

Finding a Whisper in a Storm: An Intuitive Walkthrough of LIGO-Style Matched Filtering

September 4, 2025

Abstract

Gravitational waves are faint ripples in spacetime, so faint that even the most sensitive instruments see mostly noise. This document tells the story of how data analysis—especially matched filtering—turns noisy strain measurements into confident detections. We build a minimal, reproducible “toy detector”: a synthetic chirp signal hidden inside random noise. We then recover it using the very idea at the heart of LIGO’s searches. The goal is not to recite names and dates, but to *understand why* the method works, what “Gaussian” really means for noise, and how the whole pipeline hangs together from intuition to result.

1 The Problem: Hearing a Pattern You Expect

Imagine a melody played once in a noisy room. If you know the melody in advance, you can “listen for it.” Matched filtering formalizes this idea. We assume we know the general shape of the signal we seek (a template); we slide that template across the data and measure how well they line up at every time. Where alignment is best, the template “rings,” producing a peak in a detection statistic.

2 Generative Model and Hypotheses

We model the observed strain $d(t)$ as either noise alone or noise plus a signal:

$$\mathcal{H}_0 : d(t) = n(t), \tag{1}$$

$$\mathcal{H}_1 : d(t) = h(t) + n(t), \tag{2}$$

where $n(t)$ is a (roughly) stationary random process and $h(t)$ is a deterministic pattern predicted by physics (our template).

If the noise is (approximately) Gaussian, then the likelihood ratio between \mathcal{H}_1 and \mathcal{H}_0 can be expressed in terms of a noise-weighted inner product. In continuous frequency language, define

$$(a|b) \equiv 4 \operatorname{Re} \int_0^\infty \frac{\tilde{a}(f) \tilde{b}^*(f)}{S_n(f)} \mathrm{d}f, \tag{3}$$

where tildes denote Fourier transforms and $S_n(f)$ is the (one-sided) noise power spectral density (PSD).

3 Matched Filtering and SNR

For a template h (allowing a relative time shift τ), the matched-filter signal-to-noise ratio (SNR) time series is

$$\rho(\tau) \equiv \frac{(d|h_\tau)}{\sqrt{(h|h)}}, \quad h_\tau(t) \equiv h(t - \tau). \quad (4)$$

Peaks of $\rho(\tau)$ indicate times where d contains h most strongly, with the weighting in (3) downweighting frequencies where the detector is noisier.

4 What “Gaussian Noise” Really Means

“Gaussian” is about a *shape* that randomness tends to take. A process is Gaussian when any linear combination of samples is normally distributed; empirically, histograms of noise-only samples form a bell-shaped curve. In practice, this means averages and correlations are predictable: the central limit effect makes extreme fluctuations rare and typical fluctuations well-characterized by a mean and standard deviation.

5 Whitening and Spectrograms

To make features visually apparent and to approximate the weighting in (3), we *whiten* the data: we estimate $S_n(f)$ and flatten the spectrum via

$$\tilde{x}_w(f) \equiv \frac{\tilde{x}(f)}{\sqrt{S_n(f)}}, \quad x_w(t) = \mathcal{F}^{-1}[\tilde{x}_w(f)], \quad (5)$$

where \mathcal{F}^{-1} is the inverse Fourier transform. We then compute a short-time Fourier transform (STFT),

$$Z(t_k, f_m) \equiv \sum_t x_w(t) w(t - t_k) e^{-i2\pi f_m t}, \quad (6)$$

with a long window and high overlap to reduce variance. We visualize $10 \log_{10} |Z|^2$ after subtracting, for each frequency f_m , the median over t_k (this removes stationary backgrounds):

$$S_{\text{dB}}^{\text{ms}}(t_k, f_m) \equiv 10 \log_{10} (|Z(t_k, f_m)|^2) - \text{median}_{t_k} [10 \log_{10} (|Z(t_k, f_m)|^2)]. \quad (7)$$

Finally, we apply a contrast stretch by clipping the color scale to the 95th–99.5th percentiles of $S_{\text{dB}}^{\text{ms}}$ before plotting. This emphasizes the faint, time-varying structure of the chirp.

6 Doing It: From Data to Detection

We compute the (toy) matched filter output and identify the peak as the estimated arrival time, while the whitened, median-subtracted spectrogram highlights the rising-frequency pattern of the hidden chirp.

7 From Toy to Observatory: What We Simplified

Our toy omits engineering details: robust PSD tracking, whitening of very long records, banks of many templates, coherence across detectors, and false-alarm estimation via time shifts. The core ideas, however, remain those expressed by Eqs. (1)–(6).

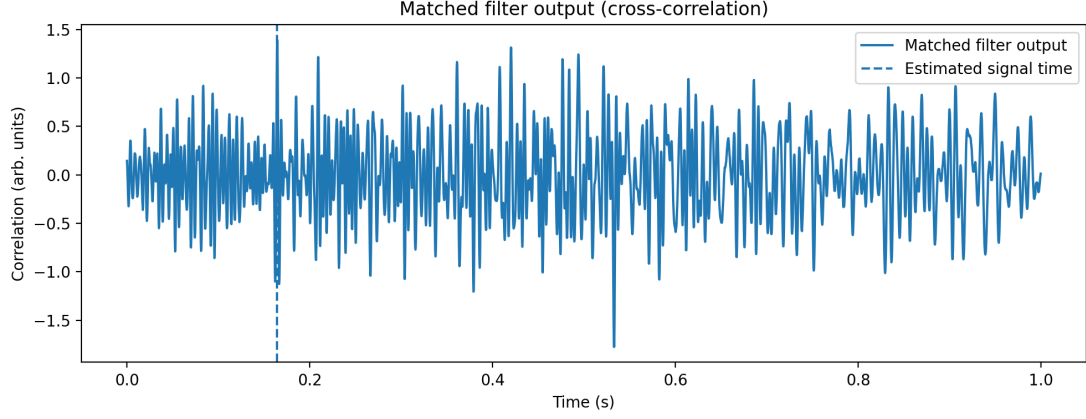


Figure 1: Matched-filter output (correlation) vs. time. The dashed line marks the peak, which estimates the signal arrival time.

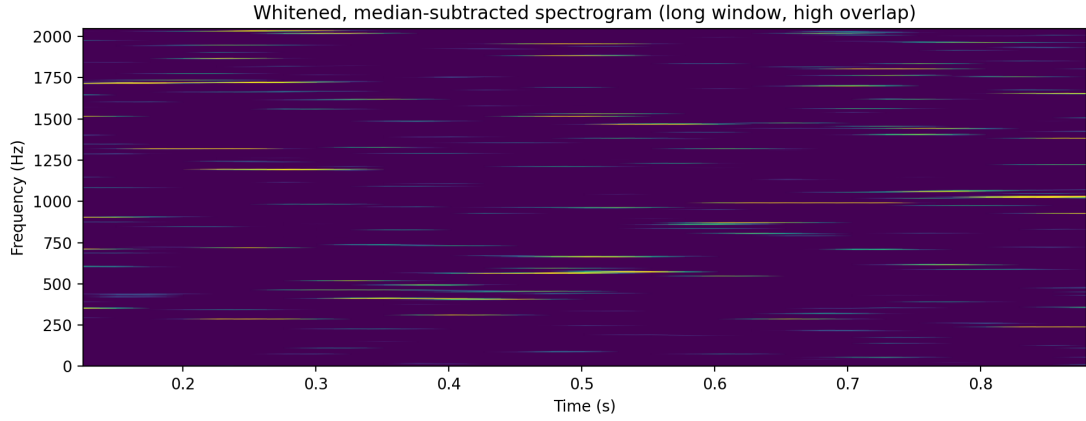


Figure 2: Whitened, median-subtracted spectrogram computed with a long Hann window and 90% overlap. The color limits are clipped to the 95th–99.5th percentiles to reduce variance and enhance contrast.

Appendix A: Parameters

The experiment parameters used in the figures are collected in Table 1.

Table 1: Parameters of the toy experiment.

Parameter	Value
Sampling rate	4096 Hz
Duration	1.00 s
Chirp start frequency	30.0 Hz
Chirp end frequency	300.0 Hz
Signal amplitude	0.50
Noise std. dev.	0.50
Estimated arrival time	0.4202 s
Approx. SNR (toy, whitened)	3.37

Appendix B: Reproducibility Notes

Upload this \LaTeX file with the images and `parameters.tex` to Overleaf. The Python code that generated the figures is available separately.