

Bài 3: Số hóa tận dụng sức mạnh của máy tính

Nội dung bài học

- Tận dụng sức mạnh của máy tính trong xử lý dữ liệu, tạo lập thông tin, và phát triển tri thức.
 - Các khái niệm cơ bản: dữ liệu, thông tin, tri thức.
- Làm thế nào để có dữ liệu số - điều kiện cần để sử dụng máy tính xử lý dữ liệu:
 - Mối quan hệ giữa dữ liệu và tín hiệu
 - Biểu diễn tín hiệu: miền thời gian và miền tần số
 - Vai trò của Biến đổi Fourier
 - Quy trình tạo dữ liệu số:
 - Rời rạc hóa (sampling): Định lý Nyquist-Shannon
 - Lượng tử hóa (quantization)
 - Mã hóa (coding)
- Các lợi ích của “số hóa”.

- Máy tính điện tử như hiện có được xây dựng dựa trên nền tảng:
 - Đại số Bool: chỉ với hai giá trị True (1) và False (0) – tương ứng trạng thái bật/tắt của một chiếc “công-tắc” – cùng với các phép logic cơ bản là có thể thực hiện biểu diễn và tính toán số học;
 - “Công-tắc”: và với cách “nối mạng, đi dây” nối các công-tắc lại là ta có thể thực hiện được việc:
 - Biểu diễn: Bật là 1, và Tắt là 0;
 - Tính toán: tạo ra các mạch AND/OR/XOR/NOT, từ đó tạo ra các bộ CỘNG, TRỪ, NHÂN, CHIA phục vụ tính toán số học;
 - Vì vậy, cần một loại “Công-tắc” bật/tắt có điều khiển và NHANH, TIN CẬY, TIÊU THỤ ÍT NĂNG LƯỢNG
=> Bóng bán dẫn, hay còn gọi là TRANSISTOR

Tóm tắt “Sức mạnh của máy tính”!!!

- **Tốc độ tính toán:** máy tính cho phép thực hiện các tác vụ rất nhanh, nhiều tác vụ cùng một lúc (đa nhiệm), và xử lý một lượng lớn dữ liệu trong thời gian ngắn.
- **Tính đa dụng:** Máy tính có thể thực hiện một cách đa dạng các tác vụ (được chỉ thị thông qua phần mềm), trong nhiều lĩnh vực, đặc biệt các tác vụ có tính “lặp đi, lặp lại” một cách rất hiệu quả. Tác vụ máy tính có thể xử lý được phụ thuộc chủ yếu và sức sáng tạo của con người.
- **Độ chính xác và tin cậy:** máy tính thực hiện công việc một cách tinh cậy và chính xác, tuân theo “chỉ thị” được thiết lập (phần mềm).
- **Khả năng lưu trữ:** máy tính có khả năng lưu trữ lượng lớn dữ liệu và sẵn sàng để xử lý bất kỳ lúc nào.

Làm thế nào có thể tận dụng được sức mạnh của MTĐT?

- Dữ liệu số, và
- Phần mềm xử lý phù hợp.



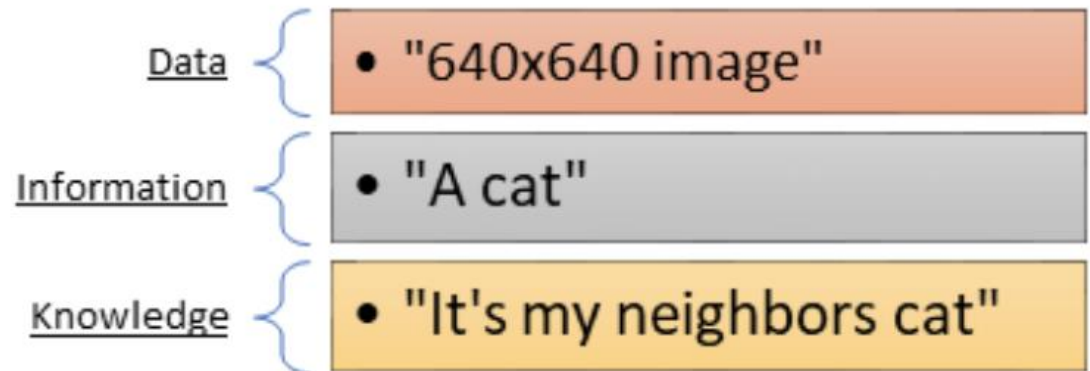
- **Dữ liệu (data):** là tập hợp các giá trị rời rạc hoặc liên tục chứa thông tin trong đó. Các giá trị này mô tả: số lượng, chất lượng, sự kiện, số liệu thống kê... về một sự vật/sự việc.
Ví dụ: nhiệt độ thế giới trong năm 2016.
- **Thông tin (information):** xử lý dữ liệu cho chúng ta thông tin, phần có ích của dữ liệu. Hay nói cách khác thông tin là kết quả thu được dưới dạng ý nghĩa (meaning), và bối cảnh (context) sau khi xử lý dữ liệu.
Ví dụ: dựa trên thống kê nhiệt độ trong năm 2016, ta tính toán trung bình và so sánh với quá khứ thì thấy: đây là một năm có nhiệt độ cao nhất từ trước đến nay.

- **Tri thức (knowledge):** thu thập, và xử lý thông tin cho chúng ta tri thức, phần hữu dụng nhất của thông tin.

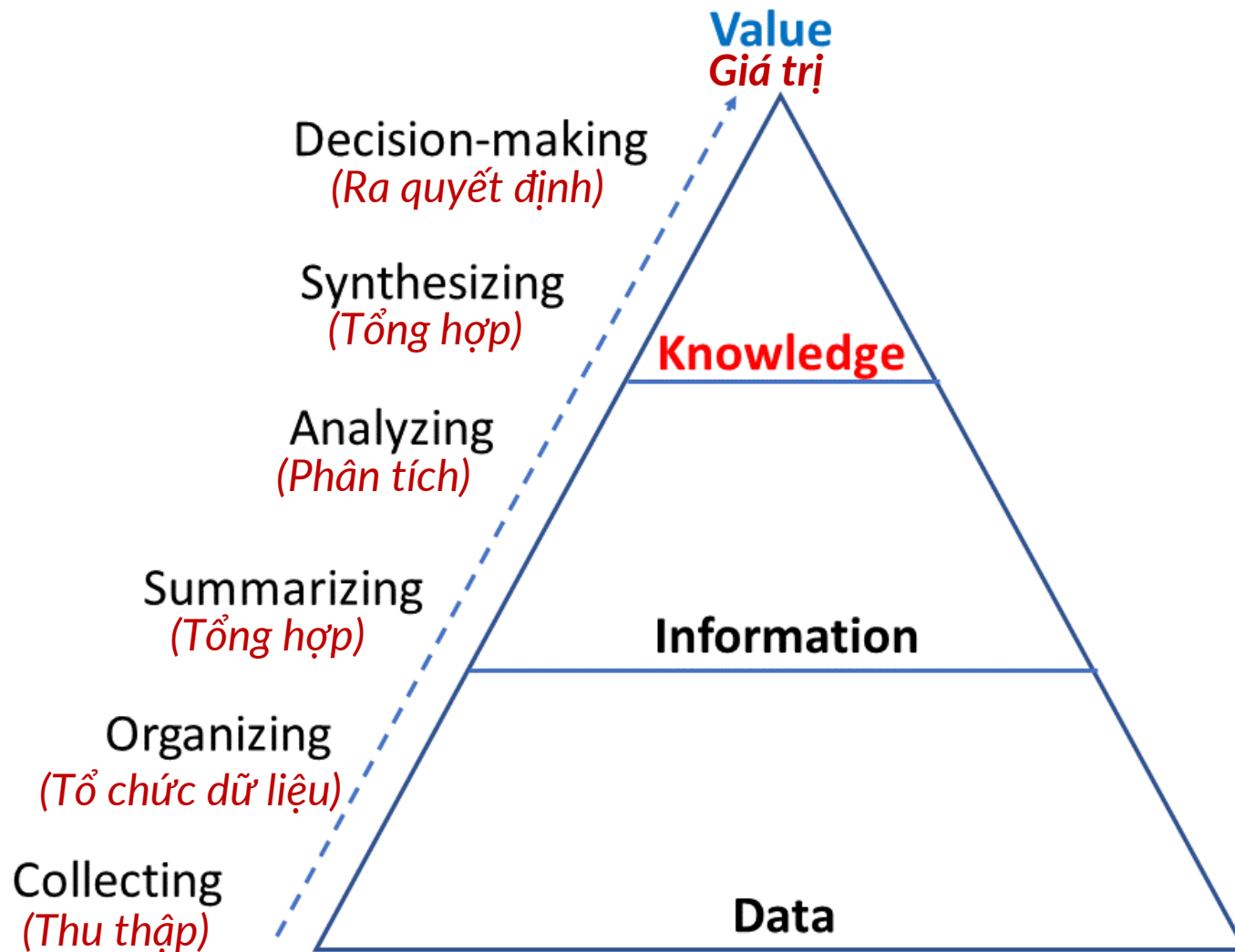
Hay nói cách khác Tri thức là sự hiểu biết về một chủ đề mà bạn có được nhờ kinh nghiệm hoặc qua quá trình học tập, nghiên cứu.

Ví dụ: năm 2016 là năm xảy ra hiện tượng El Niño, các năm trước có hiện tượng El Niño cũng nóng hơn bình thường, do vậy tri thức tạo ra: Hiện tượng El Niño làm tăng nhiệt độ trên Trái Đất.

Một ví dụ khác

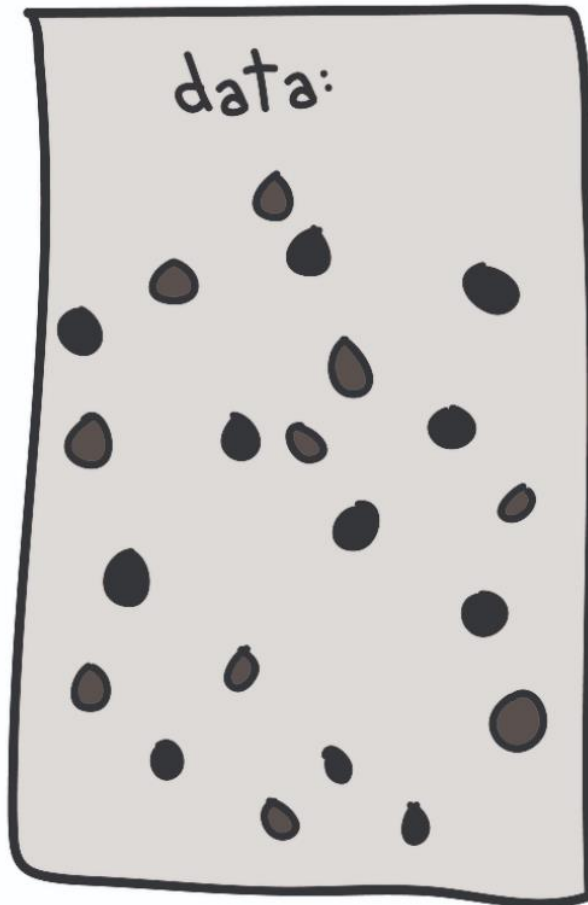


Tháp dữ liệu – thông tin – tri thức – giá trị

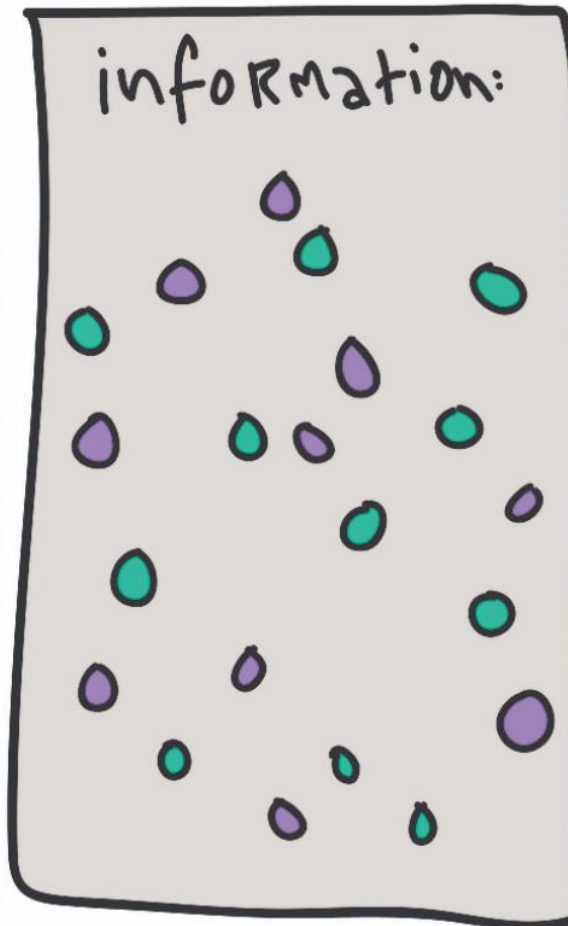


So sánh Dữ liệu – Thông tin – Tri thức

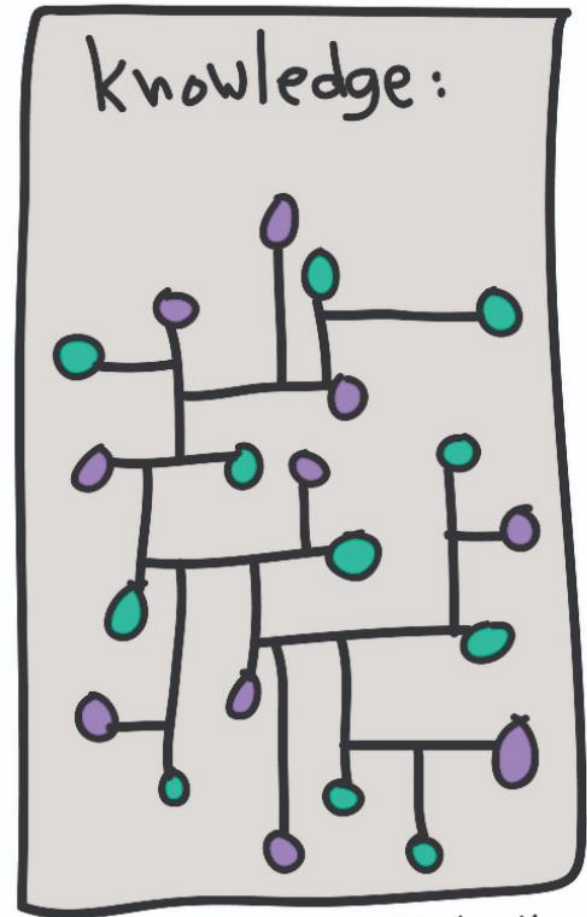
- Xuất phát điểm từ dữ liệu thô



@bestqualitycrab



@gapingvoid



@gapingvoid

THE
DATA

INFORMATION

KNOWLEDGE

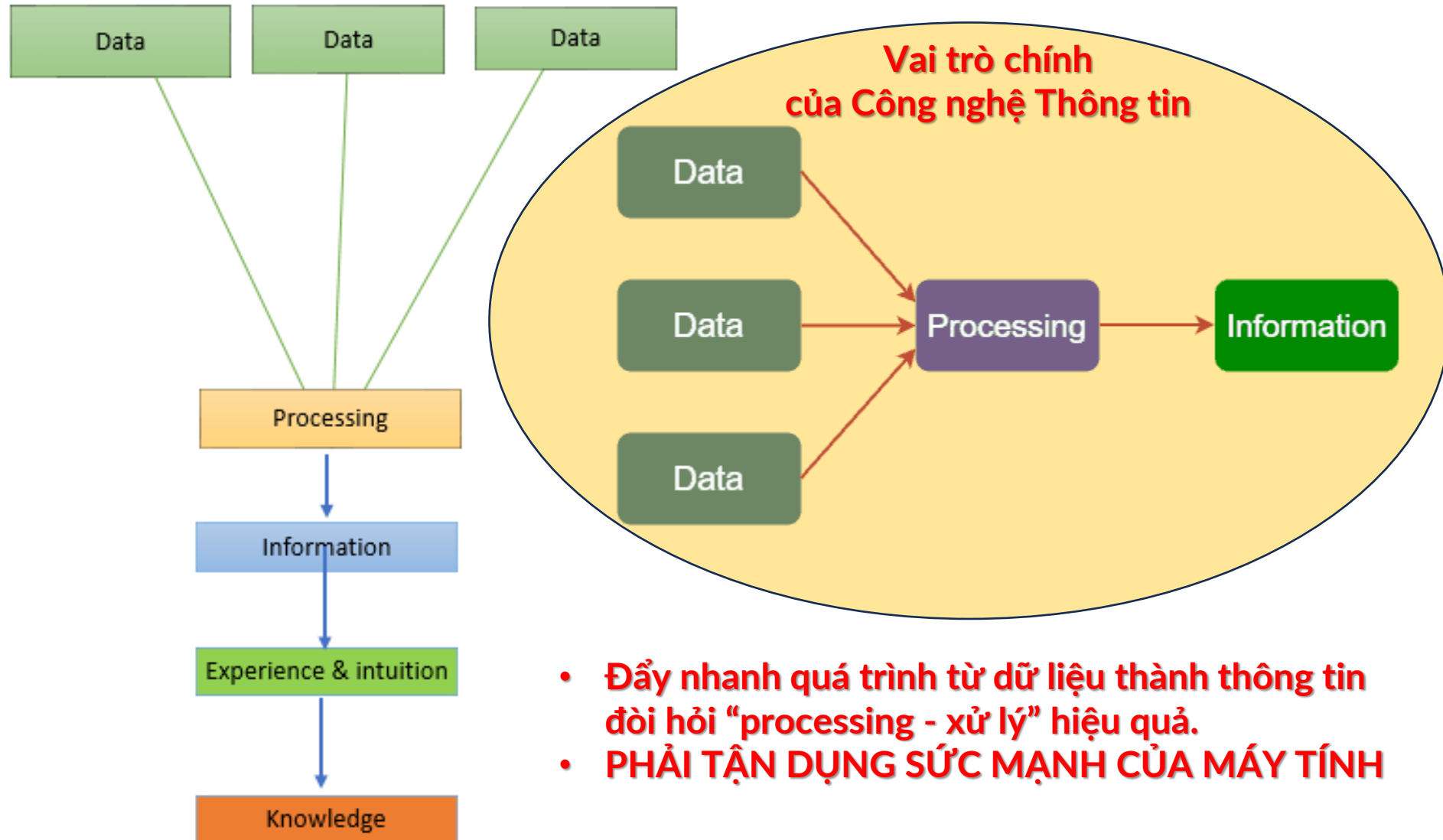
WISDOM
(Trí tuệ)

CHÀIN



sketchplanations

Con đường hình thành tri thức từ dữ liệu



- **Đẩy nhanh quá trình từ dữ liệu thành thông tin đòi hỏi “processing - xử lý” hiệu quả.**
- **PHẢI TẬN DỤNG SỨC MẠNH CỦA MÁY TÍNH**

Làm thế nào có thể tận dụng được sức mạnh của MTĐT?

(lặp lại slide)

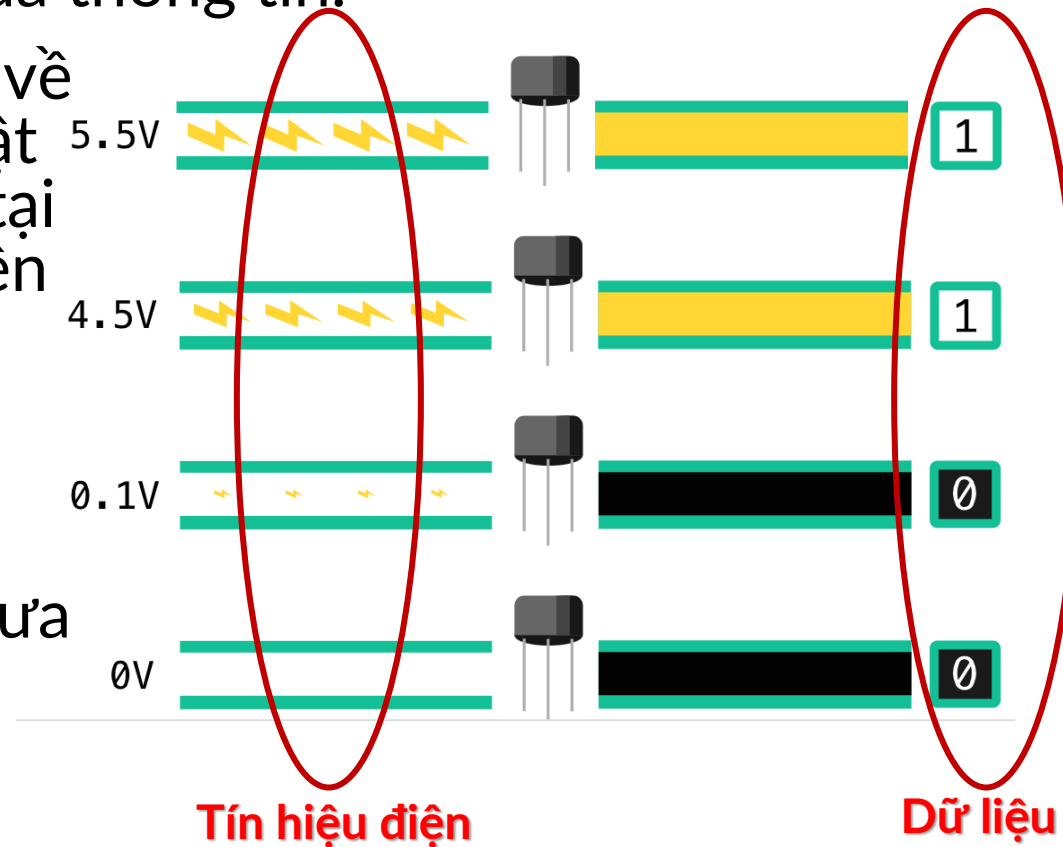
- Dữ liệu số, và
- Phần mềm xử lý phù hợp.



Làm thế nào để có dữ liệu số?

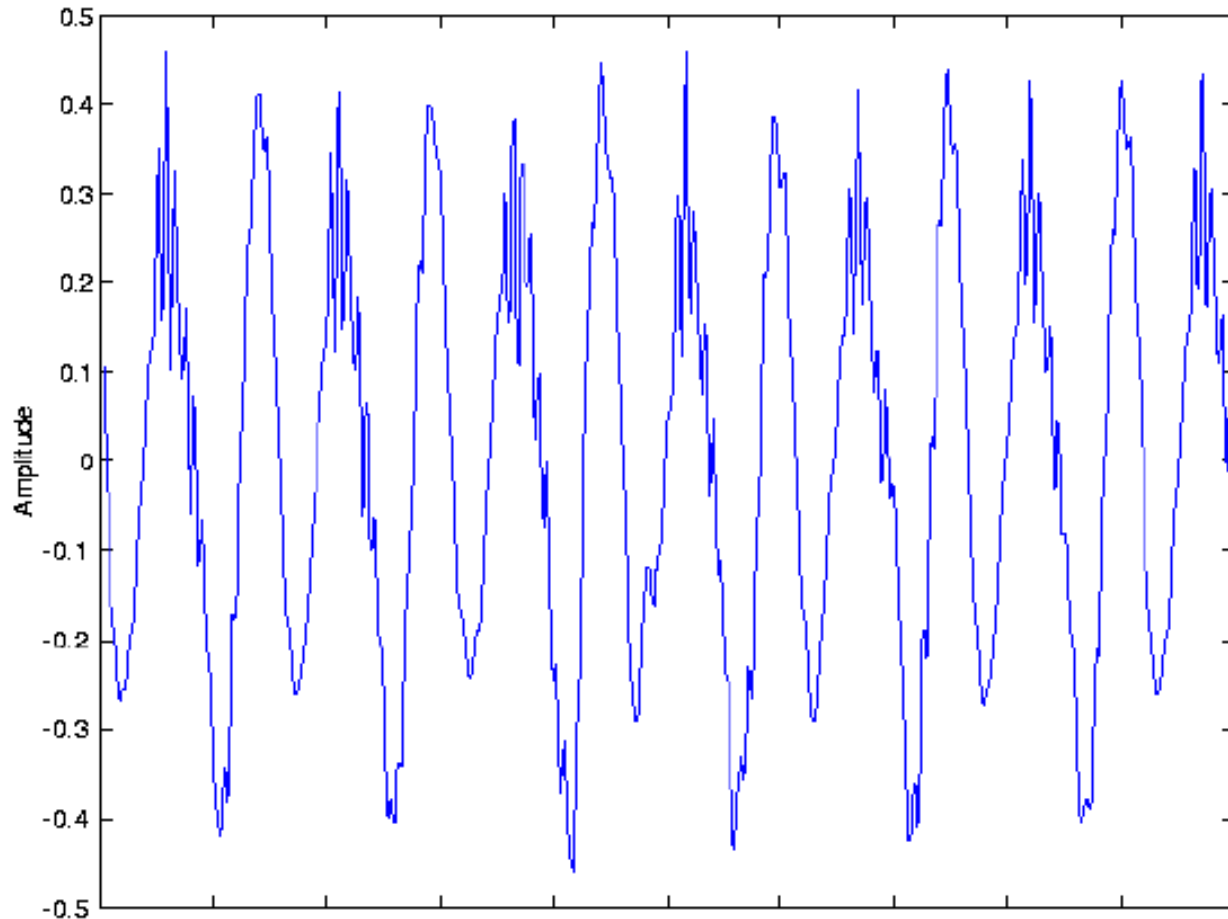
Mối quan hệ giữa dữ liệu (data) và tín hiệu (signal)

- Tín hiệu là biểu diễn về mặt vật lý của dữ liệu.
- Nếu dữ liệu liên quan đến “biểu diễn về mặt logic” của thông tin (ở dạng thô) thì (đến cuối cùng) Tín hiệu có thể coi là “biểu diễn về mặt vật lý” của thông tin.
- Các hệ thống xử lý dữ liệu về bản chất là các hệ thống vật lý, vì vậy, dữ liệu luôn tồn tại dưới dạng tín hiệu (một hiện tượng vật lý nào đó) trong các hệ thống vật lý.
- Ví dụ: các bit “biểu diễn” trong máy tính thông qua dòng điện (giá trị điện áp đưa vào transistor).



Mối quan hệ giữa dữ liệu và tín hiệu (tiếp)

- Ví dụ: Dữ liệu âm thanh thì trong thế giới vật lý, cường độ sóng âm (áp suất không khí) tại một thời điểm theo thời gian là tín hiệu biểu diễn dữ liệu này.



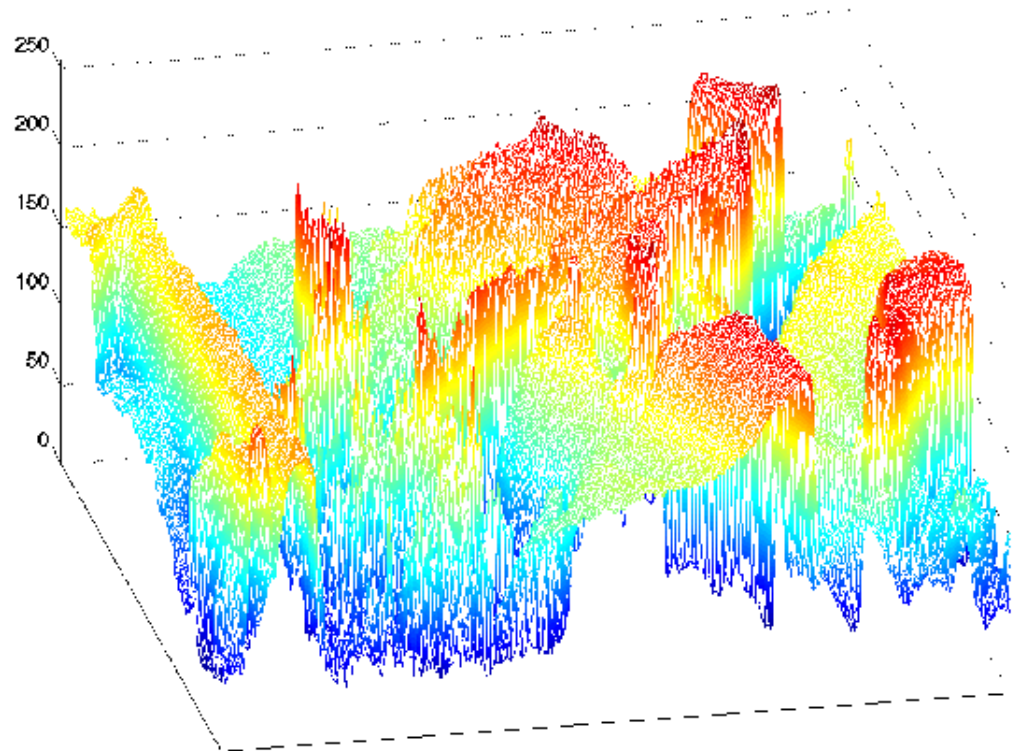
Biến đổi trong áp suất không khí (ví dụ phát âm nguyên âm “e”).

Mối quan hệ giữa dữ liệu và tín hiệu (tiếp)

- Nếu dữ liệu là hình ảnh thì trong thế giới vật lý cường độ ánh sáng tại một vị trí (2D) là tín hiệu biểu diễn dữ liệu này.



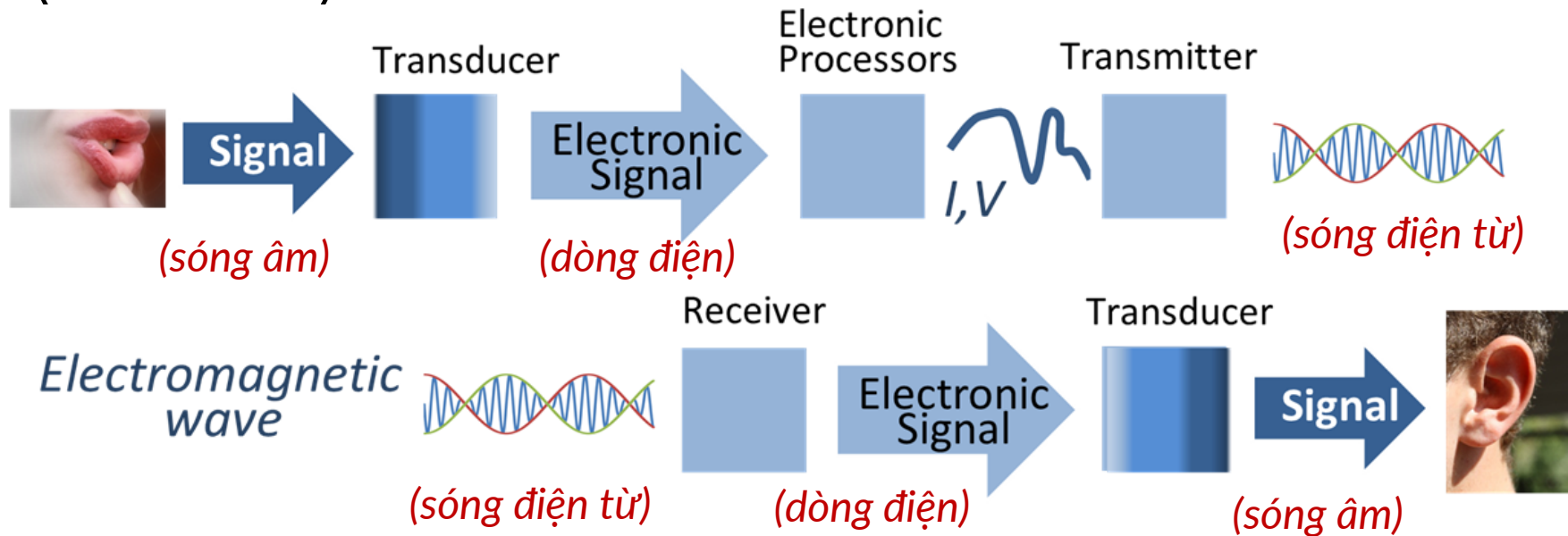
Dữ liệu



Tín hiệu (cường độ ánh sáng tại mỗi điểm ảnh)

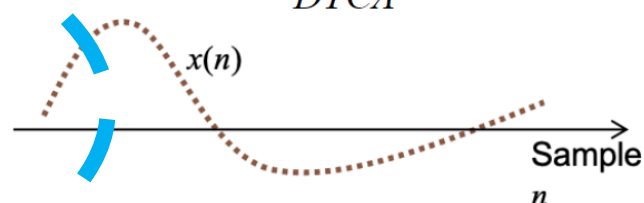
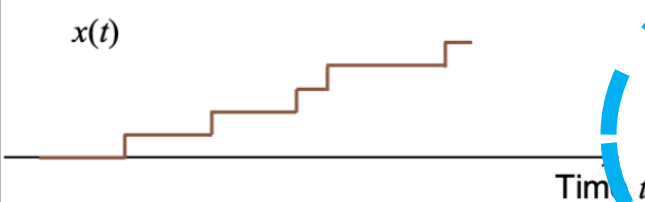
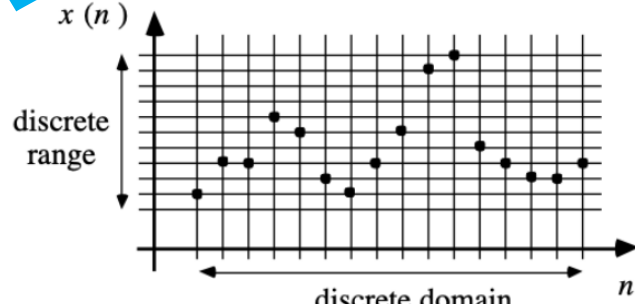
Mối quan hệ giữa dữ liệu và tín hiệu (tiếp)

- Ngoài ra dữ liệu ở dạng bất kỳ có thể được biểu diễn dưới dạng hiện tượng vật lý thông dụng, như: điện - tín hiệu điện (truyền có dây), sóng điện từ - tín hiệu điện từ (truyền không dây), hoặc sóng âm (truyền dưới nước)... thông qua các bộ chuyển đổi (transducer).



- Ánh sáng, sóng âm, điện, sóng điện từ... là các hiện tượng vật lý, khi các hiện tượng vật lý này biểu diễn dữ liệu (theo mong muốn của con người) thì ta có tín hiệu.

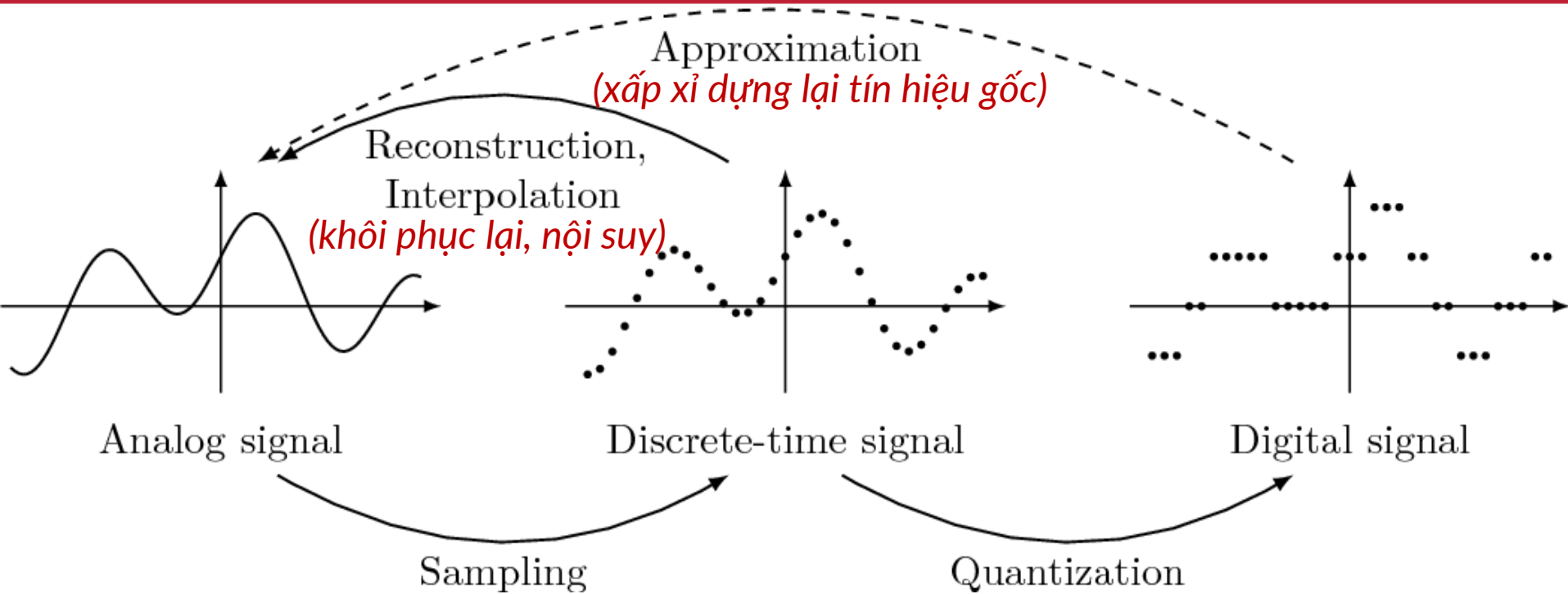
Trong thực tế tín hiệu tồn tại ở dạng nào?

	Liên tục theo thời gian Continuous-time	Rời rạc theo thời gian Discrete-time
Liên tục theo biên độ (giá trị) Continuous-amplitude	Tín hiệu tương tự Analog  Loại phổ biến trong tự nhiên	DTCA 
Rời rạc theo biên độ (giá trị) Discrete-amplitude	CTDA 	Tín hiệu số Digital  Loại xử lý được trong máy tính

1. Tín hiệu biểu diễn nhiệt độ phòng là loại nào?
2. TH biểu diễn nhiệt độ phòng theo giờ là loại nào?

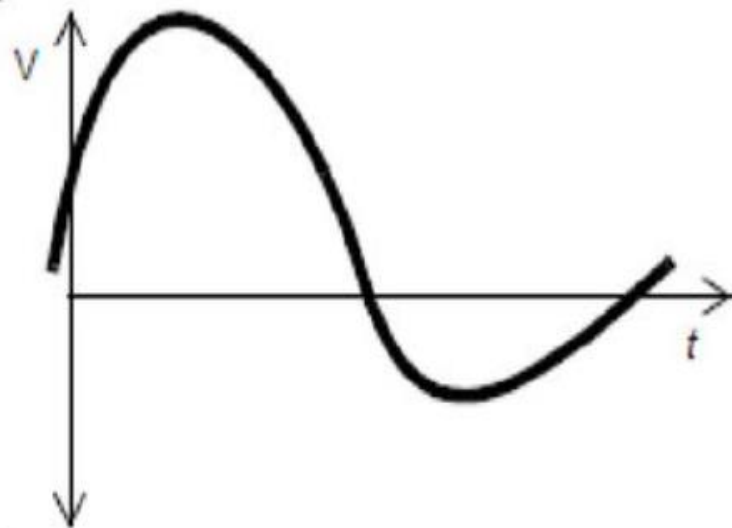
3. Dân số Việt Nam theo thời gian thực?
4. Dân số Việt Nam theo năm?

Làm thế nào để có tín hiệu số (dữ liệu số)?

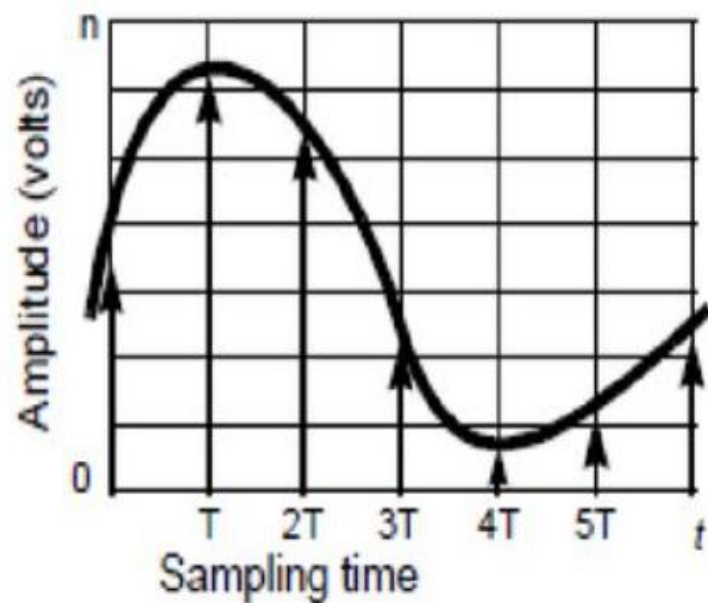


- Từ tín hiệu liên tục (tương tự) sang tín hiệu số, trải qua 2 bước:
 - Bước 1 - Lấy mẫu (sampling): rời rạc theo thời gian, rời
 - Bước 2 - Lượng tử hóa (quantization): rời rạc theo biên độ.
- Muốn sang dữ liệu số cần thêm:
 - Bước 3 – Mã hóa (coding): mã hóa các giá trị biên độ thành mã nhị phân

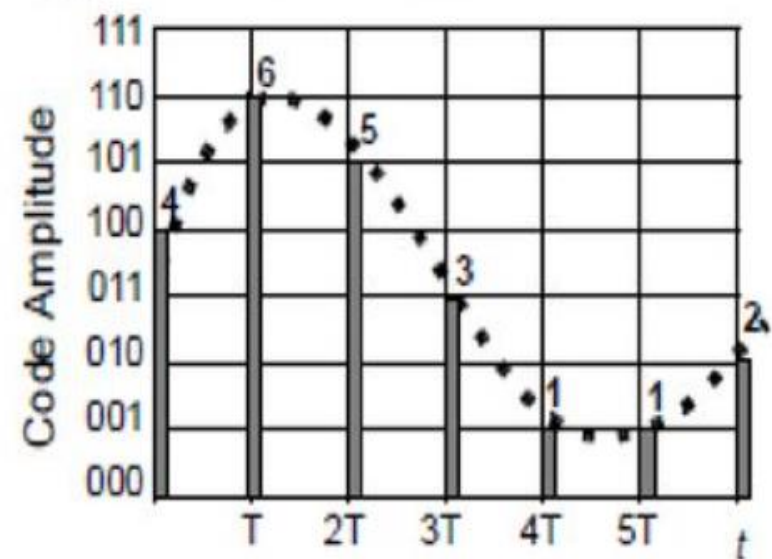
0 Analog Signal



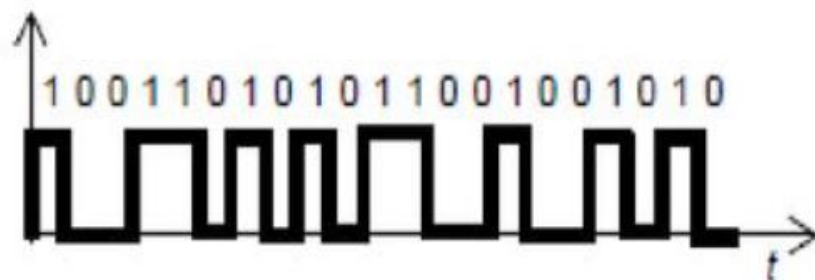
1 Sampling



2 Quantization

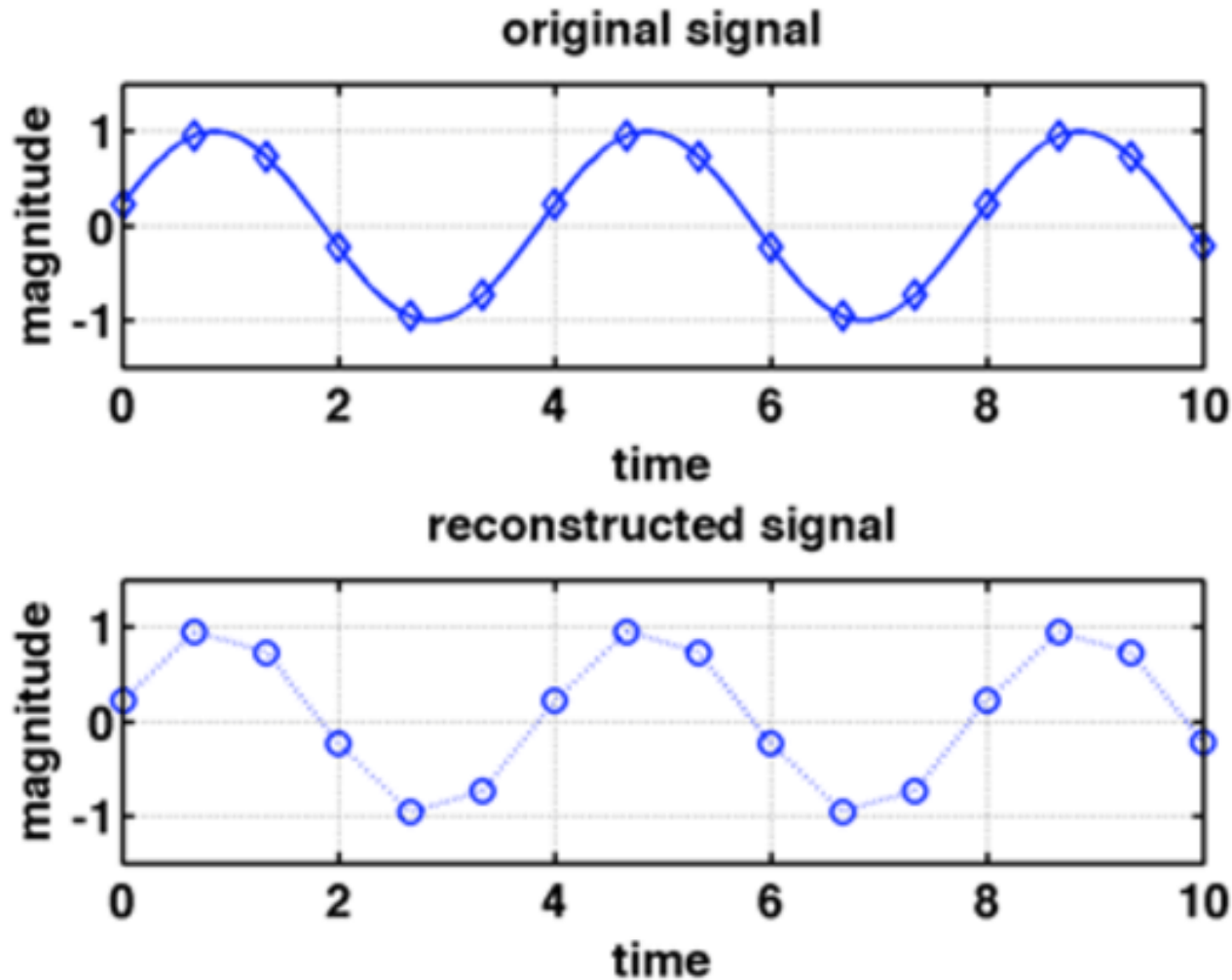


3 Coding



Quá trình lấy mẫu

- Rời rạc hóa tín hiệu theo thời gian: thông qua lấy các giá trị tại các điểm rời rạc theo thời gian, với điều kiện, sau này có thể “khôi phục lại” tín hiệu liên tục (gốc) ban đầu



Lấy mẫu sao cho đủ? Định lý lấy mẫu Nyquist-Shannon

Harry Nyquist
(1889 – 1976)



Claude Shannon
(1916 – 2001)



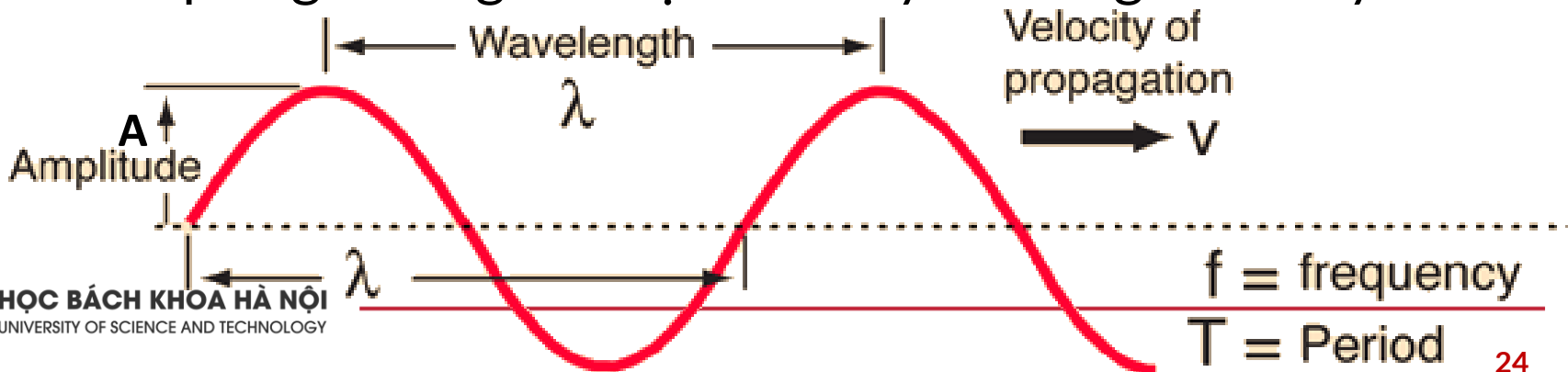
Theo Nyquist và Shannon thì: nếu tần số lấy mẫu (số mẫu lấy trong 1s) mà tối thiểu bằng 2 lần **tần số lớn nhất** của tín hiệu thì có thể xây dựng lại tín hiệu gốc từ tín hiệu rời rạc.

Tần số của tín hiệu
là gì?

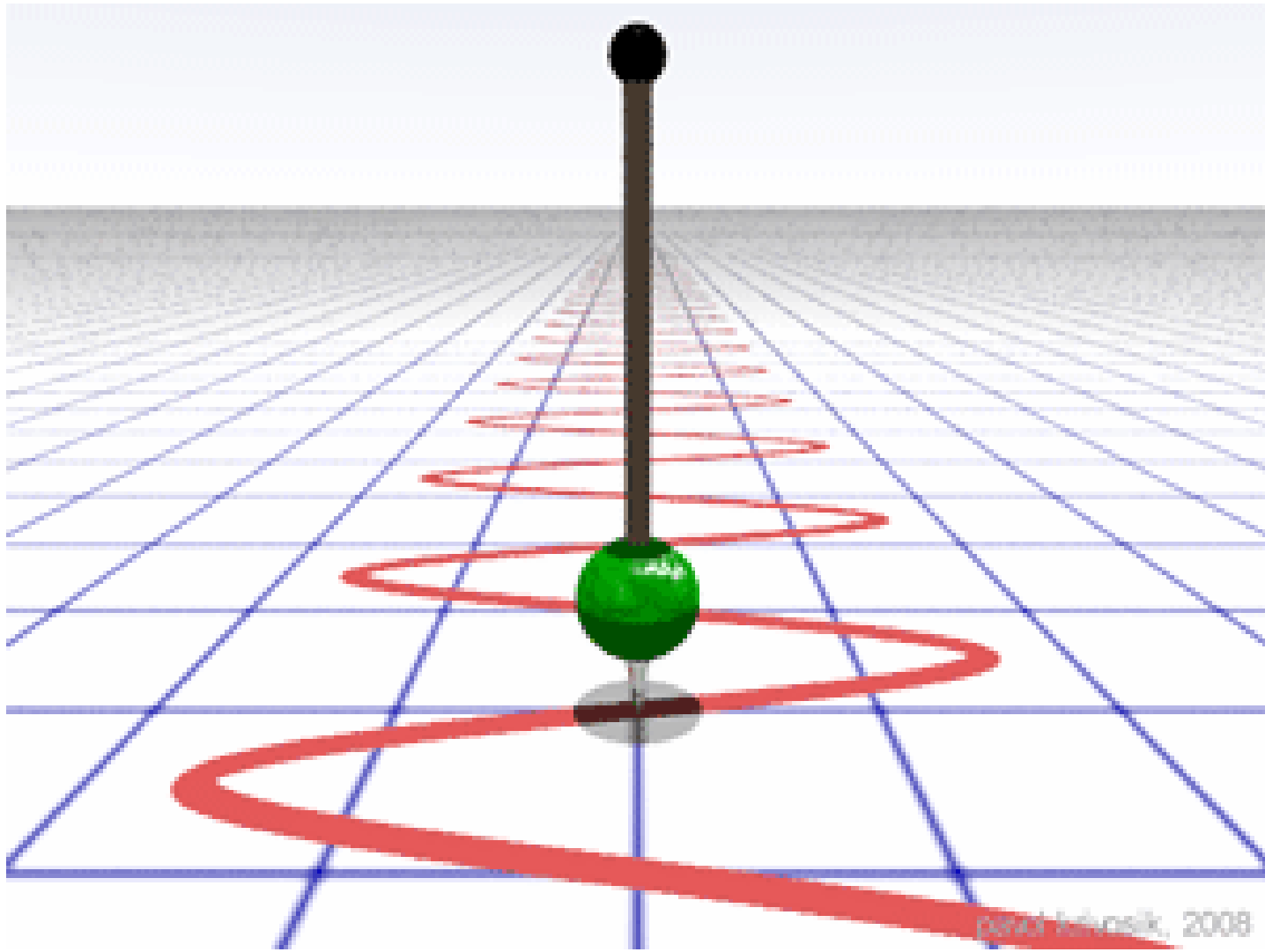
Tần số xác định
thế nào?

Tín hiệu tuần hoàn

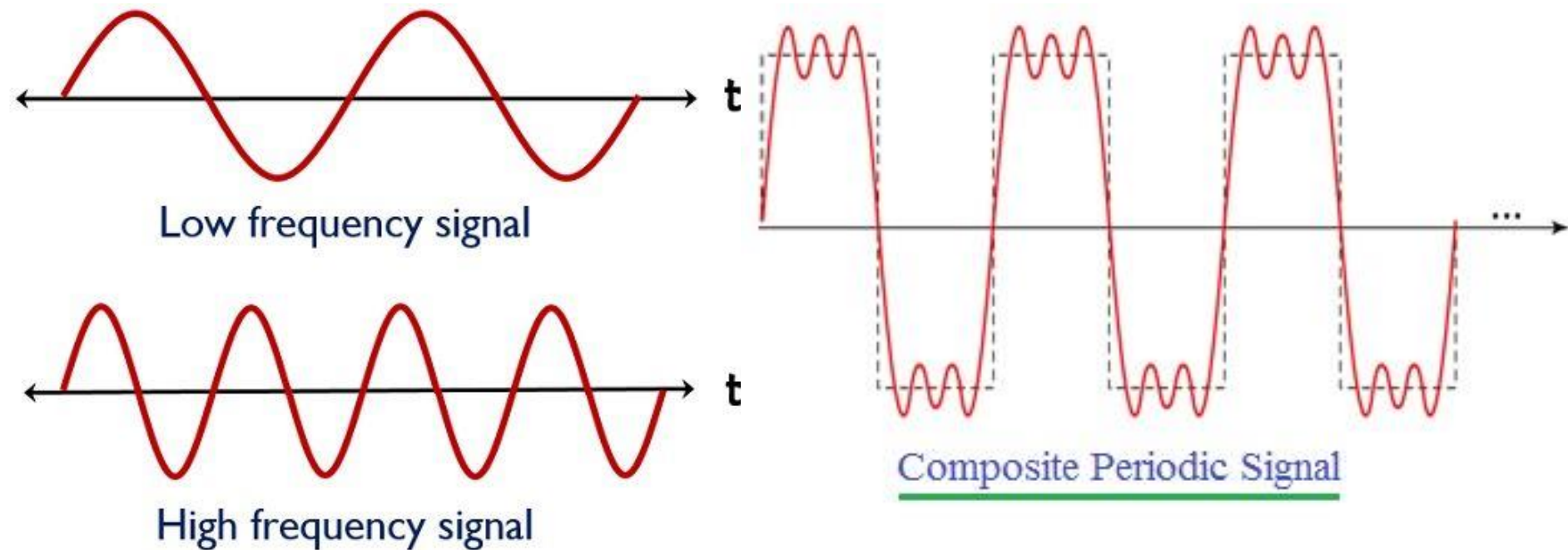
- Là tín hiệu có dạng sóng lặp lại sau mỗi chu kỳ:
 - Biểu diễn hàm số: $s(t + T) = s(t) \quad -\infty < t < +\infty$, với T được gọi là chu kỳ (s)
 - Ví dụ: tín hiệu hình sin: $\sin(t + 2\pi) = \sin(t)$
 - Tín hiệu tuần hoàn được biểu diễn bởi:
 - Biên độ (A), đơn vị theo hiện tượng vật lý, ví dụ với điện: V .
 - Tần số (f) (Hz): số lần dao động lặp lại trong 1s, với $f = 1/T$.
 - Bước sóng (λ), với $\lambda = v/f = v \times T$, với v là vận tốc lan truyền của hiện tượng vật lý, với dòng điện và sóng điện từ:
 $v = C = 3.10^8 \text{ m/s}$.
- => λ là quãng đường tín hiệu lan truyền trong 1 chu kỳ.



(Tần số và biên độ)



- Ví dụ về tín hiệu tuần hoàn

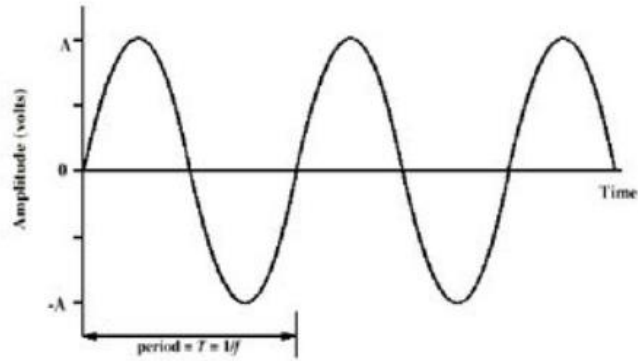


Circuit Globe

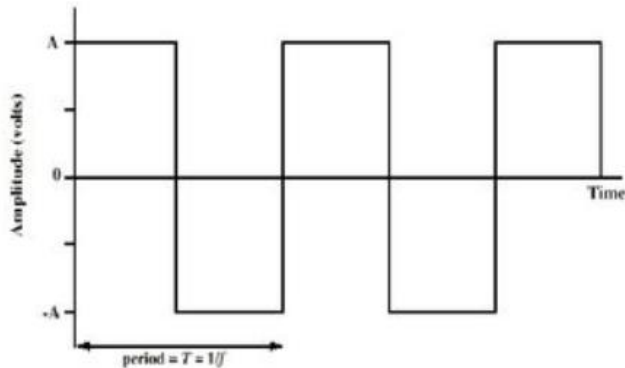
Tín hiệu không tuần hoàn (aperiodic signal)

- Ngược lại với tín hiệu tuần hoàn, tín hiệu không tuần hoàn không lặp lại theo thời gian.

Periodic signals

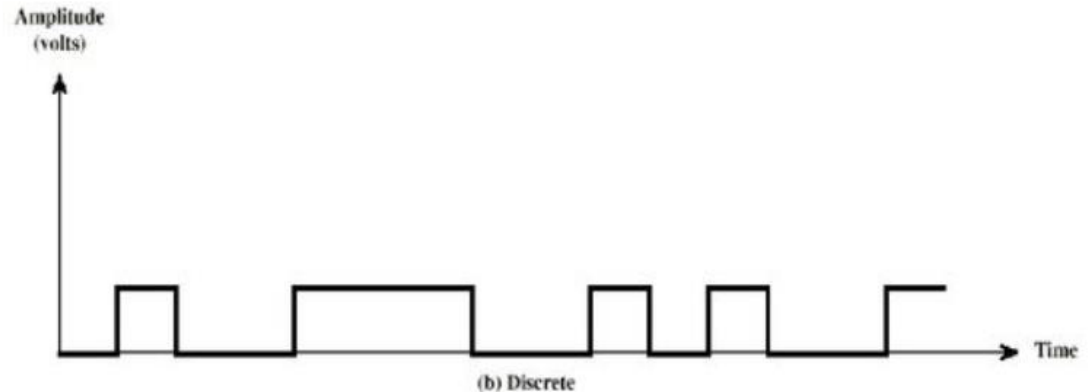
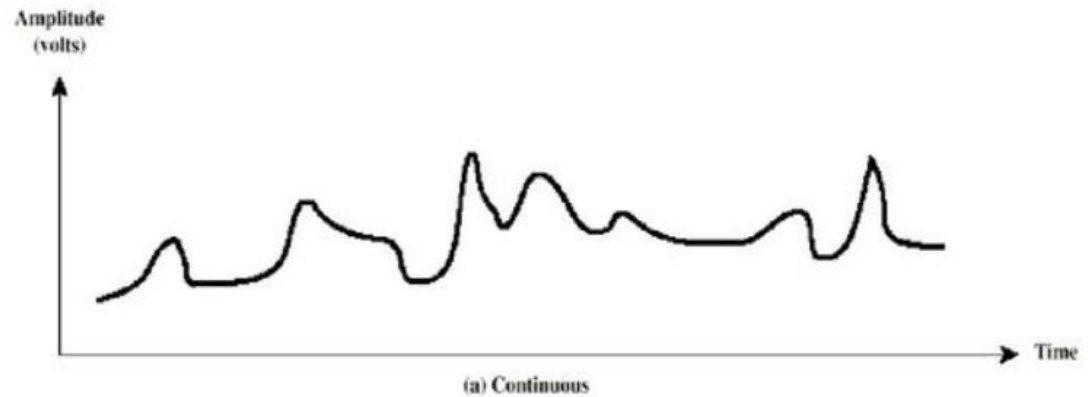


(a) Sine wave



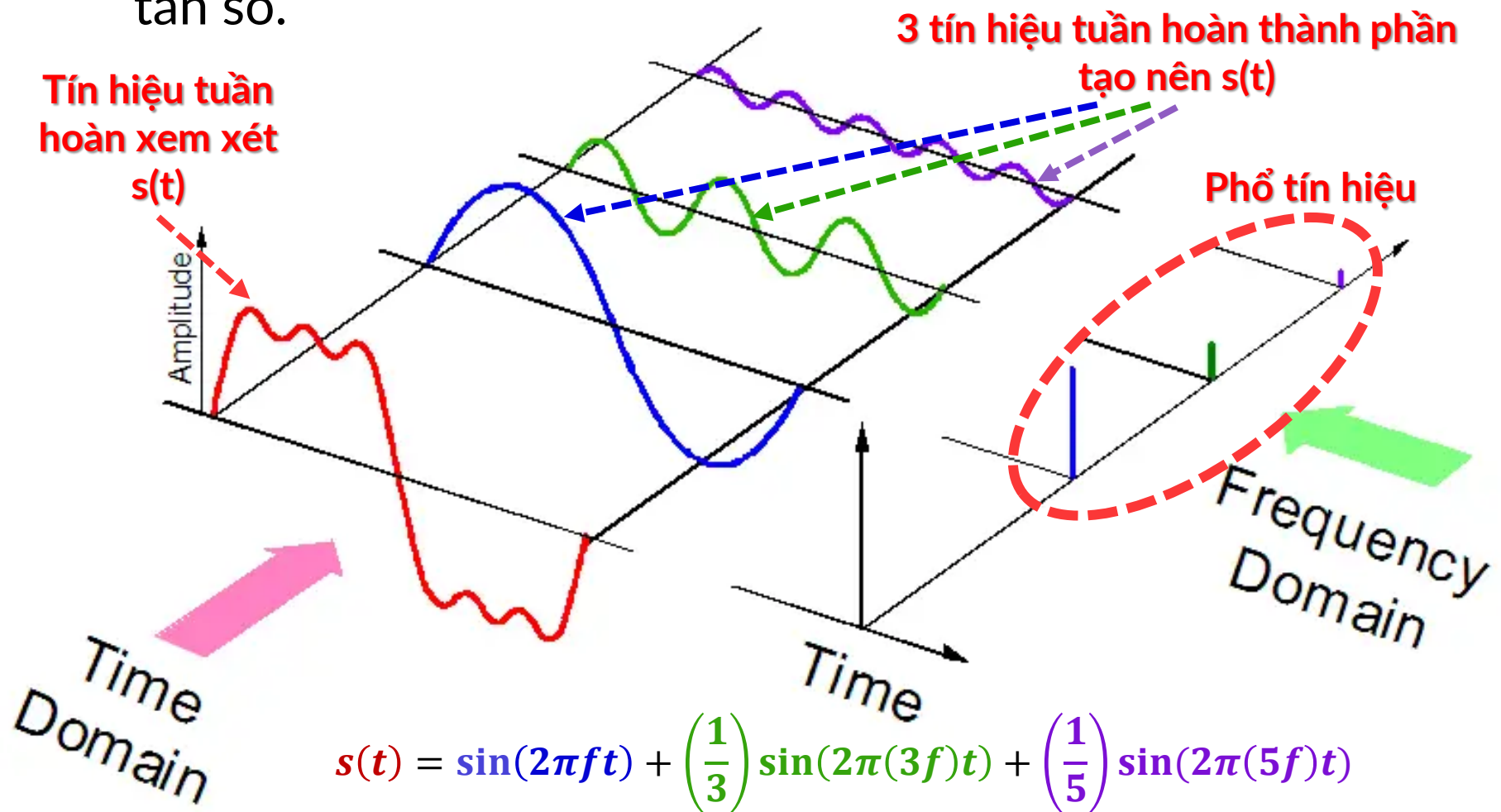
(b) Square wave

Nonperiodic signals



Phổ tín hiệu (hay biểu diễn tín hiệu theo tần số)

- Tín hiệu được biểu diễn trong:
 - Miền thời gian: sự biến thiên của biên độ theo thời gian;
 - Miền tần số: sự biến thiên của năng lượng tín hiệu theo tần số.



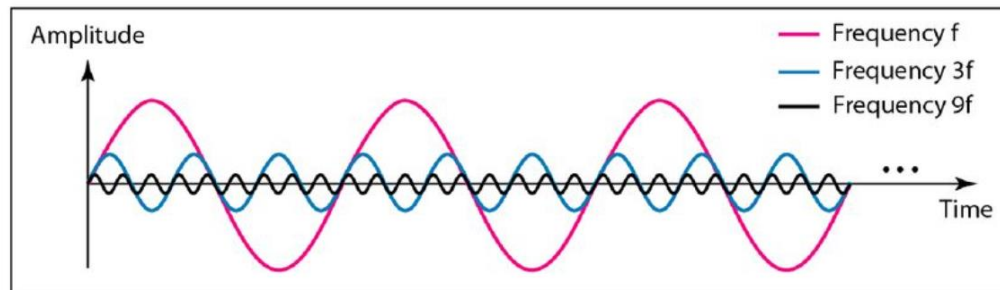
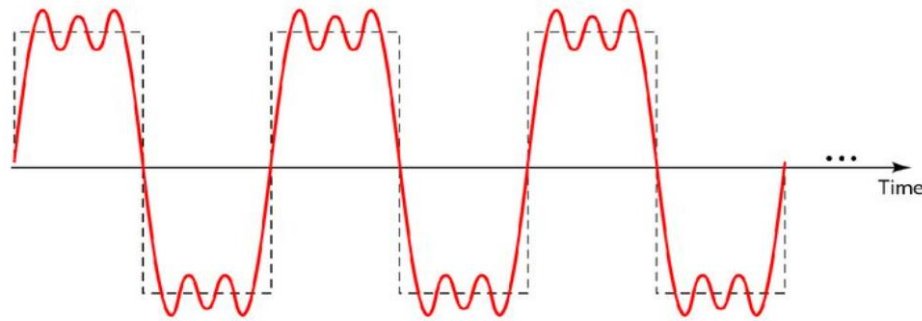
Biến đổi Fourier – Ước lượng phổ tín hiệu



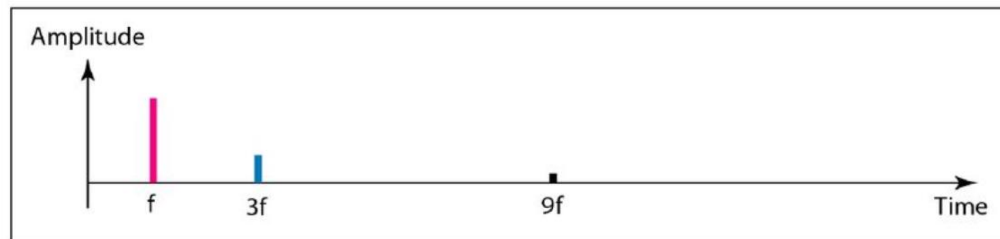
Jean Baptiste Joseph Fourier
(1768 - 1830)

Theo Fourier, bất kỳ tín hiệu nào cũng có thể được “phân rã” thành tổng hợp của các tín hiệu lượng giác tuần hoàn (hình sin hoặc cos) với tần số, biên độ và góc pha khác nhau.

- Nếu tín hiệu tổng hợp là tín hiệu tuần hoàn thì tín hiệu này được tổng hợp từ các tín hiệu tuần hoàn có tần số rời rạc.

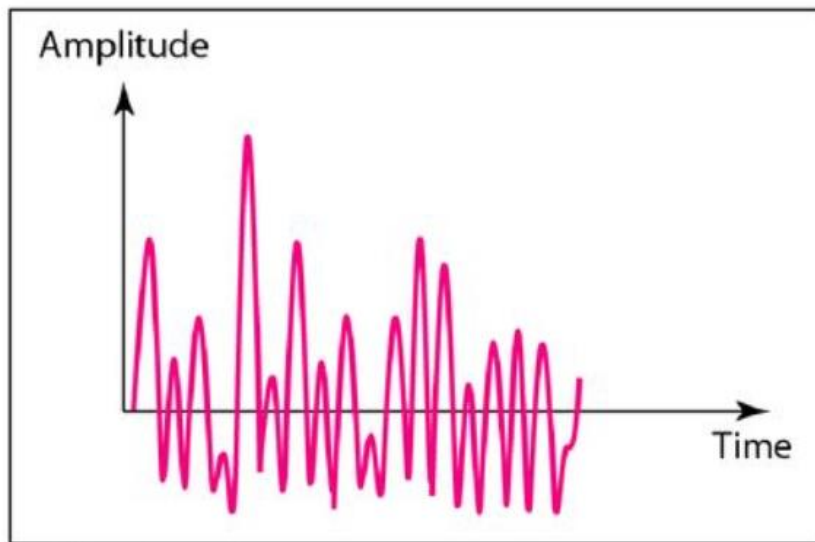


a. Time-domain decomposition of a composite signal

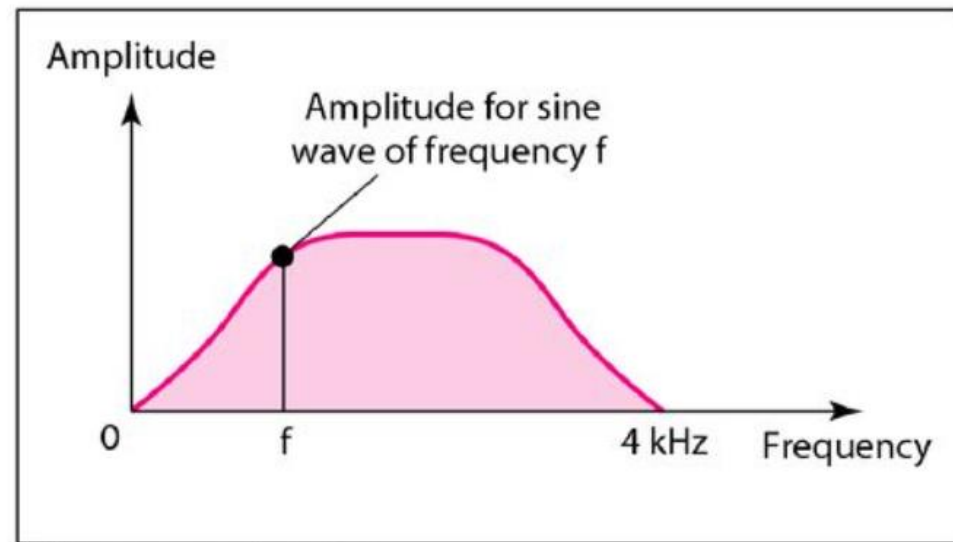


b. Frequency-domain decomposition of the composite signal

- Nếu tín hiệu tổng hợp là tín hiệu không tuần hoàn thì tín hiệu này được tổng hợp từ các tín hiệu tuần hoàn có tần số trong dải thay đổi liên tục.



a. Time domain



b. Frequency domain

Phổ tín hiệu và Lấy mẫu – Định lý lấy mẫu Nyquist-Shannon

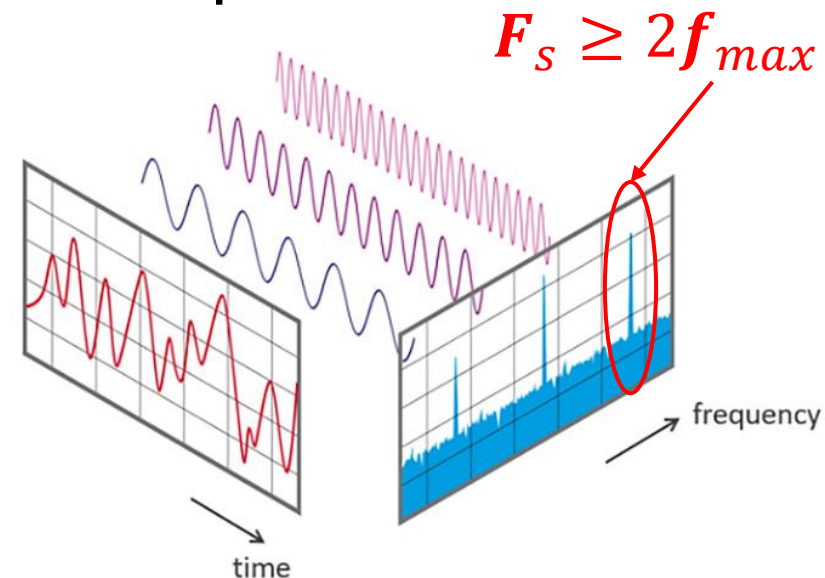
Harry Nyquist
(1889 – 1976)



Claude Shannon
(1916 – 2001)

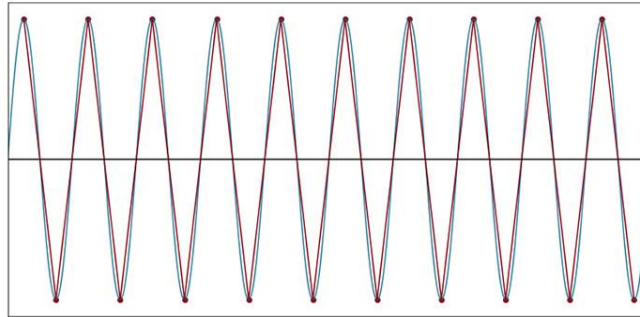


- Nhờ Biến đổi Fourier chúng ta biết được phổ của tín hiệu, và qua đó biết được tần số lớn nhất của tín hiệu.
- Theo Nyquist và Shannon thì: nếu tần số lấy mẫu (số mẫu lấy trong 1s) mà tối thiểu bằng 2 lần tần số lớn nhất của tín hiệu thì có thể xây dựng lại tín hiệu gốc từ tín hiệu rời rạc.



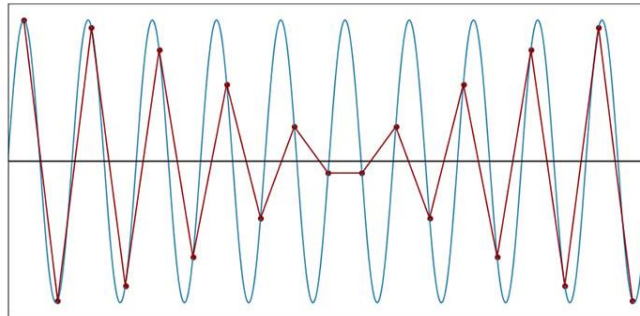
Làm thế nào để đảm bảo “khôi phục lại” tín hiệu gốc?

$$F_s = 2.0f_{signal}$$



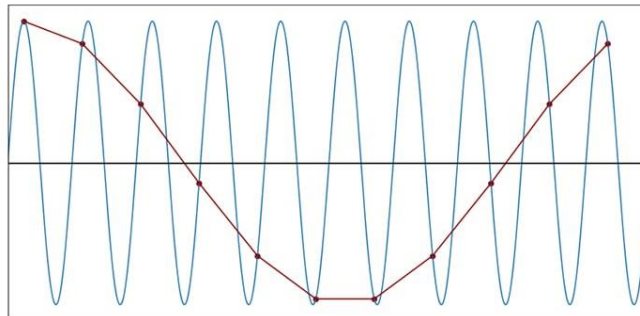
Khôi phục được
“gần đúng” tín hiệu gốc

$$F_s = 1.9f_{signal}$$



Không thể khôi phục
được tín hiệu gốc

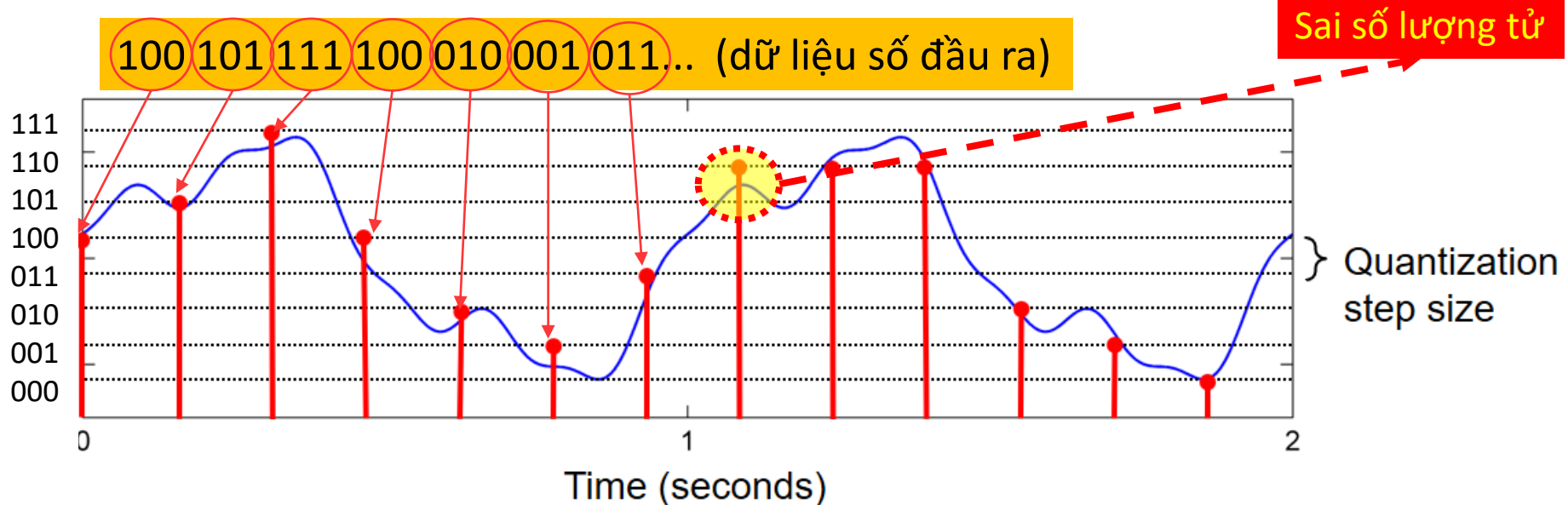
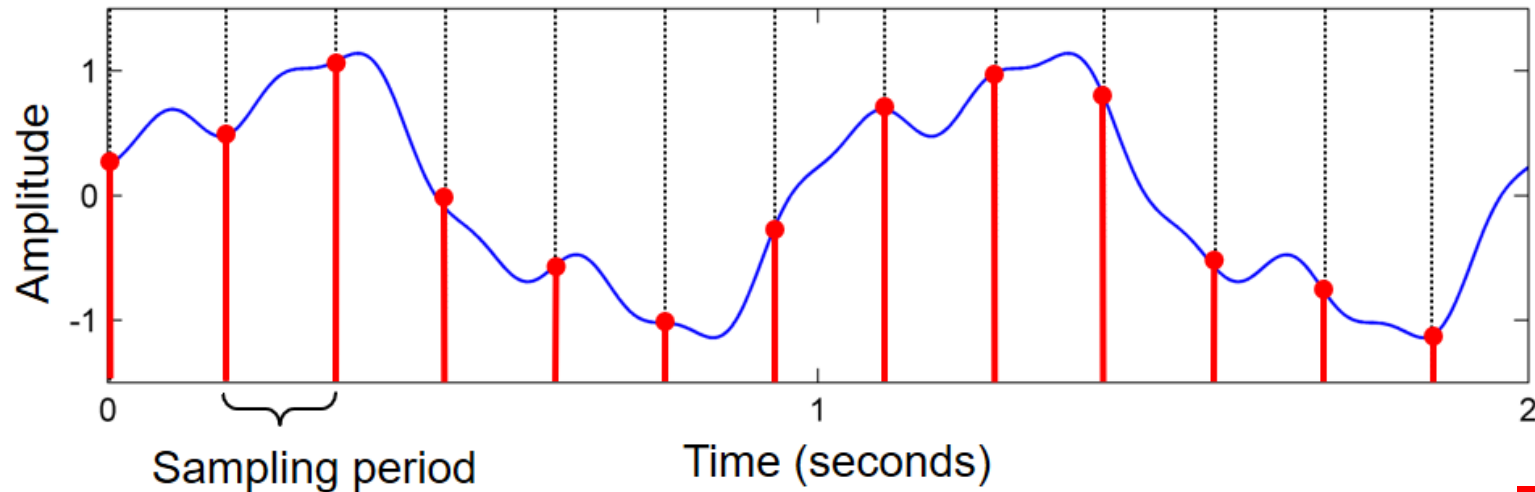
$$F_s = 1.1f_{signal}$$



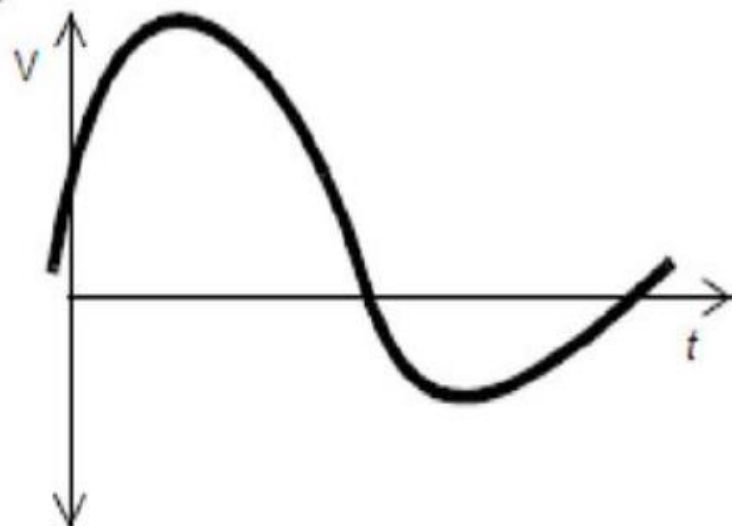
Không thể khôi phục
được tín hiệu gốc

Lượng tử hóa (rời rạc theo biên độ)

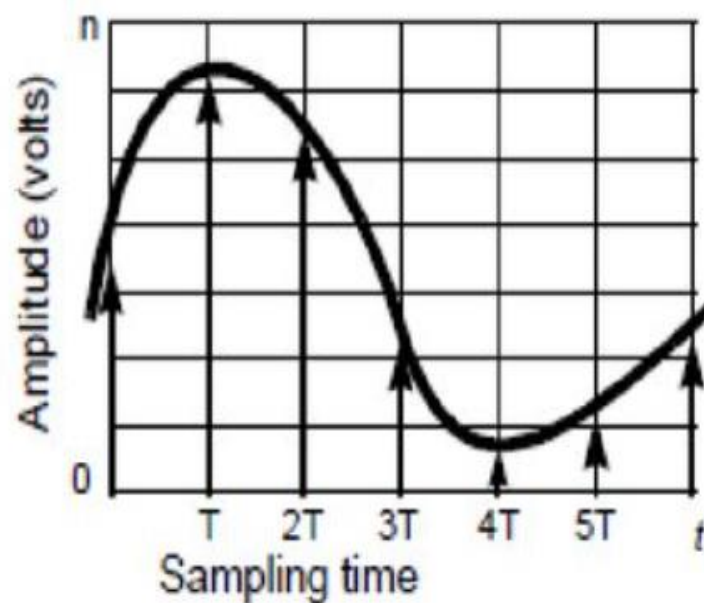
- Sau quá trình lấy mẫu, tín hiệu thu được có giá trị liên tục theo biên độ, và cần phải được “lượng tử hóa” để biến thành tín hiệu số.



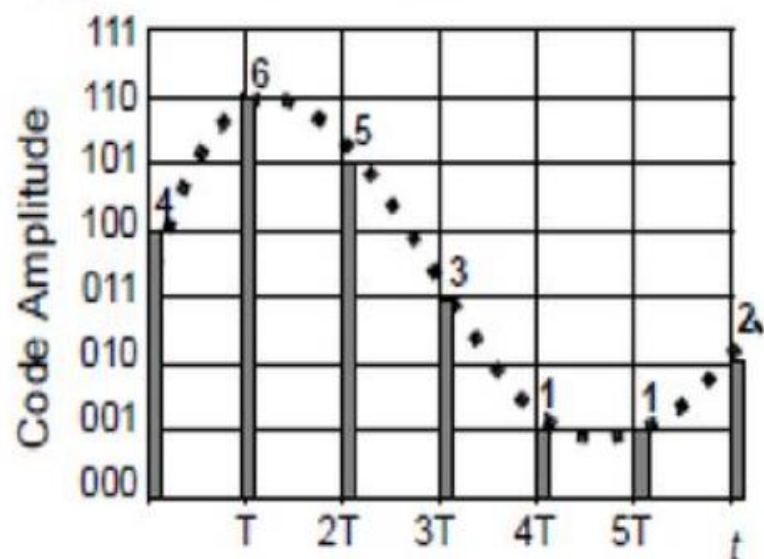
0 Analog Signal



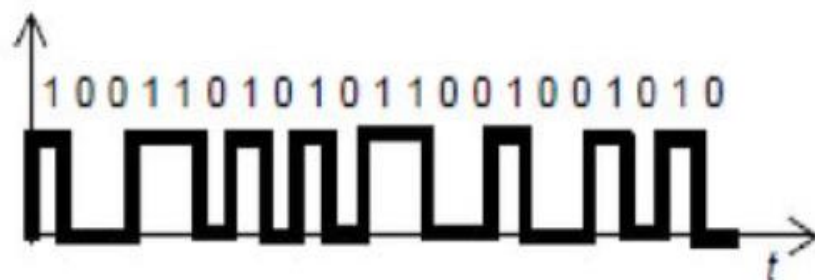
1 Sampling



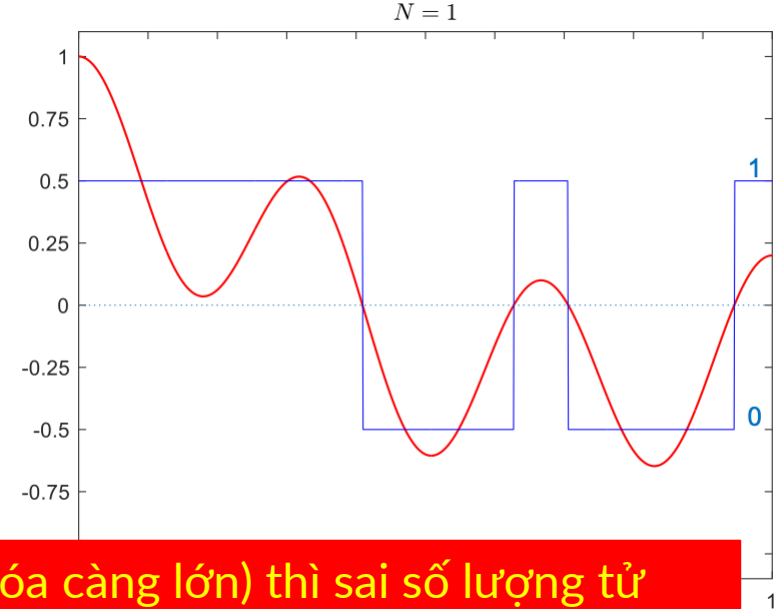
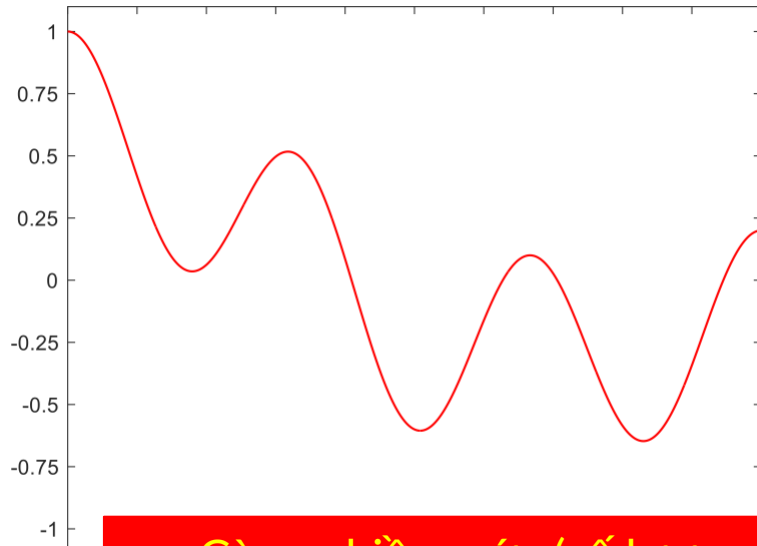
2 Quantization



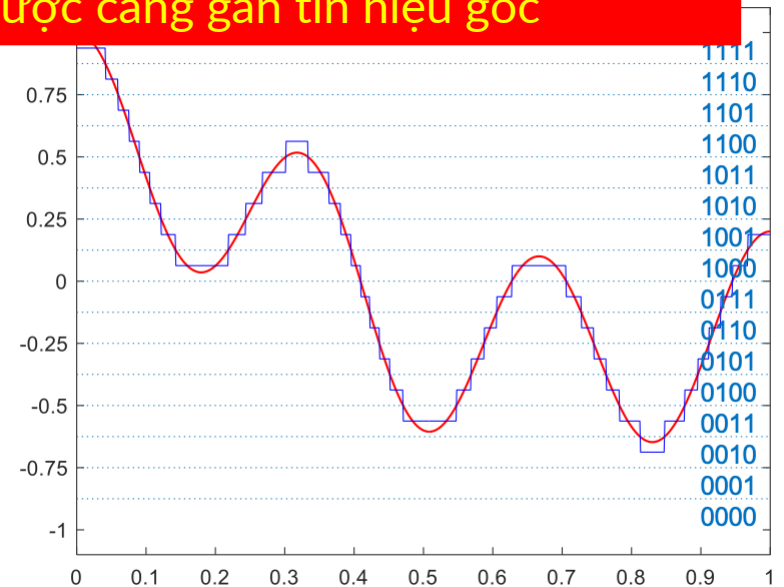
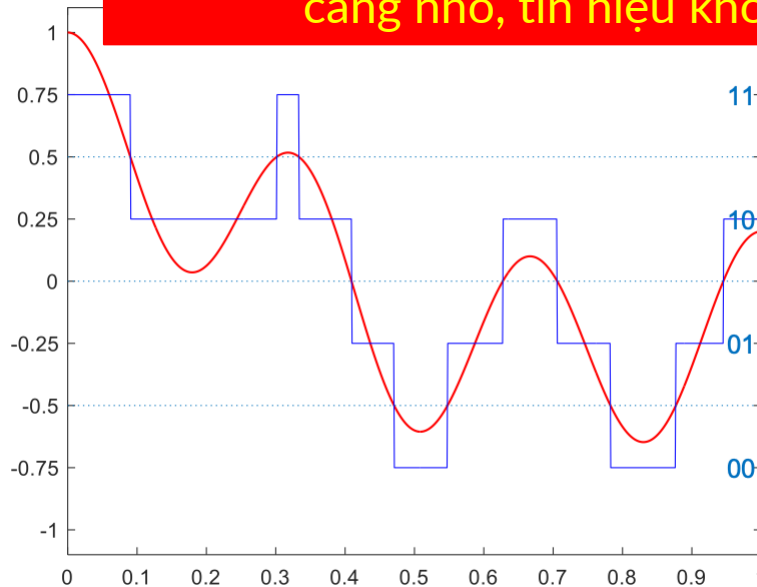
3 Coding

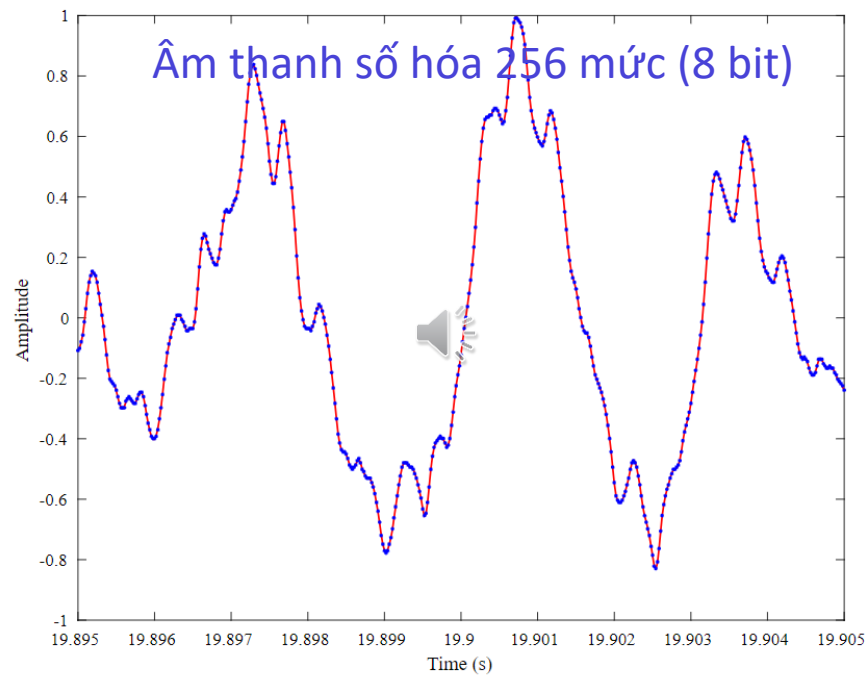
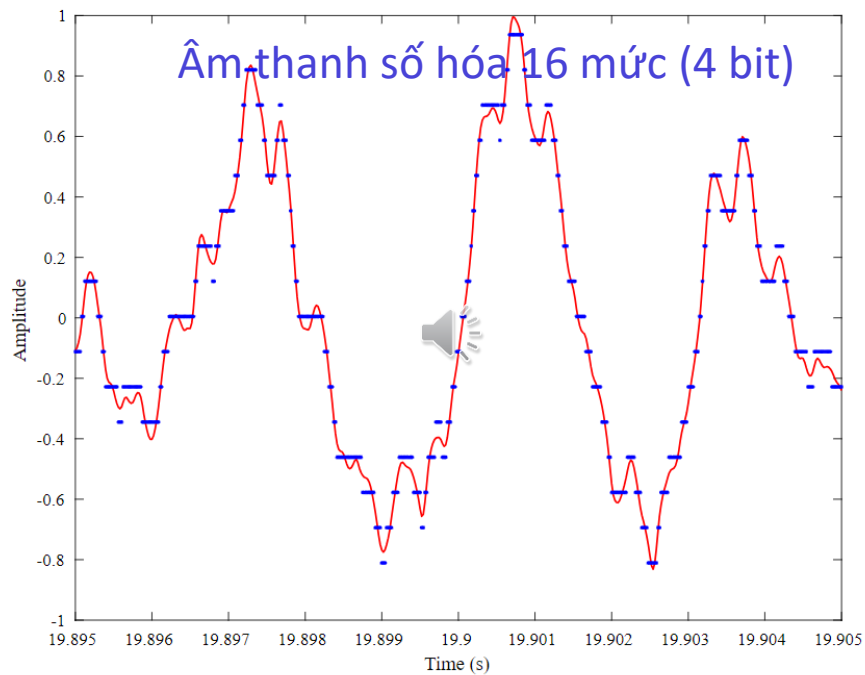
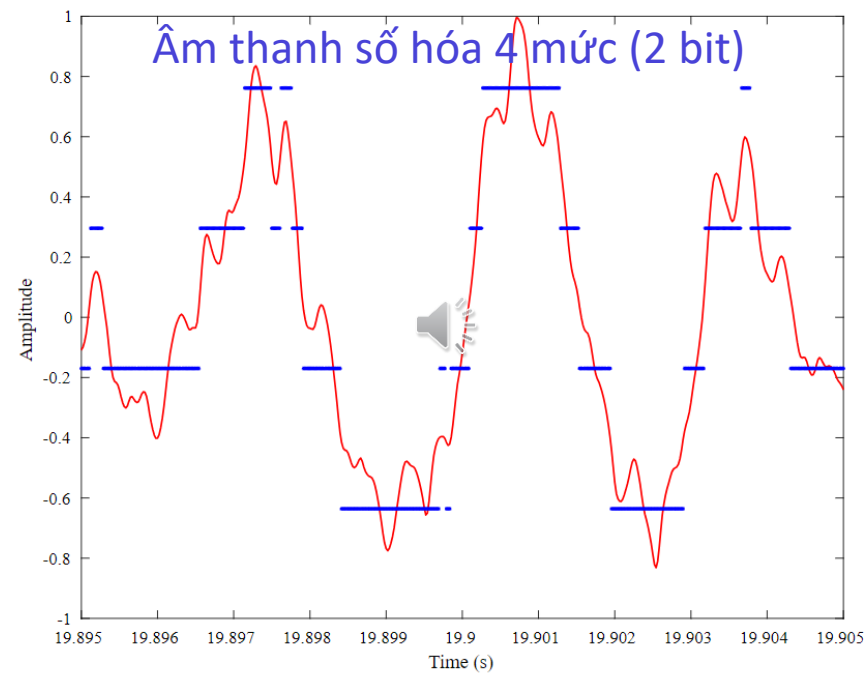
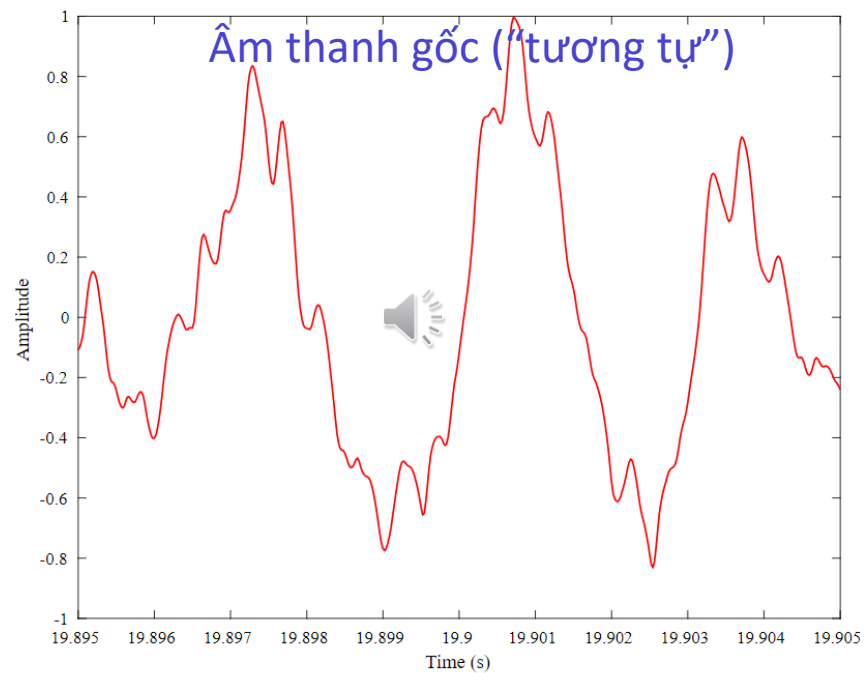


Lượng tử hóa: bao nhiêu mức là đủ



Càng nhiều mức (số lượng bit mã hóa càng lớn) thì sai số lượng tử càng nhỏ, tín hiệu khôi phục được càng gần tín hiệu gốc



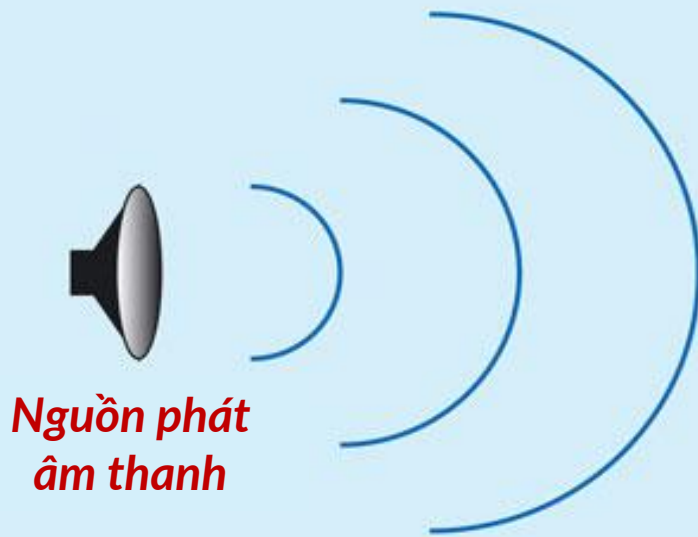


Chuẩn số hóa cơ bản (âm thanh mono – đơn kênh)

Type of Audio Signal	Sound Quality	Bandwidth	Sampling Rate (kHz)	Bit Depth	Data Rate (kbit/sec)
Speech	Phone quality speech	200 Hz – 3.2 kHz	8	12	96
Music	High-fidelity music (CD Quality)	5 Hz – 20 kHz	44.1	16	706
High-quality music	DVD quality	5 Hz – 20 kHz	192	24	4608

- Để giảm kích thước dữ liệu số có thể áp dụng các kỹ thuật nén có mất mát dữ liệu:
 - Ví dụ MP3, dựa trên “Phương pháp mã hóa theo cảm thụ”: chỉ tập trung vào các âm thanh mà tai người nghe được, và loại bỏ các âm thanh không nghe thấy, kết quả “data rate” còn 64kbit/sec (tỷ lệ nén 11:1 so với CD), hay 128kbps với hai kênh (âm thanh nổi stereo).

Tóm tắt quy trình tạo ra dữ liệu âm thanh số



Dữ liệu số, dòng bit, lưu thành file trên máy tính

0	1	1	0	1	0	0	1
1	1	0	1	1	1	0	0
0	1	0	0	1	1	0	0
0	0	1	1	0	1	0	1

Sự biến đổi áp suất không khí, qua micro được chuyển thành điện áp có giá trị thay đổi liên tục (tín hiệu điện liên tục)



Rời rạc hóa tín hiệu tương tự (liên tục)



Tổng kết tạo dữ liệu số

- Xử lý dữ liệu tạo cho ta thông tin, và việc xử lý hiệu quả nếu đem vào xử lý trong máy tính (tận dụng sức mạnh tính toán)
- Trong thực tế:
 - Dữ liệu tồn tại dưới dạng các biểu diễn hiện tượng vật lý, gọi là tín hiệu;
 - Tín hiệu “thông thường” ở dạng “không thân thiện” với máy tính (ie. không phải dạng số);
 - Do vậy phải “số hóa” tín hiệu để tạo ra tín hiệu số, sau đó mã hóa thành dữ liệu số;
 - Số hóa gồm: (1) lấy mẫu; (2) lượng tử hóa; và (3) mã hóa;
 - Lấy mẫu mà vẫn đảm bảo khôi phục lại tín hiệu gốc thì tuân theo Định luật Nyquist-Shannon;
 - Lượng tử hóa càng nhiều mức (nhiều bit sau mã hóa) thì càng gần với tín hiệu gốc (cân bằng giữa chất lượng và chi phí)

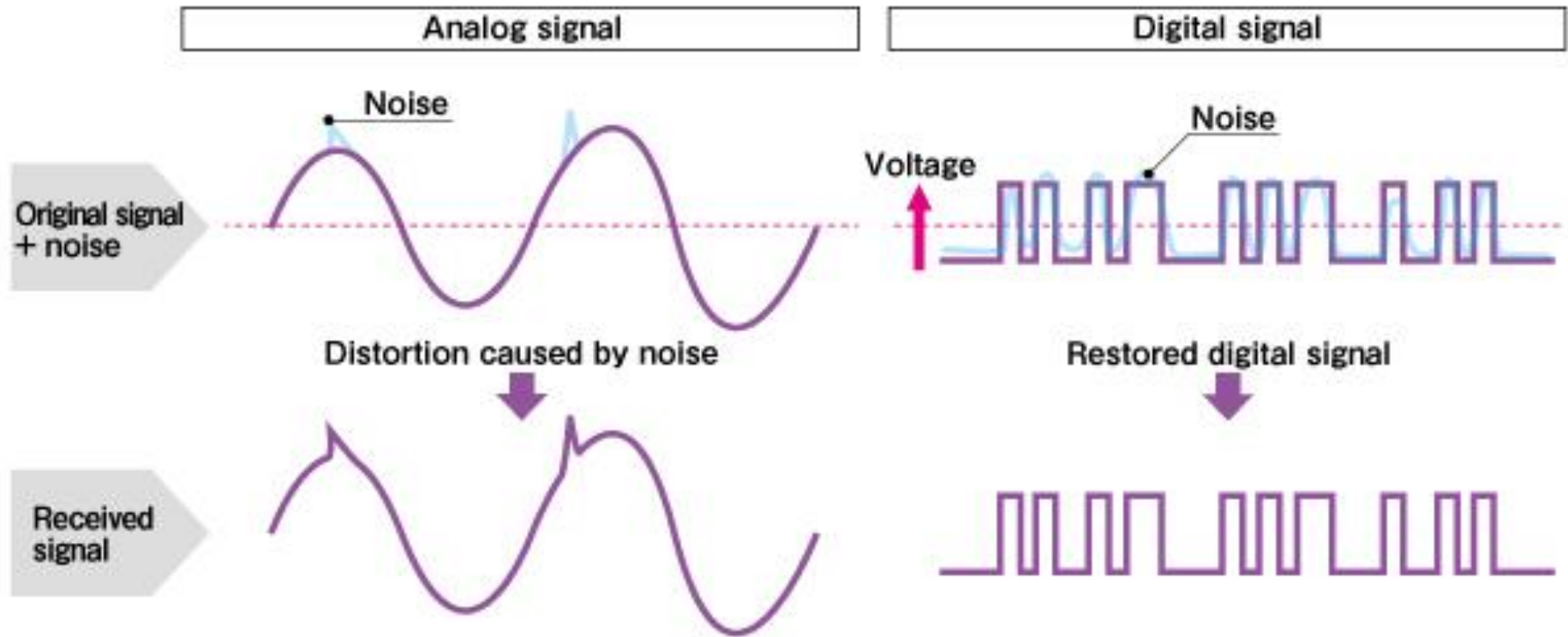
Ngoài tận dụng sức mạnh
máy tính thì chuyển sang tín hiệu
(dữ liệu) sẽ **được lợi gì nữa?**

Lợi ích của dữ liệu số

- Xử lý tính toán trong máy tính, tận dụng sức mạnh của máy tính, thuận tiện trong việc cài đặt các thuật toán phức tạp, nhưng mạnh mẽ thao tác trên con số để xử lý dữ liệu.
Ví dụ: kỹ xảo âm thanh, hình ảnh, video...

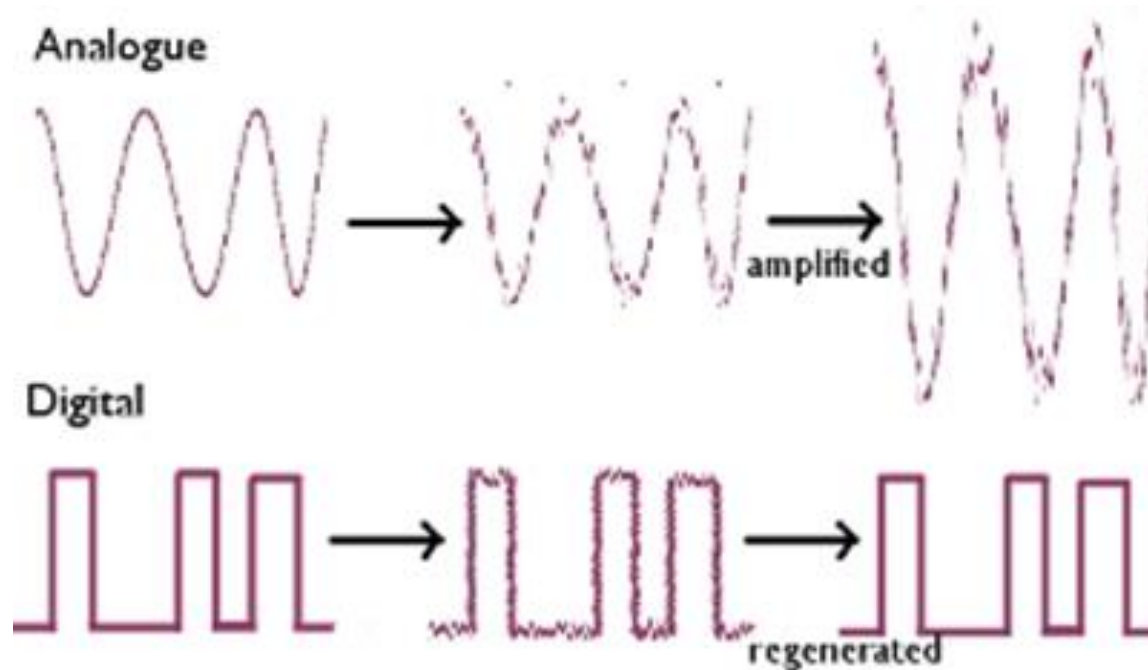


- Tin cậy (1): có khả năng chống chịu với ảnh hưởng của tạp âm / nhiễu tốt hơn tín hiệu tương tự, đặc biệt trong truyền thông.



**Tín hiệu nhận được PHẢI ĐỒNG DẠNG
với tín hiệu gốc => YÊU CẦU KHÓ**

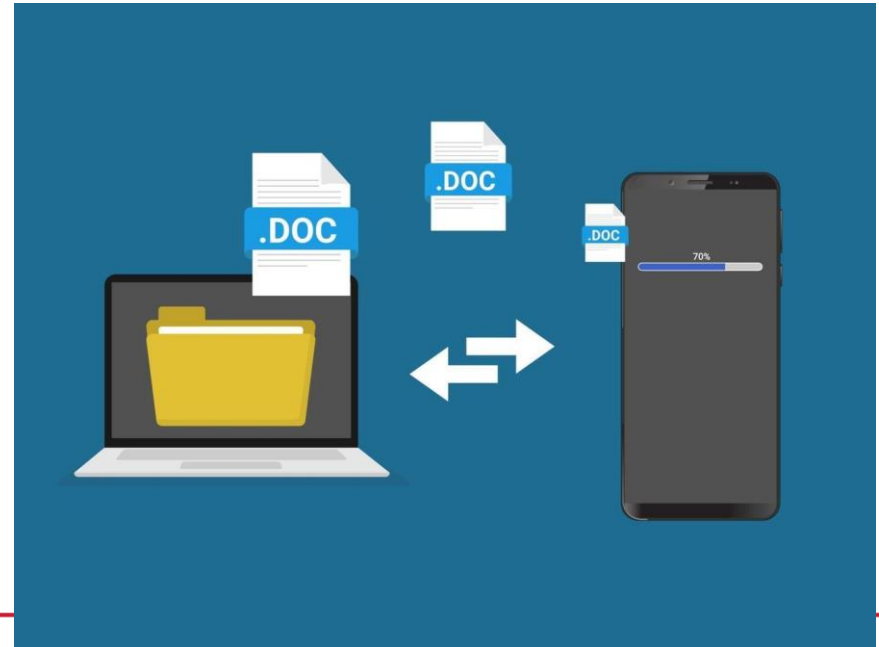
**Tín hiệu nhận được KHÔNG CÓ YÊU
CẦU ĐỒNG DẠNG, CHỈ CẦN DUY TRÌ
MỨC PHÂN BIỆT (0 VÀ 1) => DỄ HƠN**



- Tin cây (2): Sử dụng bộ lặp (repeater) hiệu quả hơn sử dụng bộ khuếch đại (amplifier) trong truyền thông tương tự.

- Thuận tiện trong lưu trữ và chia sẻ:

- Dữ liệu số không phụ thuộc vào môi trường lưu trữ và truyền thông nào cụ thể.
Ví dụ: Copy từ đĩa từ sang băng cassette (dữ liệu tương tự) cần các thiết bị đặc biệt. Trong khi với dữ liệu số thì chỉ là các bit, việc copy và lưu trữ rất đơn giản.
- Hiệu quả trong nén dữ liệu ở dạng số, ví dụ: MP3.



Yêu cầu:

- Số hóa một bài hát có thời lượng 5 phút.

Biết:

- Băng thông (dải tần số) của tín hiệu giọng nói và âm nhạc là từ 16 Hz đến 20 KHz.
- Lựa chọn số hóa với tần số lấy mẫu phù hợp tối thiểu, và lưu trữ 16 bit cho mỗi mẫu.

Hỏi:

- Ổ cứng cần tối thiểu dung lượng lưu trữ còn trống là bao nhiêu để lưu bài hát này?