

Data mining and machine learning improve gravitational-wave detector sensitivity

Gabriele Vajente¹*

LIGO Laboratory, California Institute of Technology, Pasadena, California 91101, USA



(Received 8 March 2022; accepted 10 May 2022; published 20 May 2022)

Application of data mining and machine learning techniques can significantly improve the sensitivity of current interferometric gravitational-wave detectors. Such instruments are complex multi-input single-output systems, with close-to-linear dynamics and hundreds of active feedback control loops. We show how the application of brute-force data-mining techniques allows us to discover correlations between auxiliary monitoring channels and the main gravitational-wave output channel. We also discuss the result of the application of a parametric and time-domain noise subtraction algorithm, that allows a significant improvement of the detector sensitivity at frequencies below 30 Hz.

DOI: [10.1103/PhysRevD.105.102005](https://doi.org/10.1103/PhysRevD.105.102005)

I. INTRODUCTION

Gravitational waves [1] are generated by extreme energy astrophysical and cosmological sources [2], and can be detected on earth as very small variations in the distance between test masses, at a level of 10^{-21} relative change or smaller. To detect such weak signals, laser interferometric detectors are employed [3–5], where the variations in the length of the two km-scale arms of a Michelson interferometer are measured as fluctuations in laser power. The detectors are operated as null instruments: feedback control loops [6] are employed to maintain the system near the optimal operating point, and to linearize its otherwise highly nonlinear response to gravitational-wave signals and external disturbances. The main detector output, the gravitational-wave strain channel, is reconstructed from the error and control points of the feedback loops [7]. The sensitivity of such detectors is limited by a combination of noise sources [8], some of them fundamentally related to the instrument physics and design (quantum [9] and thermal [10] noises for example), while others are due to imperfections in the system or spurious couplings of environmental disturbances [11].

A gravitational wave-detector can be described as a multi-input single-output system. The calibrated output channel contains the response to gravitational-wave events, that are typically short transient signals with durations spanning from a fraction of a seconds to minutes, and with frequency content between 10 Hz and a few kHz. The output channel is, however, polluted by a large number of noise sources, some of them unknown. To characterize and remove those noise sources, several thousands of auxiliary channels are recorded. They monitor potential sources of

spurious noise that couple into the main channel. Those can be seen as additional inputs to the system. The couplings are mostly linear, thanks to the effect of the feedback controls. However, the extreme sensitivity needed to detect gravitational-wave signals implies that even very small deviations from linearity are relevant and need to be taken into account.

The sensitivity of the current gravitational-wave detectors is continuously improved by upgrading key components and control systems. It is also important to reduce, as much as possible, the background noise in the recorded data by applying offline signal processing techniques. In the case of known linear couplings from some of the input channels to the output, one can implement noise regression and subtraction techniques, such as Wiener filters [12,13]. Recent work showed how to extend offline noise subtraction to the case of nonlinear and nonstationary noise couplings [14].

The first step in any noise subtraction scheme is the identification of candidate channels. With more than 4000 channels recorded with sampling rates varying from 256 to 16384 Hz, the amount of data available is quite large. To select the most favorable channels for the subtraction, one can rely on experience and intuition, or implement data-mining techniques that can sort through all the available channels to provide only the most promising candidate for subtraction. One of the new results presented in this paper is a method we developed to identify the auxiliary channels that could provide the best candidates for linear noise subtraction, described in Sec. II. This method is based on the brute-force computation of the frequency-domain coherence between the gravitational-wave output channel and all available auxiliary channels. Section III explains how this method can be extended to address nonstationary

*vajente@caltech.edu

noise couplings, and also discuss the difficulties involved in extending the approach to fully nonlinear couplings.

In this work we focus on techniques to identify and remove persistent noises, that are always present, although with varying properties. Another kind of disturbance that is present in gravitational wave detector output are glitches: transient signals that have short duration, typically under a few seconds [15] and that can mimic the waveform of actual astrophysical signals. Several techniques have been developed to deal with glitches and find correlations with auxiliary channels [16–18].

In Sec. IV we review the noise subtraction algorithm described in [14]. Finally, Sec. V describes the application of those methods to the data collected by the Advanced LIGO gravitational-wave detector during the O3 science run [2], showing how that same algorithm can be applied to improve the detector sensitivity to gravitational waves at frequencies below 30 Hz.

II. LINEAR NOISE COUPLINGS

In the general case of a multi-input single-output system, like a gravitational-wave detector, we have a large number of input channels $x_i(t)$ and one output channel $h(t)$. For simplicity we use a continuous time notation for all channels, although in reality they are discrete in time, possibly with different sampling rates. As a first step we want to find which ones of the input channels are contributing linearly to the output channel. One way to estimate the degree of linear correlation between two channels is to compute the frequency domain coherence:

$$C[x_i, h](f) = \frac{|S[x_i, h](f)|^2}{S[x_i, x_i](f)S[h, h](f)} \quad (1)$$

where $S[\cdot, \cdot](f)$ denotes the cross-spectral-density of two channels, that can be computed for example using a modified Welch periodogram [19]. If the output channel is simply related to the input channel via a linear filter, then the coherence is exactly one. On the other hand, if the input and output signals are uncorrelated, the coherence should be exactly zero. However, since we can only estimate the coherence from a finite amount of time, by averaging discrete Fourier transforms over multiple segments, the actual coherence will never be exactly zero. The probability distribution of the coherence as estimated by averaging N segments can be related to the F -distribution [20]. The confidence interval for the estimated coherence, at a significance level α , can be approximated by

$$C_\alpha = \frac{F_{2,2N-2}(\alpha)}{N-1 + F_{2,2N-2}(\alpha)} \quad (2)$$

where $F_{2,2N-2}$ is the inverse of the cumulative distribution function for the F -distribution. In other words, if we estimate the coherence of two uncorrelated, normally distributed

random signals, we expect values smaller than C_α with a confidence α . This provides us with a way to systematically assess if a given channel, at a frequency of interest, can be contributing to the noise in the main gravitational-wave strain channel. When considering a large number of channels, one should include a trial number correction for the coherence threshold. If the confidence level α is set to 95%, one would expect one false positive every about 20 coherence measurements. Since for every channel the coherence is computed for thousands of frequency bins, the actual number of trials can quickly become very large, forcing the coherence threshold to high values (about 0.1) that would exclude channels that are important. We find empirically that setting the confidence level at 95%–99.9% gives meaningful results, especially when the coherence is required to exceed this value in a large frequency band.

Given the large number of auxiliary channels available to analyze, a systematic approach is needed. Starting with the entire list of available channels, we remove those that are not suitable: channels sampled with too low rate; channels that are known to be related to the gravitational-wave channel [15], for example those related to the differential-arm control loop [21] and to subproducts of the calibration process [7]; and channels that are constant for the entire analysis period. We are left with several thousand channels, with sampling rates ranging from a few hundred Hz to 16 kHz. We then select a suitable duration of time for the analysis. The lower limit is given by a combination of the desired frequency resolution of the coherence computation and the minimum number of averages needed to have a good estimate. We are typically interested in resolving narrow spectral features, so a resolution of 0.1 Hz or better is often desirable. This sets the length of the segments used in the single fast Fourier transforms (FFT) computations to at least 10 s. To have a low value of the coherence confidence level, given by Eq. (2), we typically require 200 averages or more. With $N = 200$ and $\alpha = 0.95$, we have $C_\alpha = 0.015$, while for $\alpha = 0.999$ we have $C_\alpha = 0.035$. Using, for example, 1200 s of data with segments that are 12-s-long and overlap by 50% allows us to have a frequency resolution of 0.08 Hz.

The coherence of all auxiliary channels with the main channel is then computed using Eq. (1), implemented in PYTHON and using the SCIPY library. The computation can be largely parallelized on multicore machines. Typically the full analysis of about 4000 channels and 1200 s of data takes about one hour on a single server-class computer, mostly limited by the time needed to access the data from disk. Speed-ups are possible by parallelizing the computations on multiple machines and utilizing local caches for the data files. Those improvements will be necessary if the number of channels increases significantly or when considering bilinear couplings as discussed below in Sec. III.

The next step is to find a way to represent the results of all the coherence computations in an understandable way

Top 20 coherences of DCS-CALIB_STRAIN with auxiliary channels

GPS 1242441180 + 1200 s [2019/05/21 02:32:42 UTC]

Frequency [Hz]	Top channels									
9.12	OAF-CAL_XARM DQ (0.40)	CAL-CS_XARM DQ (0.40)	CAL-CS_XARM ERR_DQ (0.40)	LSC-XARM_IN1 DQ (0.40)	LSC-POP_A_RF45 I_ERR DQ (0.40)	LSC-POP_A_RF45 I_ERR 256_DQ (0.40)	LSC-POPAIR A_RF45_I_ERR 236_DQ (0.38)	LSC-POPAIR A_RF45_I_ERR DQ (0.38)	LSC-SRCLF1 OUT_DQ (0.36)	CAL-CS_SRCL ERR_DQ (0.36)
9.25	LSC-XARM_IN1 DQ (0.29)	LSC-POP_A_RF45 I_ERR DQ (0.29)	CAL-CS_XARM DQ (0.29)	CAL-CS_XARM ERR_DQ (0.29)	OAF-CAL_XARM DQ (0.29)	LSC-POP_A_RF45 I_ERR 256_DQ (0.29)	LSC-POPAIR A_RF45_I_ERR DQ (0.29)	LSC-POPAIR A_RF45_I_ERR 256_DQ (0.29)	CAL-CS_SRCL ERR_DQ (0.27)	OAF-CAL_XARM DQ (0.27)
9.38	ASC-AS_C_NSUM OUT_DQ (0.41)	ASC-AS_B_RF45 Q_PIT_OUT DQ (0.39)	ASC-AS_A_RF45 I_SUM_OUT DQ (0.30)	ASC-OMC_A_PIT OUT_DQ (0.30)	OMC-ASC_QPD A_PIT_OUT DQ (0.30)	SUS-ITMY_L3 JSCINF_P INI_DQ (0.19)	ASC-CSOFT_P OUT_DQ (0.17)	ASC-CSOFT_P SM_DQ (0.17)	LSC-POPAIR A_RF45_I_ERR 256_DQ (0.16)	LSC-POP_A_RF45 I_ERR DQ (0.16)
9.50	ASC-AS_B_RF45 Q_PIT_OUT DQ (0.73)	ASC-AS_C_NSUM OUT_DQ (0.70)	ASC-AS_A_RF45 I_SUM_OUT DQ (0.58)	PEM-EX_MAG EBAY_SUSRACK Y_DQ (0.40)	OMC-ASC_QPD A_PIT_OUT DQ (0.39)	ASC-OMC_A_PIT OUT_DQ (0.39)	ASC-X_TR_A PIT_OUT_DQ (0.31)	ASC-X_TR_B PIT_OUT_DQ (0.29)	SUS-SR2_M3 JSCINF_Y_IN1 DQ (0.25)	ASC-AS_C_NSUM OUT_DQ (0.25)
9.62	ASC-AS_C_NSUM OUT_DQ (0.67)	ASC-AS_B_RF45 Q_PIT_OUT DQ (0.57)	ASC-AS_A_RF45 I_SUM_OUT DQ (0.51)	OMC-ASC_QPD A_PIT_OUT DQ (0.37)	ASC-OMC_A_PIT OUT_DQ (0.37)	ASC-X_TR_A PIT_OUT_DQ (0.29)	PEM-EX_MAG EBAY_SUSRACK Y_DQ (0.26)	ASC-CHARD_P SM_DQ (0.18)	ASC-CHARD_P OUT_DQ (0.18)	ASC-AS_C_NSUM OUT_DQ (0.18)
9.75	ASC-AS_C_NSUM OUT_DQ (0.72)	ASC-OMC_A_PIT OUT_DQ (0.55)	OMC-ASC_QPD A_PIT_OUT DQ (0.55)	ASC-AS_A_RF45 I_SUM_OUT DQ (0.55)	ASC-AS_B_RF45 I_PIT_OUT DQ (0.31)	ASC-AS_B_RF45 Q_PIT_OUT DQ (0.23)	ASC-AS_A_DC NSUM_OUT DQ (0.17)	ASC-OMC_B_YAW OUT_DQ (0.16)	OMC-ASC_QPD B_YAW_OUT DQ (0.16)	ASC-AS_C_NSUM OUT_DQ (0.16)
9.88	ASC-AS_C_NSUM OUT_DQ (0.69)	OMC-ASC_QPD A_PIT_OUT DQ (0.55)	ASC-OMC_A_PIT OUT_DQ (0.55)	ASC-AS_A_RF45 I_SUM_OUT DQ (0.50)	ASC-AS_B_RF45 I_PIT_OUT DQ (0.34)	OMC-ASC_QPD B_PIT_OUT DQ (0.21)	ASC-OMC_B_PIT OUT_DQ (0.21)	OMC-ASC_QPD B_YAW_OUT DQ (0.17)	ASC-OMC_B_YAW OUT_DQ (0.17)	ASC-AS_C_NSUM OUT_DQ (0.17)
10.00	ASC-OMC_A_PIT OUT_DQ (0.42)	OMC-ASC_QPD A_PIT_OUT DQ (0.42)	ASC-AS_C_NSUM OUT_DQ (0.39)	ASC-Y_TR_A PIT_OUT_DQ (0.30)	ASC-CHARD_P SM_DQ (0.28)	ASC-CHARD_P OUT_DQ (0.28)	ASC-CHARD_P INI_DQ (0.27)	ASC-OMC_B_PIT OUT_DQ (0.23)	OMC-ASC_QPD B_PIT_OUT DQ (0.23)	ASC-AS_C_NSUM OUT_DQ (0.23)
10.12	ASC-AS_C_NSUM OUT_DQ (0.22)	ASC-AS_B_RF45 Q_PIT_OUT DQ (0.20)	OMC-ASC_QPD A_PIT_OUT DQ (0.17)	ASC-OMC_A_PIT OUT_DQ (0.17)	ASC-Y_TR_A PIT_OUT_DQ (0.16)	ASC-CSOFT_P SM_DQ (0.15)	ASC-CSOFT_P SM_DQ (0.15)	ASC-CSOFT_P INI_DQ (0.13)	ASC-DSOFT_P OUT_DQ (0.13)	ASC-AS_C_NSUM OUT_DQ (0.13)
10.25	ASC-AS_B_RF45 Q_PIT_OUT DQ (0.27)	ASC-AS_C_NSUM OUT_DQ (0.19)	LSC-POPAIR A_RF45_I_ERR DQ (0.16)	LSC-POP_A_RF45 I_ERR 256_DQ (0.16)	ASC-DHARD_P SM_DQ (0.16)	ASC-DHARD_P OUT_DQ (0.16)	ASC-DHARD_P INI_DQ (0.16)	CAL-CS_XARM DQ (0.16)	OAF-CAL_XARM DQ (0.16)	CAL-CS_XARM DQ (0.16)

FIG. 1. Summary table with the channels that have the highest coherence for each frequency. In this case the main gravitational-wave strain channel is called DCS-CALIB_STRAIN. The title also shows the times used in the analysis, both in UTC time and in GPS time (number of seconds elapsed since 0 hours Sunday 6 January 1980). Each row corresponds to a frequency bin, as indicated in the first column. The other columns show the coherence with a specific channel at that frequency. The name of the channel and the value of the coherence is indicated in each cell. For each row (frequency bin) the channels are listed in order of decreasing coherence. The cell background color gives a visual representation of the coherence value: red corresponds to one and white corresponds to zero.

that makes it easy to find the channels that have high coherence. For each frequency bin, we consider all the auxiliary channels that have a coherence with the output channel larger than C_α , and we sort them in order of descending values. We can then list the channels together with the corresponding coherence value. A subset of the results is shown in Fig. 1 for a typical analysis of the LIGO Hanford detector data during the science run O3 [2]. Each row of the table corresponds to a frequency, shown in the first column, while all the other cells in the row show the name of one of the auxiliary channels and the corresponding coherence value. To help with the identification of the most relevant channels, the background of each cell is color coded, with red meaning high coherence and white meaning low coherence. The reader can in this way have a glance at the full table and identify the most important channels that could contribute to the noise in the strain channel.

By clicking on any of the auxiliary channel names, it is possible to see a plot of the coherence as a function of frequency, as well as a projection of the noise contribution to the main channel (see Fig. 2). This projection is computed with the following equation, based on the definition of coherence [Eq. (1)], and assuming that there is indeed a noise coupling from the auxiliary channel to the main channel:

$$S_{\text{projection}}[x_i](f) = S[h, h](f)(1 - C[x_i, h]). \quad (3)$$

The fact that an auxiliary channel has high coherence with the output channel, and therefore its projection is close to the power spectral density of the output channel, does not necessarily imply that the channel is a good candidate for noise subtraction. This would be the case if the noise coupling is going from the auxiliary channel to the main channel. This is often the case, but it can happen that the auxiliary channel is sensing the output channel through a spurious path. Although all the channels known to be already related to the main strain channel have been removed, it is still possible that some cases of high coherence might not turn out to be good candidates for noise subtraction. Each case must be considered separately. Nevertheless, the initial brute-force approach is very effective for identifying a small number of auxiliary channels that can then be tried for offline noise subtraction. For the data analyzed in the example of Figs. 1 and 2, in the frequency region between 10 and 30 Hz, we identified several channels related to the feedback control loops of the angular sensing and control (ASC) and longitudinal sensing and control (LSC) subsystems. As we will discuss later in Sec. V, those channels performed

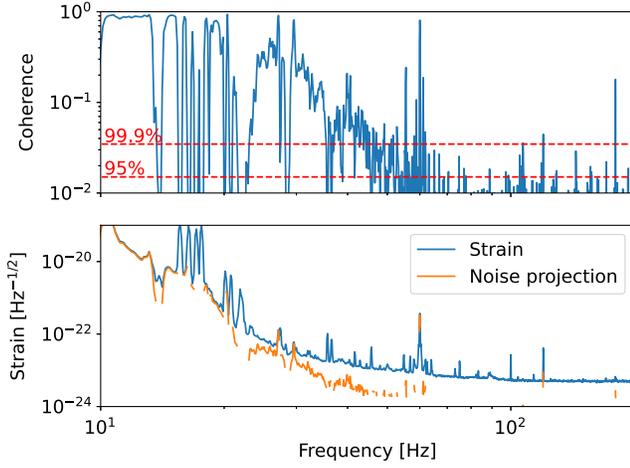


FIG. 2. Example of one of the plots generated automatically by the brute-force coherence computation. The top panel shows the coherence between the main strain channel DCS-CALIB_STRAIN and one of the angular feedback control loop channels ASC-CHARD_P_OUT_DQ. The blue trace shows the coherence as a function of frequency, and the red dashed line shows the confidence level C_α . The bottom panel shows the spectrum of the main channel, that is the square root of the power spectral density $S[h, h](f)$ and the coherence-based projection of the auxiliary channel noise.

well to reduce the noise in the gravitational wave strain channel.

The results shown in this kind of brute-force coherence do not take into account the redundancy of some channels. Often multiple channels are listed as having high coherence, but they are all related to each other. Therefore interpreting the results still needs some background knowledge of the detector inner workings. One possible way to address this issue would be to compute the cross-correlation between the auxiliary channels, and take it into account in the summary table, for example applying some form of independent component analysis. We did not investigate this approach, since it would increase significantly the computational cost of a brute-force approach, without adding much useful information.

III. NONLINEAR NOISE COUPLINGS

There is no straightforward way to extend the linear brute-force coherence-based approach to the most general nonlinear couplings from the auxiliary channels to the main output channel. However, we expect that the system should exhibit only small deviations from linearity, since there are many feedback control loops that actively maintain it near the optimal operating point.¹ We can then tentatively expand the nonlinear coupling of the auxiliary channels $\{x_i\}$ to the main channel h in terms of the lowest orders of a time-invariant Volterra series [24]:

¹One notable exception is the case of noise up-conversion due to scattered light [22,23], which we do not consider here.

$$h(t) = h_0(t) + \sum_i \int K_i^{(1)}(\tau_1) x_i(t - \tau_1) d\tau_1 + \sum_{i,j} \iint K_{i,j}^{(2)}(\tau_1, \tau_2) x_i(t - \tau_1) x_j(t - \tau_2) d\tau_1 d\tau_2 + \dots \quad (4)$$

where the $K_i^{(1)}$ are the linear coupling transfer functions discussed in the previous section, and $K_{i,j}^{(2)}$ are the Volterra kernels that describe the quadratic coupling from a pair of auxiliary channels to the main channel. The same relations can be expressed in frequency domain using the generalized frequency response functions [25]:

$$\tilde{K}^{(n)}(\omega_1, \dots, \omega_n) = \frac{1}{(2\pi)^{n/2}} \int \dots \int K^{(n)}(\tau_1, \dots, \tau_n) \times e^{-i(\omega_1 \tau_1 + \dots + \omega_n \tau_n)} d\tau_1 \dots d\tau_n$$

so that Eq. (4) becomes

$$\tilde{h}(\omega) = \tilde{h}_0(\omega) + \sum_i \tilde{K}_i^{(1)}(\omega) \tilde{x}_i(\omega) + \sum_{i,j} \int \tilde{K}_{i,j}^{(2)}(\omega_1, \omega - \omega_1) \tilde{x}_i(\omega_1) \tilde{x}_j(\omega - \omega_1) d\omega_1 \dots \quad (5)$$

where we denoted Fourier transforms with a tilde. Equation (5) makes it apparent that the quadratic terms mix frequencies. The linear coherence can be extended to the case of a quadratic coupling, obtaining the canonical bicoherence [26]. Practical applications of the bicoherence to channels from gravitational-wave detectors have met very limited success, due to the high variance of the estimate and the high computational cost. With about 4000 channels to be analyzed, a full bicoherence computation would entail processing $4000^2 > 10^7$ pairs of signals, an impractical number even for short time periods, unless the computation is parallelized on many computers.

Many of the channels recorded by gravitational wave detectors are characterized by a large dynamical range, with most of the power contained at low frequencies, below a few Hz (Fig. 3). It is quite common for the PSD of a signal to fall off by many orders of magnitude once the frequency gets above a few hertz. This behavior is a result of the use of seismic attenuation and isolation systems [27,28] to actively and passively reduce the interferometer mirror motions to less than 10^{-18} m above 10 Hz. One consequence is that a quadratic coupling would manifest like a slow modulation of an otherwise linear transfer function. This is immediately visible from Eq. (5) if we assume that one of the channels x_i has a PSD with support at frequencies ω_i much smaller than the frequencies ω_j where

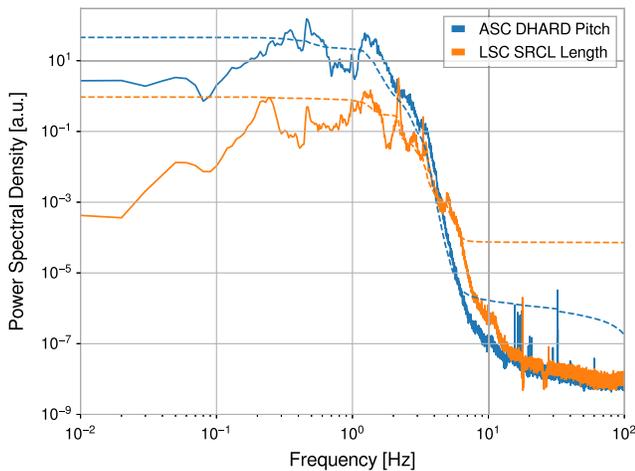


FIG. 3. Two example signals from the Advanced LIGO detector. The solid curves show the power spectral densities (PSD), while the dashed curves represent the integrated power, showing at each frequency the amount of power contained in the signal above that frequency. Note that those traces are plotted with the same vertical axis as the PSD, but are in units of signal power instead of signal power spectral density. Both signals have 99% of their power below about 3 Hz. The signals shown here are error points of two feedback control loops in the angular sensing and control (ASC) and longitudinal sensing and control (LSC) systems.

the second channel x_j has support: $\omega_i \ll \omega_j$. In this case the slow channel acts as a time-domain modulation of the linear coupling from the fast channel to the main output channel [14]. If we split the auxiliary channels into fast noise witnesses $s_i(t) \in \mathcal{N}$ and slow modulation witnesses $m_j(t) \in \mathcal{M}$ then Eq. (4) can be written as [14]:

$$h(t) = h_0(t) + \sum_{i \in \mathcal{N}} K_i^{(1)}[s_i] + \sum_{j \in \mathcal{M}} \sum_{i \in \mathcal{N}} \alpha_{i,j} [m_j(t)s_i(t)] + \dots \quad (6)$$

where $\alpha_{i,j}[\cdot]$ and $K^{(1)}[\cdot]$ represent linear and stationary transfer functions applied to the channel specified in the argument. The terms $\alpha_{i,j}[\cdot]$ are directly related to the second-order Volterra kernels. Although this simplification does not reduce the number of possible combination of channels, it provides a way to probe whether there are noise couplings that are modulated by a selected few known modulation-witness channels. We choose a small set of channels that are believed to witness the slow modulation, and multiply them by all the available channels. We then compute the frequency-domain coherence of each of those composite channels with the detector main output, like in the brute force approach described in Sec. II, and produce a table similar to Fig. 1. Each row of the table will now list the combination of fast channels and modulation channels, ordered by decreasing coherence. Instead of reproducing

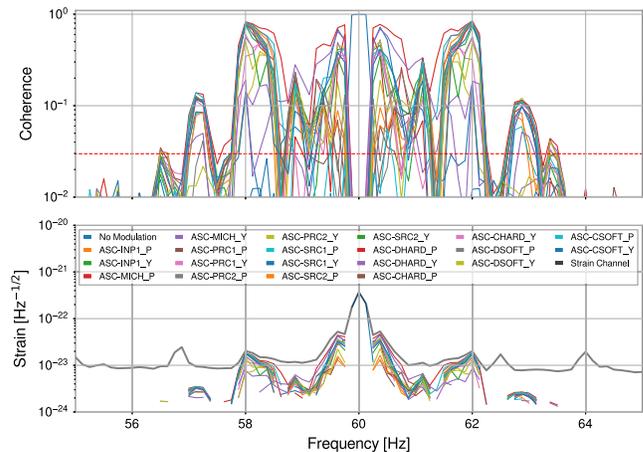


FIG. 4. An example of the plots produced by the brute-force coherence algorithm when used to look for modulated couplings. The top panel shows the coherence of the detector main output channel with a monitor of the power supply, and with the same channel modulated with angular control error channels, as in equation (6). The nonmodulated channel show high coherence only at the main power supply 60 Hz line, while the modulated channels show coherence only in correspondence of the sidebands of the main peak, as expected in the case of a nonstationary coupling. The bottom panel shows the coherence-based noise projection for each channel, as in Eq. (3).

here a portion of the coherence table, Fig. 4 shows an example of the result of this brute-force modulated-coherence analysis, using an extensions of the algorithm that produced the outputs shown in Figs. 1 and 2. In the example shown here we focus on the measured coherence around 60 Hz between the LIGO detector output and a channel monitoring the power supply main lines. As expected, there is near-one coherence for the line at 60 Hz and the not-modulated channel. However, this algorithm also shows significant coherence with the power monitor channel modulated with mirror angular sensing and control signals, at sideband frequencies around the 60 Hz line. This is an indication that the power line harmonics couple in a nonstationary way, as already pointed out in [14], where this nonstationary noise coupling was removed. This example shows that, although not the most general tool for nonlinear noise coupling discovery, this brute-force modulated-coherence search is a valuable tool to discover noises that can limit the LIGO sensitivity beyond the linear coupling case.

IV. NOISE SUBTRACTION

Once the noise couplings have been identified, there are two possible lines of action. First of all, if possible, one should try to remove the noise sources or the coupling mechanisms in the real-world detector. This is not always possible, since it might entail significant changes to the instrument hardware or control software. In other cases, the

data might be already collected, and therefore one is left with the second option, that is to implement a noise subtraction scheme. In the case of linear and stationary noise couplings, a typical approach utilizes the Wiener filter [12] in frequency domain. During the second Advanced LIGO science run O2 [29], noise from the angular motion of the input laser beam was subtracted in this way [13,30]. Similar frequency domain noise subtraction was also implemented in the O3 run [2,31] to remove known single-frequency lines, such a calibration lines and the 60 Hz power line and harmonics.

One could extend this methodology to non-stationary noise couplings like those described in Eq. (6). However, the number of channels involved increases significantly with the introduction of modulation witnesses. This in turn increases the computational complexity of the frequency-domain subtraction and makes the algorithm more prone to overfitting. Those problems motivated the development of the parametric and intrinsically causal algorithm described in [14], briefly reviewed here. The coupling of each noise-witness channel and modulation-witness channel is described using Eq. (6). The coupling transfer functions $\alpha_{i,j}$ and $K_i^{(1)}$ are described in terms of the sum of first-order poles, arranged in complex-conjugate pairs. The transfer functions can be described in the z domain [32] as:

$$\alpha_{i,j} = \sum_{k=1}^N \left[\frac{\rho_{i,j,k}}{1 - z_{i,j,k}z^{-1}} + \frac{\rho_{i,j,k}^*}{1 - z_{i,j,k}^*z^{-1}} \right] \quad (7)$$

where the star denotes complex conjugation, used to ensure that each transfer function has a real response to any real input signal. The use of a parametrization directly in the z domain ensures that the optimal subtraction parameters, once determined by fitting to the data, can be used both for offline and online noise subtraction, if desired. The optimal set of parameters $\{\rho_{i,j,k}, z_{i,j,k}\}$ can be found by minimizing a suitable cost function, that describes the improvement of the noise in the target channel $h(t)$ in the desired frequency band. This cost function can be computed directly and efficiently in frequency domain [14]:

$$c[\rho_{i,j,k}, z_{i,j,k}] = \int_{f_{\min}}^{f_{\max}} \frac{S_{r,r}(f)}{S_{h,h}(f)} df \quad (8)$$

where $S_{h,h}$ indicated the power spectral density of the target channel $h(t)$ and $S_{r,r}$ of the residual channel, that is the difference between the target and the sum of noises corresponding to the set of parameters under consideration, obtained from Eq. (6)

$$r(t) = h(t) - \sum_i K_i^{(1)}[s_i] - \sum_{j \in \mathcal{M}} \sum_{i \in \mathcal{N}} \alpha_{i,j} [m_j(t) s_i(t)] + \dots \quad (9)$$

The optimization problem involved in finding the optimal parameters is in general nonlinear and nonconvex, as it has multiple equivalent global minima (corresponding to permutations of the terms in each transfer function) and multiple local minima. As such, there is no general rule for choosing the best minimization algorithm. However, the simple analytical form of the parametrization allows us to directly compute the gradient of the cost function with respect to the parameters, and therefore gradient-based minimization strategies have been used successfully. For simpler problems like the power line harmonics subtraction, we found that using the ADAM gradient descent algorithm [33] from the `TensorFlow` library gave us good results [14]. For more complex problems, like angular and longitudinal control noise subtraction described below in Sec. V, we used the BFGS algorithm [34] or the SLSQP algorithm [35], implemented in the `scipy` library.

Regardless of the minimization algorithm, one needs to restrict the pole values to $|z_{i,j,k}| < 1$ to ensure that the resulting transfer functions are stable, meaning that their output is bounded for any bounded input and decays to zero when there is no input present. Two approaches are possible, depending on the algorithm used. For unconstrained algorithms like ADAM and BFGS, one can reparametrize the poles in terms of a magnitude and a phase $z_{i,j,k} = \zeta_{i,j,k} e^{i\phi_{i,j,k}}$ and write the magnitude as a bounded function of an additional unbounded parameter $\lambda_{i,j,k} \in \mathbb{R}$. In [14] a sigmoid function was used. For algorithms like SLSQP that allow for nonlinear constraints, one can simply parametrize the poles as the sum of real and imaginary part, and include the constraint $|z_{i,j,k}| < 1$ in the algorithm.

It is also often desirable to limit the accessible parts of the stable z plane, to avoid poles with either very low or very high frequencies, and with long ring-down times (high quality factor). This kind of constraint could be easily enforced using a reparametrization as above if the transfer function were described in the Laplace domain, as was done in [14]. One could extend this approach to the z domain using either the Tustin bilinear transform or the starred transform [32], at the price of the added nonlinearity to the dependence of the cost function on the parameters. We found that the best approach was to keep the simplest possible parametrization of pole values and residuals as real and imaginary parts, and convert the poles values in the z domain to frequency $f_{i,j,k}$ and quality factor $Q_{i,j,k}$ in the Laplace domain as follows:

$$\begin{aligned} s_{i,j,k} &= f_s \log(z_{i,j,k}) \\ f_{i,j,k} &= \frac{|s_{i,j,k}|}{2\pi} \\ Q_{i,j,k} &= \arccos \left(\frac{\text{Re}(s_{i,j,k})}{\text{Im}(s_{i,j,k})} \right) \end{aligned}$$

where f_s is the sampling frequency. The equations above can then be implemented as nonlinear constraints in an optimization algorithm such as SLSQP.

V. NOISE SUBTRACTION RESULTS

Figure 5 shows a typical improvement of the noise in the Advanced LIGO Hanford detector, obtained by applying the strategies described above. Noise was subtracted in multiple stages. First, the sidebands around the 60 Hz power supply line were removed, as already described in [14]. One of the insets in Fig. 5 shows the resulting improvement. The rest of the noise subtraction focused on the low-frequency region, below 30 Hz, where the main contributions were known to be from angular and longitudinal feedback control loop noises.

First, a series of single-frequency calibration lines are continuously injected into the angular motion of the main interferometer test masses, to aid with the angular control. Those lines are a dominant noise source between 18 and 23 Hz, and the large sidebands are an indication of the nonstationary nature of the coupling. We chose the error points of the angular control feedback systems as witnesses

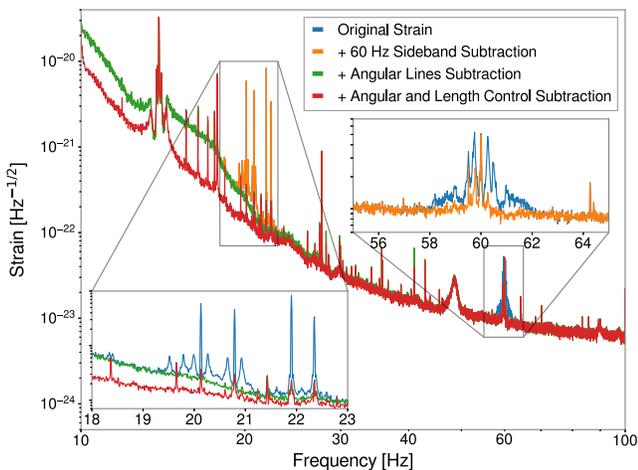


FIG. 5. Result of some stationary and nonstationary noise subtractions for the Advanced LIGO Hanford detector during the O3 observing run [2]. The two insets show zoomed-in versions of the main plot, highlighting some areas where nonstationary coupling was visible around some prominent lines (the 60 Hz power supply and some angular control calibration lines around 20 Hz). The broadband improvement at frequencies below 30 Hz is due to the subtraction of linearly coupled angular and longitudinal feedback control channels. Each trace shows the power spectral density of the detector output, calibrated in gravitational-wave strain. The blue trace is the original channel, before any subtraction. Each other trace shows the cumulative effect of additional noise subtractions: the orange trace shows the effect of removing the 60 Hz power line sidebands only; the green trace shows the effect of removing the angular calibration lines and sidebands from the data of the orange trace; finally the red trace shows the effect of all noise subtractions together.

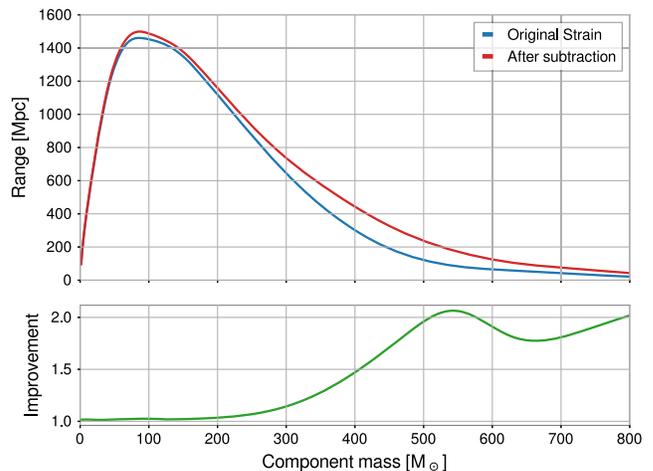


FIG. 6. Comparison of the detector range for binary black hole coalescence signals with and without the noise subtraction. The range is the maximum distance at which a signal could be detected with a signal-to-noise ratio of 8, averaged over sky position and source orientation [36]. The range is shown as a function of the component mass, for equal-mass binaries with no spin, integrating from 10 Hz. The bottom panel shows the improvement obtained with the noise subtraction. The effect is larger for the more massive binary systems.

for the slow modulation m_j , and used synthetic lines at the known calibration frequency as noise witnesses s_i . We could remove all of the calibration lines and sidebands, as shown in an inset of Fig. 5.

Secondly, the broadband improvement below 30 Hz is obtained by subtracting the longitudinal and angular feedback control channels. In this particular case no significant modulation of the coupling was detected, so a linear and stationary model was used. The detector output channel after subtraction shows negligible coherence with the noise sources, a good indication that the optimization algorithm was able to converge to a set of parameters close enough to the global minimum. Figure 6 shows the improvement in the detector range for binary black hole coalescence signals.

VI. CONCLUSIONS

Gravitational-wave detectors collect a large amount of data during their observation runs. Even though only one of the detector output channels contains the astrophysically-significant gravitational-wave strain, all the auxiliary channels have the potential of containing information that could be used to reduce the background noise in the main channel. Given the vast amount of data collected, the challenge is to determine which ones of the many auxiliary channels are useful. We described a brute-force approach that allowed us to discover many linear and stationary noise couplings from auxiliary channels to the main output channel. This tool has been routinely used in the past years to characterize the Advanced LIGO and Virgo

detector sensitivity and to discover possible schemes for noise subtraction. How to extend this method to the case of nonlinear noise couplings is not straightforward. We argued that most of the deviations from linearity in gravitational-wave detectors are likely to be small, due to the linearizing action of the feedback loops. Moreover, due to the use of sophisticated seismic isolation systems, many of the signals have content only at low frequency, reducing the quadratic coupling to an equivalent time-domain modulation of linear couplings. We therefore developed a tool to check for modulated noise couplings, where the modulation witness channels are assumed to be known. This simple extension of the linear brute-force coherence tool already proved to be useful in discovering new couplings.

We finally showed how the parametric and time-domain noise subtraction technique already presented in [14] could be used to significantly improve the low-frequency sensitivity of the Advanced LIGO detectors, by cleaning the data collected during the O3 science run. This method has been employed to remove power supply noise and angular and longitudinal control noises from the entirety of the O3 science run data. This is only the starting point, since an iterative approach is possible: the brute-force coherence algorithms can now be applied to the improved data, to discover new noise sources that can be removed.

The fact that the noise subtraction algorithm delivers directly the coefficients of time-domain infinite-impulse response filters means that it is amenable to an online implementation, where noise can be subtracted from the data as it is collected, in low latency. The advantages are multiple. On the detector side, a low-latency noise subtraction can allow for a quicker removal of noise sources

and a fast discovery of additional noise sources. From an astrophysical point of view, having lower background noise with low latency allows the gravitational-wave signal searches to detect more candidates in real time, increasing the potential of real-time multimessenger astronomy [37].

ACKNOWLEDGMENTS

This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. The authors gratefully acknowledge the support of the United States National Science Foundation (NSF) for the construction and operation of the LIGO Laboratory and Advanced LIGO as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, and the Max-Planck-Society (MPS) for support of the construction of Advanced LIGO. Additional support for Advanced LIGO was provided by the Australian Research Council. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation, and operates under cooperative agreement No. PHY-1764464. Advanced LIGO was built under Grant No. PHY-0823459. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459. This work carries LIGO Document number P2100472. We would like to thank all of the essential workers who put their health at risk during the COVID-19 pandemic, without whom we would not have been able to complete this work.

-
- [1] A. Einstein, Naherungsweise integration der feldgleichungen der gravitation, *Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl.* **1**, 688 (1916).
 - [2] R. Abbott *et al.*, GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo during the First Half of the Third Observing Run, *Phys. Rev. X* **11**, 021053 (2021).
 - [3] G. Vajente, E. K. Gustafson, and D. H. Reitze, Chapter Three—Precision interferometry for gravitational wave detection: Current status and future trends, *Advances in Atomic, Molecular, and Optical Physics* (Academic Press, New York, 2019), Vol. 68, pp. 75–148, [10.1016/bs.aamop.2019.04.002](https://doi.org/10.1016/bs.aamop.2019.04.002).
 - [4] J. Aasi *et al.*, Advanced LIGO, *Classical Quantum Gravity* **32**, 074001 (2015).
 - [5] F. Acernese *et al.*, Advanced Virgo: A second-generation interferometric gravitational wave detector, *Classical Quantum Gravity* **32**, 024001 (2015).
 - [6] A. Staley, D. Hoak, A. Effler, K. Izumi, S. Dwyer, K. Kawabe, E. J. King, M. Rakhmanov, R. L. Savage, and D. Sigg, High precision optical cavity length and width measurements using double modulation, *Opt. Express* **23**, 19417 (2015).
 - [7] A. D. Viets *et al.*, Reconstructing the calibrated strain signal in the Advanced LIGO detectors, *Classical Quantum Gravity* **35**, 095015 (2018).
 - [8] A. Buikema *et al.*, Sensitivity and performance of the Advanced LIGO detectors in the third observing run, *Phys. Rev. D* **102**, 062003 (2020).
 - [9] M. Tse *et al.*, Quantum-Enhanced Advanced LIGO Detectors in the Era of Gravitational-Wave Astronomy, *Phys. Rev. Lett.* **123**, 231107 (2019).
 - [10] T. Hong, H. Yang, E. K. Gustafson, R. X. Adhikari, and Y. Chen, Brownian thermal noise in multilayer coated mirrors, *Phys. Rev. D* **87**, 082001 (2013).
 - [11] P. Nguyen *et al.*, Environmental noise in Advanced LIGO detectors, *Classical Quantum Gravity* **38**, 145001 (2021).
 - [12] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications*

- (The MIT Press, Cambridge, MA, 1949), <https://mitpress.mit.edu/books/extrapolation-interpolation-and-smoothing-stationary-time-series>.
- [13] D. Davis, T. Massinger, A. Lundgren, J. C. Driggers, A. L. Urban, and L. Nuttall, Improving the sensitivity of Advanced LIGO using noise subtraction, *Classical Quantum Gravity* **36**, 055011 (2019).
- [14] G. Vajente, Y. Huang, M. Isi, J. C. Driggers, J. S. Kissel, M. J. Szczepańczyk, and S. Vitale, Machine-learning non-stationary noise out of gravitational-wave detectors, *Phys. Rev. D* **101**, 042003 (2020).
- [15] D. Davis *et al.*, LIGO detector characterization in the second and third observing runs, *Classical Quantum Gravity* **38**, 135014 (2021).
- [16] T. Isogai *et al.* (Ligo Scientific and Virgo Collaborations), Used percentage veto for LIGO and Virgo binary inspiral searches, *J. Phys. Conf. Ser.* **243**, 012005 (2010).
- [17] J. R. Smith, T. Abbott, E. Hirose, N. Leroy, D. MacLeod, J. McIver, P. Saulson, and P. Shawhan, A hierarchical method for vetoing noise transients in gravitational-wave detectors, *Classical Quantum Gravity* **28**, 235005 (2011).
- [18] R. Essick, P. Godwin, C. Hanna, L. Blackburn, and E. Katsavounidis, iDQ: Statistical inference of non-gaussian noise with auxiliary degrees of freedom in gravitational-wave detectors, *Mach. Learn.: Sci. Technol.*, **2**, 015004 (2020).
- [19] P. Welch, The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, *IEEE Trans. Audio Electroacoust.* **15**, 70 (1967).
- [20] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications* (Springer, New York, 2017), [10.1007/978-3-319-52452-8](https://doi.org/10.1007/978-3-319-52452-8).
- [21] G. Vajente, Readout, sensing, and control, in *Advanced Interferometers and the Search for Gravitational Waves. Astrophysics and Space Science Library*, Vol. 404, edited by M. Bassan (Springer, Cham, 2014), [10.1007/978-3-319-03792-9_6](https://doi.org/10.1007/978-3-319-03792-9_6).
- [22] B. Canuel, E. Genin, G. Vajente, and J. Marque, Displacement noise from back scattering and specular reflection of input optics in advanced gravitational wave detectors, *Opt. Express* **21**, 10546 (2013).
- [23] S. Soni *et al.*, Reducing scattered light in LIGO's third observing run, *Classical Quantum Gravity* **38**, 025016 (2021).
- [24] V. Volterra, *Sopra le funzioni che dipendono da altre funzioni. Nota II*, Opere matematiche: Memorie e note/Vito Volterra (Accademia Nazionale dei Lincei, Roma, 1887), Vol. 1, <https://rcin.org.pl/dlibra/publication/181808/sopra-le-funzioni-che-dipendono-da-altre-funzioni-nota-ii-volterra-1860-1940?language=en>.
- [25] X. Jing and Z. Lang, *Frequency Domain Analysis and Design of Nonlinear Systems based on Volterra Series Expansion* (Springer, New York, 2015), [10.1007/978-3-319-12391-2](https://doi.org/10.1007/978-3-319-12391-2).
- [26] H. X. He and D. J. Thomson, The canonical bicoherence part I: Definition, multitaper estimation, and statistics, *IEEE Trans. Signal Process.* **57**, 1273 (2009).
- [27] F. Matichard *et al.*, Seismic isolation of Advanced LIGO: Review of strategy, instrumentation and performance, *Classical Quantum Gravity* **32**, 185003 (2015).
- [28] S. M. Aston *et al.*, Update on quadruple suspension design for Advanced LIGO, *Classical Quantum Gravity* **29**, 235004 (2012).
- [29] B. P. Abbott *et al.*, GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs, *Phys. Rev. X* **9**, 031040 (2019).
- [30] J. C. Driggers *et al.*, Improving astrophysical parameter estimation via offline noise subtraction for Advanced LIGO, *Phys. Rev. D* **99**, 042001 (2019).
- [31] A. Viets and M. Wade, Subtracting narrow-band noise from LIGO strain data in the third observing run, LIGO document, T2100058, 2021, <https://dcc.ligo.org/LIGO-T2100058/public>.
- [32] C. L. Phillips, H. T. Nagle, and A. Chakraborty, *Digital Control System Analysis and Design*, 4th edition (Pearson, 2014), <https://www.pearson.com/store/p/digital-control-system-analysis-design/P100001450978/9780133496765>.
- [33] D. P. Kingma and J. Ba, ADAM: A Method for Stochastic Optimization, 2017, [10.48550/arXiv.1412.6980](https://arxiv.org/abs/1412.6980).
- [34] R. Fletcher, *Practical Methods of Optimization*, 2nd Edition (Wiley, New York, 2000), <https://www.wiley.com/en-us/Practical+Methods+of+Optimization%2C+2nd+Edition-p-9780471494638>.
- [35] J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer, New York, 2006), [10.1007/978-0-387-40065-5](https://doi.org/10.1007/978-0-387-40065-5).
- [36] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries, *Phys. Rev. D* **85**, 122006 (2012).
- [37] B. P. Abbott *et al.*, Multi-messenger observations of a binary neutron star merger, *Astrophys. J.* **848**, L12 (2017).