/Video



Nén dữ liệu

MP3

Nén audio

Định lý mã hóa kênh được ứng dụng trong:

- VCD/DVD – Mã Reed-Solomon
- Truyền thông không dây– Mã chập
- Truyền dẫn quang – Mã Reed-Solomon
- Mạng máy tính– Mã LT codes, Mã Raptor
- Truyền thông không gian

LÝ THUYẾT SHANNON THEORY CŨNG ĐƯỢC ÁP DỤNG TRONG TRUYỀN THÔNG KHÔNG GIAN

Năm 1965, Mariner 4:



Tần số = 2.3GHz (S Band)

Tốc độ dữ liệu = 8.33 bps

Không mã hóa nguồn

Mã lặp lại (2 x)

Năm 2004, Mars Exploration Rovers:



Tần số = 8.4 GHz (X Band)

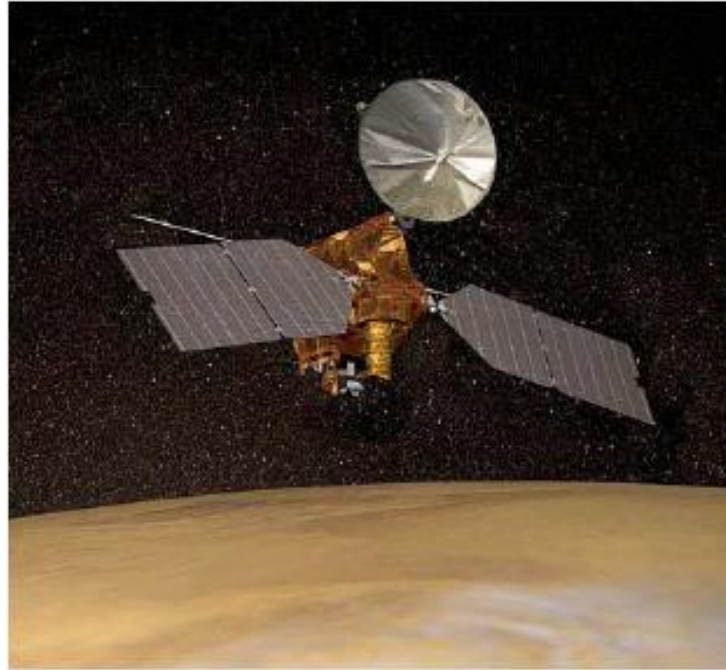
Tốc độ dữ liệu = 168K bps

Nén ICER có tổn thất

với tỷ lệ 12:1

Mã ghép nối

2006, MARS RECONNAISSANCE ORBITER



Truyền dẫn
nhanh hơn



Tần số = 8.4 GHz (X Band)
Tốc độ dữ liệu = 12 M bps
Nén FELICS không tổn thất 2:1
Mã Turbo tại khoảng cách
 2.15×10^8 Km

LÝ THUYẾT THÔNG TIN ĐÃ ĐƯỢC ỨNG DỤNG TRONG...

Tất cả các loại truyền dẫn,
Thị trường chứng khoán, kinh tế
Lý thuyết trò chơi,
Vật lý lượng tử,
Mật mã học,
....