Spectral Analysis with Windowing and FFT: Mathematical and Physical Intuition

Introduction

The purpose of short-time Fourier analysis is to extract the frequency content of a time-varying signal such as sound. Although Fourier theory provides a clean decomposition of infinite, perfectly periodic signals, real-world audio is finite, sampled, and only available in short blocks. This creates *spectral leakage*, which motivates the use of window functions such as the Hann window. This document provides both mathematical formulas and physical intuition.

Spectral leakage

Suppose we measure a sinusoid

$$x(t) = \cos(2\pi f_0 t) \tag{1}$$

for a finite time T. Multiplying by a rectangular window

$$w_T(t) = \begin{cases} 1 & 0 \leq t < T, \\ 0 & \text{otherwise}, \end{cases} \tag{2}$$

corresponds in the frequency domain to a convolution of a perfect Dirac peak at f_0 with the Fourier transform of $w_T(t)$, which is a sinc function. Physically, this means the energy of a single sinusoid "spills" into many nearby frequencies instead of appearing as one infinitely sharp line. This phenomenon is called *spectral leakage*.

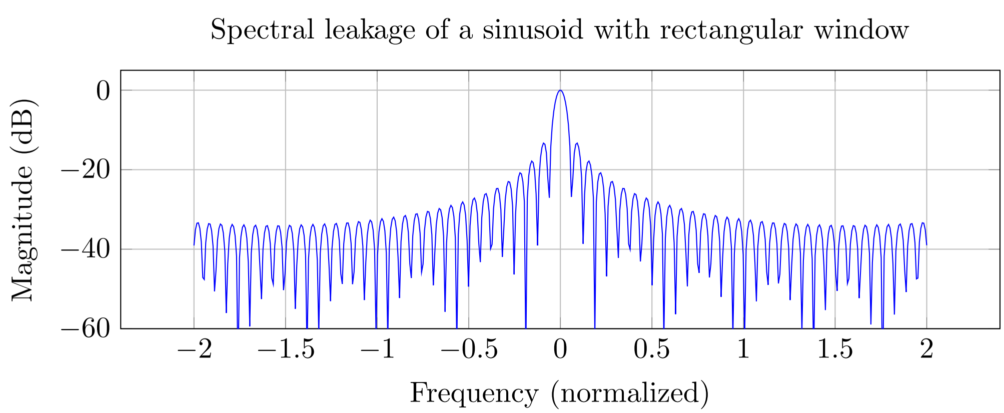

Figure 1: Sinc-shaped spectrum of a finite sinusoid. Energy leaks into distant bins.

The Hann window: definition and purpose

To reduce leakage, one applies a smooth apodization function before Fourier analysis. The Hann (Hanning) window is defined by

$$w[n] = \tfrac{1}{2}\left(1 - \cos\frac{2\pi n}{N-1}\right), \qquad 0 \le n < N, \tag{3}$$

where N is the frame length. This function smoothly ramps up from 0 to 1 and back to 0, avoiding the discontinuities of the rectangular cut-off. Its Fourier transform has significantly lower sidelobe levels (about $-31\,\mathrm{dB}$), so leakage into distant frequencies is suppressed.

Physical intuition. Applying a Hann window is analogous to tapering the edges of a finite measurement to avoid introducing sharp boundaries, much like how diffraction is reduced when an aperture is smoothly apodized instead of sharply truncated.

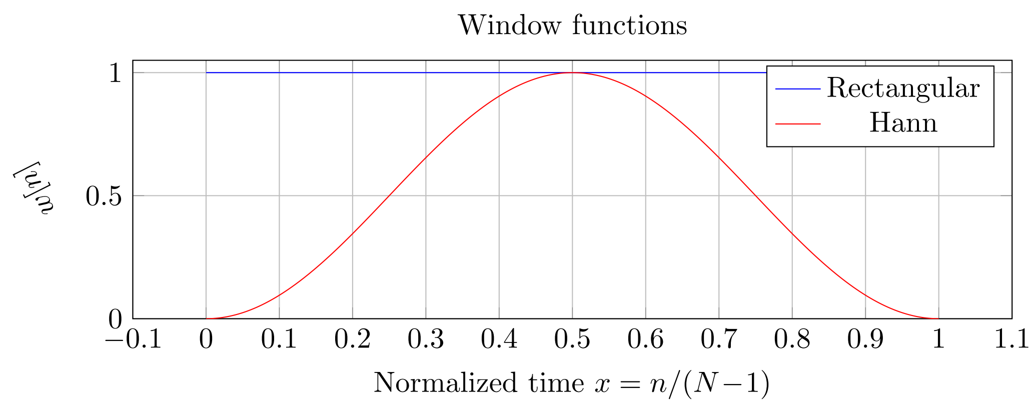

Figure 2: Rectangular vs. Hann windows. The Hann window smoothly tapers edges (raised cosine).

Short-Time Fourier Transform

Given a discrete-time signal $x[n]$, we analyze overlapping frames:

$$\tilde{x}_m[n] = w[n]\,x[n + mH], \quad 0 \le n < N, \tag{4}$$

where H is the hop size (step between frames). Each frame is Fourier transformed:

$$X_m[k] = \sum_{n=0}^{N-1} \tilde{x}_m[n]\,e^{-j2\pi kn/N}. \tag{5}$$

Here k indexes frequency bins with physical frequency $f_k = kf_s/N$, where f_s is the sampling rate.

Physics interpretation. The STFT provides the instantaneous frequency spectrum of the signal, analogous to decomposing a time-dependent wave into its Fourier modes, but restricted to a moving finite observation window. It balances time resolution (H, f_s) against frequency resolution (N).

Time–frequency trade-off

Because the frame length N fixes both duration $T = N/f_s$ and frequency resolution $\Delta f = f_s/N$, there is a trade-off: longer windows give sharper frequency peaks but blur fast temporal changes, and shorter windows give precise timing but poorer frequency discrimination.

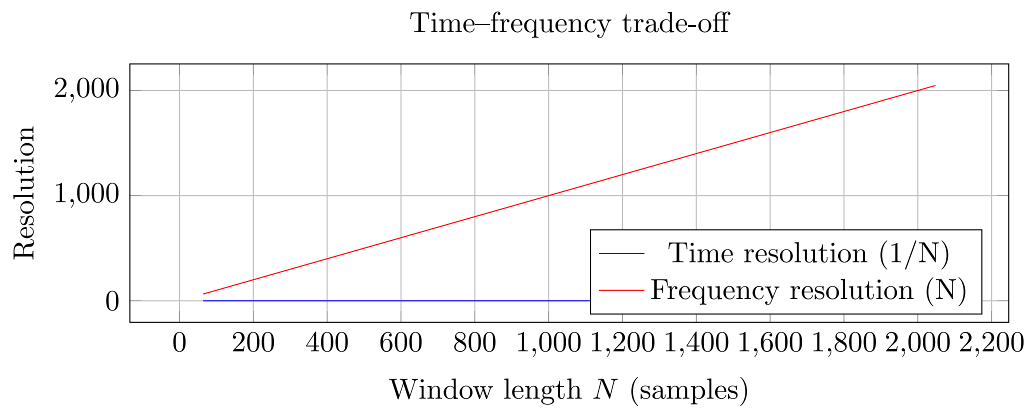

Figure 3: Increasing window length improves frequency resolution but worsens time resolution.

Quadratic interpolation of peaks

Because FFT bins are discrete, the true sinusoidal frequency f_0 may lie between bins. To estimate this, we fit a parabola through the logarithmic magnitudes at $k-1, k, k+1$ and extract the vertex offset Δ:

$$\Delta = \frac{1}{2}\frac{M[k-1] - M[k+1]}{M[k-1] - 2M[k] + M[k+1]}, \qquad \hat{f} = \frac{k+\Delta}{N} f_s. \tag{6}$$

This procedure recovers a sub-bin estimate of frequency. Physically, it amounts to estimating the location of the peak of a sinc lobe more accurately than the FFT grid allows.

Harmonic structure and timbre

For quasi-periodic signals (e.g. musical tones, voiced speech), energy appears at integer multiples of a fundamental frequency f_0. One may extract a *timbre vector*

$$\boldsymbol{t} = \big(M[k_1], M[k_2], \dots\big), \tag{7}$$

where $k_h \approx h f_0 N/f_s$. This vector characterizes the relative strength of harmonics and thus the perceptual color of the sound.

Summary

- Spectral leakage arises from the finite-time observation of signals, spreading energy into sidelobes.
- Window functions (Hann) smooth the edges, lowering sidelobe levels while widening the main lobe, trading frequency resolution for cleaner separation of tones.

- The STFT provides a time–frequency decomposition, balancing temporal and spectral resolution.
- Quadratic interpolation refines frequency estimates beyond FFT bin width.
- Harmonic vectors summarize timbre, linking physics of periodic excitation and system response to perceptual sound quality.