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

September 5, 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 hidden inside random noise. We then recover it using the core idea of LIGO's searches, with clear intuition about "Gaussian" noise and why correlation is the right question to ask of the data.

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.

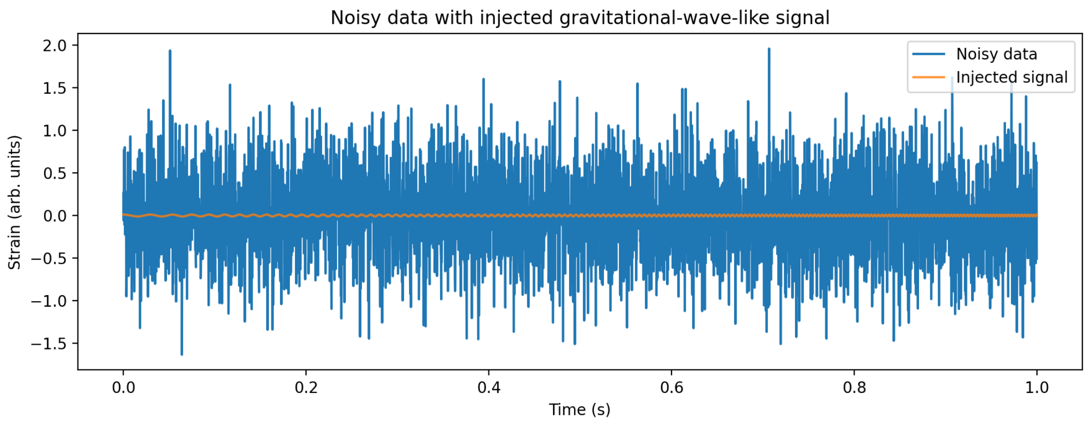

Figure 1: Signal-only, noise-only, and signal+noise, over the same time span. Seeing all three clarifies the challenge: the signal is faint and structured, the noise is random, and real data are their sum.

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 template.

If the noise is (approximately) Gaussian, the likelihood ratio can be expressed with the noise-weighted inner product:

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

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

3 Matched Filtering and SNR

For a template h with relative time shift τ, the SNR time series is

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

Peaks of $\rho(\tau)$ indicate times where d contains h most strongly, with (3) downweighting noisy bands.

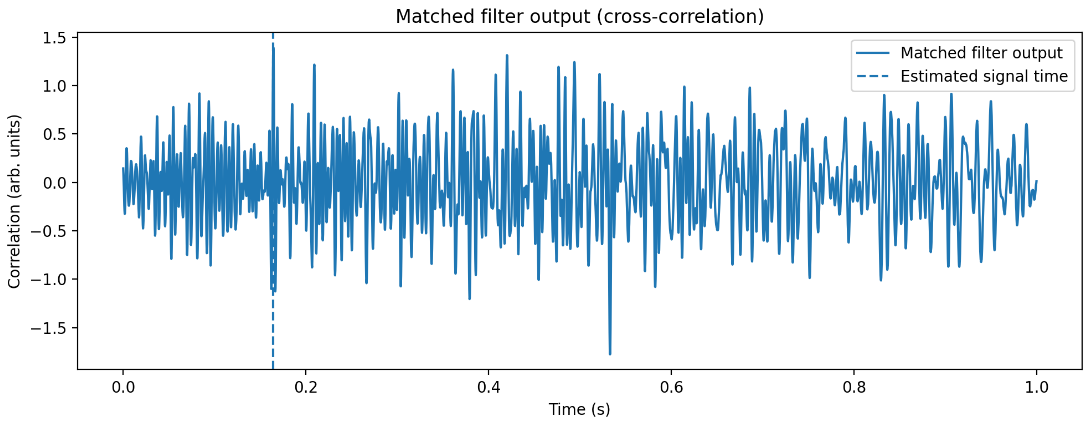

Figure 2: Matched-filter output (toy correlation) vs. time. The dashed line marks the peak (estimated arrival time).

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 visualize features and approximate the weighting in (3), we whiten the data:

$$\tilde{x}_{\mathrm{w}}(f) \equiv \frac{\tilde{x}(f)}{\sqrt{S_n(f)}}, \qquad x_{\mathrm{w}}(t) = \mathcal{F}^{-1}\big[\tilde{x}_{\mathrm{w}}(f)\big]. \tag{5}$$

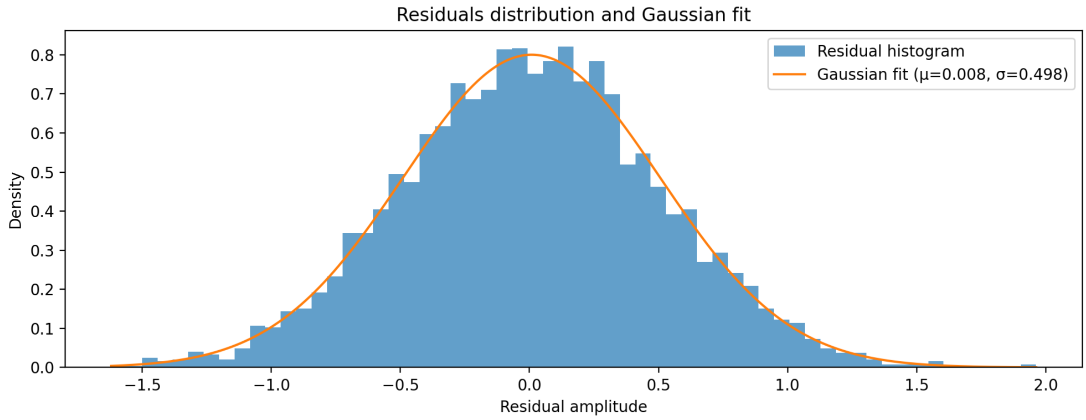

Figure 3: Residual histogram with Gaussian fit. This is the operational meaning of "Gaussian": a bell-shaped distribution with predictable tails.

We then compute an STFT with a long window and high overlap:

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

and plot $10 \log_{10} |Z|^2$ after subtracting, for each f_m, the median over t_k:

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

Finally, we clip color limits to the 95th–99.5th percentiles of $S_{\mathrm{dB}}^{\mathrm{ms}}$ to enhance contrast.

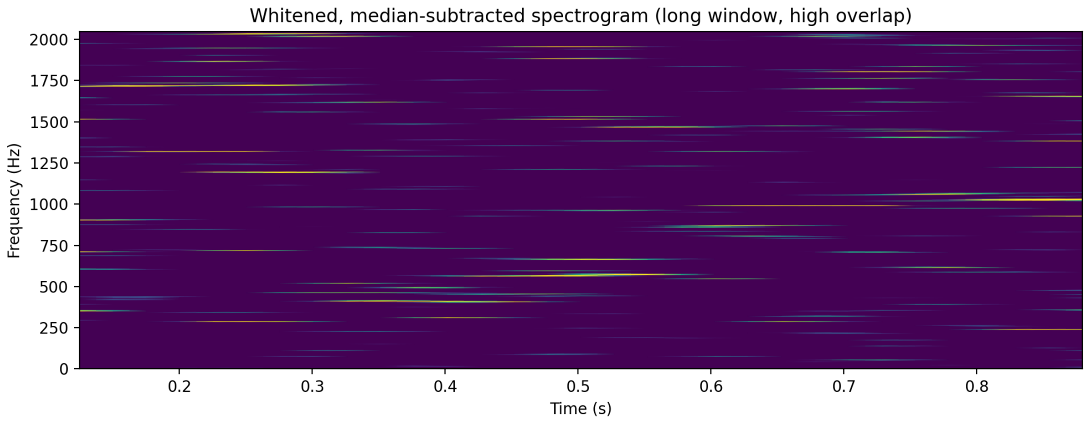

Figure 4: Whitened, median-subtracted spectrogram using a long Hann window with 90% overlap; color limits clipped to the 95th–99.5th percentiles.

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).

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	0.1643 s
Approx. SNR (toy)	3.38
Residual ,	0.008, 0.498

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.