TRƯỜNG ĐẠI HỌC XÂY DỰNG HÀ NỘI
KHOA CÔNG NGHỆ THÔNG TIN

LẬP TRÌNH ÂM THANH

CHƯƠNG 4

Cơ sở xử lý tín hiệu số

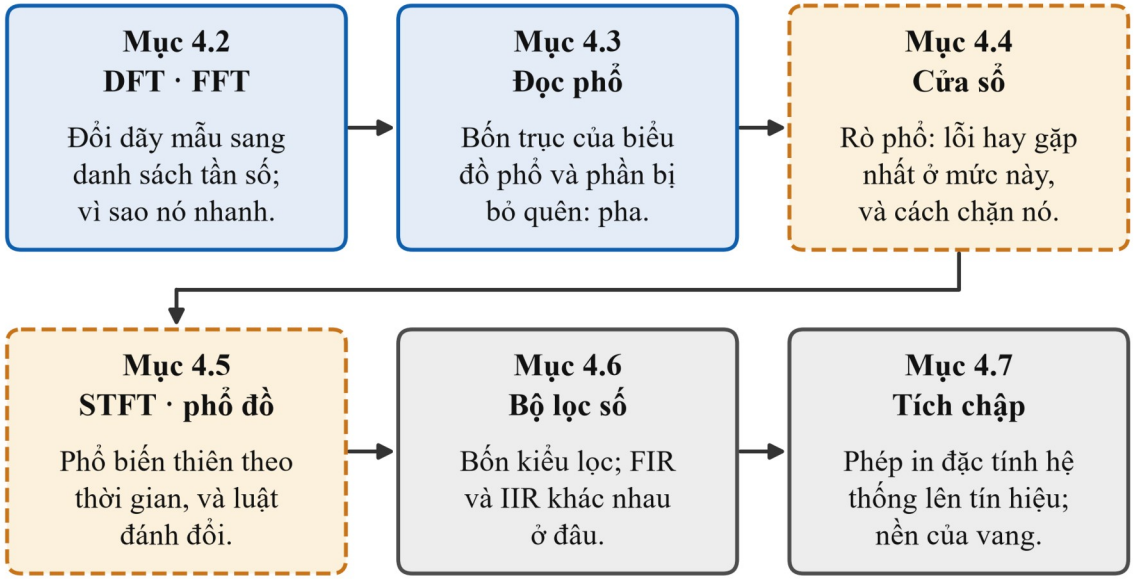
Sau chương này người học phải làm được

- Giải thích quan hệ giữa miền thời gian và miền tần số; mỗi miền trả lời loại câu hỏi nào.
- Nói bằng lời DFT (Discrete Fourier Transform – biến đổi Fourier rời rạc) làm gì không cần số phức; giải thích vì sao FFT (Fast Fourier Transform – biến đổi Fourier nhanh) nhanh hơn.
- Tính độ phân giải tần số và thời lượng khung từ N và fs; đọc đúng bốn trục của biểu đồ phổ.
- Giải thích rò phổ từ nguyên nhân gốc, chọn cửa sổ hợp lý, tính số mẫu cần để tách hai tần số.
- Đọc phổ đồ giọng nói, chọn độ dài khung theo việc cần làm, phát biểu luật đánh đổi thời gian – tần số.
- Phân biệt bốn kiểu lọc và hai họ FIR (Finite Impulse Response – đáp ứng xung hữu hạn) – IIR (Infinite Impulse Response – đáp ứng xung vô hạn); kiểm được điều kiện ổn định.
- Giải thích tích chập bằng ba động tác; chứng minh lọc FIR là tích chập; vì sao vang tích chập phải qua FFT.

MỤC 4.1

Vị trí chương học trong học phần

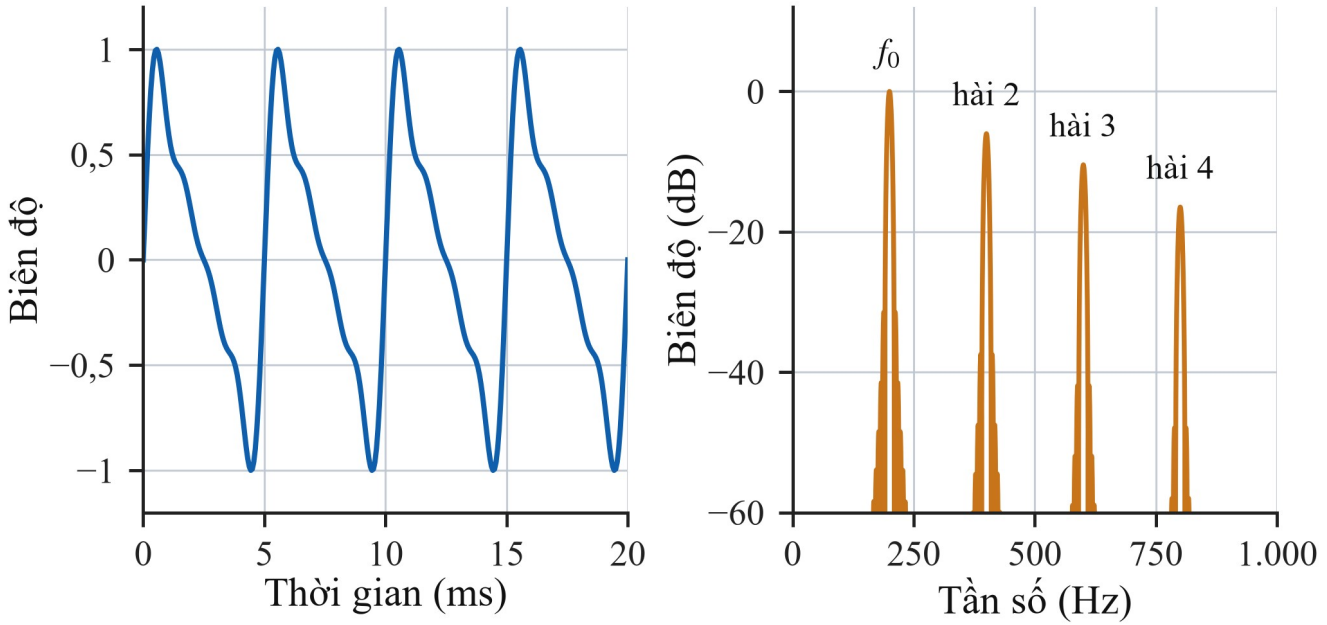
- Kết quả của ba chương đầu: một mảng số thực một chiều kèm tần số lấy mẫu.
- Từ chương này, mọi thứ diễn ra trên mảng số ấy.
- Sáu công cụ nối tiếp nhau: từ miền tần số tới tích chập.



Chương 3 kết thúc khi âm thanh đã thành một mảng số; chương 4 hỏi: làm được gì với mảng số ấy? Sáu công cụ trên là nền toán của học phần, từ chương 5 chỉ dùng lại.

Hai mô tả đầy đủ và tương đương

- Miền thời gian: tại mỗi thời điểm, tín hiệu mạnh yếu ra sao; đó là dạng sóng.
- Miền tần số: gồm những tần số nào, mỗi tần số mạnh bao nhiêu; đó là biểu đồ phổ.
- Từ bản này luôn dựng lại được bản kia mà không mất gì.
- Hỏi cách nào đúng hơn là câu hỏi vô nghĩa, như so mặt bằng với mặt đứng.



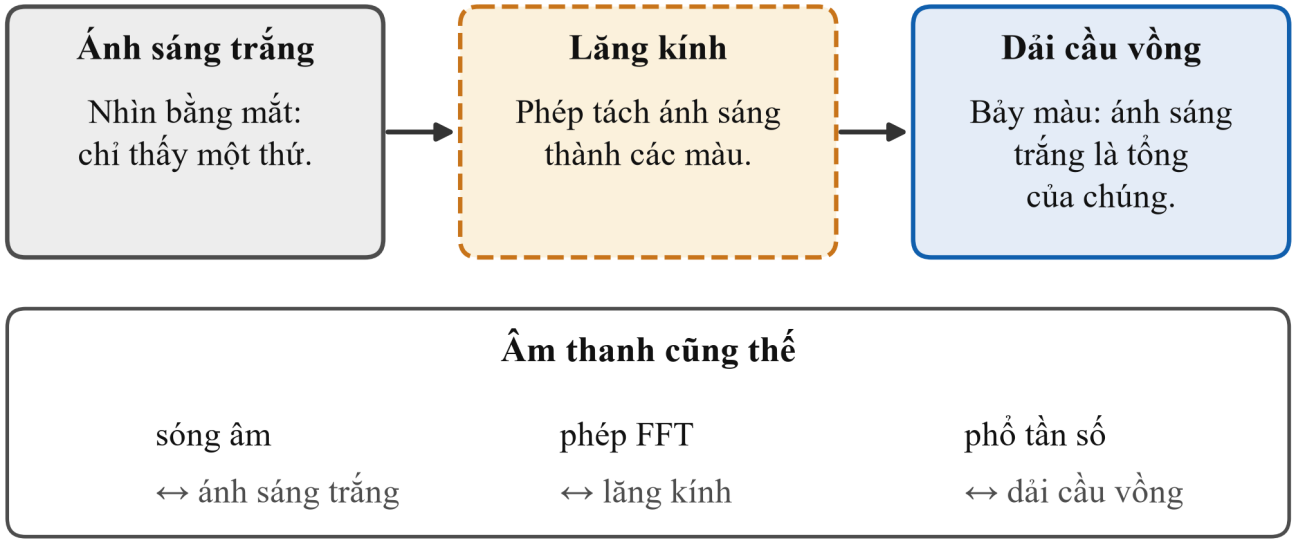
(a) Miền thời gian — *lúc nào to, lúc nào nhỏ?*

(b) Miền tần số — *gồm tần số nào, mạnh bao nhiêu?*

Hình 4.2 — Cùng một dãy số, hai cách nhìn — mỗi cách trả lời một loại câu hỏi khác nhau

Lăng kính: tách ra, không thêm vào

- Đúng: lăng kính không tạo màu nào; FFT cũng không sinh tần số nào.
- Sai: lăng kính làm việc theo thời gian thực; FFT phải nhìn một đoạn hữu hạn.
- Độ dài hữu hạn sinh ra rò phổ (mục 4.4) và luật đánh đổi (mục 4.5).
- Dùng ẩn dụ để mở bài, rồi bỏ khi bàn tới tham số.



Lăng kính không thêm gì vào ánh sáng: bảy màu vốn đã có sẵn ở đó.

FFT cũng vậy: không tạo ra tần số nào, chỉ bày ra những tần số vốn đã nằm trong dãy số.

Hình 4.3 — Ẩn dụ lăng kính: phép tách không thêm gì vào cái được tách

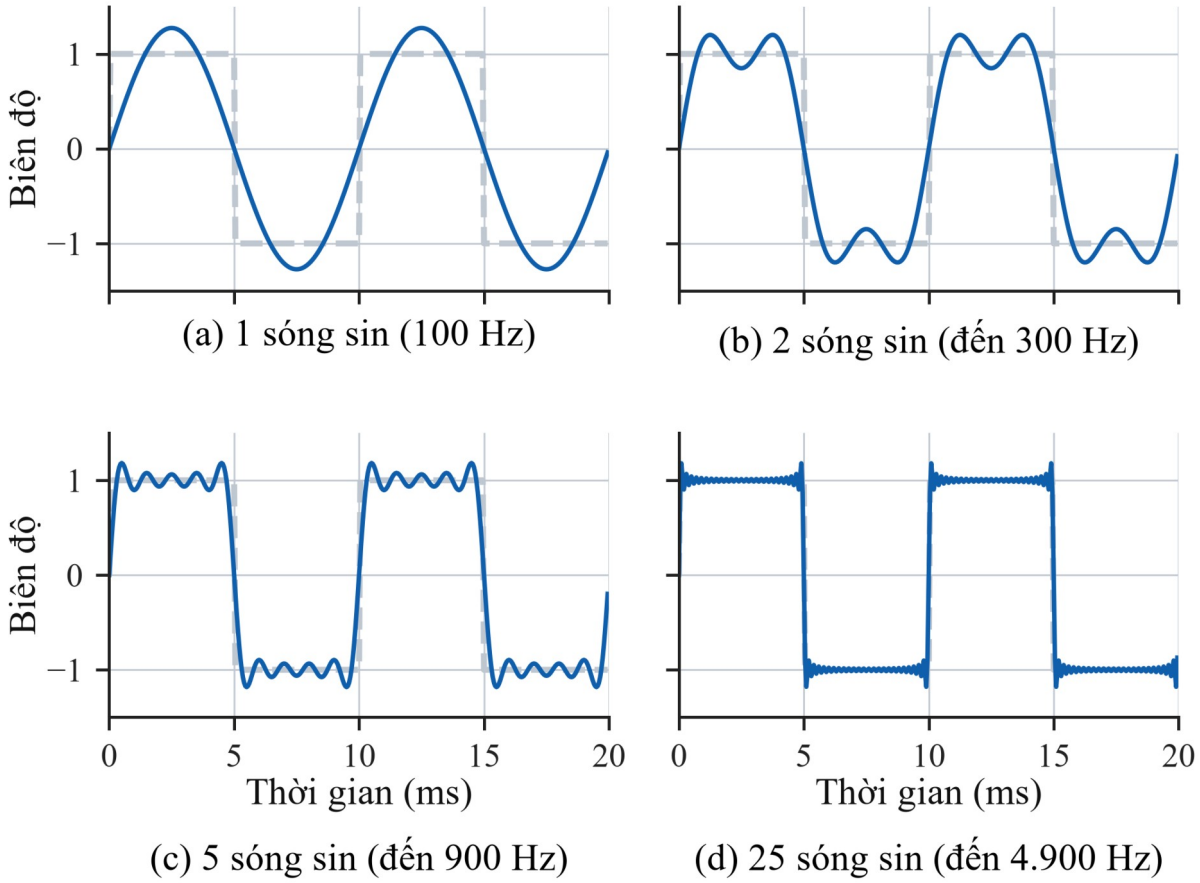
MỤC 4.2

Biến đổi Fourier rời rạc và thuật toán FFT

- Mọi tín hiệu là tổng các sóng sin, và cách tách là duy nhất.
- DFT là phép đo độ giống; FFT là cách tính nhanh cùng kết quả.
- Trục tần số, biên độ, pha, đệm số không, khung mã tối thiểu.

Cộng đủ sóng sin thì ra bất kỳ hình gì

- Cộng các hài lẻ của 100 Hz: một, hai, năm, rồi hai mươi lăm sóng sin.
- Fourier phát biểu chiều ngược lại, mạnh hơn: mọi tín hiệu đều tách được.
- Danh sách sóng sin, biên độ và độ lệch pha mô tả đầy đủ không kém dãy mẫu.



Hình 4.4 — Cộng dần các hài lẻ của 100 Hz: một, hai, năm, rồi hai mươi lăm sóng sin

Công thức DFT

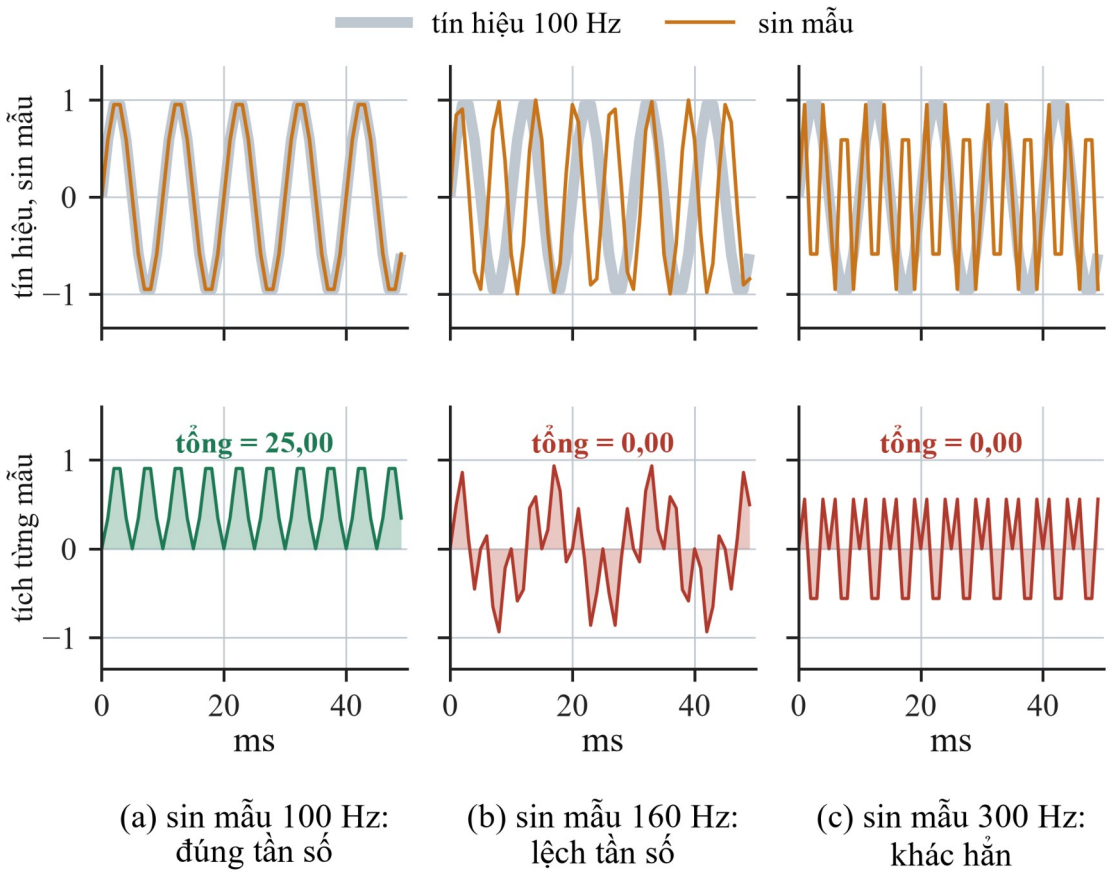
$$X[k] = \sum_{n=0}^{N-1} x[n]\, e^{-j2\pi kn/N}, \qquad k = 0, 1, \ldots, N-1$$

(4.1)

- x[n] là dãy mẫu; N là số mẫu trong khung; X[k] là kết quả ở vạch tần số thứ k.
- Thừa số mũ phức là một cặp sóng sin và cosin tần số k.
- Phép toán thật sự: nhân tín hiệu với một sóng sin mẫu rồi cộng các tích lại.

Nhân với sóng sin mẫu, rồi cộng

- Tần số trùng: tích dương ở nơi hai bên cùng dấu, tổng lớn nghĩa là giống.
- Tần số không trùng: phần dương và âm triệt tiêu, tổng gần bằng không.
- Đo với mọi tần số mẫu thì được cả phổ.



Hình 4.5 — Ý tưởng của DFT: nhân với một sóng sin mẫu rồi cộng — tổng lớn nghĩa là giống

Số phức chỉ là cách ghi chép

Thành phần lệch pha 90 độ so với sóng sin mẫu cho tổng bằng không. Cách chữa: đo hai lần, bằng sin và bằng cosin. Phần thực giữ kết quả cosin, phần ảo giữ kết quả sin. X[k] đơn giản là một cặp hai con số.

FFT trả về những tần số nào

$$f_k = k \cdot \frac{f_s}{N}; \qquad \Delta f = \frac{f_s}{N}; \qquad \Delta t = \frac{N}{f_s}$$

(4.2)

- Kết quả không phải đường cong liên tục mà là đúng N/2 + 1 con số.
- Δf = fs/N là độ phân giải tần số: khoảng cách giữa hai vạch liền nhau.
- Δt là khoảng thời gian mà khung bao phủ.

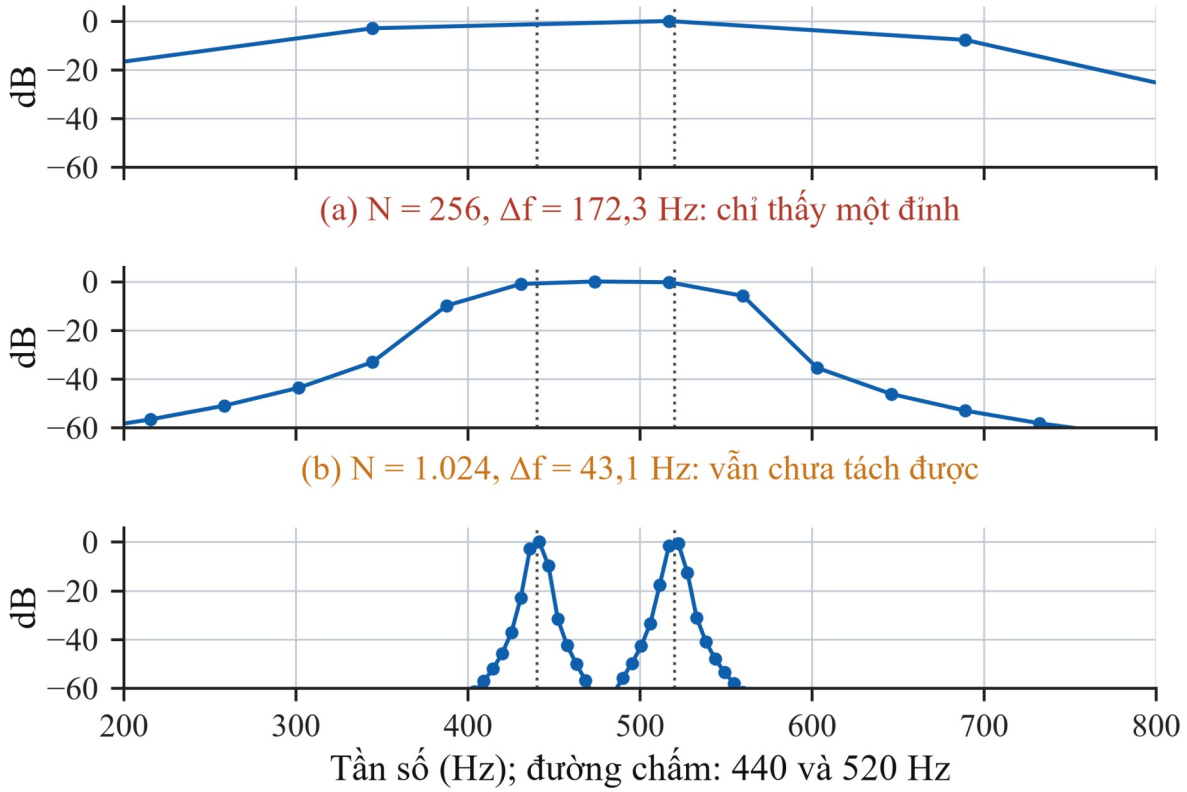
Độ phân giải tần số theo độ dài khung, ở 44.100 Hz

N (mẫu)	Δf (Hz)	Δt (ms)	Ghi chú thực hành
128	344,5	2,9	quá thô cho tần số, tốt cho tiếng gõ
512	86,1	11,6	dùng cho tiếng nói
1.024	43,1	23,2	giá trị mặc định phổ biến nhất
2.048	21,5	46,4	dùng cho nhạc, phân biệt được nốt
8.192	5,4	185,8	phân tích tĩnh, đo chính xác
16.384	2,7	371,5	chỉ dùng khi tín hiệu thật sự đứng yên

N = 1.024 không có nghĩa là tách được hai tần số cách 43 Hz

Đúng: khoảng cách giữa hai vạch là 43 Hz. Sai: hai tần số cách nhau 43 Hz chưa chắc tách thành hai đỉnh. Phải tính thêm bề rộng búp chính của cửa sổ: với Hann là bốn vạch, tức 172 Hz.

Hai tần số 440 và 520 Hz với ba độ dài khung



Hình 4.7 — Hai tần số 440 và 520 Hz với ba độ dài khung: 256, 1.024 và 8.192 mẫu

Hai nửa của kết quả X[k]

$$|X[k]| = \sqrt{\mathrm{Re}^2 + \mathrm{Im}^2}\ ; \qquad \varphi[k] = \mathrm{atan2}\,(\mathrm{Im},\ \mathrm{Re})$$

(4.3)

- Biên độ: thành phần tần số ấy mạnh bao nhiêu; mọi biểu đồ phổ vẽ ra đại lượng này.
- Pha: thành phần ấy bắt đầu ở đâu trong chu kỳ; gần như mọi biểu đồ phổ bỏ đi.
- Bỏ pha là một quyết định có lý do, không phải một sự thật (mục 4.3.4).

Chi phí: N^2 so với $N \cdot \log_2 N$

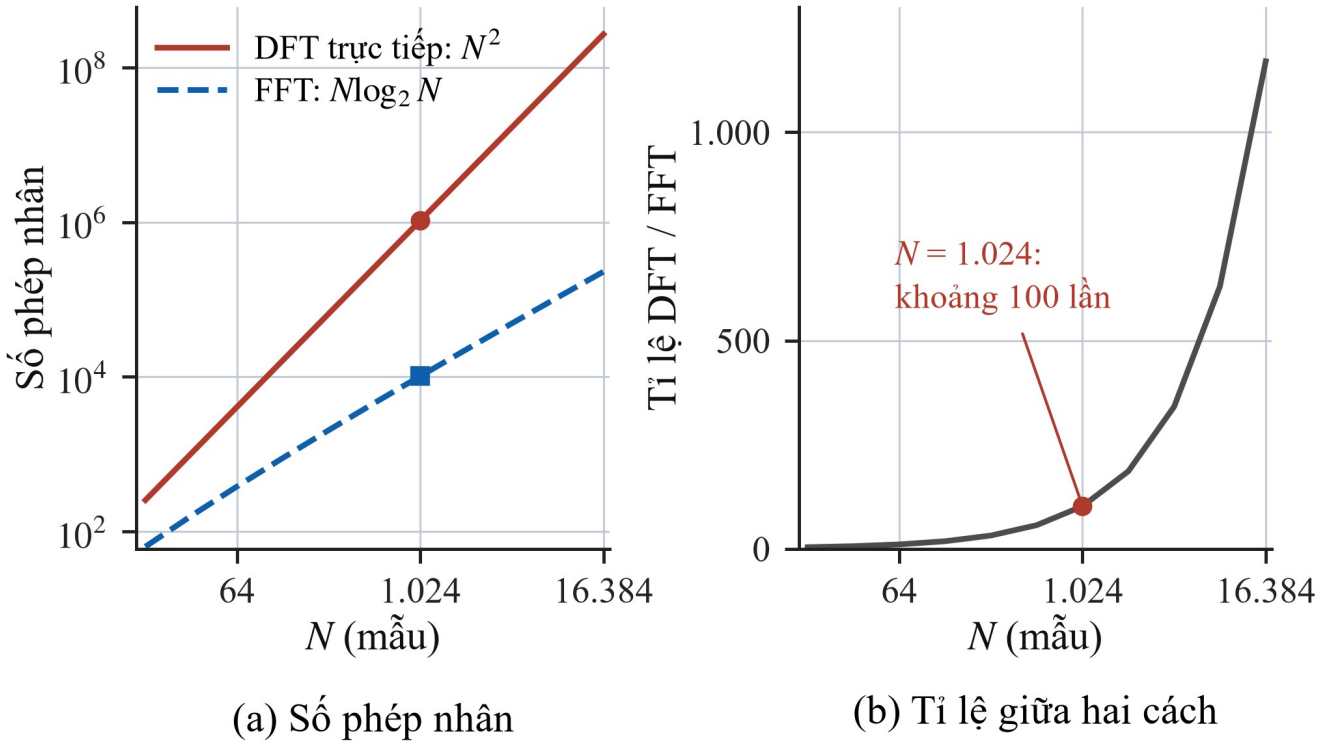
$$C_{\mathrm{DFT}} \sim N^2 ; \qquad C_{\mathrm{FFT}} \sim N \log_2 N ; \qquad \frac{C_{\mathrm{DFT}}}{C_{\mathrm{FFT}}} = \frac{N}{\log_2 N}$$

(4.4)

- DFT là định nghĩa toán học; FFT là thuật toán tính đúng kết quả ấy nhưng nhanh hơn.
- DFT theo định nghĩa: khoảng N^2 phép nhân. FFT chỉ còn khoảng $N \cdot \log_2 N$.
- Cooley và Tukey công bố năm 1965, dù ý tưởng đã có từ Gauss.

Với khung 1.024 mẫu, chênh khoảng một trăm lần

- Đó là khoảng cách giữa phân tích sau khi chờ và phân tích ngay lúc thu.
- Xử lý âm thanh thời gian thực dựng trên chênh lệch này.
- Nên chọn N trong dãy 256, 512, 1.024, 2.048, 4.096, 8.192, 16.384.
- Thiếu mẫu thì đệm số không cho đủ N.



Hình 4.8 — Chi phí tính của DFT trực tiếp và của FFT, và tỉ lệ giữa hai cách

4.2.6 · BẢNG 4.2

Đệm số không: làm gì và không làm gì

Điều người ta hy vọng	Điều thật sự xảy ra
Tách được hai đỉnh vốn dính vào nhau	Không. Hai đỉnh vẫn dính; chỉ được vẽ mịn hơn
Đo chính xác hơn tần số của một đỉnh	Có, phần nào: đỉnh nội suy gần tần số thật hơn
Nhìn rõ hình dạng của một đỉnh	Có: đây là lý do chính đáng để đệm
Làm N thành luỹ thừa của 2 cho FFT chạy nhanh	Có: lý do phổ biến nhất trong thực tế

Số không không mang tin; đệm chỉ nội suy từ thông tin đã có.

Biểu đồ phổ đúng đơn vị

- Toàn bộ mục quy về chín dòng mã; nên gõ lại bằng tay.
- Mã sinh tín hiệu ba hài rồi tìm lại: cách kiểm nhanh nhất.
- Kết quả đúng: 440 Hz ở 0 dB, 880 Hz ở khoảng −6 dB, 1.320 Hz ở khoảng −12 dB.

```
import numpy as np

x = doc_wav(...)                          # dãy mẫu, tần số lấy mẫu sr
n = 4096                                  # độ dài khung (luỹ thừa của 2)
w = np.hanning(n)                         # cửa sổ, xem mục 4.4

X = np.fft.rfft(x[:n] * w)                # chỉ lấy nửa phổ dương
f = np.fft.rfftfreq(n, 1 / sr)            # trục tần số (Hz)

mag = np.abs(X)                           # biên độ
pha = np.angle(X)                         # pha (radian)
dB = 20 * np.log10(mag / mag.max())       # chuẩn hoá về đỉnh
```

Hình 4.9 — Khung mã tối thiểu để có một biểu đồ phổ đúng đơn vị

Tính phổ và in ba tần số mạnh nhất

```
import numpy as np

def pho_dinh(x, sr, n=4096, so_dinh=3):
    """Tra ve so_dinh tan so manh nhat trong khung dau tien."""
    n = min(n, len(x))
    w = np.hanning(n)
    X = np.abs(np.fft.rfft(x[:n] * w))
    f = np.fft.rfftfreq(n, 1 / sr)
    idx = np.argsort(X)[::-1][:so_dinh]
    return [(float(f[i]), float(X[i] / X.max())) for i in idx]

# vi du: sinh mot tin hieu ba hai roi tim lai chung
sr = 44100
t  = np.arange(sr) / sr
x  = (np.sin(2*np.pi*440*t) + 0.5*np.sin(2*np.pi*880*t)
      + 0.25*np.sin(2*np.pi*1320*t))

for f, muc in pho_dinh(x, sr):
    print(f"{f:8.1f} Hz   {20*np.log10(muc):6.1f} dB")
```

MỤC 4.3

Đọc một biểu đồ phổ

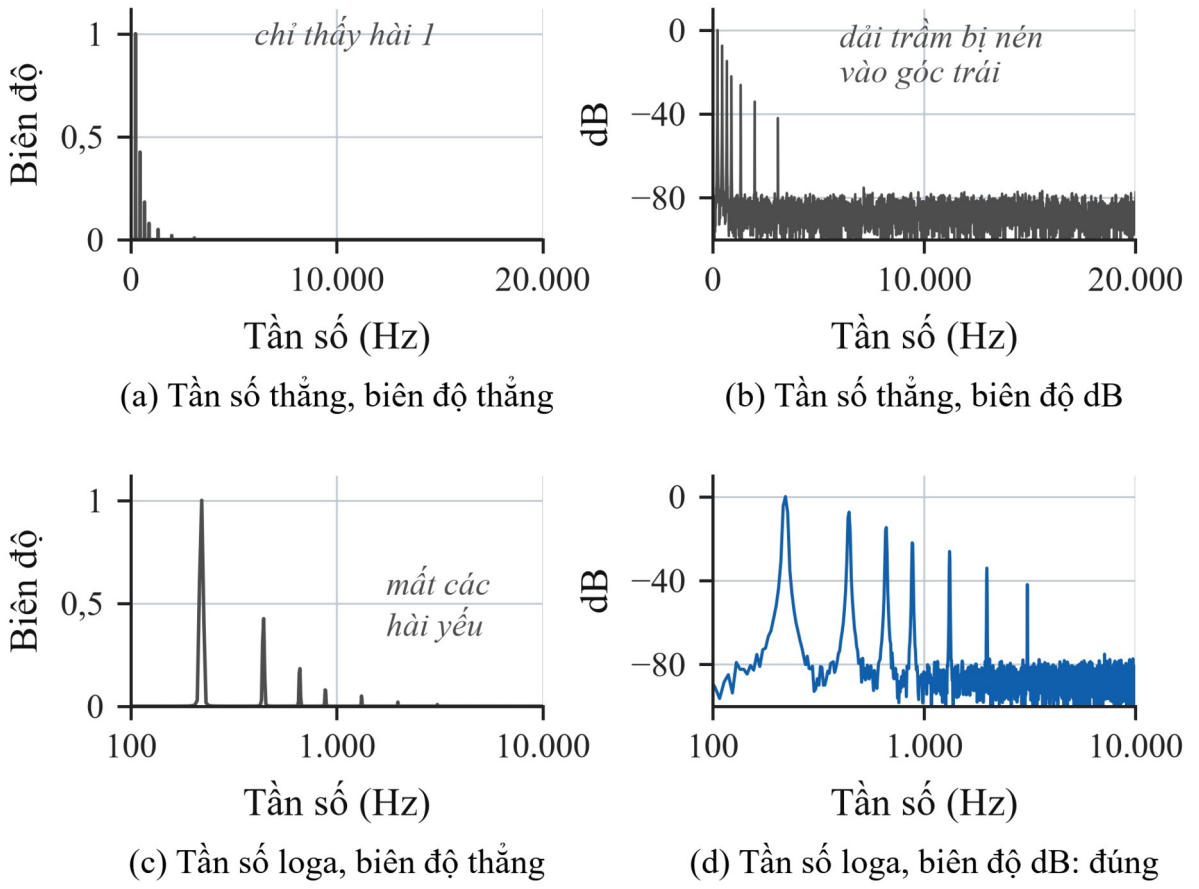
- Biểu đồ phổ là loại hình dễ bị đọc sai nhất trong ngành.
- Bốn câu hỏi về trục, vân tay phổ, sàn nhiễu, pha, con số trên trục tung.
- Gộp thành quy trình bốn bước đọc một biểu đồ lạ.

Bốn câu hỏi bắt buộc trước khi đọc biểu đồ phổ

Câu hỏi	Vì sao nó đổi cách đọc
Trục hoành thẳng hay loga?	Trục thẳng dồn dải trầm vào góc trái; trục loga dàn đều theo bội, đúng với tai
Trục tung tuyến tính hay decibel?	Tuyến tính giấu mọi thứ dưới đỉnh khoảng hai mươi decibel; dB bày ra hết
Đã chuẩn hoá về đỉnh chưa?	Nếu có, đỉnh luôn ở 0 dB; so hai biểu đồ với nhau thì không được
Khung dài bao nhiêu mẫu?	Quyết định độ phân giải tần số của biểu đồ

Cùng tín hiệu, chỉ một cách đọc được

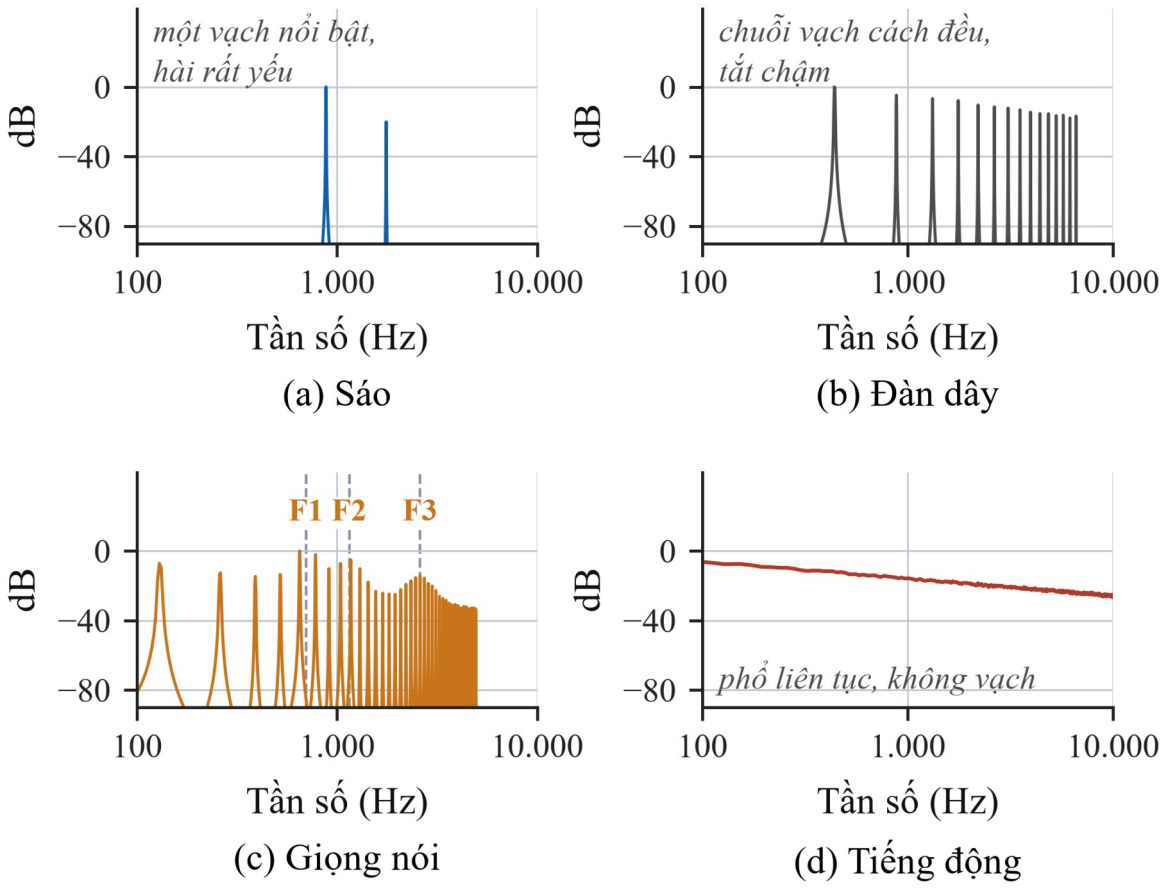
- Ô dưới bên phải là quy ước của mọi phần mềm chuyên nghiệp.
- Tai nghe tần số theo bội và nghe cường độ theo loga (chương 1).
- Đặt trục như vậy là vẽ cái tai nghe, không phải cái phép toán trả về.



Hình 4.10 — Cùng một tín hiệu, bốn cách đặt trục — và chỉ một cách đọc được

Nhận nguồn âm bằng mắt

- Kỹ năng đáng luyện nhất của chương, chuyển thẳng sang việc làm nghề.
- Bốn nguồn âm điển hình có bốn dạng phổ rất khác nhau.
- Khoảng cách giữa các vạch cho biết tần số cơ bản, tức cao độ.
- Chuỗi vạch tắt nhanh thì nghe tối và mượt; tắt chậm thì sáng và gắt.



Hình 4.11 — Bốn nguồn âm điển hình và bốn dạng phổ đặc trưng của chúng

Bốn dấu hiệu cần nhận ra trên biểu đồ phổ

Dấu hiệu nhìn thấy	Kết luận về nguồn
Các vạch nhọn, cách đều nhau	Nguồn tuần hoàn có cao độ: nhạc cụ, giọng hát, máy quay đều
Một vạch nổi bật, các vạch sau rất yếu	Âm gần thuần: sáo, âm chuẩn của bộ đo, sóng sin
Chuỗi vạch cách đều, vài vùng gồ lên nằm ngang	Giọng nói hoặc nhạc cụ có khoang cộng hưởng; vùng gồ là formant
Phổ liên tục, không có vạch	Tiếng động, gió, nhiễu nền: không có cao độ

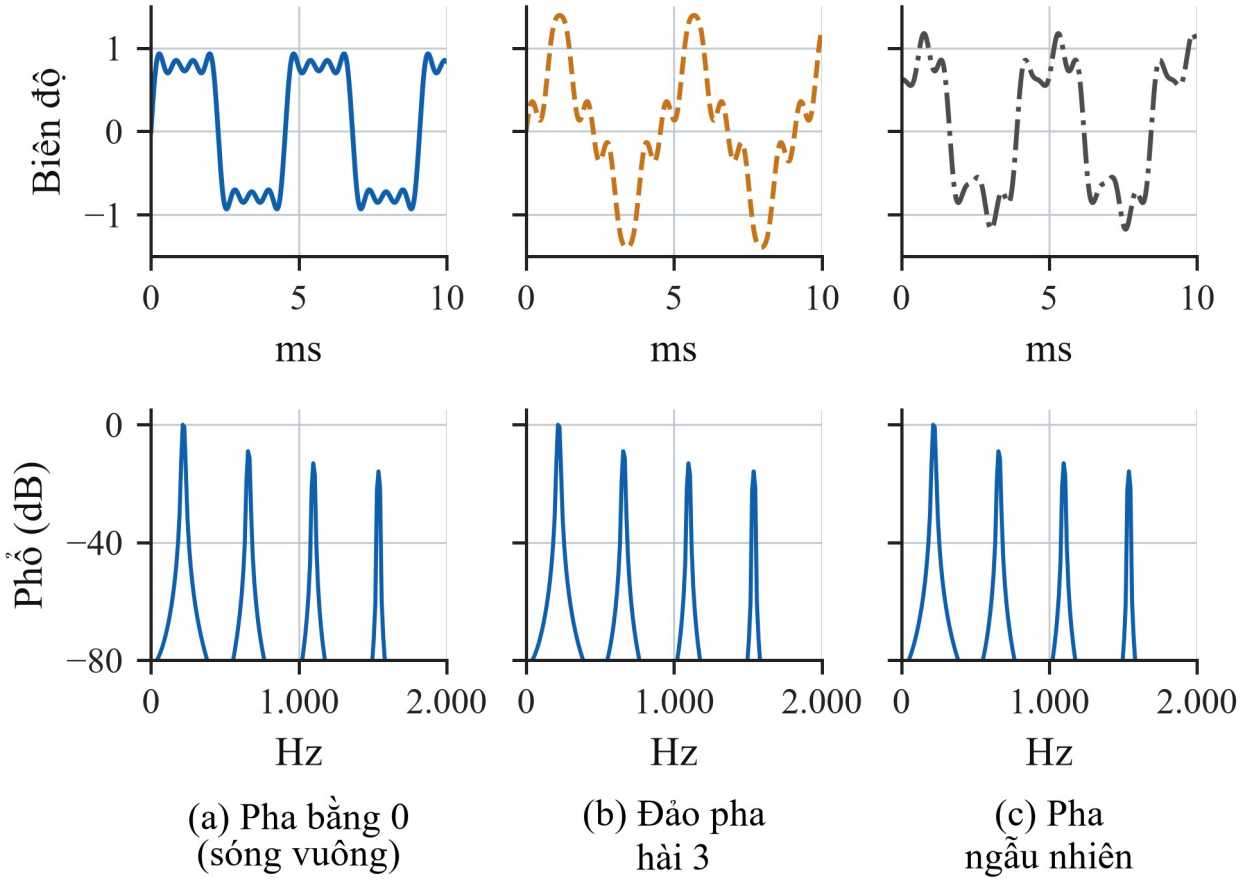
Phần dưới của biểu đồ cũng có nghĩa

- Sàn nhiễu cách đỉnh bao nhiêu cho biết tỉ số tín hiệu trên nhiễu (chương 2).
- Sàn phẳng là nhiễu trắng của mạch điện; sàn nghiêng xuống về phía cao là nhiễu hồng hoặc ồn môi trường.
- Vạch nhô lên ở 50 Hz và các bội là nhiễu ù lưới điện: vấn đề phần cứng.

Bẫy: sàn do rò phổ trông giống sàn nhiễu. Đổi độ dài khung: sàn thật gần như không đổi, sàn rò đổi theo.

Ba tín hiệu cùng phổ biên độ, dạng sóng khác hẳn

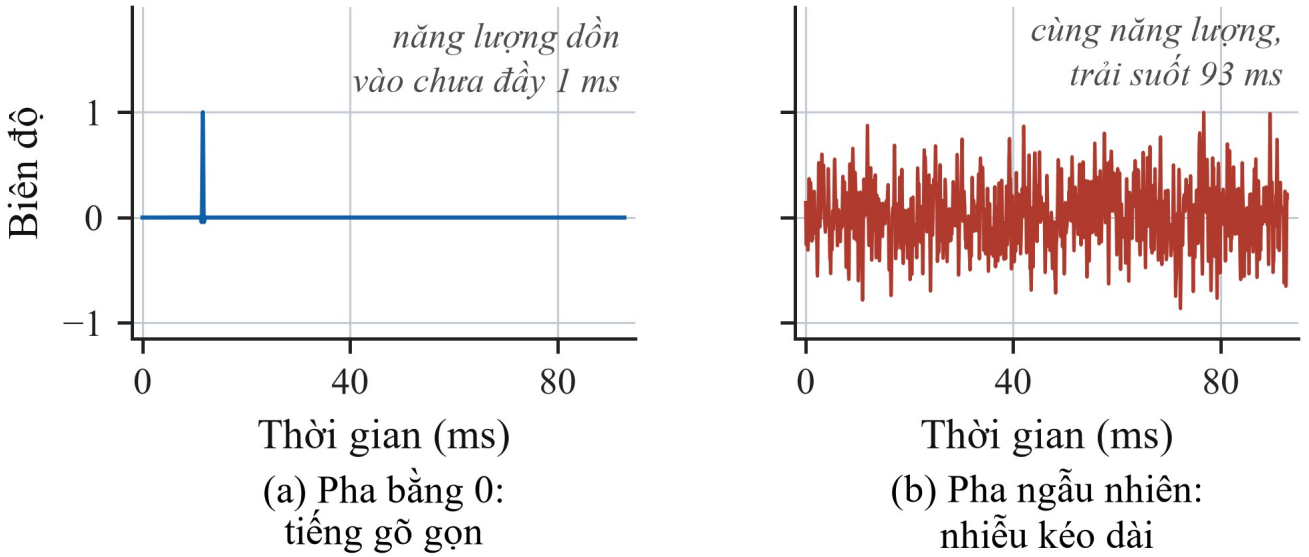
- Ba dạng sóng hàng trên khác nhau; ba phổ hàng dưới giống từng vạch.
- Toàn bộ khác biệt nằm ở pha.
- Ở âm ổn định, tai gần như không phân biệt được.



Hình 4.12 — Ba tín hiệu có phổ biên độ giống hệt nhau nhưng dạng sóng khác hẳn

Với âm chuyển tiếp, pha quyết định tất cả

- Âm chuyển tiếp: tiếng gõ, tiếng đập, phụ âm bật.
- Pha bằng không cho tiếng gõ sắc gọn; pha ngẫu nhiên cho tiếng nhiễu nhoè.
- Vì vậy co giãn thời gian và dịch cao độ gặp khó với tiếng trống (chương 10).



Hình 4.13 — Hai tín hiệu cùng phổ biên độ: pha bằng không cho tiếng gõ, pha ngẫu nhiên cho tiếng nhiễu

Ba cách trình bày trục tung

Cách trình bày	Cách tính	Đọc được gì
Chuẩn hoá về đỉnh	chia cho max của chính nó	Chỉ đọc hình dạng phổ; không so được hai biểu đồ
Biên độ thật	nhân 2, chia N, chia hệ số cửa sổ	Biên độ từng thành phần đúng như trong tín hiệu
dBFS	từ biên độ thật, áp công thức (3.7)	So được với mức toàn thang; biết có cắt đỉnh không

Cùng sóng sin biên độ 1: $|X[k]|$ ở N = 1.024 lớn gấp bốn lần ở N = 256; cửa sổ Hann làm nhỏ đi khoảng một nửa.

Bốn bước đọc một biểu đồ phổ lạ

- Bước 1. Đọc hai trục: thẳng hay loga, dB hay tuyến tính, đã chuẩn hoá chưa, khung dài bao nhiêu.
- Bước 2. Tìm cấu trúc: có vạch không, cách đều không; khoảng cách vạch là tần số cơ bản.
- Bước 3. Nhìn sàn: cách đỉnh bao nhiêu dB, phẳng hay nghiêng, có vạch ù không, là sàn thật hay rò phổ.
- Bước 4. Hỏi ngược: với khung ấy biểu đồ phân biệt được gì.

Khung 512 mẫu mà kết luận hai đỉnh cách 30 Hz là hai nguồn thì không có cơ sở.

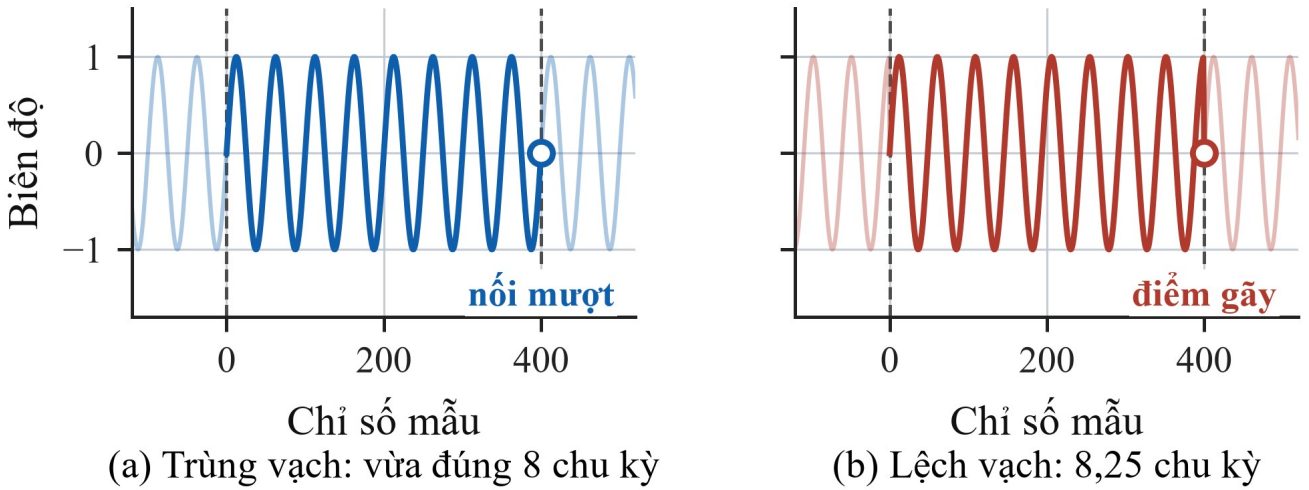
MỤC 4.4

Cửa sổ và rò phổ

- Lỗi hay gặp nhất ở người mới phân tích phổ, và là lỗi không báo.
- Nguyên nhân gốc, đo thử mức rò, cửa sổ chữa bằng cách nào.
- Hai hệ quả: thành phần yếu bị che, và số mẫu cần để tách hai tần số.

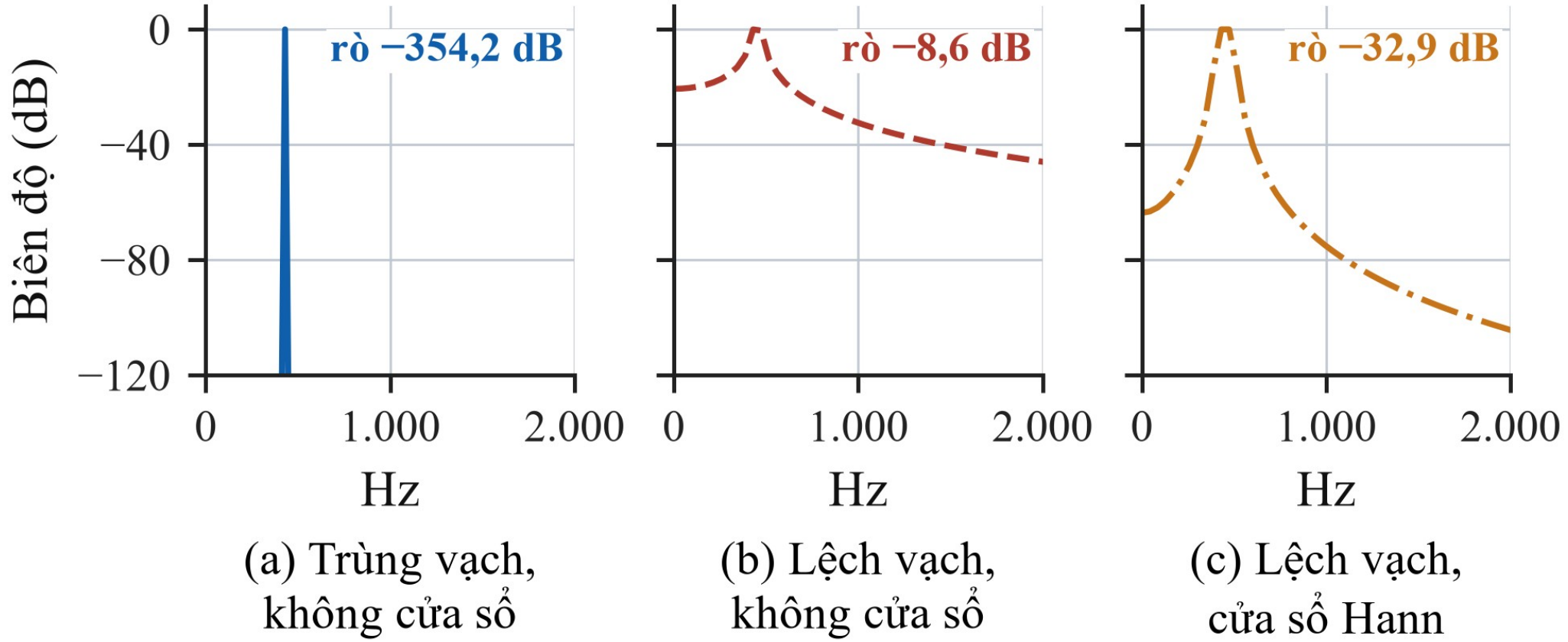
FFT ngầm nối đầu khung với cuối khung

- DFT chỉ nhìn N mẫu và ngầm giả định khung lặp lại vô hạn.
- Số chu kỳ nguyên: chỗ nối mượt, một vạch sắc nét.
- Không nguyên: chỗ nối gãy, chứa vô số tần số không có trong tín hiệu. Đó là rò phổ.



Hình 4.14 — Chỗ nối giữa hai bản lặp: mượt khi số chu kỳ chẵn, gãy khi không chẵn

Cùng một sóng sin sạch: trùng vạch, lệch vạch, lệch vạch có Hann



Hình 4.15 — Cùng một sóng sin sạch: trùng vạch, lệch vạch, và lệch vạch có cửa sổ Hann

Năng lượng rò đo trên sóng sin sạch, khung 1.024 mẫu

Trường hợp	Cửa sổ	Năng lượng ngoài búp chính
trùng đúng tâm vạch (430,7 Hz)	không	−354,2 dB
lệch nửa vạch (452,2 Hz)	không	−8,6 dB
lệch nửa vạch (452,2 Hz)	Hann	−32,9 dB
lệch nửa vạch (452,2 Hz)	Blackman	−54,8 dB

Dòng đầu chỉ còn là sai số làm tròn của máy; dòng hai là điều xảy ra nếu không làm gì.

Nhân cửa sổ trước khi đưa vào FFT

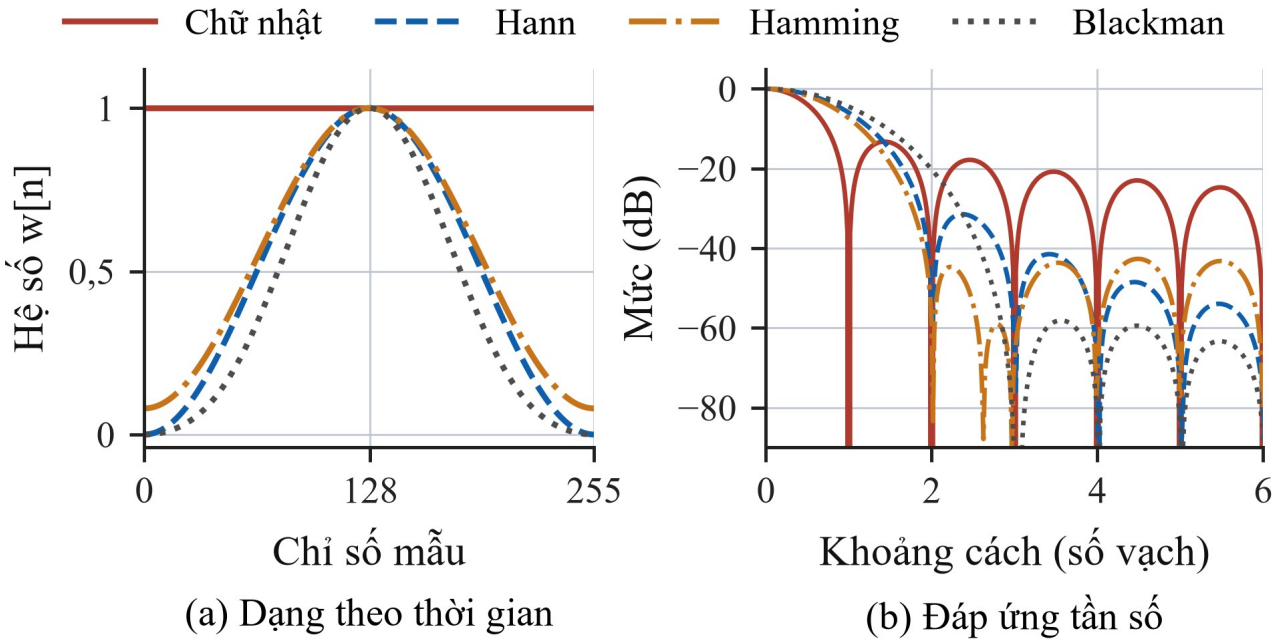
$$x_{\mathrm{w}}[n] = w[n] \cdot x[n]\,; \qquad w_{\mathrm{hann}}[n] = \frac{1}{2}\left[1 - \cos\frac{2\pi n}{N-1}\right]$$

(4.6)

- Cửa sổ là dãy hệ số cùng độ dài khung, nhân vào từng phần tử của tín hiệu.
- Dạng chuông: bằng 1 ở giữa, thon dần về 0 ở hai mép.
- Hai mép về không thì chỗ nối không còn điểm gãy, bất kể tần số tín hiệu.

Dạng theo thời gian và đáp ứng tần số

- Bề rộng búp chính: một sóng sin hiện thành búp rộng chừng ấy vạch; càng rộng càng khó tách hai tần số.
- Mức búp bên: phần năng lượng ngoài búp chính; càng thấp càng ít rò.
- Hai đặc trưng đối nghịch nhau.



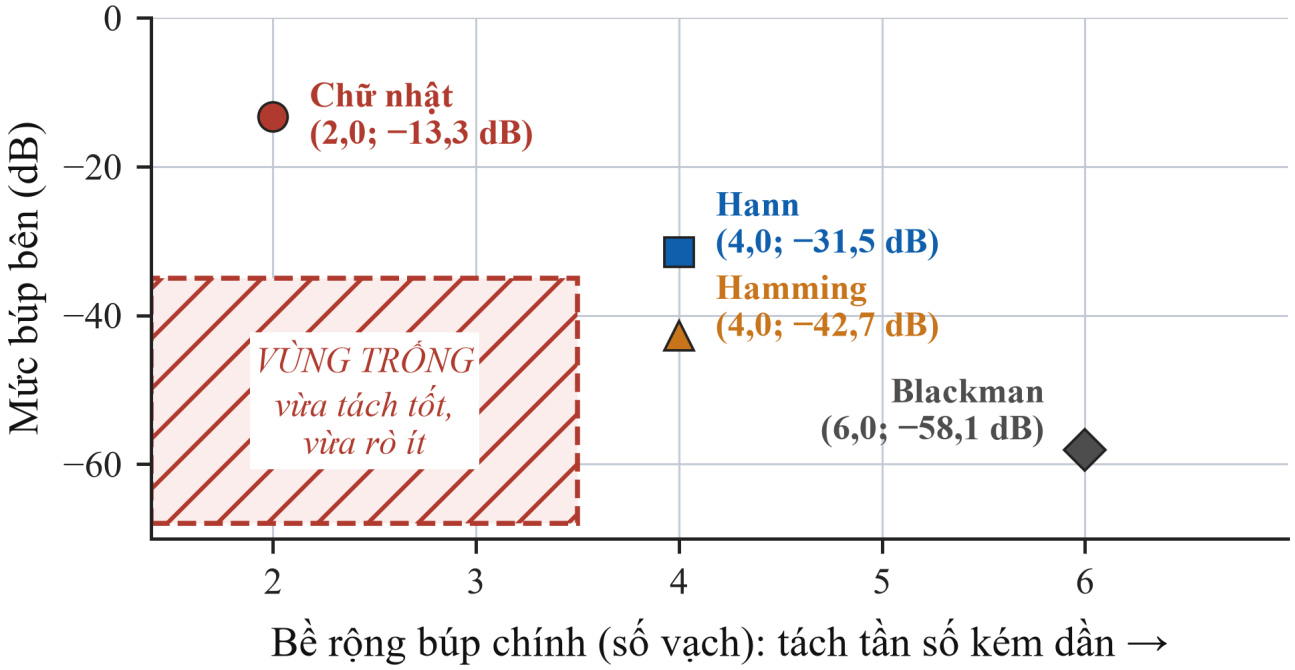
Hình 4.16 — Bốn cửa sổ thông dụng: dạng theo thời gian và đáp ứng tần số

Bốn cửa sổ thông dụng và đặc trưng

Cửa sổ	Búp chính (vạch)	Búp bên (dB)	Dùng khi nào
Chữ nhật	2,0	−13,3	tín hiệu chắc chắn trùng vạch, hoặc xung ngắn nằm gọn trong khung
Hann	4,0	−31,5	mặc định, hợp với gần hết mọi việc
Hamming	4,0	−42,7	cần thấy thành phần yếu bên cạnh đỉnh mạnh
Blackman	6,0	−58,1	đo chính xác mức một đỉnh, soi méo hài rất nhỏ

Không có cửa sổ nào tốt nhất

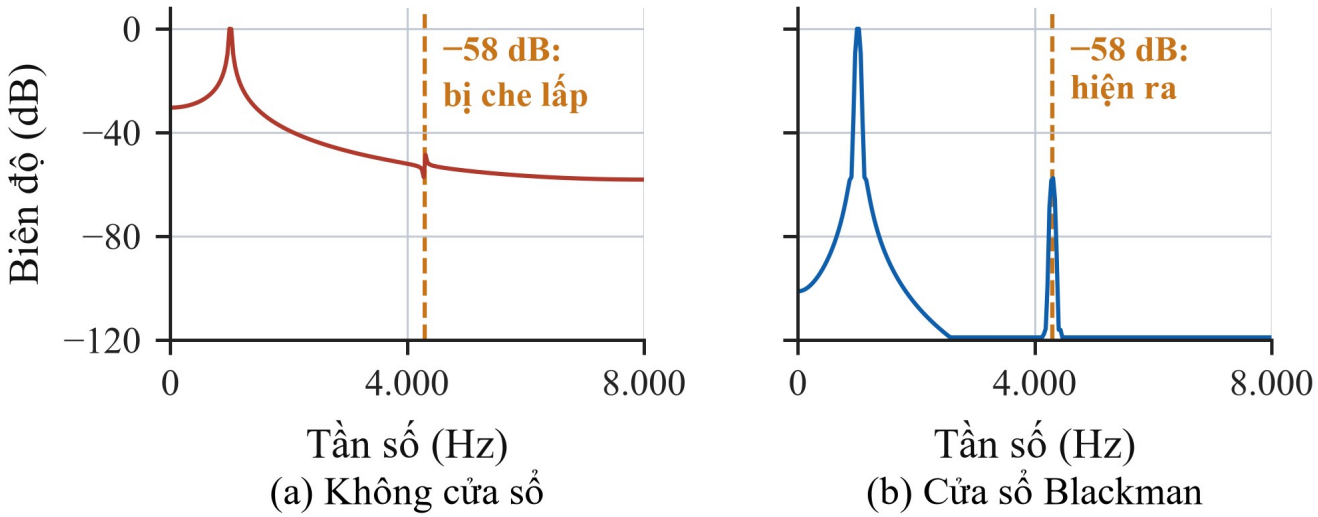
- Vùng vừa tách tốt vừa rò ít là vùng trống, sẽ không bao giờ có cửa sổ ở đó.
- Hàm càng thon trong miền thời gian thì càng bè trong miền tần số.
- Câu hỏi đúng: việc này sợ gì hơn, dính hai đỉnh hay che mất thành phần yếu.



Hình 4.17 — Chọn cửa sổ là chọn một điểm trên đường đánh đổi — góc tốt nhất thì trống

Thành phần yếu −58 dB bị sàn rò che lấp

- Phần rò của đỉnh mạnh tạo sàn giả ở hai bên.
- Chữ nhật: sàn giả quanh −13,3 dB; Blackman: xuống −58,1 dB.
- Thành phần yếu hơn sàn giả thì chìm hẳn, không còn cách nhìn thấy.
- Gặp khi đo méo hài, tìm tiếng ù nhỏ, kiểm sàn nhiễu micro.



Hình 4.18 — Một thành phần yếu −58 dB bị sàn rò che lấp, và hiện ra khi đổi cửa sổ

Số mẫu cần để tách hai tần số

$$\Delta f_{\text{tách}} \geq B_{\text{w}} \cdot \frac{f_{\text{s}}}{N} \iff N \geq B_{\text{w}} \cdot \frac{f_{\text{s}}}{\Delta f_{\text{tách}}}$$

(4.7)

- Bw là bề rộng búp chính của cửa sổ, tính bằng số vạch (Bảng 4.7).
- Vế phải dùng được hơn: biết khoảng cách tần số cần phân biệt, tính ngược ra số mẫu phải thu.
- Δf của công thức (4.2) không phải khoảng cách tối thiểu tách được.

Số mẫu cần để tách hai tần số, Hann, 44.100 Hz

Khoảng cách	Số mẫu tối thiểu	Thời lượng thu	Tình huống
500 Hz	353	8 ms	tách hai nhạc cụ ở dải trung
100 Hz	1.764	40 ms	hai nốt cách một cung ở quãng tám giữa
20 Hz	8.820	200 ms	tách hài của hai giọng hát cùng lúc
5 Hz	35.280	800 ms	đo độ lạc dây
1 Hz	176.400	4.000 ms	đo trôi tần số của máy phát

Tách hai tần số cách 5 Hz với Hann: 35.280 mẫu, tức 0,8 giây

Hai sóng lệch 5 Hz phải chờ 0,2 giây mới lệch pha đủ để phân biệt; Hann cần bốn lần khoảng ấy. Phân giải tần số cao thì bắt buộc là bài toán trên tín hiệu ổn định.

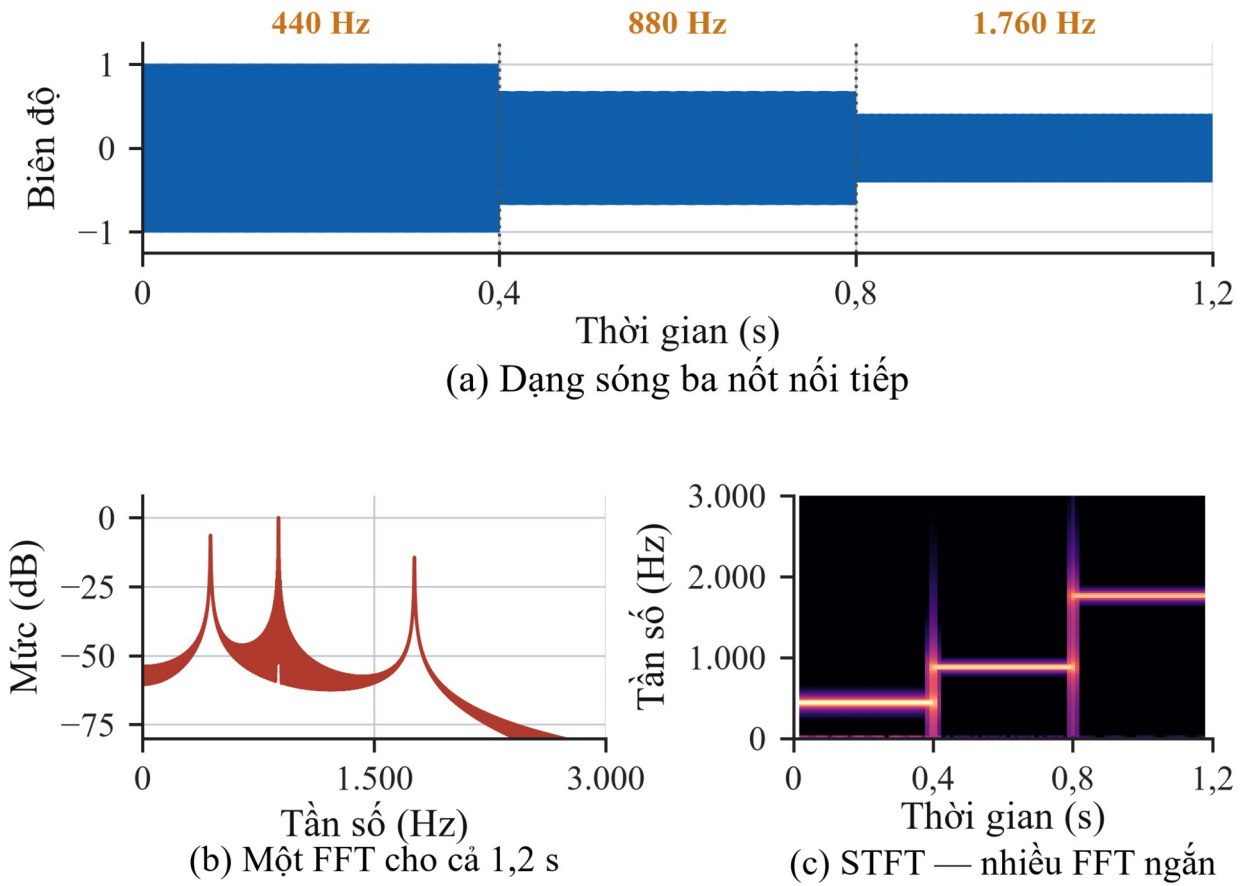
MỤC 4.5

Biến đổi Fourier thời gian ngắn và phổ đồ

- Một phép FFT giả định tín hiệu đứng yên trong khung.
- STFT (Short-Time Fourier Transform – biến đổi Fourier thời gian ngắn): cắt khung chồng lấn, làm FFT từng khung, xếp theo thời gian.
- Đọc phổ đồ và luật đánh đổi thời gian – tần số.

Đủ ba tần số nhưng mất thứ tự thời gian

- Phổ hoàn toàn đúng, nhưng không nói được ba tần số đến lần lượt.
- Giả định ngầm: tín hiệu đứng yên trong khung.
- Âm nhạc, giọng nói thay đổi theo thời gian; giả định ấy sai hoàn toàn.



Hình 4.19 — Một FFT cho cả tín hiệu thấy đủ ba tần số nhưng mất sạch thứ tự thời gian

Công thức STFT

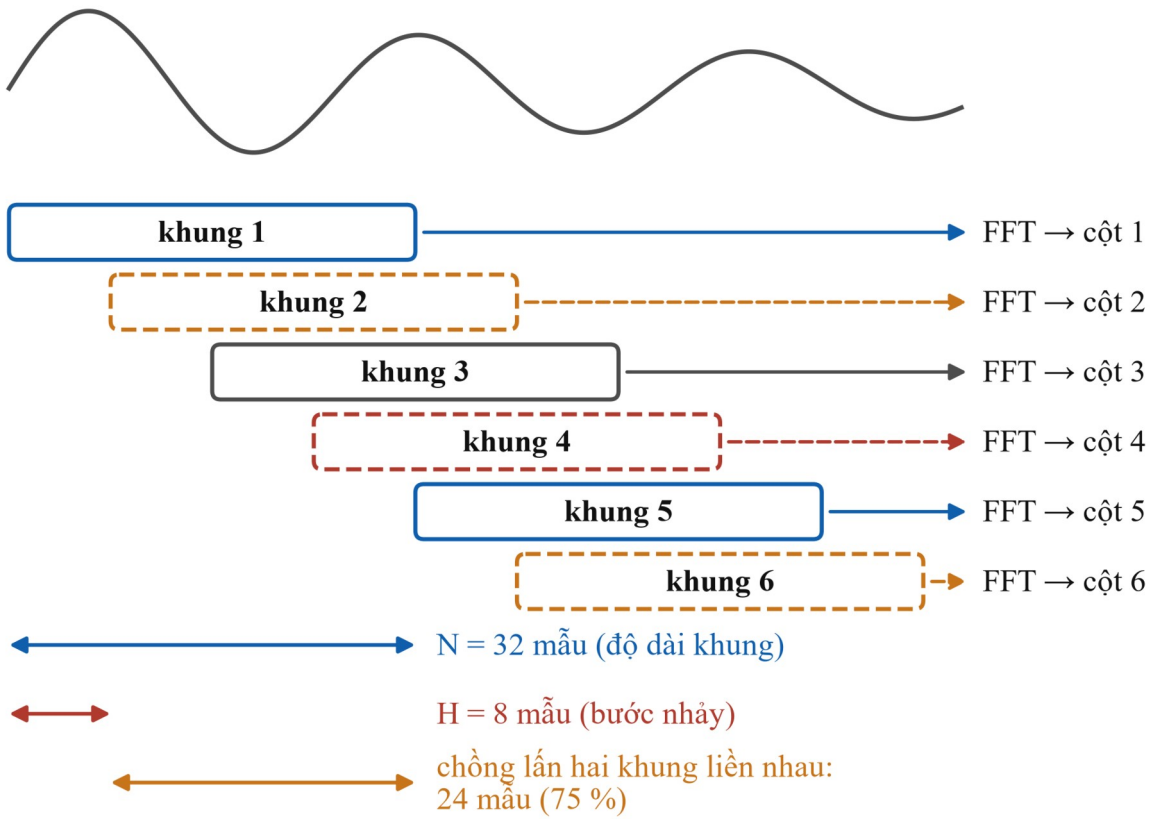
$$X[m,k] = \sum_{n=0}^{N-1} w[n]\, x[n + mH]\, e^{-j2\pi kn/N}$$

(4.8)

- m là chỉ số khung; H là bước nhảy giữa hai khung liền nhau; w là cửa sổ.
- So với công thức (4.1): chỉ thêm chỉ số m và phép dịch mH.
- STFT không phải phép toán mới, chỉ là FFT lặp lại ở nhiều vị trí.

Khung dài N trượt dọc tín hiệu, mỗi lần nhảy H mẫu

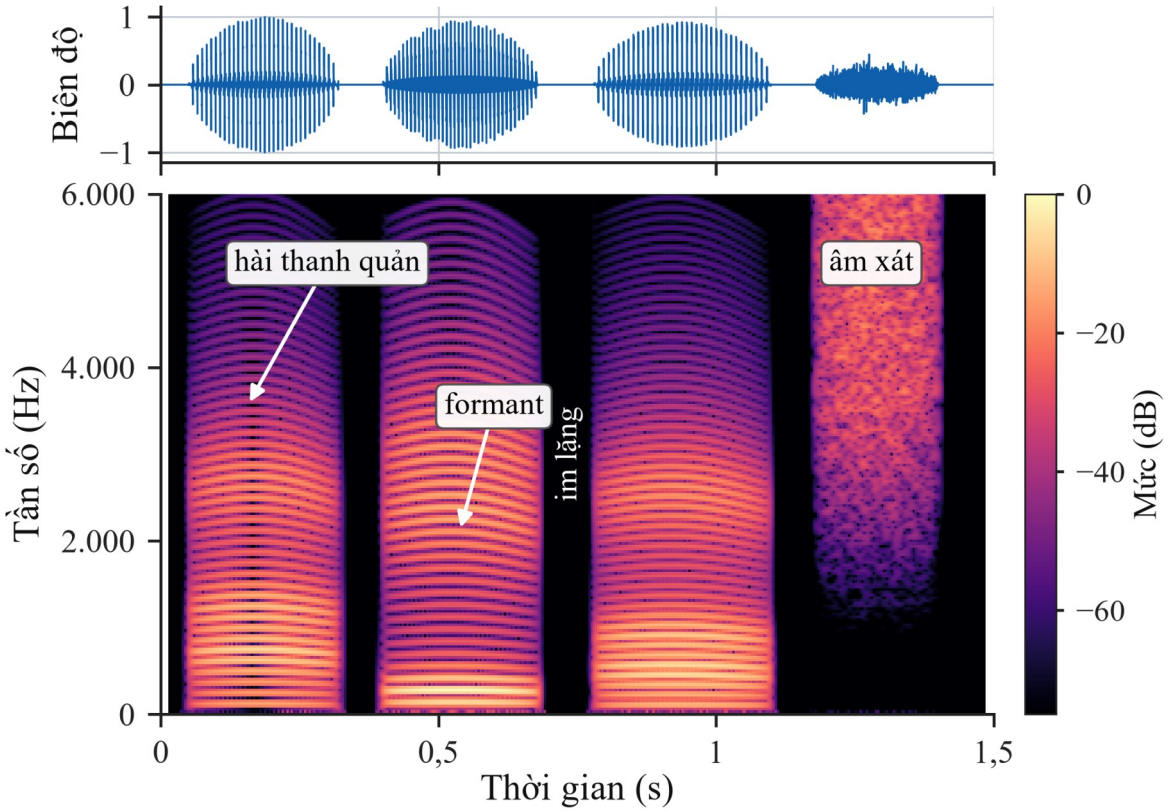
- Kết quả là bảng hai chiều: hàng là tần số, cột là thời gian, ô là mức.
- Vẽ bảng ấy bằng màu thì gọi là phổ đồ.
- Chồng lấn cần thiết vì cửa sổ làm yếu mép khung.



Hình 4.20 — Cơ chế STFT: khung dài N trượt dọc tín hiệu, mỗi lần nhảy H mẫu

Phổ đồ có ba trục: thời gian, tần số, màu

- Trục hoành: thời gian, đọc như dạng sóng.
- Trục tung: tần số, thấp ở dưới, cao ở trên.
- Màu: mức năng lượng, thường tối là yếu, sáng là mạnh; thang tính bằng decibel.



(a) Dạng sóng (trên) · (b) Phổ đồ cùng đoạn ấy (dưới)

Hình 4.21 — Phổ đồ một đoạn giọng nói: bốn thứ đọc được và ý nghĩa của chúng

Bốn dấu hiệu trên phổ đồ giọng nói

Nhìn thấy	Nghĩa là
Vạch ngang cách đều, chồng lên nhau	Hài của thanh quản; khoảng cách là tần số cơ bản, tức cao độ
Dải sáng nằm ngang, rộng, ở vài tần số cố định	Formant, cộng hưởng của khoang miệng; quyết định nguyên âm
Vùng tối, không có gì	Im lặng hoặc phụ âm câm
Mảng sáng ở dải cao, không có vạch	Âm xát "s", "x", "h": nhiễu băng rộng, không có cao độ

Dùng để soát lỗi: tìm tiếng ù, xác định đoạn bị cắt, nhận ra bản thu nén có tổn hao.

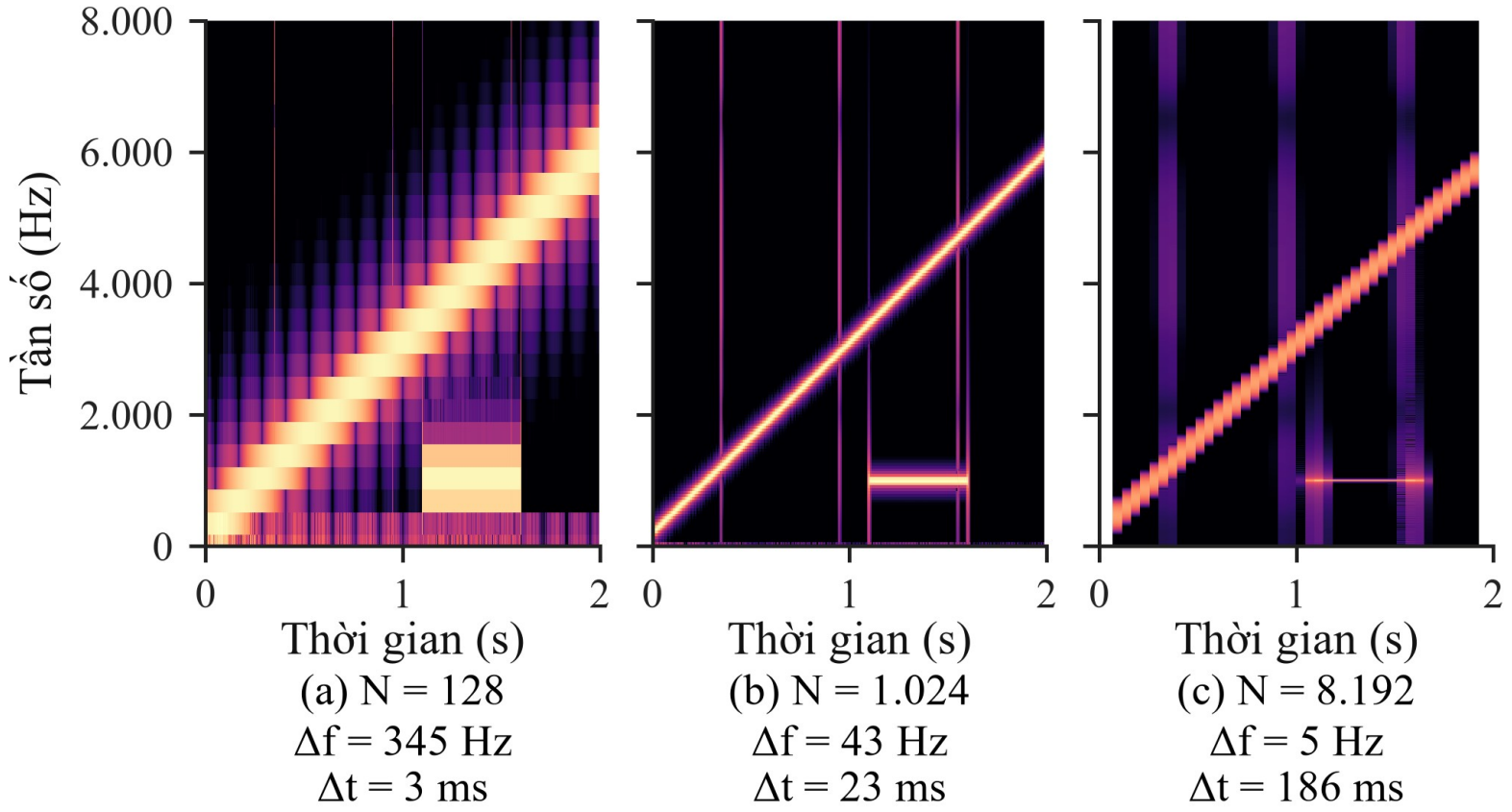
Δf × Δt = 1

$$\Delta f \cdot \Delta t = \frac{f_s}{N} \cdot \frac{N}{f_s} = 1 \quad [\text{Hz} \cdot \text{s}]$$

(4.9)

- Tích của độ phân giải tần số và độ dài khung luôn bằng 1.
- Không phải xấp xỉ: Δf = fs/N, Δt = N/fs, nhân vào thì N triệt tiêu.
- Mịn về tần số thì nhoè về thời gian, và ngược lại.

Cùng một tín hiệu, ba độ dài khung: 128, 1.024 và 8.192 mẫu



Hình 4.23 — Cùng một tín hiệu, ba độ dài khung: 128, 1.024 và 8.192 mẫu

Chọn độ dài khung theo việc cần làm, 44.100 Hz

N	Δf	Δt	Việc hợp với nó
128	345 Hz	3 ms	định vị chính xác tiếng gõ, tiếng trống
512	86 Hz	12 ms	phân tích tiếng nói, bắt được phụ âm
1.024	43 Hz	23 ms	mặc định cho việc chung
2.048	22 Hz	46 ms	phân tích nhạc, phân biệt các nốt
8.192	5 Hz	186 ms	đo tần số chính xác trên tín hiệu ổn định

Đây là một định luật, không phải hạn chế kỹ thuật

Không thuật toán nào vượt qua được, kể cả phương pháp hiện đại nhất; chúng chỉ phân bố sai số khác đi. Cùng một định luật với nguyên lý bất định trong vật lý lượng tử. Khi một công cụ hứa vừa mịn tần số vừa chính xác thời gian, hãy hỏi nó đánh đổi bằng cái gì.

Dựng và vẽ phổ đồ bằng scipy

```
import numpy as np
import matplotlib.pyplot as plt
from scipy import signal

n_fft, hop = 1024, 256          # chong lan 75 %

f, t, Z = signal.stft(x, fs=sr, window="hann",
                      nperseg=n_fft, noverlap=n_fft - hop)

A  = np.abs(Z)                           # bo pha di
dB = 20 * np.log10(np.maximum(A / A.max(), 1e-6))

plt.pcolormesh(t, f, dB, vmin=-80, vmax=0,
               shading="auto", cmap="magma")
plt.ylim(0, 8000)                # chi nhin dai co noi dung
plt.xlabel("Thoi gian (s)"); plt.ylabel("Tan so (Hz)")
plt.colorbar(label="dB")
```

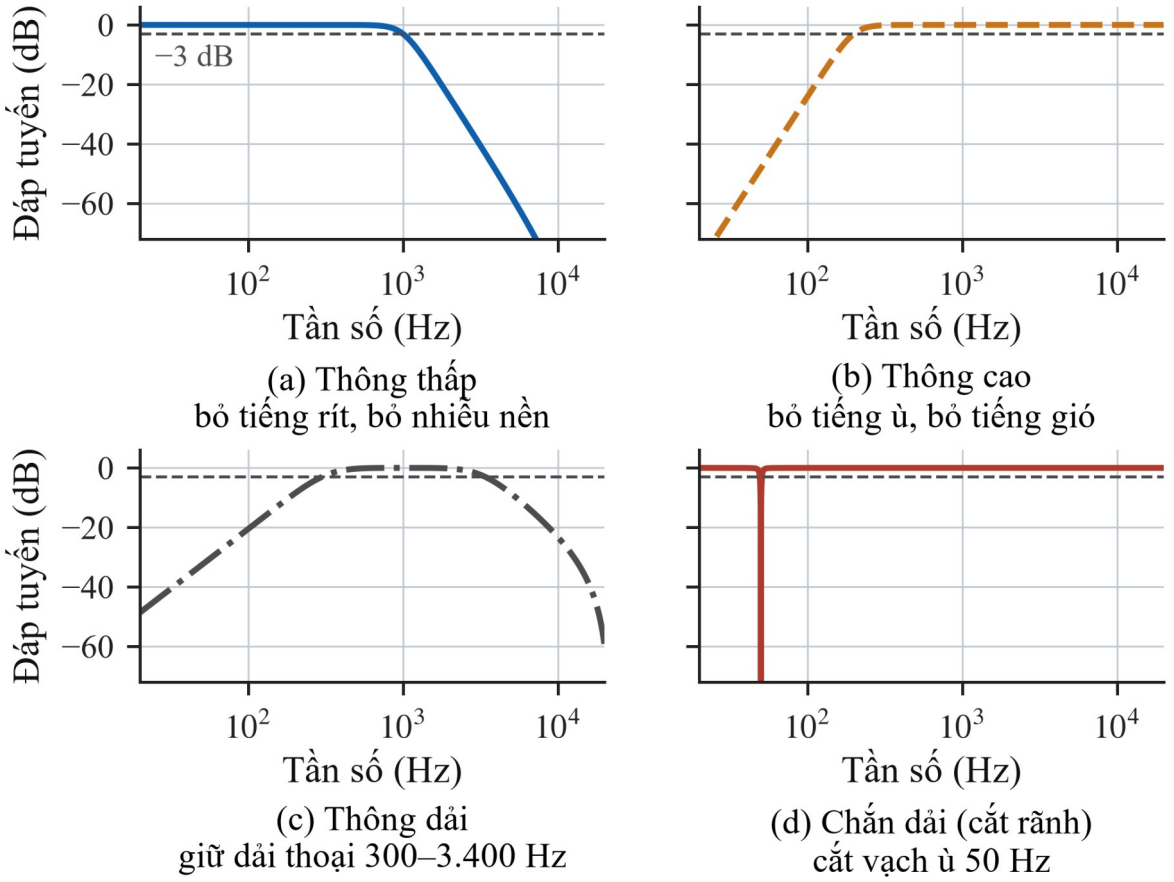
MỤC 4.6

Bộ lọc số

- Bốn kiểu lọc là bảng chữ cái của mọi phép xử lý tần số.
- Một phương trình sai phân sinh ra tất cả bộ lọc.
- FIR và IIR: mỗi họ có một cái giá riêng.

Đáp tuyến của bốn kiểu lọc số cơ bản

- Mọi bộ chỉnh âm, khử tạp, tách dải đều ghép từ bốn kiểu này.
- Đường đứt ở −3 dB là tần số cắt theo quy ước.
- Ở tần số cắt, tín hiệu còn khoảng 70 % biên độ, vẫn nghe rõ.
- Độ dốc do bậc của bộ lọc quyết định.



Hình 4.24 — Đáp tuyến tần số của bốn kiểu lọc số cơ bản, dựng từ bộ lọc thật

Bốn kiểu lọc và việc chúng làm trong phòng thu

Kiểu lọc	Giữ lại	Việc điển hình
Thông thấp	Phần dưới tần số cắt	Bỏ tiếng rít cao tần, nhiễu nền mạch, làm mềm nguồn gắt
Thông cao	Phần trên tần số cắt	Bỏ ù lưới điện, tiếng gió vào micro, rung nền sàn
Thông dải	Một dải ở giữa	Giữ dải thoại 300 – 3.400 Hz, tách một nhạc cụ
Chắn dải (cắt rãnh)	Mọi thứ trừ một dải hẹp	Cắt đúng vạch ù 50 Hz, triệt một cộng hưởng của phòng

Một công thức cho mọi bộ lọc số

$$y[n] = \sum_{i=0}^{P} b_i x[n-i] - \sum_{j=1}^{Q} a_j y[n-j]$$

(4.10)

- Đầu ra tại mẫu n: tổ hợp có trọng số của vài mẫu vào gần đây, trừ tổ hợp có trọng số của vài mẫu ra gần đây.
- Dãy b là hệ số thuận; dãy a là hệ số hồi tiếp.
- Chọn hai dãy khác nhau thì được bộ lọc khác nhau: đó là nghề thiết kế bộ lọc.

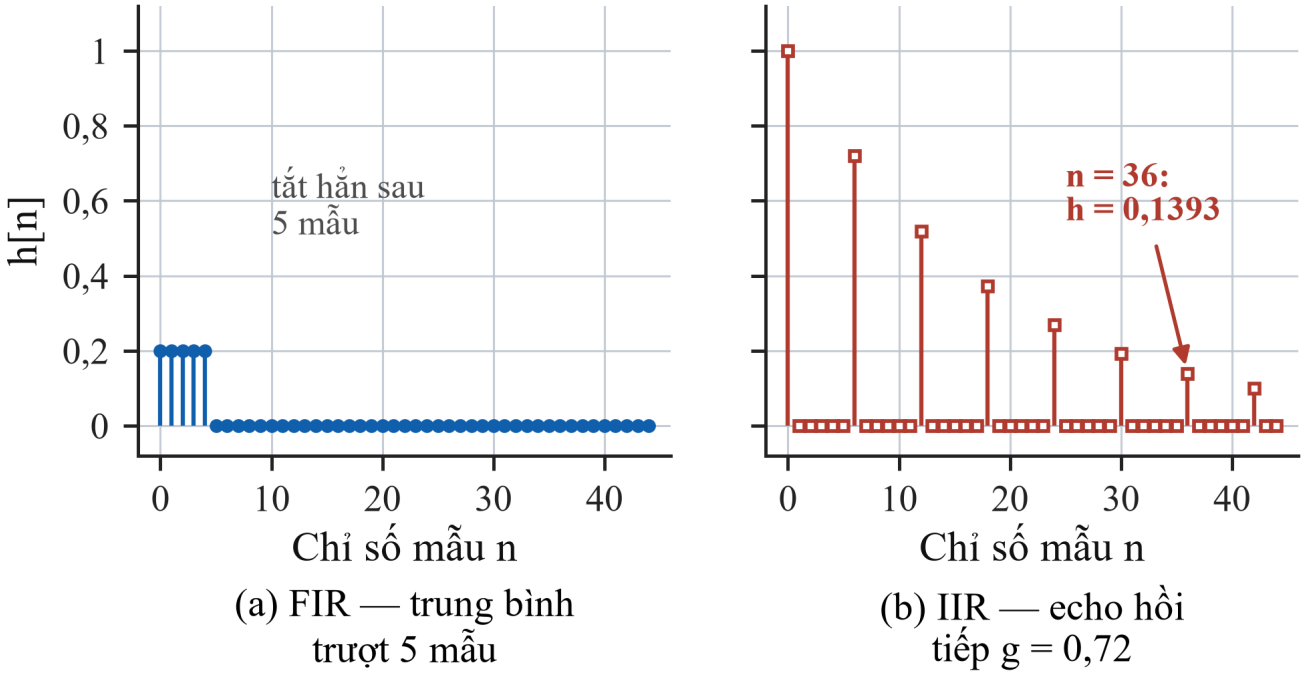
Phương trình sai phân viết thành vòng lặp

```
def loc(b, a, x):
    """Phuong trinh sai phan (4.10) viet ra tuong minh."""
    y = [0.0] * len(x)
    for n in range(len(x)):
        acc = 0.0
        for i in range(len(b)):            # phan thuan
            if n - i >= 0:
                acc += b[i] * x[n - i]
        for j in range(1, len(a)):         # phan hoi tiep
            if n - j >= 0:
                acc -= a[j] * y[n - j]
        y[n] = acc / a[0]
    return y
```

Chậm hơn hàm thư viện vài trăm lần; chỉ để thấy trong scipy.signal.lfilter chỉ có hai vòng lặp lồng nhau.

Đáp ứng xung: hữu hạn và vô hạn

- Mảng a chỉ có một phần tử bằng 1: đầu ra chỉ phụ thuộc đầu vào, đó là FIR.
- Mảng a dài hơn: đầu ra quay lại làm đầu vào, đó là IIR.
- firwin trả về một mảng (FIR); butter trả về hai mảng (IIR).
- FIR cho chuỗi hữu hạn rồi tắt; IIR cho chuỗi vô hạn nhỏ dần.



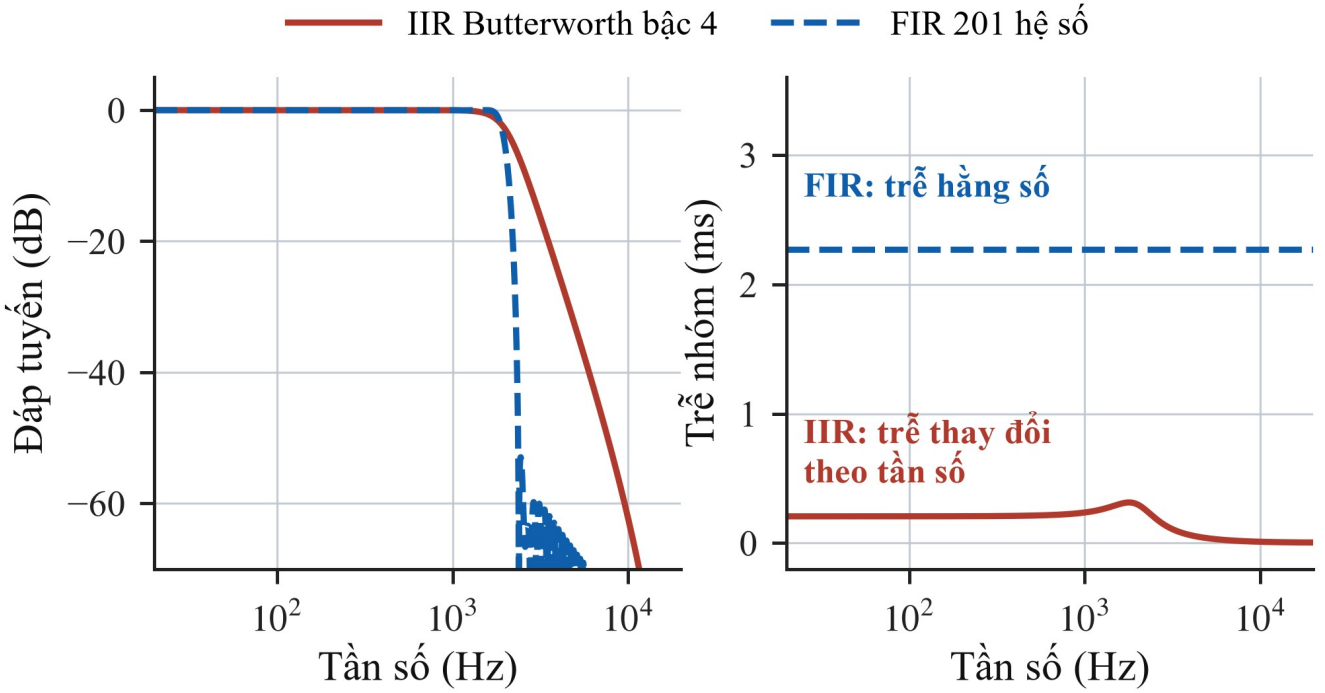
Hình 4.25 — Đáp ứng xung của một bộ lọc FIR và một bộ lọc IIR

FIR và IIR so sánh theo năm mặt

Mặt	FIR (không hồi tiếp)	IIR (có hồi tiếp)
Đáp ứng xung	Hữu hạn, tắt sau P mẫu	Vô hạn, tắt dần mãi
Pha	Tuyến tính được: mọi tần số trễ bằng nhau	Không tuyến tính: mỗi tần số trễ một kiểu
Ổn định	Luôn ổn định, không cần kiểm	Có thể phân kỳ; phải kiểm trước khi chạy
Chi phí	Cần nhiều hệ số cho đáp tuyến dốc	Ít hệ số, hiệu quả hơn nhiều
Độ trễ	Bằng nửa số hệ số, có thể đáng kể	Rất nhỏ

Cùng đáp tuyến biên độ, hai kiểu trễ khác nhau

- Ô trái: hai bộ lọc cắt gần như giống nhau.
- FIR đối xứng trễ mọi tần số bằng nhau: tín hiệu chỉ dời chỗ, không đổi hình.
- IIR trễ mỗi tần số một khác: các thành phần xô lệch, dạng sóng biến dạng.



(a) Đáp tuyến biên độ, gần như giống nhau

(b) Trễ theo tần số, khác hẳn nhau

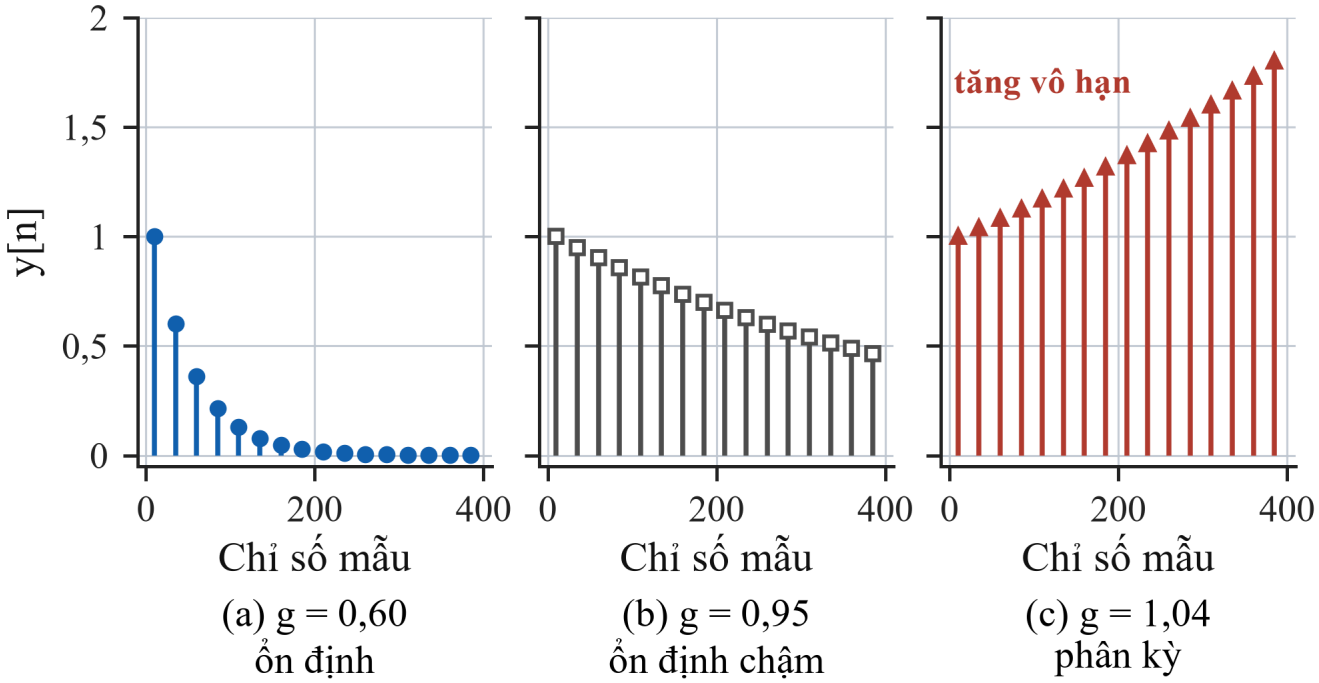
Hình 4.26 — Cùng một đáp tuyến biên độ, hai kiểu trễ theo tần số hoàn toàn khác nhau

Ba câu hỏi khi chọn họ bộ lọc

- Có sợ méo dạng sóng không? Xử lý hàng loạt ở hậu kỳ, nhất là tiếng trống và âm chuyển tiếp: chọn FIR.
- Có phải chạy thời gian thực với độ trễ rất nhỏ không? Theo dõi lúc thu, vi điều khiển: chọn IIR.
- Không chắc: dùng IIR bậc thấp, mặc định của gần hết phần mềm; méo pha thường không nghe thấy.

Bộ echo hồi tiếp: ổn định, ổn định chậm, phân kỳ

- Đầu ra quay lại làm đầu vào nên vòng lặp có thể tự khuếch đại mãi.
- Biên độ tăng không giới hạn, tín hiệu ra vô nghĩa.
- Giống tiếng hú khi micro hướng vào loa, chỉ khác là xảy ra trong mảng số.
- Lỗi hay gặp khi tự viết vang: đặt hệ số hồi tiếp quá cao.



Hình 4.27 — Cùng một bộ echo hồi tiếp với ba hệ số: ổn định, ổn định chậm, và phân kỳ

Kiểm điều kiện ổn định trước khi chạy bộ lọc

```
import numpy as np

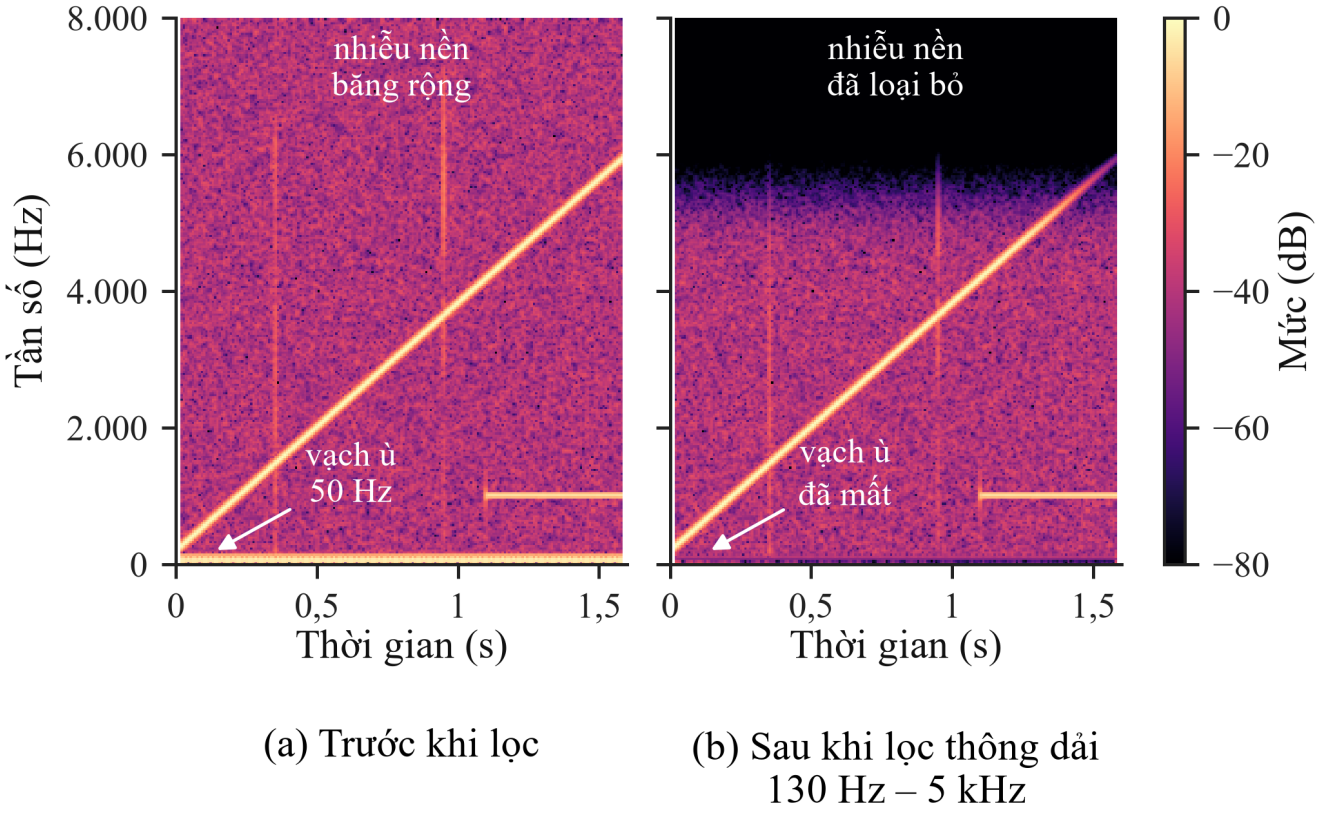
def on_dinh(a):
    """He IIR on dinh khi moi nghiem cua a nam trong duong tron"""
    if len(a) <= 1:
        return True                # FIR: luon on dinh
    nghiem = np.roots(a)
    return bool(np.all(np.abs(nghiem) < 1.0))

if not on_dinh(a):
    raise ValueError("bo loc khong on dinh — se vo")
```

Gọi một lần mỗi khi hệ số thay đổi; tốn vài micro giây.

Phổ đồ trước và sau lọc, cùng một thang màu

- Nghe chỉ nói được "sạch hơn"; phổ đồ nói được vạch ù 50 Hz đã hết nhưng dải trên 5 kHz cũng mất.
- Mỗi phổ đồ tự chuẩn hoá thì hai hình trông giống nhau dù bộ lọc cắt rất mạnh.
- Tính mốc chuẩn từ tín hiệu trước khi lọc, dùng cho cả hai hình.



Hình 4.28 — Phổ đồ trước và sau khi lọc thông dải, vẽ trên cùng một thang màu

Bốn bước soát lỗi bộ lọc

- Bước 1. Vẽ phổ đồ tín hiệu gốc, chỉ ra vấn đề và tần số ước lượng.
- Bước 2. Chọn kiểu lọc và tần số cắt từ con số đã đọc; kiểm ổn định nếu dùng IIR.
- Bước 3. Lọc, vẽ lại phổ đồ cùng thang màu, đặt cạnh hình cũ.
- Bước 4. Cái cần bỏ đã bỏ chưa; cái không được đụng có bị đụng không. Xong mới nghe để xác nhận.

Nhìn trước, chỉnh theo con số, rồi mới nghe: tai quen rất nhanh với thứ vừa nghe.

MỤC 4.7

Phép toán tích chập

- Định nghĩa đơn giản nhất của chương: ba động tác lặp lại.
- Lọc FIR chính là tích chập; vang tích chập là ứng dụng nổi tiếng.
- Định lý tích chập đưa phép chập về FFT.

Định nghĩa tích chập

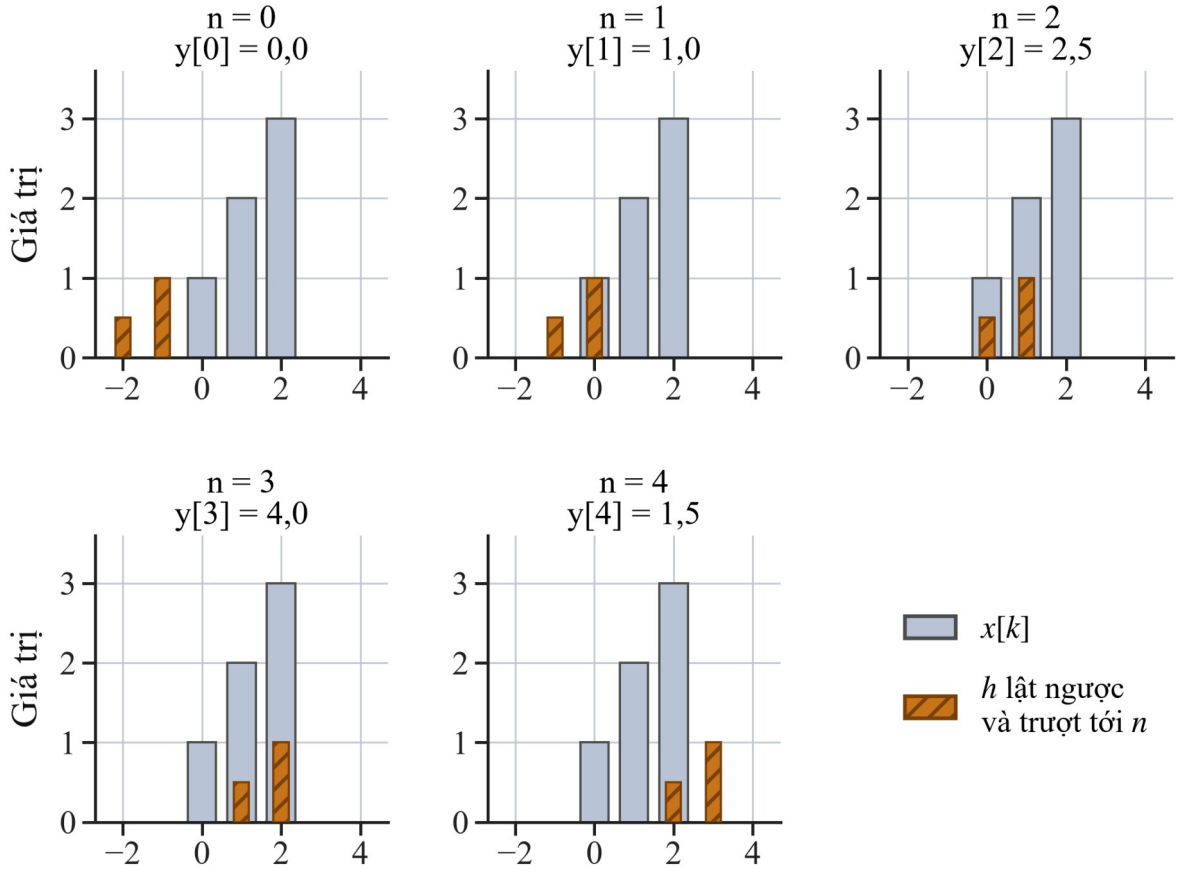
$$y[n] = (x * h)[n] = \sum_{k=0}^{L_h - 1} h[k]\, x[n-k]$$

(4.11)

- Lật ngược dãy thứ hai.
- Trượt nó tới vị trí đang xét.
- Nhân từng cặp trùng nhau rồi cộng lại.

Chập [1; 2; 3] với [0; 1; 0,5] từng bước

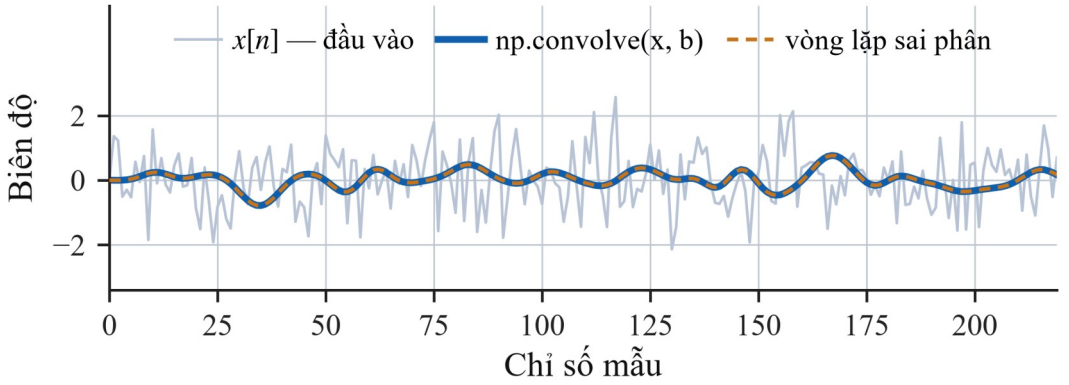
- Kết quả dài 5 phần tử, đúng bằng 3 + 3 − 1.
- Quy tắc chung: dãy dài Lx chập dãy dài Lh cho dãy dài Lx + Lh − 1.
- Nếu h là phòng vang hai giây thì tín hiệu ra dài hơn đúng hai giây: phần đuôi vang.



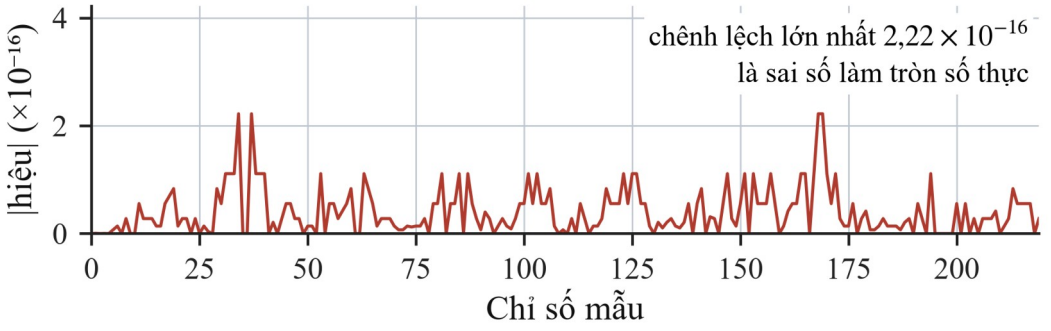
Hình 4.29 — Tích chập của [1; 2; 3] với [0; 1; 0,5] làm từng bước một

Vòng lặp sai phân và np.convolve trùng khít

- Đặt công thức (4.11) cạnh (4.10) với dãy hồi tiếp rỗng thì chúng là một.
- Hệ số b của bộ lọc FIR chính là đáp ứng xung của nó.
- Lọc bằng dãy ấy chính là chập tín hiệu với đáp ứng xung.



(a) Hai cách tính cho hai đường trùng khít



(b) Chênh lệch giữa hai cách tính

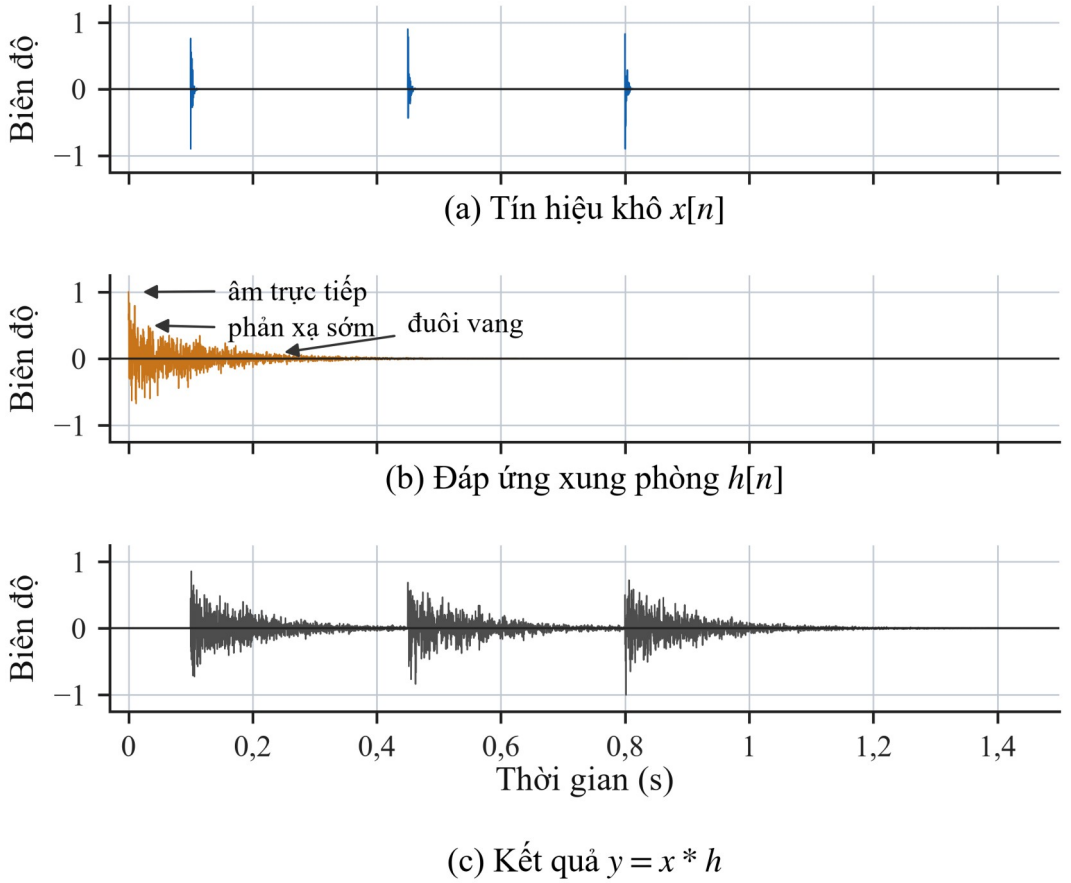
Hình 4.30 — Lọc bằng vòng lặp sai phân và chập bằng np.convolve cho kết quả trùng khít

Lọc là chập

Chập với đáp ứng xung của bộ lọc thì được hiệu ứng lọc; của một căn phòng thì được hiệu ứng vang; của một cái loa cũ thì được âm sắc loa ấy. Mọi hệ tuyến tính bất biến theo thời gian đều được mô tả đầy đủ bởi đáp ứng xung.

Tín hiệu khô chập với đáp ứng xung của phòng

- Bước một: ghi đáp ứng xung của không gian thật bằng tiếng nổ ngắn hoặc tiếng quét tần số.
- Bước hai: chập giọng thu khô với tệp ấy; kết quả nghe như thu trong chính không gian đó.
- Không phải xấp xỉ, nếu phòng là hệ tuyến tính bất biến.



Hình 4.31 — Tín hiệu khô chập với đáp ứng xung của phòng cho tín hiệu mang đặc tính phòng ấy

Chập ở miền thời gian là nhân ở miền tần số

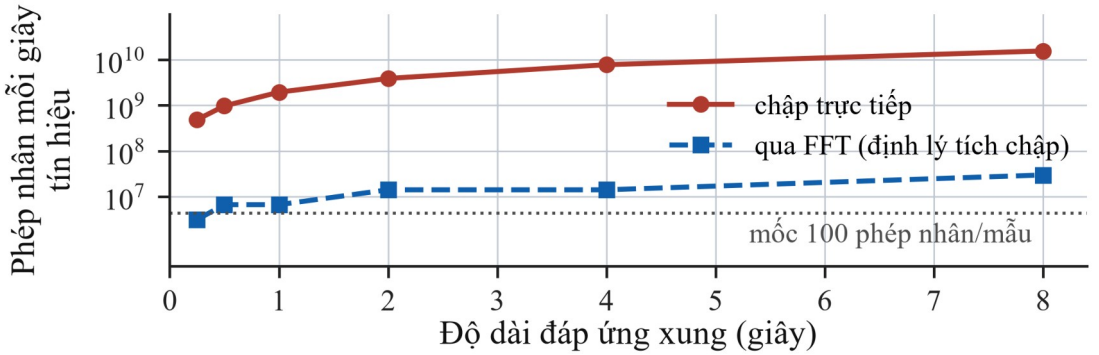
$$y = x * h \iff Y[k] = X[k] \cdot H[k]\,; \qquad L_y = L_x + L_h - 1$$

(4.12)

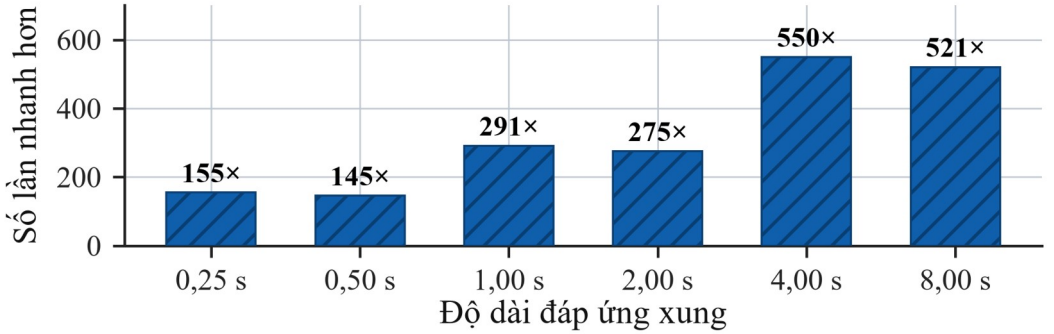
- Chập trực tiếp tốn Lx × Lh phép nhân.
- Đáp ứng xung hai giây ở 44.100 Hz, xử lý một giây tín hiệu: khoảng bốn tỉ phép nhân.
- Cách làm: FFT cả hai vế, nhân từng phần tử, FFT ngược; ba bước đều rẻ.

Chập trực tiếp và chập qua FFT

- Với đáp ứng xung hai giây, cách qua FFT nhanh hơn khoảng 275 lần.
- Đó là khoảng cách giữa chờ kết xuất và nghe ngay lúc dựng.
- Phần mềm vang tích chập đều làm thế, thường kèm kỹ thuật chia khối để giảm độ trễ.



(a) Chi phí xử lý một giây âm thanh



(b) Tỉ lệ giữa hai cách

Hình 4.33 — Chi phí của chập trực tiếp và của chập qua FFT theo độ dài đáp ứng xung

MỤC 4.8

Thực hành và bài tập

- Ba bài thực hành trong hai tiết cuối, bài sau xây trên bài trước.
- Bảy bài tập; mọi đáp số kiểm được bằng sigtools.py.
- Ba lỗi hay gặp nhất của chương.

Ba bài thực hành và tiêu chí hoàn thành

Bài	Việc phải làm	Coi là xong khi
TH1 — Phổ	Đọc WAV, vẽ dạng sóng và phổ 4.096 mẫu, có cửa sổ Hann, trục hoành đúng hertz	Chỉ ra ba tần số mạnh nhất và nói có phải hài của nhau không
TH2 — Cửa sổ	Vẽ lại phổ ấy không có cửa sổ, đặt hai hình cạnh nhau	Chỉ ra phần rò phổ và giải thích nguồn gốc
TH3 — Phổ đồ và lọc	Vẽ phổ đồ cả tệp với ba giá trị N, rồi lọc thông dải và vẽ lại	Nói mỗi N hợp với câu hỏi nào; chỉ ra chỗ bộ lọc đã cắt

Bảy bài tập của chương

- Bài 1–2: Δf, thời lượng khung, vạch thứ 100; số mẫu để tách hai nốt cách 12 Hz bằng cửa sổ Hann.
- Bài 3: sóng sin 1.000 Hz sạch mà năng lượng trải rộng; nêu ba nguyên nhân và cách kiểm.
- Bài 4: STFT với N = 4.096, H = 1.024: độ phân giải, thời lượng khung, chồng lấn, số khung.
- Bài 5–6: FIR hay IIR, đáp ứng xung; chập tay x = [2; 0; −1] với h = [1; 1; 1].
- Bài 7: quy trình xử lý bản thu ngoài trời có ù 50 Hz, gió trầm và tiếng rít 15 kHz.

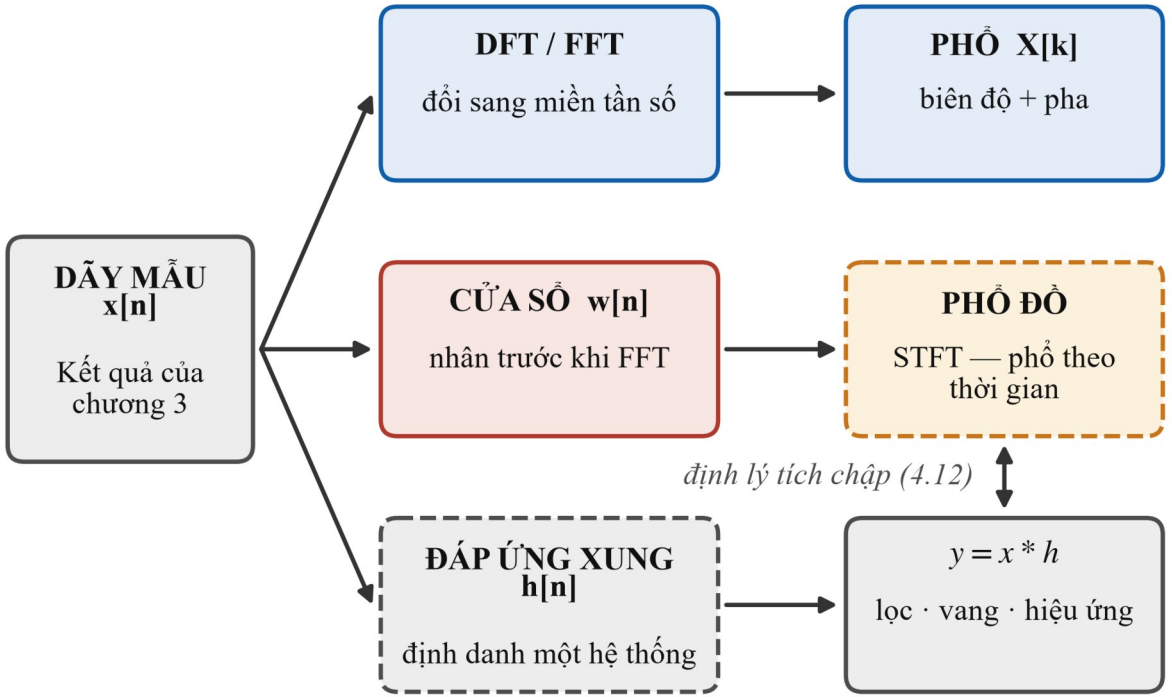
Trình bày: công thức dùng (ghi số hiệu), thay số, kết quả kèm đơn vị. Bài 3 và 7 không có đáp số duy nhất.

MỤC 4.9

Tổng kết chương học

- Bảy khái niệm của chương và cách chúng nối với nhau.
- Chọn công cụ theo câu hỏi.
- Năm điều phải nhớ; tám câu tự kiểm; chương sau.

Bảy ô và các đường nối của chương 4



Hình 4.34 — Bản đồ khái niệm của chương 4 — bảy ô và các đường nối

Câu hỏi nào thì dùng công cụ nào

Câu hỏi	Công cụ nên dùng
1. "Lúc nào thì to?"	Dạng sóng — không cần FFT.
2. "Có những tần số nào?"	Một FFT với cửa sổ Hann.
3. "Tần số ấy xuất hiện lúc nào?"	Phổ đồ STFT.
4. "Hai tần số gần nhau có tách được không?"	Tính N cần thiết theo công thức (4.7) trước khi đo.
5. "Bỏ bớt một dải tần"	Bộ lọc — FIR nếu cần tránh méo pha, IIR nếu cần tiết kiệm tính toán.
6. "In đặc tính một không gian lên tiếng thu"	Tích chập với đáp ứng xung, làm qua FFT.

Chọn sai công cụ là lỗi tốn thời gian nhất — và luôn bắt đầu bằng việc chưa xác định rõ câu hỏi.

Hình 4.35 — Bảng quyết định: câu hỏi nào thì dùng công cụ nào

Tám câu nên trả lời không cần mở tài liệu

- DFT làm gì, không dùng chữ số phức? (4.2.2) Δf với N = 2.048 ở 44.100 Hz? (4.2.3)
- Vì sao sóng sin sạch hiện như tín hiệu bẩn? (4.4.1) Hann đánh đổi gì lấy gì? (4.4.3)
- Tách hai tần số cách 10 Hz cần thu bao lâu? (4.4.5) Formant trên phổ đồ trông thế nào? (4.5.3)
- FIR và IIR khác nhau ở ba điểm nào? (4.6.3) Vì sao vang tích chập phải qua FFT? (4.7.4)

Những điều cần nhớ

1. Một tín hiệu có hai mô tả tương đương: miền thời gian trả lời "lúc nào", miền tần số trả lời "cái gì".
2. FFT cho N/2 + 1 con số; khoảng cách vạch là fs/N; muốn vạch dày hơn chỉ có cách lấy thêm mẫu.
3. Luôn nhân cửa sổ, Hann là mặc định; không nhân là lỗi không báo, kết quả sai.
4. $\Delta f \times \Delta t = 1$, không thể vượt qua: mịn về tần số thì nhoè về thời gian.
5. Bốn kiểu lọc, hai họ FIR – IIR; IIR phải kiểm ổn định trước khi chạy.
6. Lọc là chập; mọi hệ tuyến tính bất biến được mô tả đủ bởi đáp ứng xung.
7. Chập dài chạy qua FFT nhờ định lý tích chập; từ chương 5 chỉ dùng lại các công cụ này.