Subtraction-noise projection in gravitational-wave detector networks

Jan Harms, Christoph Mahrdt, Markus Otto, and Malte Prieß

Institut für Gravitationsphysik, Universität Hannover and Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut),
Callinstraße 38, 30167 Hannover, Germany

In this paper, we present a successful implementation of a subtraction-noise projection method into a simple, simulated data analysis pipeline of a gravitational-wave search. We investigate the problem to reveal a weak stochastic background signal which is covered by a strong foreground of compact-binary coalescences. The foreground, which is estimated by matched filters, has to be subtracted from the data. Even an optimal analysis of foreground signals will leave subtraction noise due to estimation errors of template parameters which may corrupt the measurement of the background signal. The subtraction noise can be removed by a noise projection. We apply our analysis pipeline to the proposed future-generation space-borne Big Bang Observer mission which seeks for a stochastic background of primordial gravitational waves in the frequency range ∼0.1 Hz—1 Hz covered by a foreground of black-hole and neutron-star binaries. Our analysis is based on a simulation code which provides a dynamical model of a time-delay interferometer network. It generates the data as time series and incorporates the analysis pipeline together with the noise projection. Our results confirm previous *ad hoc* predictions which say that the Big Bang Observer will be sensitive to backgrounds with fractional energy densities below $\Omega = 10^{-16}$.

PACS numbers: 95.85.Sz, 02.50.Sk, 04.80.Nn

I. INTRODUCTION

Currently, the first generation of large-scale laser interferometers is being operated to make a direct detection of a gravitational wave (GW) [1–3]. The primary targets of these detectors are compact-binary coalescences (CBCs), pulsars, and supernovae. However, the predicted event rate is so small that a detection of GWs by these instruments is highly unlikely. Their scienctific goals are to push technological developments towards a next generation of detectors and to place upper limits on GW amplitudes, thereby deriving, at least to some degree, restrictions on astrophysical processes. These limits either refer to deterministic sources like pulsars and binaries [4,5] or to stochastic backgrounds of GWs which may have an astrophysical or cosmological origin [6]. Although limits on stochastic backgrounds already become (weakly) scientifically relevant, the current upper limit of the background energy is about 10 orders of magnitude above a likely value for the cosmological background, assuming standard inflationary models. At this stage, confidence in a detection event would be significantly increased by combining the data of many different detectors [7]. This technique has become a standard tool in GW data analysis, and it is the only method to coherently detect stochastic signals. Within ten years, the first generation of ground-based detectors will be joined or replaced by advanced LIGO—a second-generation ground-based detector—and LISA, which will be the first space-borne laser-interferometric GW detector [8,9]. In contrast to first-generation detectors and advanced LIGO, LISA has to cope with a totally different data analysis problem. LISA will be sensitive to many sources which combine to form a GW signal foreground [10,11]. This foreground is formed by millions of galactic white-dwarf (WD) binaries and cannot be resolved completely. The unresolved, residual foreground acts as Gaussian confusion noise which impairs the detection and analysis of other signals. Any future detector will have to take the source-confusion problem into account and find a way to solve or circumvent it.

Even if a signal foreground can be resolved, the estimated signal waveforms will deviate somewhat from the true signals due to instrumental or confusion noise. If the estimated waveforms are subtracted from the data, then a residual signal spectrum, the subtraction noise, remains. Recently, a method was proposed to remove the residual foreground under certain conditions [12]. This method is based on a geometrical interpretation of signal analysis. It allows one to access a weak target like a stochastic GW background, irrespective of the fact that the residual foreground—in that case, resulting from inaccurate fitting of waveforms from binary neutron stars (NS) and black holes (BH)—may be much stronger. The conditions which have to be fulfilled are that (1) an accurate model exists for the waveform of individual foreground signals, (2) the overlap between foreground and background signals is negligible or irrelevant, (3) there are not too many foreground sources compared to the amount of data being collected (this will be specified in Sec. VI), and (4) the data are taken with a network consisting of at least 2 detectors. If the second condition is not fulfilled, then the removal of subtraction noise may deteriorate the waveforms of background signals in the data. The noise-removal algorithm, which is geometrically defined as a noise projection, comes with

many numerical challenges which could not be addressed in [12]. The purpose of this paper is to present a detailed discussion of the noise projection and to show how it can be implemented into a data analysis pipeline of a simulated future-generation detector network.

The network model of our simulation is based on a design draft for a future mission, the Big Bang Observer (BBO) [13]. Its primary target is to measure the stochastic GW background with cosmological origin (CGWB), which was generated shortly after the big bang, presumably during the inflationary phase [14–16]. For nonexotic (likely) models of the CGWB, the detector should be designed with peak sensitivity at the lowest possible frequencies, since sensitivity towards stochastic backgrounds increases with decreasing signal frequency. This background will be overlaid at all frequencies by a foreground of CBCs which needs to be subtracted. At this point, one has to take into account the confusion-noise problem. The galactic WD/WD foreground poses an intractable barrier even for future detectors. Consequently, its spectrum, which reaches out to 0.25 Hz [17], sets a lower boundary on the BBO's detection band. Lower frequencies beyond the WD/WD barrier are excluded by a foreground of merging supermassive black holes, and too little data would be collected at these frequencies. Above 0.25 Hz, a remaining foreground of 10^5–10^6 NS and BH binaries has to be subtracted from the data. For the BBO, the NS/NS mergers are the weakest foreground signals, and they are the most difficult to analyze and to subtract. Estimates for the number of merging NS/NS are highly uncertain. Extrapolating predictions of the galactic merger rate to the whole observable universe, one obtains values around 10^5 for the NS/NS mergers per year [18–20]. As was explicitly shown in [12], the BBO is sensitive to virtually all NS and BH CBCs in the entire observable universe. Not surprisingly, detection and analysis of CBCs build the secondary target of the BBO.

At an early stage of creating the simulation, it became clear that it would not be possible to demonstrate the projection method on a realistic foreground with 10^5 or more events even if the search for the signals was excluded from the pipeline. Therefore, we chose to test the algorithm on a much smaller foreground consisting of 100 injected

NS/NS systems. And even then it was necessary to shorten the observation time from the mission lifetime of 3 yrs down to 10^5 s. In the end, our results have to be extrapolated to the full observation time of 10^8 s in order to derive a prediction of the BBO's sensitivity towards the CGWB.

Our paper is organized as follows. In Sec. II, we describe in detail the fully dynamical model of the detector network which underlies the simulation. In Sec. III, we give an overview of the simulation pipeline and highlight that the network design of the BBO is tightly linked to the demands of the data analysis pipeline. A brief description of the signal model which determines the CBC waveforms and an introduction to the Fisher matrix, which is one of the basic quantities of the projection method, are given in Sec. IV. In Sec. V, we present a general framework for how to simulate a stochastic signal in a network of space-borne detectors. The geometrical interpretation of statistics is outlined in Sec. VI, including a description of the subtraction-noise projection. The optimal cross-correlation scheme for the BBO is explained and investigated in Sec. VII. Results are given in Sec. VIII together with an extrapolation of the BBO's sensitivity to an observation time which is equal to the BBO's lifetime.

II. THE NETWORK MODEL

The BBO consists of four independent detectors which orbit the sun at 1 AU. Each detector is formed by three spacecrafts in a nearly equilateral triangular configuration (Fig. 1). The nominal distance between spacecrafts is 50 000 km, which entails them following slightly eccentric orbits with $e \sim 9.65 \times 10^{-5}$. Each detector performs a cartwheel motion on the orbital path, completing one rotation in one year. All triangles are tilted against the orbital plane by 60°. In addition, the relative distances between spacecrafts change by small amounts (0.01%–0.02%) during one year—the so-called breathing motion—and therefore the detectors cannot be treated as rigid objects. The motion of each detector can be described in a power series of the orbital eccentricity e [21]. Expanding the exact orbital equations (which can be found in [22]) up to second order, the position vectors read

$$\vec{r}_{ij}(t) = \text{AU} \cdot \begin{pmatrix} \cos(\alpha_i(t)) \\ \sin(\alpha_i(t)) \\ 0 \end{pmatrix} + e \cdot \text{AU} \cdot \begin{pmatrix} \frac{1}{2}\cos(2\alpha_i(t) - \beta_{ij}) - \frac{3}{2}\cos(\beta_{ij}) \\ \frac{1}{2}\sin(2\alpha_i(t) - \beta_{ij}) - \frac{3}{2}\sin(\beta_{ij}) \\ -\sqrt{3}\cos(\alpha_i(t) - \beta_{ij}) \end{pmatrix} + e^2 \cdot \text{AU}$$

$$\cdot \begin{pmatrix} \frac{3}{8}\cos(3\alpha_i(t) - 2\beta_{ij}) - \frac{5}{8}\cos(\alpha_i(t) - 2\beta_{ij}) - \frac{5}{4}\cos(\alpha_i(t)) \\ \frac{3}{8}\sin(3\alpha_i(t) - 2\beta_{ij}) + \frac{5}{8}\sin(\alpha_i(t) - 2\beta_{ij}) - \frac{5}{4}\sin(\alpha_i(t)) \\ -\frac{1}{2}\sqrt{3}[\cos(2\alpha_i(t) - 2\beta_{ij}) - 3] \end{pmatrix} \tag{1}$$

where the first-order correction of the orbital path corresponds to the detector's cartwheel motion and the second-order correction describes the small relative motion of the spacecrafts. The angle $\alpha_i(t) = 2\pi f_{\text{orb}}t + \alpha_i(0)$ with $f_{\text{orb}} = 1/\text{yr}$ determines the location of the detectors $i = 1, \ldots, 4$ on the orbit of the Earth and $\beta_{ij} = 2(j-1)\pi/3 + \xi_i$ fixes the
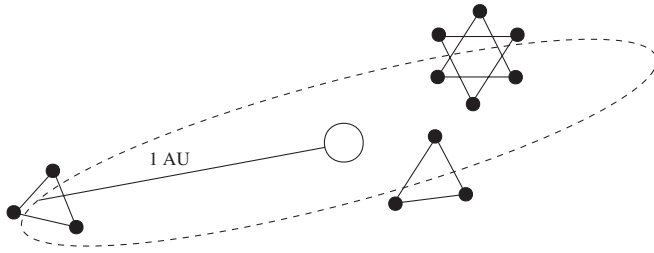
FIG. 1.   The BBO network of LISA-type detectors.

FIG. 2.   Triangular BBO detector configuration.

position of spacecrafts $j = 1, 2, 3$ in each detector i. The constants ξ_i govern the relative cartwheel phases of the detectors. In the following we will identify a spacecraft ij by a single index j assuming that all formulas are independently valid for each detector. The initial detector positions are $\vec{\alpha}(0) = (0, 0, 2\pi/3, 4\pi/3)$, and the internal configuration for each detector is given by $\vec{\xi} = (0, \pi, 0, 0)$. A π difference of the first two components of $\vec{\xi}$ puts the first two collocated detectors in a Star-of-David configuration. Otherwise, components can be chosen arbitrarily. A BBO spacecraft will certainly have a different design than a LISA spacecraft. However, it should be clear that the optimal sensitivity, which is quantum-noise limited at high frequencies and acceleration-noise limited at low frequencies, does not depend on the topologies of the optical benches. Therefore, we assume a LISA-type optical-bench design of the BBO spacecrafts and make use of well-known LISA results to evaluate the BBO's instrumental noise and GW response. For LISA, a minimum number of four photocurrents per spacecraft has to be included in a detector simulation. Two of them measure frequency fluctuations y_l of the light coming from a neighboring spacecraft via the detector arm l, and the other two measure intraspacecraft signals which are denoted by z_l where the photodiode (which records) z_l is found on the same optical bench as the diode y_l (there are two optical benches, one for each link to a neighboring spacecraft). The interspacecraft signals y_l are sensitive to GWs. The link index l assumes positive and negative values to discriminate between the two light-traveling directions $\hat{n}_l$ of the detector arm (see Fig. 2). Now, the noise spectrum of each photocurrent will be dominated by laser-frequency fluctuations and optical-bench noise [23]. The situation is different for ground-based detectors where laser noise interferes destructively at a beam splitter towards the output port and suspension systems and isolation schemes attenuate the equivalent of optical-bench noise. The solution is to establish destructive interference electronically by appropriately adding and subtracting photocurrents in each detector. Some photocurrents have to be added to others with certain time delays,

$$D_d y_l \equiv y_{l,d}(t) \equiv y_l(t - L_d(t)/c),$$

$$D_{d_2} D_{d_1} y_l \equiv y_{l;d_1 d_2}(t)$$
$$\equiv y_l(t - L_{d_2}(t)/c - L_{d_1}(t - L_{d_2}(t)/c)/c), \quad (2)$$

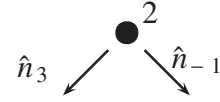where $L_d(t)$ is the length of the optical path of link d which was travelled by light being detected at time t, and c is the speed of light. That is why the electronic interference scheme is known as time-delay interferometry (TDI). We mention that one has to take into account that the light propagation directions $\hat{n}_l(t)$, $-\hat{n}_{-l}(t)$ differ predominantly due to the detector's cartwheel motion. Also, the relative spacecraft velocities $\dot{L}_d$ lead to minor corrections of the TDI combinations predominantly through relations of the form

$$L_{d_1}(t + L_{d_2}(t)/c) \approx L_{d_1}(t) + \dot{L}_{d_1}(t) \cdot L_{d_2}(t)/c \quad (3)$$

with typical relative speeds $\dot{L}_d \approx 10e^2 \mathrm{AU} f_{\mathrm{orb}} \approx 5 \times 10^{-4}\,\mathrm{m/s}$ (a smaller contribution comes from a term which is proportional to $e^2 \mathrm{AU}^2 f_{\mathrm{orb}}^2/c$; see Appendix A in [24] for details). The assumption for Eq. (3) is that the distance between spacecrafts does not change much during the light-traveling time L/c. Henceforth, to make our descriptions more readable, we will not make explicit reference to the optical-bench noise. It should be automatically included into the argument whenever we mention laser-frequency noise. Algebraically, it is always possible to treat optical-bench noise effectively as additional laser noise.

Previous investigations led to the introduction of three generations of TDI combinations which cancel laser-frequency noise based on various assumptions [24,25]. The *first-generation* combinations are defined to cancel laser noise of a detector which does not have cartwheel or relative spacecraft motion. If they were used to analyze realistic data, then residual laser noise would contribute to the total instrumental noise. The same is true for the *modified first-generation* variables which are based on the assumption that the detector is a rigid object which may perform cartwheel motion. However, residual laser noise will be much weaker in this case, since it is exclusively caused by the relative motion of spacecrafts, which is a second-order effect in terms of the orbital eccentricity e. Finally, the *second-generation* combinations take relative motions of the spacecrafts into account and cancel

noise contributions which depend linearly on $\dot{L}_d/c \propto e^2$. We claim that choosing second-generation instead of modified first-generation TDI variables has a much weaker influence in the case of the BBO than for LISA. The reason is that the relative motion of BBO spacecrafts compared to LISA spacecrafts is a factor 10^4 smaller. Investigating residual laser noise spectra in modified first-generation and second-generation combinations of BBO will show which generation has to be implemented. However, at least for the purpose of this paper, we just need to introduce the TDI combinations in their modified first-generation form.

In order to obtain a concise expression of the TDI combinations, we define new Doppler variables where a certain combination of intraspacecraft links z is added to interspacecraft links y,

$$y_l' \equiv y_l + \tfrac{1}{2}(z_{-l,l} - z_l). \quad (4)$$

In terms of these quantities, the laser-noise-free TDI combination X_1 can be cast into the form [26]

$$X_1(t) \equiv [y_{-3,32-2}' + y_{3,2-2}' + y_{2,-2}' + y_{-2}'] \\ - [y_{2,-2-33}' + y_{-2,-33}' + y_{-3,3}' + y_3']. \quad (5)$$

Time delays commute in first-generation variables, and therefore, semicolons in Eq. (2) have been substituted by commas. TDI X_1 mimics an unequal-arm Michelson interferometer centered at spacecraft 1. Cyclic permutation of all indices leads to the definitions of X_2 and X_3 which represent interferometers centered at spacecrafts 2 and 3. Each of the two sets of square brackets in Eq. (5) comprise
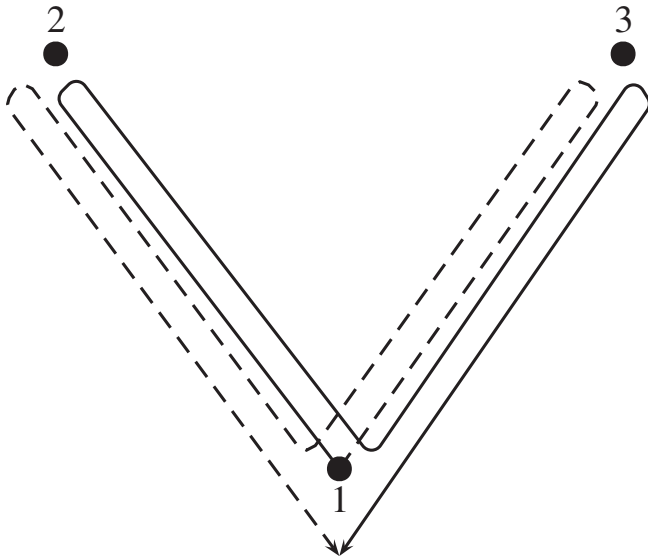

FIG. 3. Graphical representation of the unequal-arm Michelson TDI combination X_1 [27]. This combination of photo currents mimics the subtraction of two counter-propagating beams.

terms which represent a complete round trip of light in clockwise and counterclockwise directions. These two beams are then subtracted from each other to form the unequal-arm Michelson, which can be represented geometrically as shown in Fig. 3. The instrumental noise of the three channels X_i is correlated. It is more convenient to use channels with uncorrelated instrumental noise, especially if information from all channels is combined to provide optimal sensitivity with respect to GWs. These channels are known by the names A, E, T and can be defined in terms of the basis vectors X_i,

$$A = \frac{X_3 - X_1}{\sqrt{2}}, \qquad E = \frac{X_1 - 2X_2 + X_3}{\sqrt{6}},$$
$$T = \frac{X_1 + X_2 + X_3}{\sqrt{3}}. \quad (6)$$

Each of these variables can be seen as one detector, and so, in principle, each LISA-type detector has to be treated as a network which consists of three independent detectors. It turns out that these channels have quite different sensitivities to GWs and, also, correlation measurements between them do not yield the same profit as one may naively expect from a detector network. We will come back to this in a later section.

III. OVERVIEW OF THE SIMULATION

The simulation is organized according to the pipeline shown in Fig. 4. The first step is to generate a time series for the various Doppler streams (12 per detector, 48 in total). These data contain the instrumental noise and contributions from 100 CBCs. It is fairly simple to derive time-domain models of the test-mass noise ($S_{\text{tm}} \propto 1/f^2$ in units of Doppler shift) and the shot noise ($S_{\text{shot}} \propto f^2$) [28]. For the GW signal, we use a time-domain post-Newtonian (pN) approximation [Eq. (8)]. The data will depend on spacecraft motion and are generated consistently throughout the
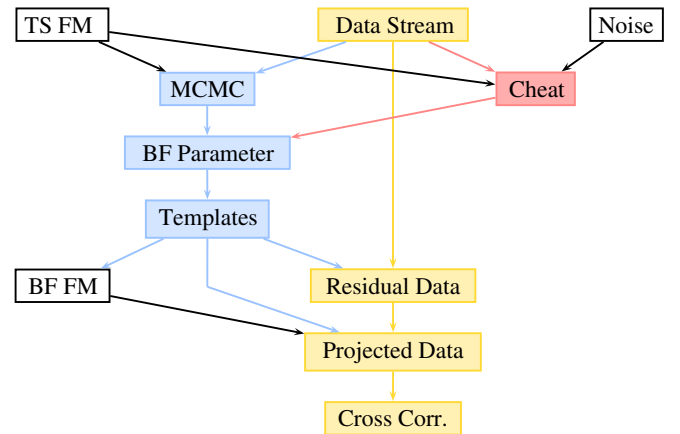

FIG. 4 (color online). FM: Fisher Matrix. TS: True signal. BF: Best-fit.

network by evaluating the GW phase at retarded time [Eq. (11)]. In contrast, the CGWB has to be generated directly in the frequency domain [Eq. (21)]. Assuming a Gaussian model, the frequency-domain representation of the CGWB is completely determined by a function called the overlap-reduction function, which essentially characterizes correlations between different output channels of the detector network (see Sec. V B).

The second step is to search the data for the CBCs. For that purpose, one has to exploit features of the network in a distinct order. The CBCs are analyzed by coherently combining the data of all 12 independent channels. The detector arrangement significantly improves parameter estimation of signals which cannot be integrated over long enough times. Any poorly fitted broadband signal could have a devastating effect on the mission goal: it may be that the respective residual noise, even after applying the noise projection method, is stronger than the CGWB spectrum. Initially, we implemented a Markov-chain Monte-Carlo (MCMC) algorithm, which searched for the maximum of the posterior distribution determined by our signal models and simulated data [29,30]. However, our computational resources were not sufficient to perform a realistic search. Therefore, we decided to *calculate* the best-fit parameter values. This is a trick (the "Cheat" box in Fig. 4) to avoid the CBC analysis. The idea is to calculate the noise vector on the template manifold, which points from the true signal to the best fit. Vectors are defined in tangent spaces, so the best fit has to lie in close proximity to the true signal (high signal-to-noise ratio). Further details can be found in Secs. IV B and VI B. The best-fit waveforms are subtracted from the data, which then consist of instrumental noise, the CGWB, and the subtraction noise.

The third step is to carry out the projection method to remove the subtraction noise. The projection operator is defined in Sec. VI C. This step is required for the following reason. A final correlation measurement of data streams of the two collocated detectors is supposed to lift the CGWB, which is correlated to some degree in different channels above any other contribution to the data. Now, this only works if correlations between channels of the instrumental noise and the subtraction noise can be neglected. This is true for the instrumental noise, but it is not for the subtraction noise which is highly correlated. Remember, the subtraction noise corresponds to the difference of the true signal and the estimated signal, which is a single quantity for the whole network (modulo detector transfer functions).

Finally, as described in Sec. VII, we use cross-correlation results to obtain a signal-to-noise ratio (SNR) for the CGWB with a given energy density. Knowing how the SNR scales with observation time, we derive a prediction for the BBO's sensitivity towards the CGWB based on its full mission lifetime.

IV. THE FISHER MATRIX IN A TDI FRAMEWORK

A. The signal model and its derivatives

For decades, people have been developing a geometrical interpretation of data analysis. These models usually consist of a distribution carried by a certain model manifold [31–33]. If one considers Gaussian distributions, the metric of the statistical manifold is given by the Fisher-information matrix $\Gamma_{\alpha\beta}$. To calculate it, one needs a noise model and a signal model. The model $\mathcal{T}^m$, which determines the signal inside a TDI combination $\mathcal{T}$, depends on parameters λ^α. Concerning the noise model, one assumes complete knowledge of its (double-sided) spectral density S^n. In practice, the model for the noise spectral density itself would depend on a few parameters which would have to be estimated before searching the data for certain signals.

In general, the Fisher matrix $\Gamma_{\alpha\beta}$ is associated with a TDI variable $\mathcal{T}$ and it is defined as a scalar product of derivatives of the signal model $\mathcal{T}^m$ with respect to the model parameters $\vec{\lambda}$. Defining $\partial_\alpha \mathcal{T}^m \equiv \partial \mathcal{T}^m / \partial \lambda^\alpha$, the Fisher matrix assumes the form

$$\Gamma_{\alpha\beta} = \langle \partial_\alpha \mathcal{T}^m | \partial_\beta \mathcal{T}^m \rangle$$
$$\equiv 2 \int_0^\infty df \frac{\text{Re}(\partial_\alpha \tilde{\mathcal{T}}^m(f) \partial_\beta \tilde{\mathcal{T}}^{m*}(f))}{S^n(f)}. \quad (7)$$

The model which determines a single CBC signal inside our simulation depends on 5 extrinsic parameters ($\lambda^0 =$ luminosity distance r of the source to the detector, $\lambda^1 =$ declination θ, $\lambda^2 =$ right ascension ϕ, $\lambda^3 =$ polarization angle ψ, $\lambda^4 =$ inclination angle ι of the binary orbit with respect to the line of sight) and 4 intrinsic parameters ($\lambda^5 =$ orbital phase ϕ_c, $\lambda^6 =$ coalescence time t_c, $\lambda^7 =$ total mass $M = M_1 + M_2$ of the binary system, $\lambda^8 =$ reduced mass $\mu = M_1 M_2 / (M_1 + M_2)$). In other words, we neglect the spin of the two binaries and assume zero eccentricity of their orbit. If two signals have to be parametrized, then λ^9 would be the distance parameter of the second signal, etc. A simulation is completely based on signal models (there are no true data). Still, we have to distinguish between the model for the simulated signal $\mathcal{T}^h$, which adds to the noise $\mathcal{T}^n$ to form the total data $\mathcal{T}^s$, and the model $\mathcal{T}^m$, which is used to analyze the data. In general, one may choose different models to generate and analyze data. Once generated, $\mathcal{T}^h$ is not considered as a function that depends on parameters, but as a set of fixed numbers. Here, we use the same model to generate and analyze data, and therefore the next paragraph gives a description of both.

Since Eq. (5) tells us that a TDI variable is a linear combination of the Doppler signals y_l, derivatives of a TDI variable with respect to certain model parameters can be expressed as a sum of derivatives of our signal

model y_l^{m} for a single Doppler signal. Therefore, it suffices to calculate and present the derivatives $\partial_\alpha y_l^{\mathrm{m}}$. The Doppler signal y_l^{m} is a projection of the GW tensor onto the light-traveling direction $\hat{n}_l$. In transverse-traceless coordinates, the matrix representation of the GW tensor contains the two GW polarizations h_+, $h_\times$ which are functions of the distance r and all intrinsic parameters:

$$
h_+(t) = \frac{c}{2r}\left[5\frac{(\mathcal{M}_c)^5}{t_c - t}\right]^{1/4} \cdot (1 + \cos^2(\iota)) \cdot \cos(\phi(t) + \phi_c),
$$

$$
h_\times(t) = \frac{c}{2r}\left[5\frac{(\mathcal{M}_c)^5}{t_c - t}\right]^{1/4} \cdot 2\cos(\iota) \cdot \sin(\phi(t) + \phi_c). \quad (8)
$$

We implement the restricted waveform, which neglects all harmonics higher than twice the orbital frequency and whose amplitude is determined by the chirp mass $\mathcal{M}_c \equiv GM^{2/5}\mu^{3/5}/c^3$. The evolution of the GW phase is given by a 3.5 post-Newtonian expansion

$$
\phi(t) = -\frac{2}{\eta}\sum_{k=0}^{7} p_k \tau^{(5-k)/8} \quad (9)
$$

with $\tau \equiv (\eta c^3(t_c - t))/(5GM)$ and expansion coefficients [34]

$$
p_0 = 1, \qquad p_1 = 0, \qquad p_2 = \frac{3715}{8064} + \frac{55}{96}\eta, \qquad p_3 = -\frac{3}{4}\pi, \qquad p_4 = \frac{9\,275\,495}{14\,450\,688} + \frac{284\,875}{258\,048}\eta + \frac{1855}{2048}\eta^2,
$$

$$
p_5 = \left(-\frac{38\,645}{172\,032} - \frac{15}{2048}\eta\right)\pi \log\left(\frac{\tau(t)}{\tau(0)}\right),
$$

$$
p_6 = \frac{831\,032\,450\,749\,357}{57\,682\,522\,275\,840} - \frac{53}{40}\pi^2 - \frac{107}{56}C + \frac{107}{448}\log\left(\frac{\tau(t)}{256}\right) + \left(-\frac{126\,510\,089\,885}{4\,161\,798\,144} + \frac{2255}{2048}\pi^2\right)\eta + \frac{154\,565}{1\,835\,008}\eta^2
$$

$$
-\frac{1\,179\,625}{1\,769\,472}\eta^3 p_7
$$

$$
= \left(\frac{188\,516\,689}{173\,408\,256} + \frac{140\,495}{114\,688}\eta - \frac{122\,659}{516\,096}\eta^2\right)\pi, \quad (10)
$$

which are most suitably expressed in terms of the symmetric mass ratio $\eta \equiv \mu/M$, and $C = 0.577\,215\,66\ldots$ is Euler's constant. In total, the GW phase depends on the mass parameters M, μ and the chirp time t_c. At some point we had to choose a convenient mass parametrization. Obviously, it would have been possible to use η instead of μ, and indeed in many situations this could be a good choice. However, for comparable-mass binaries like neutron-star binary systems (which are the only kind of signals we included in our simulation), the mass ratio η has the odd property of being close to its maximum value $\eta_{\max} = 0.25$, which holds for equal-mass binaries. Consequently, in our case, probability distributions for η will not be Gaussian and the distribution of other parameters may also exhibit non-Gaussian features through parameter correlations. So, without further investigations, we decided to use the reduced mass μ as a second mass parameter. As we will show later, the distributions for M and μ are highly correlated even for strong signals, which complicates the calculation of the inverse of the Fisher matrix. It would be interesting to investigate whether a different mass parametrization behaves better in this respect.

Now, projecting the GW tensor, we arrive at the following form for the Doppler signal [21]:

$$
y_l^{\mathrm{m}}(t) = \frac{1}{2(1 - \vec{k}(\theta, \phi) \cdot \vec{n}_l(t))} \cdot \vec{n}_l^{\mathsf{T}}(t)
$$

$$
\cdot \left(\sum_{I=+,\times} \mathbf{e}_I(\theta, \phi, \psi) \cdot h_I(t_s - \vec{k}(\theta, \phi) \cdot \vec{r}_s(t_s)/c)\right.
$$

$$
\left. - \sum_{I=+,\times} \mathbf{e}_I(\theta, \phi, \psi) \cdot h_I(t - \vec{k}(\theta, \phi) \cdot \vec{r}_r(t)/c)\right)
$$

$$
\cdot \vec{n}_l(t). \quad (11)
$$

The Doppler signal y_l^{m} is the GW induced frequency change of light which is sent at $t_s = t - L_l(t)/c$ from a spacecraft at position $\vec{r}_s(t_s)$ and received at time t by its neighbor at position $\vec{r}_r(t)$. "T" denotes a transposition. The two polarization matrices $\mathbf{e}_+$, $\mathbf{e}_\times$ are derived from their simple form in the GW propagation frame by a rotation $\mathcal{D}(\theta, \phi, \psi)$ into the solar barycentric frame whose coordinates are used to describe the spacecraft positions. The rotation matrix is shown explicitly in the Appendix of [24]. The propagation direction of the GW is given by $\vec{k} = -(\cos(\theta)\cos(\phi), \cos(\theta)\sin(\phi), \sin(\theta))$. To calculate the Fisher matrix, we generate a time series of derivatives $\partial_\alpha \mathcal{T}^{\mathrm{m}}$ with $\mathcal{T} \in \{A, E, T\}$ and subsequently apply a fast-Fourier transform (FFT) to obtain the amplitudes which govern the Fisher-matrix components. Taking derivatives

with respect to r and ϕ_c is trivial. The same is true for all other extrinsic parameters, although the result is rather complicated due to the waveform's dependence on the angles θ, ϕ, ψ [see Eq. (11)]. However, we think that it is instructive to present derivatives of the GW phase, since for all three nonzero derivatives of the phase, the result can be cast into a form which resembles the pN expansion Eq. (9) of the phase. The corresponding expressions can be found in the Appendix.

Finally, one has to specify models for the noise spectral densities $S^n(\mathcal{T})$. The spectral densities for the uncorrelated channels read [26]

$$S^n(A) = S^n(E)$$
$$= 16\sin^2(2\pi fL/c) \cdot (3 + 2\cos(2\pi fL/c)$$
$$+ \cos(4\pi fL/c))S^{tm} + 8\sin^2(2\pi fL/c)$$
$$\times (2 + \cos(2\pi fL/c))S^{shot}, \qquad (12)$$
$$S^n(T) = 128\sin^2(2\pi fL/c)\sin^4(\pi fL/c)S^{tm}$$
$$+ 16(1 - \cos(2\pi fL/c))\sin^2(2\pi fL/c)S^{shot},$$

with test-mass noise $S^{tm} = S^{acc}/(2\pi fc)^2$ and shot noise $S^{shot} = \hbar\omega_0/P_{rec}(2\pi f/\omega_0)^2$ in terms of double-sided spectral densities. The standard design of the BBO provides a spectral density of test-mass acceleration $S^{acc} = 9 \times 10^{-34} m^2/s^4/Hz$, which in our simulation is assumed to be equal for all test masses and light power $P_{rec} = 9$ W which is received by a spacecraft from one of its neighbors. The carrier frequency of the laser is $\omega_0 = 5.31 \times 10^{15} s^{-1}$. It is sufficient to express the noise models in terms of the nominal arm length $L = 50\,000$ km, because estimation errors of the noise will probably exceed the systematic errors due to the implementation of a simplified model. This is certainly true in our simulation, where an estimation of the low-frequency test-mass noise spectrum would be based on a few frequency bins. In contrast, the simulated noise $\mathcal{T}^n$ is based on a combination of individual Doppler signals, which then depends on detector motion and asymmetry.

B. Numerical evaluation of Fisher matrices

The complete simulated CBC foreground is composed of 100 NS/NS systems which occupy frequencies between 52 mHz and 2.2 Hz. Restricted by computational power of single notebooks (a cluster version of the code is being developed), we could simulate data with an observation time $T = 10^5$ s and a sampling frequency of $f_s = 5.242\,88$ Hz, which essentially fix the frequency range of the injected binaries. Parameter values for t_c, r and the two mass parameters of the CBCs are drawn from nonuniform priors. Values for M and μ are derived from normal distributions of the individual masses M_1, M_2 which are centered at 1.4 $M_\odot$, distance values are restricted to yield sensible signal-to-noise ratios, and values for the chirp time are determined by assuming a certain distribution of

signals over frequency bins. Assuming a Newtonian evolution of the orbital frequency of the binaries, the number of signals per frequency bin has to obey a distribution $N(f) \propto 1/f^{11/3}$ near BBO frequencies [35]. However, in our simulation, we draw initial frequencies from a $N(f) \propto 1/f$ distribution which yields a few systems at higher frequencies, but otherwise has no significant effect on our analysis. The frequency distribution of the 100 NS/NS signals is displayed in Fig. 5. The reason for taking greater care in frequency priors is that the signal distribution in frequency space has a significant impact on correlation values between parameter distributions of different CBCs, especially since the sky resolution of the detector network is comparatively poor for short observation times. The systems with highest frequencies have chirp times t_c which are of the order of a few T, and therefore the signal spectrum which is shown in Fig. 6 exhibits multiple quasimonochromatic peaks at low frequencies and a few chirp "plateaux" at higher frequencies.

Before being able to numerically evaluate the Fisher matrix, one has to search the simulated data for the injected binaries and estimate their parameter values. The signal derivatives in Eq. (7) have to be evaluated at the estimated parameter values $\hat{\vec{\lambda}}$. Implementing uniform priors of parameter distributions at this point, an optimal analysis is performed by searching the likelihood function $\mathcal{L}(\vec{\lambda}) \propto \exp(-1/2\sum_i\langle\mathcal{T}_i^s - \mathcal{T}_i^m(\vec{\lambda})|\mathcal{T}_i^s - \mathcal{T}_i^m(\vec{\lambda})\rangle)$ for its global maximum [36,37]. Here the sum has to be taken over all independent network channels (the BBO network furnishes 12 independent channels, 3 per detector). As was argued in [12], no existing computer or network of computers could accomplish that search for a realistic foreground formed by 10^5–10^6 CBCs. Even searching simulated BBO data for 100 signals, including the estimation of parameters, is a difficult task which optimally requires a high-end cluster. Our work is not intended to make any propositions for how to perform that search, let alone to carry it out. So we have
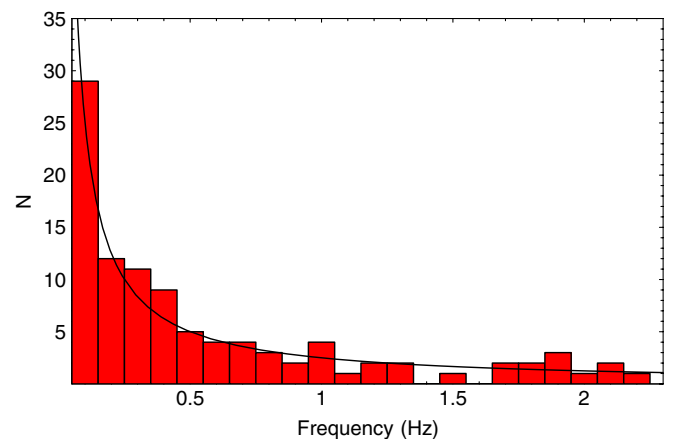

FIG. 5 (color online). Initial distribution of NS/NS signals over frequency bins.
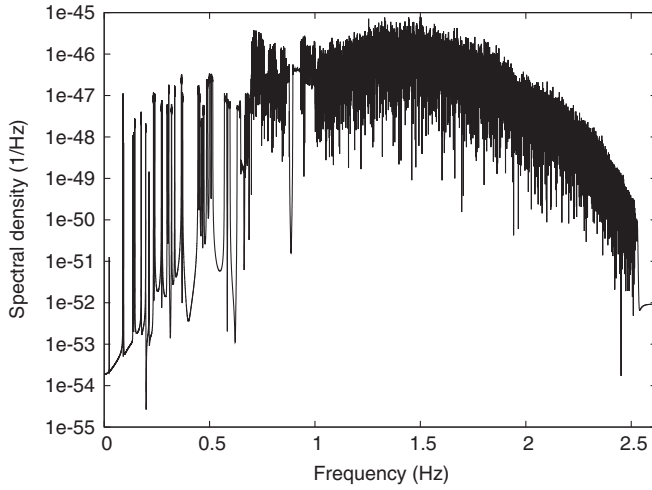
FIG. 6. Signal spectrum of the A channel of the first detector. At low frequencies up to 0.7 Hz, the spectrum features distinguishable, mildly chirping signals. At high frequencies, many signals overlap to form a plateau of chirps. The signals' merger times were chosen such that, within $T = 10^5$ s, signal frequencies never become greater than 2.6 Hz, which is half the sampling frequency. Note that at frequencies above 0.5 Hz, only every 20th frequency bin is plotted to reduce figure size.

to work around the problem. The idea is that knowing the realization of the instrumental noise in a simulation run—which we do, since we generate the noise $\mathcal{T}^n$ and add it to the signals $\mathcal{T}^h$—and assuming Gaussian distributions for the signal parameters with small estimation errors (high SNR), one can use the following equation to calculate the estimation errors $\delta\lambda^\alpha$ (the difference between the maximum likelihood values $\hat{\lambda}^\alpha$ and the true parameter values λ^α of the signals) [12,38]:

$$\delta\lambda^\alpha = \Gamma^{\alpha\mu}\sum_i\langle\mathcal{T}_i^n|\partial_\mu\mathcal{T}_i^m\rangle. \qquad (13)$$

Again, the sum has to be taken over all independent channels $\mathcal{T}_i$ of the detector network. The parameter errors depend on the inverse of the network Fisher matrix $(\Gamma^{\alpha\beta}) \equiv (\sum_i\Gamma_{\alpha\beta}^i)^{-1}$ with $\Gamma_{\alpha\beta}^i \equiv \langle\partial_\alpha\mathcal{T}_i^m|\partial_\beta\mathcal{T}_i^m\rangle$. The parameter estimation errors are added to the true parameter values of the signals injected into the simulation pipeline, and the Fisher matrix can finally be evaluated. A brief introduction into the geometric interpretation of the Fisher matrix and how to make use of it can be found in Sec. VI, which also explains Eq. (13). The reader may have worked with a close relative of Eq. (13) in another context. It is a generalization of the $\mathcal{F}$-statistic equation to obtain best fits of its 4 amplitude parameters [39]. The $\mathcal{F}$ statistic is based on templates which are linearized with respect to r, ι, ψ, and ϕ_c. It is straightforward to show that if one substitutes the complete data $\mathcal{T}_i^s$ for the noise $\mathcal{T}_i^n$ in Eq. (13), then the equation directly yields the best fit of any parameter which enters linearly into the definition of

the template. Our model does not have linear parameters, but many alternative template models do.

Notice that by calculating the best fits, we neglect the detection problem, i.e. we assume that all binaries in the data are detected. We should also mention that our method to calculate the best fits represents an optimal analysis scheme. The optimal scheme is to search simultaneously for all signals. A more realistic search which requires much less computational power is the hierarchical search. There, one detects signals one by one, starting with the highest SNR and "digging" down the signal spectrum until the last binary is identified and measured. Unresolved binaries act as confusion noise. Identified binaries are subtracted from the data so that, during the hierarchical search, parameters of already detected binaries are constantly refined as the confusion noise decreases. This scheme has been studied by means of a self-consistent recursive evaluation in [12]. In contrast, the optimal search is not corrupted by confusion noise. Correlations between different signals, which lead to confusion noise in the hierarchical search, are incorporated into the signal model of the optimal search. The only possible shortcoming of an optimal search is that it may fail to accurately estimate parameter values of the model (including finding the right template-manifold dimension which depends on the number of detected signals).

C. Multisignal templates and Fisher-matrix inversion

Given 100 signals which each depend on 9 parameters, the total Fisher matrix $\Gamma_{\alpha\beta}^i$ for each channel i becomes a 900×900 matrix. It turns out that inverting the Fisher matrix is a highly nontrivial task. In our simulation, inverse Fisher matrices are used in two different ways. First, we need the network matrix evaluated at the true parameter values to compute the estimation errors by means of Eq. (13). Second, the inverse Fisher matrices of individual channels evaluated at the estimated parameter values have to be computed to define the subtraction-noise projector in Eq. (42). To start with, we outline a generic inversion scheme which, in the end, does not solve all problems. However, this method still forms the foundation of the complete solution. In the next part of this section, we omit the channel index i, since the described method is used in the same way to invert channel and network Fisher matrices.

The inversion procedure starts with the computation of a new matrix $\Gamma_{\alpha\beta}' = \Gamma_{\alpha\beta}/\sqrt{\Gamma_{\alpha\alpha}\Gamma_{\beta\beta}}$ such that $\Gamma_{\alpha\alpha}' = 1$ and all off-diagonal components have an absolute value smaller than 1. This step is necessary since in our simulation the numerical range of Fisher-matrix components is 10^{-50}–10^5. Consequently, ratios of different eigenvalues of the Fisher matrix can assume large values. Such a matrix is called *ill conditioned* and is known to be hard to invert numerically, because tiny inaccuracies of a few matrix

components may have a great effect on the eigenvalues or the components of the inverted matrix. These inaccuracies are unavoidable due to limited machine precision. It turns out that $\Gamma'_{\alpha\beta}$ is still ill conditioned and cannot be inverted using standard double-precision variables. At this point, one has to implement a multiprecision package into the code. We found that CLN (Class Library for Numbers) [40] provides all required functions. Using a 50-digit precision, the scaled matrix can be inverted following its LU decomposition. Next, the inverted matrix is scaled back to form the inverted Fisher matrix $\Gamma^{\alpha\beta} = \Gamma'^{\alpha\beta}/\sqrt{\Gamma_{\alpha\alpha}\Gamma_{\beta\beta}}$. We expect that the degree of ill conditioning decreases significantly once much longer observation times can be simulated. To explain this, we need to have a look at the correlation matrix of a single binary. The correlation matrix is derived from the covariance matrix $\Gamma^{\alpha\beta}$ in the same way as $\Gamma'_{\alpha\beta}$ was derived from $\Gamma_{\alpha\beta}$. As can be seen in Table I, correlation is especially strong between $r \leftrightarrow \iota$, $\phi_c \leftrightarrow t_c$, and $M \leftrightarrow \mu$. Strong correlations indicate that by changing one parameter from its best-fit value, the loss in accuracy of the waveform fit can be compensated by changing the value of the other parameter of the correlation pair. Therefore, at first sight, it seems to be obvious that these pairs may be strongly correlated. The question is, under which circumstances do these correlations become weaker. The pair $M \leftrightarrow \mu$ decorrelates once a considerable amount of the chirp is observed and the signal-frequency change accelerates. It is well known that the low-frequency evolution of CBC waveforms is completely determined by a single mass parameter, the binary's chirp mass $\mathcal{M}_c$. For those waveforms, implementing a model which needs two mass parameters must exhibit maximal correlations between them. A similar argument can be invoked for the pair $\phi_c \leftrightarrow t_c$. Consequently, correlation matrices of signals with higher frequencies have lower correlation coefficients for these pairs. A decorrelation of $r \leftrightarrow \iota$ (and weakening of many other correlation coefficients) is observed as soon as the orbital motion of the detectors leads to a measurable amplitude and phase modulations of the signal. With maximal observation times of $T = 10^5$ s, we are not able to study the impact of the Doppler shift on parameter estimations. However, as we are not particularly interested in the quality of best fits, but accept any quality as long as a projection of subtraction noise can be carried out successfully, we do not investigate parameter correlations further in this paper. It turns out that the best fits are accurate enough for this purpose.

As mentioned in the beginning of this paper, the inversion algorithm as presented in the last paragraph does not provide a complete solution of the inversion problem. The reason is that a 900×900 matrix determined by multiprecision components needs too much memory, and even if it can be kept in memory during runtime (e.g. by implementing specifically designed inversion schemes [41]), the inversion would take too much time. To solve this problem, we have to understand a little more about Fisher matrices. Consider a matrix which includes N copies of single CBC templates. Each CBC is determined by P parameters. In our case, the number of templates is $N = 100$ and the number of parameters is $P = 9$. Let us introduce the "confused" Fisher matrix

$$\Gamma^0 = \begin{pmatrix} g^1 & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & 0 & \\ 0 & \dots & 0 & g^N \end{pmatrix}. \tag{14}$$

It is a block matrix which contains the Fisher matrices $g^k_{\alpha\beta}$ with $\alpha, \beta \in \{1, \dots, P\}$ of N CBCs on its diagonal. In other words, it differs from the total Fisher matrix by neglecting correlations between different signals. We call it confused, because whenever this matrix is applied instead of the total matrix, it is like that our knowledge of correlations between different signals is ignored. Correlation coefficients become random variables in the analysis pipeline leading to confusion noise. We claim that it is legitimate to use the block matrix when calculating estimation errors by means of Eq. (13). To support this claim, one has to investigate the impact of the correlation coefficients on the eigenvalues of Fisher matrices of individual CBCs. Fisher matrices (their inverses to be precise) define a multivariate Gaussian

TABLE I. The network correlation matrix for a single NS/NS at 0.57 Hz. Some correlation coefficients strongly depend on the observation time, which is $T = 10^5$ s in this case.

$\vec{\lambda}$	r	θ	ϕ	ψ	ι	ϕ_c	t_c	M	μ
r	1	-0.287	-0.038	-0.049	-0.999	-0.066	-0.066	0.050	-0.048
θ		1	0.162	0.118	0.287	-0.189	-0.189	-0.062	0.068
ϕ			1	-0.055	0.038	0.043	0.043	0.045	-0.046
ψ				1	0.049	0.014	0.014	-0.084	0.084
ι					1	0.066	0.066	-0.050	0.048
ϕ_c						1	0.999	-0.017	-0.011
t_c							1	-0.014	-0.014
M								1	-0.999
μ									1

distribution in parameter space. Their eigenvalues correspond to the variances of the distribution along its principal axes. The question is, what happens to the distribution defined by a matrix $g^k_{\alpha\beta}$ when the correlations between the signal k and other signals are incorporated into the model. This problem can be treated with perturbation theory, similar to perturbations of a Hamiltonian (here, Γ^0) which is weakly perturbed by interactions,

$$
C = \begin{pmatrix} 0 & g^{12} & \cdots & g^{1N} \\ g^{12,\top} & \ddots & & \vdots \\ \vdots & & & g^{N-1,N} \\ g^{1N,\top} & \cdots & g^{N-1,N,\top} & 0 \end{pmatrix} \quad (15)
$$

where $g^{ij}_{\alpha\beta}$ are the correlation coefficients between signals $i, j \in \{1, \ldots, N\}$ and " $\top$ " denotes the transposition of a matrix. The proper condition to justify the perturbation approach is that correlations g^{ij} have to be small compared to differences of eigenvalues of g^i and g^j, which is the case for any combination of signals in our simulation. For this particular form of perturbation C, theory tells us that the lth eigenvalue ξ^0_{kl} of the Fisher matrix g^k ($l \in \{1, \ldots, P\}$) is perturbed at second order in C according to

$$
\xi_{kl} = \xi^0_{kl} + \sum_{i \neq k}^N \sum_{j=1}^P \frac{|\langle kl^0|g^{ki}_{lj}|ij^0\rangle|^2}{\xi^0_{kl} - \xi^0_{ij}} + \mathcal{O}(C^4) \quad (16)
$$

where $|kl^0\rangle$ are the P eigenvectors of the Fisher matrix g^k with eigenvalues ξ^0_{kl}. The perturbation of the eigenvectors reads

$$
|kl\rangle = |kl^0\rangle + \sum_{i \neq k}^N \sum_{j=1}^P \frac{\langle kl^0|g^{ki}_{lj}|ij^0\rangle}{\xi^0_{kl} - \xi^0_{ij}} \cdot |ij^0\rangle + \mathcal{O}(C^2). \quad (17)
$$

Therefore, up to first order in C, one can say that the multivariate Gaussian does not change the lengths of its major axes. Instead, the distribution is rotated and the small rotation angles are given by the fraction in Eq. (17). This means that, when using the block matrix Γ^0 instead of the total Fisher matrix Γ, the coordinate basis $|\partial_\mu \mathcal{T}^m_i\rangle$ in Eq. (13) is misaligned with respect to the inverted Fisher matrix $(\Gamma^0)^{-1}$, and that the parameter errors $\delta\lambda^\alpha$ lie in false "directions," but give rise to a comparable accuracy of the waveform fit up to order $\mathcal{O}(C)$. That is the reason why we may use the block matrix at this point. The projection method described later does not depend on these rotations of the parameter space. The benefit is that we can easily invert the confused Fisher matrix by inverting Fisher matrices of each signal. In principle, we could even correct the misalignment by rotating the inverted matrix with rotation angles $\langle kl^0|g^{ki}_{lj}|ij^0\rangle/(\xi^0_{kl} - \xi^0_{ij})$ provided that we also calculate the eigenvalues and eigenvectors of all Fisher matrices g^k. As a corollary, we add that correlations between different signals always lead to a loss of Fisher information represented by the determinant of the Fisher

matrix,

$$
\det\Gamma = \det\Gamma^0\left(1 - \sum_k^N \sum_{i \neq k}^N \sum_{l,j=1}^P \frac{|\langle kl^0|g^{ki}_{lj}|ij^0\rangle|^2}{\xi^0_{kl}\xi^0_{ij}}\right) + \mathcal{O}(C^4). \quad (18)
$$

The second term in round brackets is always positive and well defined, because Fisher matrices are positive definite. So, the decrease of the determinant is a second-order effect in correlation coefficients which is further suppressed by the Fisher information of particular template parameters (i.e. the respective eigenvalue ξ^0_{kl}), and therefore, especially in the high SNR regime, one may neglect information loss.

Unfortunately, we cannot make use of the same simplification when dealing with Fisher matrices which define the projection operator. There, directions reflect the amount of correlations between the template derivatives and actual subtraction noise in the data. These very correlations have to be exploited to facilitate removal of the subtraction noise. Our strategy in this case is to reduce the dimension of the template manifold by projecting a subset of all signals. Namely, we project all signals which possess power in the frequency range that contributes most of the SNR to the final correlation measurements of the CGWB. More details can be found in Sec. VIII A.

V. THE STOCHASTIC BACKGROUND

In this section we sketch out how to simulate the CGWB. The stochastic data are generated directly in the frequency domain starting with one channel and then taking correlations between channels into account to derive data for other channels. The function which describes the correlations is called the overlap-reduction function (ORF) $\gamma_{ab}(f)$ between channels a and b of the same detector or different detectors. As we are going to learn in Sec. V B, correlations between channels A, E, T of the same detector are negligible. We introduce a new definition for the ORF which does not make any attempt to factor out a channel's transfer function. This is the most convenient approach based on a dynamical detector model where relative detector and satellite motion has to be taken into account. Henceforth, since the T channel does not furnish significant sensitivity with respect to a measurement of the CGWB (see [42] and Sec. V B), it will be excluded from the branch of the pipeline which processes the CGWB.

A. Generation of stochastic backgrounds in detector networks

A zero-mean Gaussian background is completely characterized by its second-order moments, i.e. the autocorrelations and cross correlations of TDI channels. In our case, we just have to include channels of the two collocated detectors in our investigations since correlations between other detectors are negligible (correlations fall off rapidly

if the distance between detectors becomes much larger than the length of the GW). Therefore, based on correlation properties of a simulated stochastic background signal $\mathcal{T}^{\mathrm{b}}$ in detector channels a and b, notably

$$\frac{1}{T}\mathrm{Re}\{\langle \tilde{\mathcal{T}}^{\mathrm{b}}_{a}(f)[\tilde{\mathcal{T}}^{\mathrm{b}}_{b}(f)]^{*}\rangle\} = S^{\mathrm{b}}(f)\gamma_{ab}(f), \qquad (19)$$

we have to find an algorithm to calculate the stochastic TDI signals $\mathcal{T}^{\mathrm{b}}(f)$ produced by all channels of the two collocated detectors. We define the ORF γ_{ab} as a real-valued function, which essentially establishes a convention for how correlations between channels are evaluated [see Eq. (46)]. The ORF governs the strength of cross correlations ($a \neq b$) and autocorrelations ($a = b$) in the frequency domain. Here, the observation time T is used to convert amplitude squares into spectral densities and the (double-sided) background spectral density $S^{\mathrm{b}}(f)$ of the GW amplitude is related to the fractional energy density Ω by

$$S^{\mathrm{b}}(f) = \frac{3H_0^2}{4\pi^2} \cdot \frac{\Omega(f)}{f^3}. \qquad (20)$$

The background is assumed to be isotropic and to have a white energy spectrum with the fiducial value $\Omega(f) = 10^{-15}$. In our simulation, the value of the Hubble constant is $H_0 = 72$ km/s/Mpc.

Now, the idea is to generate a stochastic signal with the correct spectrum in one channel and then to proceed with other channels by taking mutual correlations into account. The equations used to generate the background are [6]

$$\tilde{\mathcal{T}}_1(f) = \frac{1}{\sqrt{2}}\sqrt{S^{\mathrm{b}}(f)T}\sqrt{\gamma_{11}(f)}(a_1(f) + \mathrm{i}b_1(f)),$$

$$\tilde{\mathcal{T}}_2(f) = \tilde{\mathcal{T}}_1(f)\frac{\gamma_{12}(f)}{\gamma_{11}(f)} + \frac{1}{\sqrt{2}}\sqrt{S^{\mathrm{b}}(f)T\left(\gamma_{22}(f) - \frac{\gamma_{12}^2(f)}{\gamma_{11}(f)}\right)}$$
$$\times (a_2(f) + \mathrm{i}b_2(f)). \qquad (21)$$

For each frequency, random values $a_1(f)$, $a_2(f)$, $b_1(f)$, and $b_2(f)$ are drawn from a normal distribution $\mathcal{N}(0, 1)$. It is not necessary to extend Eq. (21) to three or more channels, because in the end, correlations are evaluated between independent pairs $A_0 \leftrightarrow A_1$ and $E_0 \leftrightarrow E_1$ of the two collocated detectors. In the next section, we show how to obtain the overlap-reduction function $\gamma_{ab}(f)$ in a TDI network.

B. Overlap-reduction function

As already mentioned, we need a procedure to calculate the ORF between arbitrary detector channels at all frequencies. In previous publications, the ORF has been defined by [43]

$$\gamma_{ab}^{\mathrm{pre}}(f) = \frac{5}{8\pi}\int_{S^2} d\hat{\Omega}\mathrm{e}^{2\pi i f\hat{\Omega}\Delta\vec{x}/c}(F_a^+ F_b^+ + F_a^\times F_b^\times). \qquad (22)$$

Here $\Delta\vec{x}$ denotes the separation vector between the detec-

tors, $\hat{\Omega}$ is a unit-length vector on the two-sphere S^2, and $F_a^{+,\times}$ are the response functions of the detector a to the $+$ or $\times$ polarization. This function is normalized such that its value is unity for two coincident, aligned Michelson detectors with perpendicular arms. The motivation for this definition was to separate the optical properties of the detectors from their geometric properties which determine the response functions. The ORF $\gamma_{ab}^{\mathrm{pre}}$ is used to calculate correlations between projected GW amplitudes in two detectors, and then optical transfer functions can be used to derive the correlations of the detector outputs.

There are a few reasons why the ORF in Eq. (22) cannot be used under general circumstances. First, it assumes that the response functions are constant in time. Obviously, this is not the case for space-borne detectors, where test masses which are on individual orbits around the sun move relative to each other. Also, the separation vector between different detectors does not have to be constant. Beyond the long-wavelength limit which demands that the length of a GW is much larger than the dimension of detectors, it is in any case difficult to agree upon a detector position. In other words, light-traveling times between test masses are neglected in $\gamma_{ab}^{\mathrm{pre}}$. Therefore, we propose a slightly different definition of the ORF which is not normalized and which is a direct representation of the correlation strength between channels. We will not make any attempts to separate optics and geometry, since they are tightly linked in TDI detectors. The complete detector dynamics can be incorporated into a frequency-domain correlation function in the following way:

(1) Inject a time-domain delta signal $\delta(t, t') = f_{\mathrm{s}} \cdot \mathrm{sinc}(\pi f_{\mathrm{s}}(t - t'))$ (f_{s} being the sampling frequency) associated with a polarization and sky direction into each detector of the network. Propagate it from a common origin so that relative phase shifts of the GW at different detectors are automatically taken into account. Make sure that t' is larger than any light-traveling times between detectors; otherwise the peak does not appear in all data streams.

(2) Record the outputs $\mathcal{T}_a^{\mathrm{w}+,\times}(\theta, \phi; t)$ of all TDI channels.

(3) Apply an FFT to the data and thereby obtain the complex-valued transfer functions $\tilde{\mathcal{T}}_a^{\mathrm{w}+,\times}(\theta, \phi; f)$ of the TDI channels. In our case, the transfer function can be used to map GW amplitudes to TDI Doppler outputs in the frequency domain.

(4) Multiply the transfer functions of different channels and average the product over many sky directions and polarizations to obtain the ORF.

In summary, the ORF for a TDI network is defined by

$$\gamma_{ab}(f) = \left\langle \sum_{I=+,\times} \mathrm{Re}\{\tilde{\mathcal{T}}_a^{\mathrm{w}I}(f, \theta, \varphi)[\tilde{\mathcal{T}}_b^{\mathrm{w}I}(f, \theta, \varphi)]^*\}\right\rangle_{\mathrm{s.a.}}. \qquad (23)$$

In our case the phase factor $e^{2\pi i f \hat{\Omega} \Delta \vec{x}/c}$ arising from the time delay between the detectors is already included in the Doppler signal outputs of the TDI. We found that averaging over 200 random sky directions provides very accurate results. Values for the right ascension ϕ are drawn from a uniform distribution $\mathcal{U}(0, 2\pi)$, whereas values for the declination θ are obtained by calculating the arcsin of values drawn from a uniform distribution $\mathcal{U}(-1, 1)$, which entails an isotropic distribution of corresponding sky directions.

The set of curves displayed in Fig. 7 shows the ORFs between channels A, E, and T of the two collocated detectors in comparison to the channels' sky-averaged squared transfer functions (STF).

$$\gamma_{aa} = |T_a(f)|^2 = \left\langle \sum_{I=+,\times} |\tilde{\mathcal{T}}_a^{wI}(f, \theta, \varphi)|^2 \right\rangle_{\text{s.a.}}. \quad (24)$$

The response of the T channel lies well below the response of channels A, E at frequencies up to 2 Hz. More specifically, within the correlation band 0.1–0.4 Hz (see Sec. VII) of the BBO, the T channel response (i.e. expressed as STFs or ORF) is smaller by a factor $\sim 2 \times 10^5 (0.4 \text{ Hz}/f)^6$ than the response of the other two channels. This again is the reason why our correlation analysis will not include the T channel.

In this paper, the ORFs are defined as correlation functions between TDI Doppler channels which are—at low frequencies—proportional to the square of the second (e.g. A, E) or even higher derivative (e.g. T) of the GW ampli-
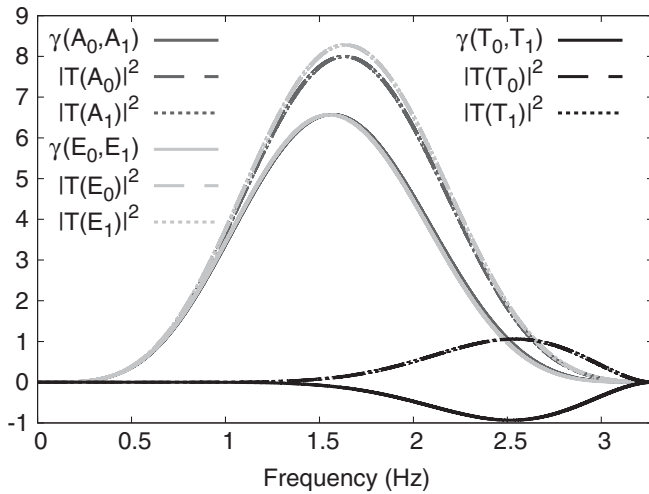
tude. To find a better measure of stochastic GW background correlations (as projection/combination onto two different TDI channels), one has to compare the ORF with the channel responses to GWs. In Fig. 8, the ORF between the two A channels of the collocated detectors is shown normalized by the geometrical mean of the two respective STFs:

$$\gamma_{A_0 A_1}^{\text{N}} \equiv \frac{\gamma_{A_0 A_1}}{\sqrt{\gamma_{A_0 A_0} \gamma_{A_1 A_1}}}. \quad (25)$$

At frequencies up to 0.5 Hz, the ORF and the STFs show identical response. This region is called the long-wavelength regime where the length of GWs is much larger than the dimension of the detectors. Beyond the long-wavelength limit, the ORF becomes weaker at higher frequencies compared to the STFs. As we expound in Sec. VII, correlation measurements with BBO detectors will be dominated by contributions of frequencies in the long-wavelength regime.

The graph in Fig. 9 allows one to draw another interesting and important conclusion. It shows the normalized ORF between channels A, E of the same detector. Obviously, cross correlating A channels of two different (collocated) detectors provides much more sensitivity than cross correlating independent channels of the same detector. In fact, one can show that if the detectors were equilateral triangles, then the ORF between channels A and E of the same detector would vanish [44,45]. Sky averages of

FIG. 7. The figure shows the overlap-reduction functions γ between channels A, E, T of the collocated detectors compared with the corresponding squared transfer functions. The sky average was taken over 400 random sky directions. The squared transfer functions of each channel type are identical in both detectors. At frequencies below 0.1 Hz, all displayed curves related to channels A, E are proportional to f^4, and the T channel curves are proportional to f^{10}. The response of the T channel to GWs is very poor below 2 Hz.
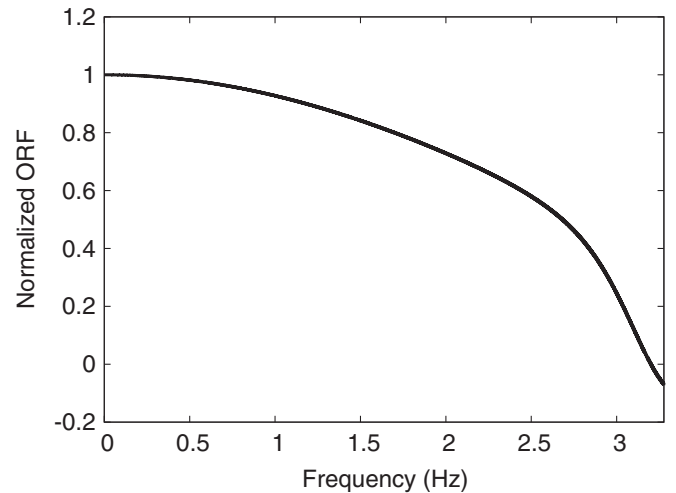

FIG. 8. The figure displays the normalized ORF between channels $A_0 \leftrightarrow A_1$ of the two collocated detectors. The normalized ORF has a maximum at $f = 0$ Hz and then oscillates around zero with constantly decreasing amplitude towards higher frequencies. The ORF as defined in Eq. (23) governs the correlation of TDI outputs. In contrast, the normalized ORF is a better representation of the correlation of a stochastic GW signal as input to the detectors. However, one has to keep in mind that all frequency-domain functions are obtained via FFT from a dynamical model, and therefore it is not possible in a simple way to deduce GW correlations from measured TDI outputs.
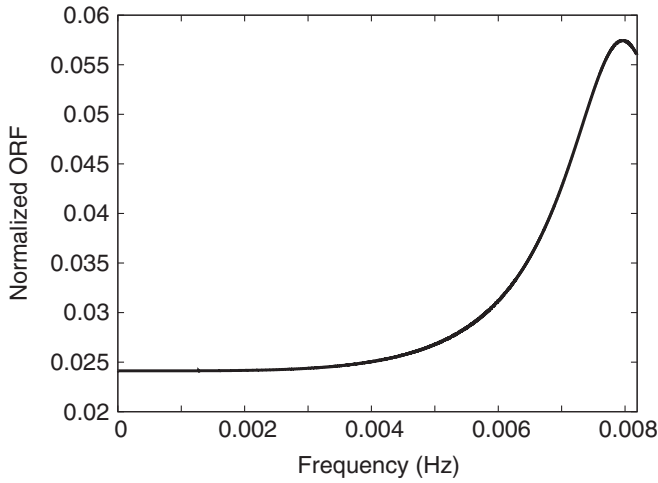
FIG. 9.   The figure shows the normalized ORF between channels $A_0 \leftrightarrow E_0$ of the same detector. Correlations between $A_0 \leftrightarrow E_0$ are much weaker than correlations between $A_0 \leftrightarrow A_1$. Still, the residual correlation could lead to an increased sensitivity of a single detector to isotropic stochastic backgrounds. Here, the data are based on a model detector with orbital eccentricity $e = 0.04$ to make the effect of a residual response due to asymmetries stronger than artificial anisotropies resulting from a sky average over a finite number of sky directions.

these two channels are orthogonal to each other in terms of their response to an isotropic stochastic GW background, but they do permit nonvanishing correlations arising from higher moments of anisotropic backgrounds (e.g. the hexadecapole moment). In our simulation, asymmetries in the triangular detector shape are responsible for a residual sky-averaged correlation strength. The detector asymmetry is of the order of the orbital eccentricity of the satellites. To make this effect stronger than artificial anisotropies resulting from a sky average over a finite number of sky directions, we chose a model detector with increased eccentricity value ($e = 0.04$) to generate Fig. 9. Since the BBO has collocated detectors, the residual correlation will not be exploited, because in that case one can form more efficient correlation schemes based on channels of different detectors. Also, when generating the CGWB data, we may neglect correlations between channels A and E which justifies the two-channel approach in Eq. (21) to the BBO network.

In terms of the calculated ORFs $\gamma_{A_0A_1}(f)$, $\gamma_{E_0E_1}(f)$ and STFs $\gamma_{A_0A_0}(f)$, $\gamma_{A_1A_1}(f)$, $\gamma_{E_0E_0}(f)$, and $\gamma_{E_1E_1}(f)$, we are able to calculate the stochastic background $A_{0,1}^{b}(f)$ and $E_{0,1}^{b}(f)$ by means of Eq. (21). In Fig. 10, its spectral density in channel A_0 is shown together with the instrumental-noise spectral density. One can directly infer from the graph that the SNR is about 0.1. A correlation measurement has to raise this value to at least 5. By simple arguments, we can determine the conditions under which the CGWB becomes detectable by means of correlation measurements. As a first approximation one can say that the
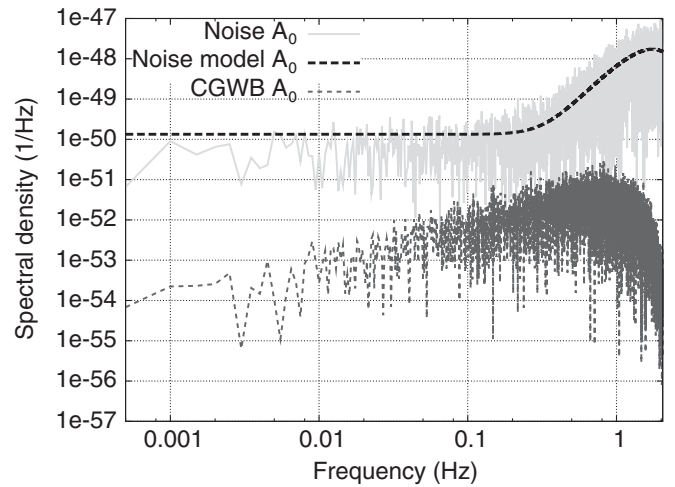

FIG. 10.   Simulated CGWB with fractional energy density $\Omega = 10^{-15}$ in comparison to the instrumental noise. A smaller observation time, $T = 2 \times 10^3$ s, was chosen to generate the curve, but the spectra shown are independent of the observation time. The noise model is exclusively used to compute Fisher matrices, projection operators, and the SNR (see next section).

correlation measurement effectively shifts the CGWB curve in Fig. 10 upwards by a factor $\sqrt{T}$. Adding the SNRs obtained from two independent correlation pairs, we find that a CGWB with energy density $\Omega = 10^{-15}$ can be detected if correlation times exceed 10^4 s. Section VII provides more details of the correlation measurement, and we are going to show in Sec. VIII that our first guess gives the right order of magnitude for the minimal correlation time.

VI. TEMPLATE-BASED PROJECTIONS OF SUBTRACTION NOISE

In this section we introduce a differential geometric point of view based, in principle, on the work of S.-I. Amari [32] but adapted to the needs of data analysis of gravitational-wave signals as is shown in [46]. The main focus is on presenting methods to deal with the inevitably occurring errors when subtracting the best-fit waveforms from the data stream. The residual errors have comparable spectral densities to the instrumental noise, because, roughly speaking, the subtraction of the best fit reduces the signal spectral density by a factor of $1/\text{SNR}^2$. This is true for broadband and narrow-band signals. Therefore, residual errors are too large to allow a measurement of an inflation generated background of gravitational waves. Fortunately, though arising from the presence of instrumental noise, the subtraction errors are not completely random but mostly confined to the tangent space of the template manifold at the point of the best fit. This restriction can be used to define a projection operator on the tangent space that cancels out all parts tangential to the manifold and hence most parts of the residual error [12].

In Sec. VI A, we give a short introduction to matched filtering and briefly point out the connection to differential geometry. Section VI B shows the derivation of the first-order approximation of the maximum likelihood estimator in the case of a high signal-to-noise ratio, which provides important formulas and justifies the use of differential geometry. In the last section, we present the actual method of projection.

A. Matched filtering and differential geometry

Most of the NS/NS signals observed by the BBO will have amplitudes roughly 2 orders of magnitude smaller than the amplitude of the instrumental noise. Therefore some technique of filtering must be used to extract the information from the noisy data. Post-Newtonian expansions up to order 3.5 of the equations of motions of stellar objects, such as compact-binary systems, yield very accurate waveforms throughout the BBO detection band, which can be employed to search for CBCs in the data streams. The fact that the shape of the signals is known to high accuracy is the reason why one can use *matched filtering*, which is also known as *optimal filtering* since it provides the highest SNR of all linear filters [47].

The detector outputs can, in two ways, be regarded as a vector. The first way which will be introduced in this section helps us find an expression for the SNR. In this case the detector output is a vector whose components are the outputs of the different detectors in different TDI channels. For LISA-type detectors the output vector could be $\vec{\mathcal{T}}^{s}(t) = (\mathcal{T}^{s}_{A}(t), \mathcal{T}^{s}_{E}(t), \mathcal{T}^{s}_{T}(t))$, which is the one-detector case, and for the BBO the vector has 12 components, respectively. In the case of a successful detection of a GW, the detector output is the sum of the signal $\mathcal{T}^{h}_{i}(t)$ and additive, stationary, Gaussian noise $\mathcal{T}^{n}_{i}(t)$, which are vectors in the same manner as described above. Optimal filtering of the data stream now means folding the output of the detector with a filter function $\vec{k}(t)$, and normalizing it with the correlation of the instrumental noise. Henceforth, for ease of notation, we will drop the tilde " ˜ " over frequency-domain functions and distinguish between time- and frequency-domain functions by means of their arguments. Then, the multichannel SNR can be defined in terms of scalar products between channel vectors:

$$\text{SNR} = \frac{\int \mathrm{d}f \vec{k}(f) \cdot \vec{\mathcal{T}}^{s}(f)}{\text{rms} \int \mathrm{d}f \vec{k}(f) \cdot \vec{\mathcal{T}}^{n}(f)}. \qquad (26)$$

The optimal filter function $\vec{k}(f)$ can be cast into the form

$$\vec{k}(f) = [\vec{\mathcal{T}}^{m}(\vec{\lambda}, f)]^{\dagger} (S^{n})^{-1}(f) \qquad (27)$$

where $\vec{\mathcal{T}}^{m}(\vec{\lambda}, f)$ is the Fourier transform of the GW signal parametrized by a set of parameters denoted by $\vec{\lambda}$, and the dagger is the conjugate transpose. $S^{n}(f)$ is the Fourier transform of the noise covariance matrix and its diagonal

elements are the (double-sided) power spectral densities of the noise in the corresponding TDI channel. Hence it follows that

$$\text{SNR}(\vec{\lambda}) = \frac{\int \mathrm{d}f [\vec{\mathcal{T}}^{m}(\vec{\lambda}, f)]^{\dagger} \cdot (S^{n})^{-1}(f) \cdot \vec{\mathcal{T}}^{s}(f)}{\text{rms} \int \mathrm{d}f [\vec{\mathcal{T}}^{m}(\vec{\lambda}, f)]^{\dagger} \cdot (S^{n})^{-1}(f) \cdot \vec{\mathcal{T}}^{n}(f)}, \qquad (28)$$

where $\vec{\mathcal{T}}^{s}(f)$ is the Fourier transform of the detector output, and $\vec{\mathcal{T}}^{n}(f)$ the noise in Fourier space.

In the following we assume, as in the previous sections, that the noise of different detectors and channels is uncorrelated, and that we use the optimal TDI configuration. The optimal channels, typically called *A*, *E*, and *T* [48], are all statistically independent and, hence, have uncorrelated noise contributions. In this case the correlation matrix of the noise is diagonal, and with a new optimal filter function $k'_{i}(\vec{\lambda}, f) = [\mathcal{T}^{m}_{i}(\vec{\lambda}, f)]^{*}/S^{n}_{i}(f)$, the SNR can be rewritten in the form

$$\text{SNR}(\vec{\lambda}) = \frac{\sum_{i} \int \mathrm{d}f k'_{i}(\vec{\lambda}, f) \mathcal{T}^{s}_{i}(f)}{\text{rms} \sum_{i} \int \mathrm{d}f k'_{i}(\vec{\lambda}, f) \mathcal{T}^{n}_{i}(f)}. \qquad (29)$$

Here $S^{n}_{i}(f)$ is the noise spectral density in the ith optimal TDI channel. The index i runs over all detectors and channels.

However, at this point it is advantageous to regard the data of each channel $\mathcal{T}^{s}_{i}$ as vectors themselves. Since a detector will sample data at a fixed frequency f_{s} (≈ 10 Hz for the BBO), each data stream will comprise $N = f_{s} \cdot T$ measuring points, where T is the total observation time (typically 10^{8} s, which is the BBO's lifetime). Each measuring point at time $t_{k} = k/f_{s}$ can be seen as a component of an N dimensional vector $\mathcal{T}^{s}_{i} = (\mathcal{T}^{s}_{i}(t_{1}), \mathcal{T}^{s}_{i}(t_{2}), \dots, \mathcal{T}^{s}_{i}(T))$ in the vector space $\mathcal{V}_{i}$ of all detector outputs of channel i. In the first place, it is this definition of a data vector which underlies the geometrical interpretation presented in the following paragraphs, not necessarily the gathering of detector outputs into a channel vector. The outputs of all channels form a $12N$-dimensional vector space $\mathcal{V}$ which formally can be thought of as a direct sum of the $\mathcal{V}_{i}$. The instrumental noise $\mathcal{T}^{n}_{i}$ and the gravitational-wave signal $\mathcal{T}^{h}_{i}$ are vectors in the same manner. Because of the fact that each binary signal is described by a set of 9 parameters (neglecting spin and eccentricity), the complete signal formed by about 10^{4}–10^{5} NS/NS with redshifts $z < 8$ will be parametrized by $N_{P} = 10^{5}$–10^{6} parameters and hence will lie on a sub-manifold $\mathcal{M}$ in $\mathcal{V}$ with dimension N_{P}.

In case of stationary, Gaussian, instrumental noise, the matched filter induces an inner product on $\mathcal{V}$ that is defined by

$$\langle \vec{g} | \vec{h} \rangle \equiv \sum_i \int_0^\infty \mathrm{d}f \frac{g_i^*(f) h_i(f) + g_i(f) h_i^*(f)}{S_i^{\mathrm{n}}(f)}. \quad (30)$$

It is straightforward to show that the ensemble average of $\langle \vec{g} | \vec{\mathcal{T}}^{\mathrm{n}} \rangle \langle \vec{\mathcal{T}}^{\mathrm{n}} | \vec{h} \rangle$ is equal to $\langle \vec{g} | \vec{h} \rangle$ for an ensemble of realizations of instrumental noise, which can be proved by using $\overline{\mathcal{T}^{\mathrm{n}}(f) \mathcal{T}^{\mathrm{n}}(f')} = \delta(f - f') S^{\mathrm{n}}(f)$ (see also [12,46]). In terms of this inner product, the optimal signal-to-noise ratio, with the help of the above property, can easily be written as

$$\mathrm{SNR}\,(\vec{\lambda}) = \langle \vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda}) | \vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda}) \rangle^{1/2}. \quad (31)$$

The inner product defined in Eq. (30) has the same features as the scalar product in Euclidean vector spaces; i.e. it is positive definite, and can therefore be used as a measure of distance and angles within the vector space. The length in Euclidean space corresponds to the total SNR Eq. (31) collected by all channels. Angles quantify the correlation of two outputs, e.g. two outputs are orthogonal if their correlation vanishes. Thus the definition of an inner product enables one to establish a geometrical description of the problem of filtering.

In the high SNR limit, the best-fit parameters are found by maximizing the likelihood function, see Eq. (36), which is equivalent to minimizing the inner product $\langle \vec{\mathcal{T}}^{\mathrm{s}} - \vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda}) | \vec{\mathcal{T}}^{\mathrm{s}} - \vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda}) \rangle$, which can be considered as the distance of the detector output to the template manifold. Hence the best-fit template waveform is the one which has the least separation from the detector output. In other words, geometrically, the best fit corresponds to the projection of the output $\vec{\mathcal{T}}^{\mathrm{s}}$ onto the submanifold $\mathcal{M}$; for more details see Secs. VI B and VI C.

To finish this section it should be said that generally the template manifold is not flat but rather has a curvature varying with the values of the parameters of the true signal. Like in general relativity, this can be addressed by introducing a metric on the manifold. One possible choice for the metric is the covariance matrix of the parameter errors, which also is the inverse of the Fisher-information matrix (see Sec. II), defined by

$$\Gamma_{\alpha\beta} = \left\langle \frac{\partial \vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda})}{\partial \lambda^\alpha} \middle| \frac{\partial \vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda})}{\partial \lambda^\beta} \right\rangle. \quad (32)$$

This statement will not be proven in this paper but the interested reader shall be referred to [46].

B. Maximum likelihood estimator in the high signal-to-noise ratio limit

In this section we briefly derive the deviation of the best-fit parameters $\hat{\vec{\lambda}}$ from the true ones $\vec{\lambda}_0$ in the limit of high SNR, assuming that the signal model is accurate. The outcome will give the desired results in terms of the geometrical quantities introduced in the previous section. To obtain the results, we use the Bayesian estimator and write the exponent of the posterior distribution as a Taylor series around the best-fit parameters. For more information and different ways of deriving the parameter errors, see [38].

First consider the Bayesian estimator of the parameter error which we name $\vec{\epsilon} = \vec{\lambda} - \vec{\lambda}_0$ and which is defined by

$$\langle \epsilon^\alpha \rangle = \int \mathrm{d}^{N_{\mathrm{P}}} \epsilon \, \epsilon^\alpha p(\vec{\epsilon} | \vec{\mathcal{T}}^{\mathrm{s}}). \quad (33)$$

The function $p(\vec{\epsilon} | \vec{\mathcal{T}}^{\mathrm{s}})$ is the *posterior* distribution which determines the probability of a model determined by parameter-value deviations $\vec{\epsilon}$ from the true signal, given a measurement $\vec{\mathcal{T}}^{\mathrm{s}}$. The connection between the posterior and the likelihood function is the following:

$$p(\vec{\epsilon} | \vec{\mathcal{T}}^{\mathrm{s}}) = \mathcal{N} p_0(\vec{\epsilon}) p(\vec{\mathcal{T}}^{\mathrm{s}} | \vec{\epsilon}), \quad (34)$$

where $\mathcal{N}$ is a normalization constant, $p_0(\vec{\epsilon})$ comprises the *prior* knowledge of parameter values, and $p(\vec{\mathcal{T}}^{\mathrm{s}} | \vec{\epsilon})$ is the likelihood of data $\vec{\mathcal{T}}^{\mathrm{s}}$ given a model $\vec{\epsilon}$. At the moment, only flat *a priori* probabilities are put into our simulation, which means $p_0(\vec{\epsilon}) \equiv 1$, and the estimator can be written as

$$\langle \epsilon^\alpha \rangle = \mathcal{N} \int \mathrm{d}^{N_{\mathrm{P}}} \epsilon \, \epsilon^\alpha p(\vec{\mathcal{T}}^{\mathrm{s}} | \vec{\epsilon}), \quad (35)$$

depending only on the likelihood function. In this case the normalization constant is the inverse of the integral over the likelihood over the whole parameter space.

As shown in [36], the likelihood can be expressed with the help of Eq. (30) as

$$p(\vec{\mathcal{T}}^{\mathrm{s}} | \vec{\lambda}) \propto \exp(-\tfrac{1}{2} \langle \vec{\mathcal{T}}^{\mathrm{s}} - \vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda}) | \vec{\mathcal{T}}^{\mathrm{s}} - \vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda}) \rangle). \quad (36)$$

Because of the fact that the data are the sum of instrumental noise and a signal, one can rewrite the argument within the brackets of Eq. (36) as $\vec{\mathcal{T}}^{\mathrm{n}} + \delta\vec{\mathcal{T}}^{\mathrm{m}}$, where $\delta\vec{\mathcal{T}}^{\mathrm{m}} = \vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda}_0) - \vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda})$. Here, $\vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda}_0)$ represents the true signal $\vec{\mathcal{T}}^{\mathrm{h}}$ in terms of the accurate model. Now, the difference in the waveforms can be expanded in a Taylor series as

$$- \delta\vec{\mathcal{T}}^{\mathrm{m}} = \partial_\alpha \vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda}_0) \epsilon^\alpha + \tfrac{1}{2} \partial_\alpha \partial_\beta \vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda}_0) \epsilon^\alpha \epsilon^\beta + \mathcal{O}(\epsilon^3). \quad (37)$$

Inserting Eq. (37) into Eq. (36) and introducing normalized waveforms $\vec{\mathcal{N}}^{\mathrm{m}} \equiv \vec{\mathcal{T}}^{\mathrm{m}}/A$ and $\vec{\mathcal{N}}_\alpha^{\mathrm{m}} \equiv \partial_\alpha \vec{\mathcal{T}}^{\mathrm{m}}/A$ with $A = \langle \vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda}_0) | \vec{\mathcal{T}}^{\mathrm{m}}(\vec{\lambda}_0) \rangle^{1/2}$, which helps to highlight the dependence of the equations on the SNR, yields the following approximation of the likelihood:

$$p(\vec{\mathcal{T}}^s|\vec{\epsilon}) \propto \exp\left(-\frac{1}{2}\left[\langle\vec{\mathcal{T}}^n|\vec{\mathcal{T}}^n\rangle - 2A\langle\vec{\mathcal{T}}^n|\vec{\mathcal{N}}^m_\alpha(\vec{\lambda}_0)\rangle\epsilon^\alpha\right.\right.$$
$$+ A^2\left(\langle\vec{\mathcal{N}}^m_\alpha|\vec{\mathcal{N}}^m_\beta\rangle - \frac{1}{A}\langle\vec{\mathcal{T}}^n|\vec{\mathcal{N}}^m_{\alpha\beta}\rangle\right)$$
$$\left.\left.\times\,\epsilon^\alpha\epsilon^\beta + \mathcal{O}(\epsilon^3)\right]\right). \tag{38}$$

The expansion can be cut off at order ϵ^3 because higher order corrections in the exponent would correspond to higher order corrections to the estimator and we are only interested in the first-order terms. One sees that for high SNR the likelihood is approximated by a multivariate normal distribution. The first summand is just a constant and will be absorbed into the normalization constant $\mathcal{N}$, the second term shifts away the maximum of the distribution from the true parameters due to instrumental noise, whereas the third term contains, in round brackets, the inverse of the covariance matrix of the errors and mainly determines the width of the distribution. The correlation of the noise with the second derivatives of the signal gives a first correction to the Fisher matrix as the inverse of the covariance, but as will be clear from Eq. (39), one can neglect this correction for high SNR since it scales with $1/A$ compared to the Fisher matrix.

With all that at hand, one can compute the Bayesian estimator by solving the integral over the likelihood, which is a lengthy but straightforward calculation. Here we present just the result,

$$\langle\epsilon^\alpha\rangle = \frac{1}{A}\left[\langle\vec{\mathcal{N}}^m_\alpha|\vec{\mathcal{N}}^m_\beta\rangle - \frac{1}{A}\langle\vec{\mathcal{T}}^n|\vec{\mathcal{N}}^m_{\alpha\beta}\rangle\right]^{-1}\langle\vec{\mathcal{T}}^n|\vec{\mathcal{N}}^m_\beta\rangle$$
$$\approx \Gamma^{\alpha\beta}\langle\vec{\mathcal{T}}^n|\vec{\mathcal{T}}^m_\beta\rangle. \tag{39}$$

This equation reveals that, as promised, the deviations from the true parameters are completely determined by the template-manifold metric and the length of the projection of the noise vector onto the tangent space of $\mathcal{M}$ at the point of the best fit. Also, the parameter errors decrease with 1/SNR. We stress again that this result is obtained as a first-order SNR approximation, which is supposed to be sufficient for the BBO as the expected SNRs are high enough to justify this approach. Anyway, higher order expansions and the influence of *prior* information on the estimator can be looked up in [37,38].

C. The projection operator

In the last two sections, we outlined a strong connection between differential geometry and methods used in data analysis of gravitational waves from compact-binary objects such as neutron-star neutron-star binaries. Investigations of matched filtering of data led to a definition of a scalar product on the vector space $\mathcal{V}$ of detector outputs, and modeling the waveforms by post-Newtonian templates smoothly depending on a set of parameters describing the physical properties of the binaries, the de-

tector as well as their relative motion, confines the possible outputs generated by a gravitational wave to a submanifold $\mathcal{M}$ within $\mathcal{V}$. The errors occurring at the estimation of the signal parameters are then completely given in terms of geometrical quantities such as projections onto and within the tangent space of $\mathcal{M}$ at the best-fit parameters.

In this section, we further exploit the geometrical restrictions on the waveforms and present a successful implementation of a projection method to cancel the subtraction noise $\delta\vec{\mathcal{T}}^m$ occurring by subtracting the best fit from the data stream. Equation (39) showed that the expected parameter errors are proportional to 1/SNR, which can be used to see how the amplitude of the subtraction noise depends on the SNR. The following equation provides a Taylor expansion of the template waveform around the true signal $\vec{\mathcal{T}}^m(\vec{\lambda}_0)$ evaluated at the expected error:

$$\vec{\mathcal{T}}^m(\langle\vec{\epsilon}\rangle) = \vec{\mathcal{T}}^m(\vec{\lambda}_0) + \vec{\mathcal{T}}^m_\alpha(\vec{\lambda}_0)\langle\epsilon^\alpha\rangle + \vec{\mathcal{T}}^m_{\alpha\beta}(\vec{\lambda}_0)\langle\epsilon^\alpha\rangle\langle\epsilon^\beta\rangle$$
$$+ \mathcal{O}(\mathrm{SNR}^{-2}). \tag{40}$$

Making use of the fact that $\vec{\mathcal{T}}^m(\vec{\lambda}_0)$ and all derivatives of $\vec{\mathcal{T}}^m$ depend linearly on the SNR, one finds that the first term in this expansion is proportional to SNR; the second, which is the leading term of the subtraction noise, is independent of the SNR, which can be seen, too, by computing the mean norm of the subtraction noise to leading order in the SNR,

$$\overline{\langle\delta\vec{\mathcal{T}}^m(\vec{\epsilon})|\delta\vec{\mathcal{T}}^m(\vec{\epsilon})\rangle} = \langle\vec{\mathcal{T}}^m_\alpha(\vec{\lambda}_0)|\vec{\mathcal{T}}^m_\beta(\vec{\lambda}_0)\rangle\overline{\epsilon^\alpha\epsilon^\beta} = N_\mathrm{P}. \tag{41}$$

Again, N_P is the number of parameters describing the total signal or in other words, the dimension of the template manifold. The mean spectral density of the subtraction noise in each channel is most suitably expressed in terms of the ratio $S(\delta\mathcal{T}^m_i; f)/S^n(f)$ of spectral densities of the subtraction and the instrumental noise. In the simplest possible case, we could consider a signal with a constant ratio $S(\delta\mathcal{T}^m_i; f)/S^n(f)$ for each frequency within a given signal bandwidth and negligible ratios outside the bandwidth. The integral which yields the scalar product in Eq. (41) is converted into a sum of these ratios over all frequency bins at $f = 1/T, \ldots, f_s/2$ within the signal bandwidth. For the BBO which furnishes data from 8 channels that are sensitive to GWs and which permits a total observation time of $T = 10^8$ s, and assuming a total signal of 10^5 CBC s($N_\mathrm{P} \sim 10^6$) is contained within a bandwidth of 1 Hz, this would lead to $S(\delta\mathcal{T}^m_i; f)/S^n(f) = 0.5N_\mathrm{P}/(8 \cdot 1\,\mathrm{Hz} \cdot T) \sim 10^{-3}$. A cosmic gravitational-wave background with energy density $\Omega \gtrsim 10^{-17}$ would have a spectral density which is more than 1 order of magnitude less than the weakest possible subtraction-noise level. This level is based on an optimal search of the CBCs

which cannot be performed even if a steady development of computational facilities over two or three decades in accordance with Moore's law is assumed. So the subtraction noise, if remaining within the data, would prohibit the detection of an inflation-generated background with high certainty.

Equation (40) also shows that the third term is proportional to $1/\mathrm{SNR}$. So in deleting the zeroth- and first-order terms from the data stream, one reduces the signal strength by a factor of $1/\mathrm{SNR}^2$. The first-order term is a linear combination of first derivatives of the signal or, in other words, a vector lying in the tangent space of $\mathcal{M}$ at the true parameter values. Since the expected parameter errors scale with $1/\mathrm{SNR}$, it follows that the two tangent spaces at the best fit and the true signal can be regarded as nearly identical. So from now on, all derivatives are taken at the best-fit parameters which are obtained by searching the data for signals and estimating model parameters. The leading term of the subtraction noise is taken as a vector in the tangent space at the best fit.

Figure 11 schematically shows the template manifold $\mathcal{M}_i$ of signals contained in the data stream of channel i, the true signal $\mathcal{T}_i^{\mathrm{h}}$, the best fit $\mathcal{T}_i^{\mathrm{m}}(\hat{\vec{\lambda}})$, and the parallel and perpendicular parts of the instrumental noise. In a specific TDI channel, the projection of instrumental noise generally does not coincide with the difference between the best-fit template and the true signal, as it is the case in the complete vector space $\mathcal{V}$. That is due to the fact that by using information from all channels to find the best fit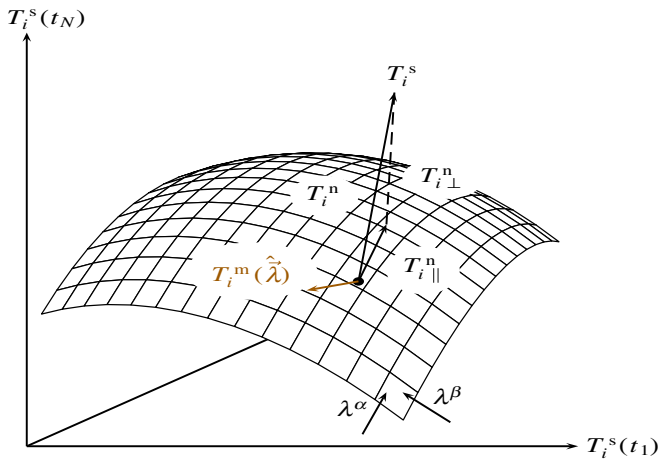, the estimate will be somewhat better than if one had determined the parameters just with one data stream. It is important to note that, because the best fit subtracted from the data in each channel is determined by the same best-fit parameters, the subtraction noise will be correlated in all channels and will not be deleted by a cross-correlation measurement between channels; see Sec. VII. But the good news is that most of the correlated error is restricted to the tangent space of the manifolds $\mathcal{M}_i$ at the best fit, and projection out tangential directions of the residual data will delete the correlations. The tangential part of the CGWB which gets projected out is negligible [12].

To perform this task, one can define a projection operating on a specific channel i, which removes the tangential parts:

$$\mathsf{P}_i = \mathbb{1} - \Gamma_i^{\alpha\beta}|\partial_\alpha \mathcal{T}_i^{\mathrm{m}}\rangle\langle\partial_\beta \mathcal{T}_i^{\mathrm{m}}| \qquad (42)$$

where $\Gamma_i^{\alpha\beta}$ denotes the inverse of the Fisher matrix $\Gamma_{\alpha\beta}^i = \langle\partial_\alpha \mathcal{T}_i^{\mathrm{m}}(\hat{\vec{\lambda}})|\partial_\beta \mathcal{T}_i^{\mathrm{m}}(\hat{\vec{\lambda}})\rangle$. The projected data stream can be calculated, e.g. in the frequency domain, according to

$$\mathsf{P}_i \mathcal{T}_i^{\mathrm{res}}(f) = \mathcal{T}_i^{\mathrm{res}}(f) - \Gamma_i^{\alpha\beta}\langle\partial_\beta \mathcal{T}_i^{\mathrm{m}}|\mathcal{T}_i^{\mathrm{res}}\rangle\partial_\alpha \mathcal{T}_i^{\mathrm{m}}(f). \qquad (43)$$

Here, $\mathcal{T}_i^{\mathrm{res}}$ denotes the residual data which remains after subtraction of the best fits. If the initial data stream contains a GW foreground $\mathcal{T}_i^{\mathrm{h}}$ generated by CBCs, a CGWB $\mathcal{T}_i^{\mathrm{b}}$, and instrumental noise $\mathcal{T}_i^{\mathrm{n}}$, then after subtracting the best fit and projecting the data, what is left is

$$\mathsf{P}(\mathcal{T}_i^{\mathrm{res}}) = \mathcal{T}_{i\perp}^{\mathrm{n}} + \mathcal{T}_{i\perp}^{\mathrm{b}} + \mathcal{O}(\mathcal{T}_i^{\mathrm{h}}/\mathrm{SNR}^2), \qquad (44)$$

the perpendicular parts of the instrumental noise and the cosmic GW background. The loss of power in $\mathcal{T}_i^{\mathrm{b}}$ is negligible as already mentioned, and the remnants of the instrumental noise will be removed by cross correlating the data of the two collocated detectors in a Star-of-David configuration.

We want to point out another interesting property of the subtraction-noise projection. For arbitrary template manifolds, the best fit has to lie close to the true signal. If that is not the case, then tangential planes at the best fit and the true signal do not coincide. This could entail a poor performance of the projection operator Eq. (42) in removing residual power of the signal spectrum. Now, imagine a flat template manifold. In that case, the best fit does not have to be accurate, since tangential planes coincide at all points on the manifold. In fact, since the vanishing signal is always an element of the model (e.g. coming from a source which is far away), the projection method works without having to estimate the signals at all. The only information which is needed is the number of signals. This information determines the right dimension of the manifold. Although a completely flat metric does not represent most analysis problems, we can already see in (signal → subtraction noise → projected signal) spectra shown in

FIG. 11 (color online). The figure represents the template manifold $\mathcal{M}_i$ of a specific TDI channel i. Since the best-fit parameters $\hat{\vec{\lambda}}$ are estimated with information from all channels, the projection of data $\mathcal{T}_i^{\mathrm{s}}$ in channel i does not yield the best-fit waveform $\mathcal{T}_i^{\mathrm{m}}(\hat{\vec{\lambda}})$. It is assumed that the true signal lies on the template manifold at the marked point. The noise vector $\mathcal{T}_i^{\mathrm{n}}$, which points from the true signal to the measured data, is split into its components parallel and perpendicular to the manifold at the true signal. The (tiny) vector of the CGWB is not shown.

Sec. VIII A that the power of the projected spectrum is very low at all frequencies independent of the power of the subtraction noise. Obviously in our simulation, some parameters of some signals were poorly estimated, but their subtraction noise is removed with high accuracy. This can just be explained with a sort of "partial" flatness of the template manifold. Certainly, this feature needs to be investigated in the future.

VII. CROSS CORRELATION OF TDI CHANNELS

In Sec. V B we have seen that the CGWB is completely covered by the detector noise. One needs to reduce the difference between noise and background spectral densities by a factor of 10^2 in the case of a CGWB with $\Omega = 10^{-15}$. In this section we will describe how that reduction can be done by cross correlating the TDI streams of the collocated detectors. Increasing the SNR of the background is achieved by increasing the observation time until the correlation output is dominated by contributions from the CGWB. One can study the correlation measurement for small observation times T and then extrapolate the output for higher T by making use of a simple scaling law of the SNR with observation time. We will see that performing a correlation measurement with data gathered over 3 years, which is the BBO's proposed mission lifetime, it is possible to detect a CGWB with energy densities below $\Omega = 10^{-16}$. These results are based on the assumption that the subtraction noise from the CBC foreground can be removed with sufficient accuracy. Our results, which are presented in Sec. VIII, show that this is indeed the case (at least for 100 NS/NS).

We consider the TDI channel output $\mathcal{T}_i^s(f)$ as a sum of the CGWB and of the detector noise,

$$\mathcal{T}_i^s(f) = \mathcal{T}_i^b(f) + \mathcal{T}_i^n(f), \tag{45}$$

where the noise is assumed to be Gaussian, stationary, and uncorrelated between different channels, and the foreground signal is subtracted below the CGWB. Unlike the instrumental noise, the CGWB will be correlated in different channels to some degree, which can be predicted by the ORF. In the frequency domain, the expected outcome of a correlation C_{ij} between channels i, j is an integral over all frequencies of the data-stream product

$$C_{ij} = \int df \mathrm{Re}\{\mathcal{T}_i^s(f)[\mathcal{T}_j^s(f)]^*\}Q_{ij}(f)$$
$$= \sum_f \frac{\mathrm{Re}\{\mathcal{T}_i^s(f)[\mathcal{T}_j^s(f)]^*\}}{T} Q_{ij}(f) \tag{46}$$

where a channel-dependent filter function Q_{ij} is used to suppress contributions from frequencies with strong instrumental noise or weak (expected) CGWB. According to Eq. (19), the expectation value $\langle C_{ij} \rangle$ of the correlation

measurement for sufficiently long observation times is

$$\langle C_{ij} \rangle = \sum_f \gamma_{ij}(f)S^b(f)Q_{ij}(f) \tag{47}$$

where S^b is the GW strain spectral density. The variance of correlation noise, which is dominated by contributions from the instrumental noise, is given by

$$\langle (\Delta C_{ij})^2 \rangle = \sum_f S_i^n(f)S_j^n(f)Q_{ij}^2(f). \tag{48}$$

Now, we can understand how, in general, the correlation signal-to-noise ratio $\mathrm{SNR}_{ij} = \langle C_{ij} \rangle / \sqrt{\langle (\Delta C_{ij})^2 \rangle}$ scales with observation time T. The number of frequencies (frequency bins) summed over in Eq. (47) increases proportional to T, whereas the standard deviation, which is the square root of Eq. (48), scales with $\sqrt{T}$. Therefore, increasing the observation time, one eventually raises contributions from the CGWB above the expected deviations. If the range of frequencies Δf contained in the sums is small enough, then the functions within the summands can be taken as constants and the two equations, evaluated at a fiducial frequency f_0 which lies within the bandwidth, become

$$\langle C_{ij} \rangle \approx (T \cdot \Delta f)\gamma_{ij}(f_0)S^b(f_0)Q_{ij}(f_0)\sqrt{\langle (\Delta C_{ij})^2 \rangle}$$
$$\approx \sqrt{T \cdot \Delta f}\sqrt{S_i^n(f_0)S_j^n(f_0)}Q_{ij}(f_0). \tag{49}$$

So, in this small-bandwidth approximation, the SNR is independent of the (constant) filter function Q_{ij},

$$\mathrm{SNR}_{ij} = \sqrt{T \cdot \Delta f} \frac{\gamma_{ij}(f_0)S^b(f_0)}{\sqrt{S_i^n(f_0)S_j^n(f_0)}}. \tag{50}$$

If the small-bandwidth, flat spectrum approximation does not hold, then there exists an optimal filter which is based on models for the noise spectral densities in the two channels, the ORF and the spectral density of the stochastic background. Its purpose is to suppress contributions from frequencies which would contribute strongly to the instrumental noise in C_{ij}, but weakly to the GW correlations. Accordingly, the optimal filter is given by [14]

$$Q_{ij}(f) = \frac{\gamma_{ij}(f)S^b(f)}{S_i^n(f)S_j^n(f)}. \tag{51}$$

The filter function between channels A_0 and A_1 based on a flat $\Omega = 10^{-15}$ model for the CGWB is shown in Fig. 12. In addition, one should keep in mind that the WD/WD barrier enforces a lower boundary on correlation frequencies. The filter maximum lies at 0.2 Hz which is halfway
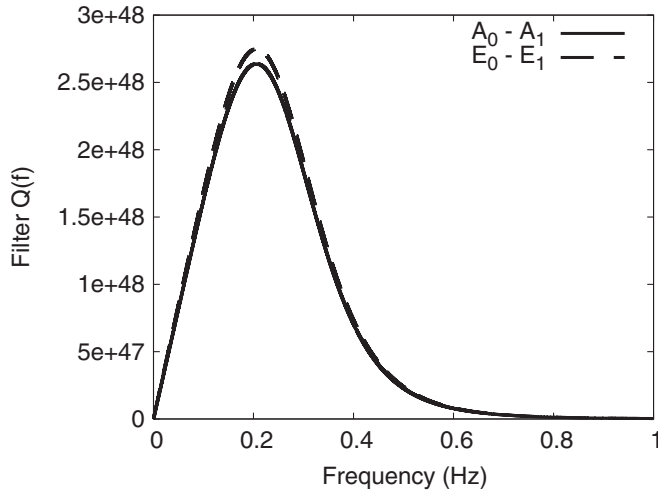
FIG. 12. The graph displays the filter functions $Q_{A_0 A_1}(f)$ and $Q_{E_0 E_1}(f)$. It contains models for the instrumental-noise spectral densities [given in Eq. (12)] and is based on a flat-Ω model of the CGWB.

inside the WD/WD spectrum. In fact, most models presented in [17] predict a cosmological distribution of WD/WD which gives rise to an energy density of $\Omega \sim 10^{-14}$–10^{-13} at 0.2 Hz, which would make detection of a cosmological background impossible at these frequencies unless the WD/WD signals were resolvable. Assuming that the WD/WD foreground at frequencies 0.2 Hz cannot be analyzed, one has to find out if the maximum of the filter does determine the most efficient correlation frequencies. This is not the case, but as we will see, efficient frequencies are not much greater than suggested by the filter. Optimally, one had to design the instrument such that the most efficient frequencies lie above the WD/WD barrier. To find a definite answer to this problem, one has to calculate the contribution from certain frequencies to the SNR expected from a specific model of the CGWB. The SNR for the optimal filter assumes the form

$$\mathrm{SNR}_{ij} = \left(\sum_f \frac{\gamma_{ij}^2(f)[S^{\mathrm{b}}(f)]^2}{S_i^{\mathrm{n}}(f) S_j^{\mathrm{n}}(f)} \right)^{1/2}. \qquad (52)$$

For a network of detectors, one would sum over all independent correlation pairs (ij) to obtain the total network SNR

$$\mathrm{SNR}_{\mathrm{tot}} = \left(\sum_{(ij)} \mathrm{SNR}_{ij}^2 \right)^{1/2} = \left(\sum_f \mathrm{SNR}_{\mathrm{tot}}^2(f) \right)^{1/2}$$

$$= \left(\sum_f \sum_{(ij)} \frac{\gamma_{ij}^2(f)[S^{\mathrm{b}}(f)]^2}{S_i^{\mathrm{n}}(f) S_j^{\mathrm{n}}(f)} \right)^{1/2}. \qquad (53)$$

Figure 13 shows $\mathrm{SNR}_{\mathrm{tot}}^2(f)$ including the two correlation pairs $A_0 \leftrightarrow A_1, E_0 \leftrightarrow E_1$ of the BBO. The maximum of this curve is shifted towards higher frequencies with respect to
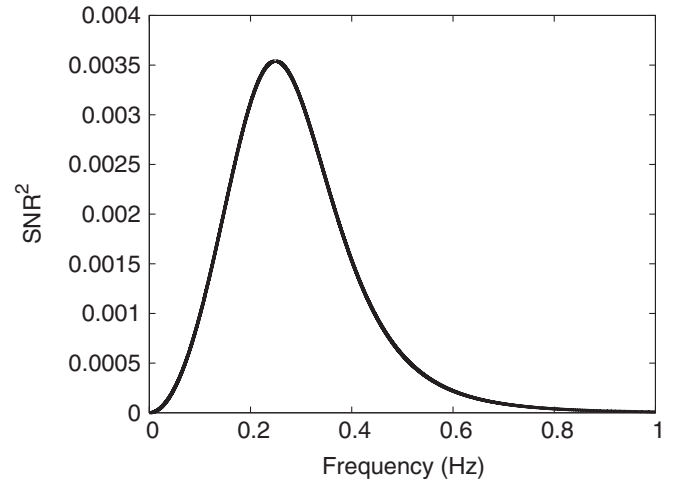

FIG. 13. The curve shows $\mathrm{SNR}_{\mathrm{tot}}^2(f)$ as defined in Eq. (53). It includes contributions from the two statistically independent pairs $A_0 \leftrightarrow A_1$ and $E_0 \leftrightarrow E_1$. This curve does not depend on the observation time T. The total SNR^2 can be calculated by adding values of $\mathrm{SNR}_{\mathrm{tot}}^2(f)$ for each frequency bin $f_i = i/T$ within a chosen correlation bandwidth (a lower boundary for this band is set by the WD/WD barrier; the upper boundary is ultimately set by half of the sampling frequency f_{s}). For example, neglecting the WD/WD barrier and approximating the area under the curve by a rectangle $(0.3 \text{ Hz}) \times (0.003)$, a $\mathrm{SNR}^2 = 25$ would be obtained after $T = 25/(0.3 \times 0.003)\mathrm{s} \approx 3 \times 10^4$ s.

the filter maximum, because the additional γ_{ij} brings in a factor f^4 at frequencies below 1 Hz and the additional background spectrum a factor f^{-3}. So, in total, the filter spectrum is multiplied by a factor f. Most of the SNR is collected at frequencies near 0.23 Hz which may still be a bit too low. Certainly, this issue needs to be investigated in the future.

VII. RESULTS

Our results are presented in two ways. In the first part of this section, we show subtraction noise and projected signal spectra and compare them in total power. In the second part, the outcome of correlation measurements between the two collocated detectors is summarized in Table II, and we compare contributions from instrumental noise, CGWB, and CBCs. Also, the decrease of CBC correlations by subtraction of best fits and noise projection is investigated. Sensitivity of the BBO to stochastic backgrounds is derived and extrapolated to an observation time of 3 years.

A. Projection

In this section, we focus on two salient features of the projection results, which are predicted by theory. First, the projected spectra are compared with subtraction-noise

spectra to find that the residual signal power is sufficiently small to enable CGWB detection within a certain frequency range. Second, we show that the projection operates selectively on subtraction noise and leaves stochastic signals like the instrumental noise and the CGWB more or less unaffected.

In Fig. 14, the CBC spectrum together with the subtraction noise and its projection are displayed from 0.1 Hz to 1 Hz. With a few exceptions, the subtraction noise is weaker than the signal spectrum. The projection is applied to 17 out of 100 signals to remove the subtraction noise between 0.2 Hz and 0.5 Hz. The peaks in the subtraction-noise spectrum are removed. The residual spectrum lies below a CGWB with $\Omega = 10^{-15}$ (see Fig. 10). Remarkably, the projection works accurately although some of the CBCs in the correlation band are poorly estimated, which can be seen by comparing the subtraction noise with the signal spectrum: the weaker the subtraction noise, the better the best fit. Simulating longer observation times would significantly improve the gain of the noise projection. We conclude that the subtraction noise will pose no problem for future-generation detectors as long as all foreground signals can be detected and modeled accurately.

Projecting the data means removing all contributions which are correlated with certain template derivatives. These derivatives are associated with directions in a sam-

pling space. A stochastic process has the property of distributing its energy randomly along all directions of the sampling space with a given mean value of the energy per direction, which is the geometrical interpretation of the fact that a stationary process has constant variance. If we project out N_p directions of the sampling space, then, on average, we will remove a fraction N_p/N of the power of any stochastic, stationary process, where N is the dimension of the sampling space which is the number of samples. As long as the number of projected directions is much smaller than N, the loss in power will be negligible. In our case, we project along 17×9 (17 CBCs with 9 parameters each) out of $10^5 \times 5.242\,88$ (observation time multiplied by sampling frequency) directions. The mean predicted power loss of stationary processes like the CGWB or the instrumental noise is 0.03%, which is indeed insignificant. In Fig. 15, we plot the spectrum of the total
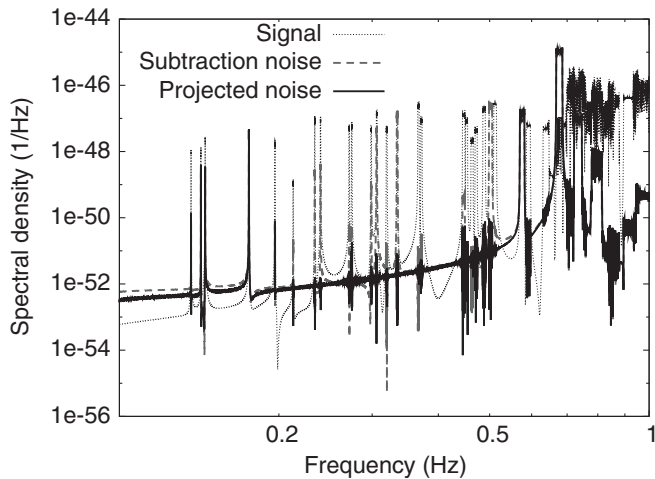

FIG. 14. The figure shows the recorded signal spectrum (dotted line), the subtraction-noise spectrum (dashed line), and the projected subtraction-noise spectrum (solid line) between 0.1 Hz and 1 Hz. The subtraction noise is comparatively high at low frequencies, since pN waveforms are determined by strongly correlated parameters. These correlations decay at higher frequencies where a considerable part of the phase evolution is observed, leading to better waveform estimates. Subtraction noise within the correlation band 0.2 Hz–0.5 Hz is projected. As one can see, all peaks of the subtraction noise are removed. The residual noise is negligible compared with a CGWB spectrum with fractional energy density $\Omega = 10^{-15}$.
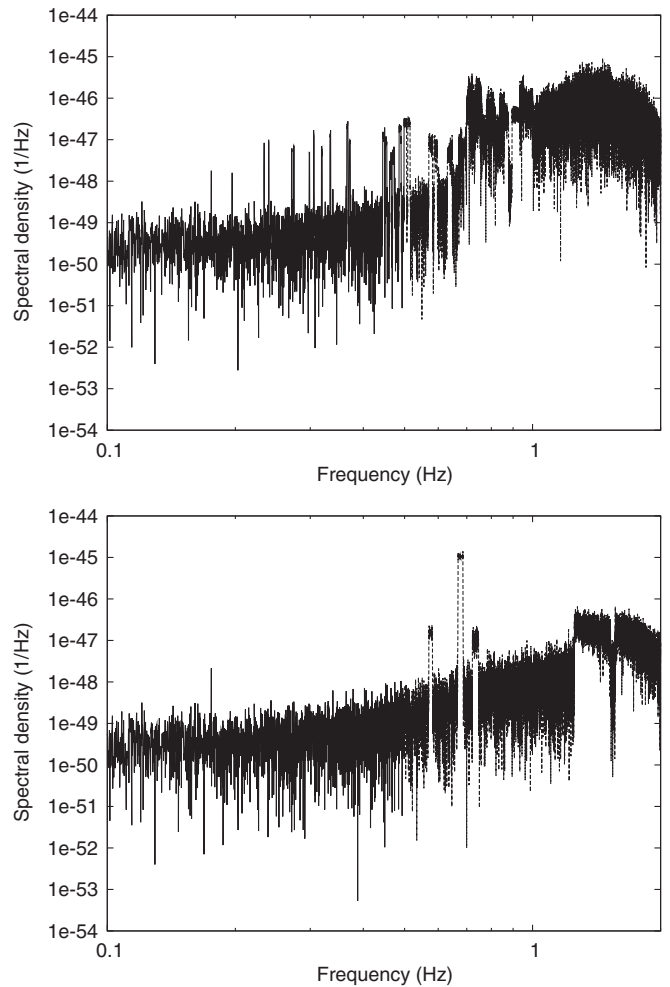

FIG. 15. The upper graph displays the spectrum of the total data before it is analyzed, whereas the lower graph displays the spectrum after best-fit subtraction and noise projection. Some of the signal peaks in Fig. 14 can be identified in these two spectra. There is no visible change of the instrumental-noise level. Here, every 20th frequency bin is plotted to reduce the amount of data.

data before (upper figure) and after (lower figure) data analysis, including the noise projection. There is no visible change in the instrumental-noise spectrum, but all signal peaks between 0.2 Hz and 0.5 Hz are removed. More quantitative results can be found in the next section which presents correlation values of all components of the total data.

B. Correlation

The correlation measurement is evaluated in two ways. At first, we compute statistics for different observation times to make it possible to extrapolate our results to a 3 yr observation time with certain confidence. These statistics are based on simulations without the CBC foreground. The reason is that the computation would take too long, and also, the projection performance is insufficient at observation times much shorter than $T = 10^5$ s because the Fisher matrices are highly ill conditioned and parameter correlations are too strong to allow a meaningful estimation of signal parameters by means of Eq. (13). Second, we investigate the correlation including the CBCs and noise projection for an observation time $T = 10^5$ s.

As mentioned in earlier sections, the correlation measurements are carried out between channels $A_0 \leftrightarrow A_1$ and $E_0 \leftrightarrow E_1$, and subsequently, the two values are added to form the total correlation SNR. Each correlation value corresponds to an integral over the correlation band 0.2 Hz–0.5 Hz. The results are shown in Table II. The first column contains values of four different observation times used to obtain the correlation statistics. The second column shows the respective correlation outcome of the instrumental noise, the third column shows that of the CGWB with $\Omega = 10^{-15}$, and the last column shows that of the total data, which, in this case, is the sum of the instrumental noise and the CGWB. Each correlation value is based on 20 measurements. In summary, the evolution of the mean

TABLE II. The mean values and standard deviations of correlation measurements between channels $A_0 \leftrightarrow A_1$ and $E_0 \leftrightarrow E_1$ for different observation times. Values in each row are based on 20 measurements. Here, the total data are the sum of the CGWB and the instrumental noise. Whereas the mean value of the correlation C_{tot} of the total data is dominated by contributions from the CGWB, its standard deviation is determined by contributions from the instrumental noise. The mean value of C_{tot} is supposed to increase linearly with T, whereas its standard deviation—predicted to be the square root of the mean value—is supposed to increase with $\sqrt{T}$.

T	C_{n}	C_{cgwb}	C_{tot}
1×10^4 s	0.00 ± 2.60	5.90 ± 0.07	5.90 ± 2.54
2×10^4 s	0.83 ± 5.74	14.72 ± 0.12	15.34 ± 5.87
4×10^4 s	-0.93 ± 6.54	26.28 ± 0.19	25.56 ± 6.42
8×10^4 s	1.73 ± 9.87	53.80 ± 0.29	55.20 ± 10.12

values and standard deviations of the correlation with observation time confirms theoretical predictions. The mean value of C_{tot} increases approximately linearly with observation time T, and its standard deviation increases with the square root of T. Furthermore, the standard deviations of C_{tot} are always close to the square root of the mean value which would ideally hold, since the optimal filter defined in Eq. (51) is applied. The fact that the measured standard deviation is somewhat higher than the predicted value means that our noise model which governs the correlation filter is weaker than the actually measured spectrum of the instrumental noise. This discrepancy can be explained since the (equal-arm) noise model is not used to generate a time series of the noise in our simulation. In reality, one would take greater care in choosing the right noise model such that variances are equal to mean values, and the SNR is faithfully calculated for a single measurement. Alternatively, one may use the total spectrum after projection as a noise model. In our simulation, we obtain the SNR by averaging over many measurements, and the quality of the noise model does not have to be very high.

Before extrapolating to higher observation times, we have to make sure that subtraction noise can be removed from the data. Therefore, we show correlation values for our longest run with $T = 10^5$ s related to the CBC:

$$C_{\text{cbc}} = 19\,000, \qquad C_{\text{sub}} = 3560, \qquad C_{\text{proj}} = 2.34. \tag{54}$$

The correlation value of the CBC without best-fit subtraction and projection is C_{cbc}. When the best fits are subtracted, this value is reduced to C_{sub}. Consequently, a CGWB with $\Omega = 10^{-15}$ could not be detected without noise projection, since the CGWB correlation is around 55. The correlation of the projected noise is further reduced to C_{proj}, which lies well below the CGWB value. The background becomes detectable. A remaining problem is that the true data will contain about 100 to 1000 times more signals, which leads to a similar increase of the projected spectrum, but two effects are working for the good of the mission. Theoretically, the projected spectrum should decrease with $1/T$, which is due to an improvement of the accuracy of the estimated signal parameters. In addition, we believe that our projection results would be much better, if a different template class was used or if we had implemented the F statistics to maximize over nuisance parameters, since our waveforms are governed by templates with highly correlated parameters. These issues will have to be scrutinized in the future.

Now, assuming that the foreground subtraction noise can be projected out, we extrapolate correlation values to an observation time $T = 10^8$ s. Our measurements confirm that the SNR scales with the square root of the observation time, and so we extrapolate our results in Table II accordingly for each observation time, and average over the four different predictions to obtain the BBO's SNR correspond-

ing to a flat stochastic background (i.e. constant Ω):

$$\text{SNR} \, (3 \text{ years}) \approx 200 \cdot \frac{\Omega}{10^{-15}}. \tag{87}$$

One has to keep in mind that this prediction is based on a restricted correlation band, 0.2 Hz–0.5 Hz. The SNR would double if the entire detection band, especially towards lower frequencies, was chosen as the correlation band. The lower boundary is understood as the cutoff frequency due to the WD/WD foreground which may not be analyzable below 0.2 Hz. The upper boundary has no significant effect since high-frequency contributions to the SNR of the cosmic background are negligible. If one demands a minimal SNR of 5, then the BBO's sensitivity is

$$\Omega_{\min} \approx 2.5 \times 10^{-17}. \tag{56}$$

Given a lower bound of 0.2 Hz for correlation measurements, there should exist an optimal arm length of the BBO which is slightly smaller than 50 000 km, and which leads to a slightly improved sensitivity. It may also turn out that our choice for the lower boundary was too pessimistic and that the BBO is more sensitive even without modifications.

IX. CONCLUSION

The analysis presented in this paper serves three main purposes. The primary intention is to demonstrate the application of a subtraction-noise projection on a simulated data-analysis problem. The method is explained by invoking a geometrical interpretation of optimal signal detection and parameter estimation in the presence of additive, Gaussian noise. Second, we showed how to construct an analysis pipeline for a time-delay interferometer network, which seeks for a stochastic gravitational-wave background in the presence of a foreground built of CBCs. The detector network is simulated dynamically, thereby automatically including detector motion and time-varying response functions. Therefore, the definition of the overlap-reduction function has to be generalized so that it directly quantifies the total instrumental influence on correlation strengths of detector outputs, instead of quantifying the correlation of projected gravitational-wave induced strains in terms of somehow normalized quasistationary detector response functions. Third, since our simulation is based on a design proposal for the BBO detector network, we derive a prediction of its sensitivity towards stochastic backgrounds, which should be more robust than values obtained from previous investigations.

The simulation creates a time series of the instrumental noise and compact binaries for each photodiode in the network. During observation, response functions and detector positions change. We use equations of the orbital motion of each satellite which are specified up to second order in the orbital eccentricity. Consequently, each triangular configuration of satellites is simulated with cartwheel and breathing motion. The isotropic stochastic background

with a given spectral density is directly generated in the frequency domain, by first computing transfer functions and overlap-reduction functions of and between each network output. These frequency-domain functions are obtained via FFT of the detectors' impulse responses.

Based on the assumption that all foreground signals in the data from compact binaries are detected, we found, as expected, that by removing the estimated foreground from the data, the residual foreground spectrum due to inaccurate fitting of waveforms covers the spectrum of the cosmological background and makes it impossible for the BBO to detect it, unless its fractional energy density assumes very high values. We implemented a subtraction-noise projection into the pipeline, which allows us to accurately remove the subtraction noise. In simulation runs with $T = 10^5$ s and operating on data with 100 CBCs, the total power of the projected signal is 4 orders of magnitude weaker than the original signal spectrum, which is much better than previously expected since the analysis is based on a comparatively short observation time which entails strong correlations between different parameters of the signal model. This outcome can just be explained by assuming that the template manifold is approximately flat, at least along certain directions. The smaller the curvature of the manifold in the vicinity of the true signal, the weaker the requirements on the accuracy of waveform fitting. A detailed analysis of the template-manifold curvature will have to be carried out in the future to understand how precise a best fit has to be such that the respective subtraction noise can be projected out.

If one wants to extrapolate this result to higher observation times and more CBCs, then one needs to separate the detection and estimation problem from the projection problem. Certainly, it will be much more difficult to detect all signals of a realistic foreground, and for a more realistic analysis scheme, like the hierarchical search, one has to study the influence of confusion noise on the quality of the waveform estimates. The final answer depends on how close one will come to an optimal CBC analysis. We do not intend to make any predictions here concerning this point. However, assuming that all CBCs are detected and accurately analyzed, we claim that the projected subtraction-noise spectrum of a realistic foreground observed over 3 years will be negligible. First, we argued in our paper that confusion noise is not an issue in the subtraction-noise projection, since the projection is based on the total Fisher matrix of all projected signals, including mutual correlations, into the model. Only if excluded, correlations between signals manifest themselves as confusion noise, which may deteriorate the projection. Second, Fisher matrices which characterize 3 years of data will be much less ill conditioned and, for that reason, much easier to handle numerically. That is the main reason why we consider our projection results very promising: we

obtained good results choosing templates with strong parameter correlations. Third, and more fundamentally, the projected spectral power is supposed to decrease with $1/T$, which further reduces the projected spectrum by a factor 1000.

Now, applying simple, numerically confirmed scaling laws of the correlation SNR, we extrapolate correlation results obtained for a few runs with observation times below $T = 10^5$ s to the full BBO mission lifetime of $T = 10^8$ s. In conclusion, provided that the subtraction noise can be removed, the extrapolated sensitivity of the BBO to a stochastic background is $\Omega_{\mathrm{min}} \approx 2.5 \times 10^{-17}$. However, one condition for this result is that the cosmological WD/WD foreground of above 0.2 Hz can be resolved and fitted like the foreground of NSs and BHs. Conversely, it may turn out that the lower boundary of the correlation band is too pessimistic, which would entail a better sensitivity of the BBO (up to a factor of 2). Ultimately, the problem of the WD barrier can be ameliorated by making the BBO arms shorter by a small factor to increase the frequency of optimal sensitivity towards stochastic backgrounds.

ACKNOWLEDGMENTS

APPENDIX: PARAMETER DERIVATIVES OF GW PHASE

In this appendix, we elaborate on some details concerning the Fisher matrix of comparable-mass, compact-binary inspirals within the pN formalism. The general framework is set in Sec. IV.

The GW phase evolution depends on the three parameters t_c. M, μ, and therefore the respective Fisher-matrix components depend on the derivative of the GW phase with respect to these parameters. We found that the result can be cast into a form which resembles the pN expansion Eq. (9). In the following, these results will be presented. We start with the derivative of the phase with respect to the chirp time. It can be written as a sum

$$\partial_{t_c}\phi = -\frac{c^3}{4GM} \sum_{k=0}^{7} p_k^{t_c} \tau^{-(3+k)/8} \qquad (A1)$$

with expansion coefficients

$$p_0^{t_c} = 1, \qquad p_1^{t_c} = 0, \qquad p_2^{t_c} = \frac{743}{2688} + \frac{11}{32}\eta, \qquad p_3^{t_c} = -\frac{3}{10}\pi,$$

$$p_4^{t_c} = \frac{1\,855\,099}{14\,450\,688} + \frac{56\,975}{258\,048}\eta + \frac{371}{2048}\eta^2, \qquad p_5^{t_c} = -\left(\frac{7729}{21\,504} + \frac{3}{256}\eta\right)\pi,$$

$$p_6^{t_c} = -\frac{720\,817\,631\,400\,877}{288\,412\,611\,379\,200} + \frac{107}{280}C + \frac{53}{200}\pi^2 - \frac{107}{2240}\log\left(\frac{\tau(t)}{256}\right) + \left(\frac{25\,302\,017\,977}{4\,161\,798\,144} - \frac{451}{2048}\pi^2\right)\eta \qquad (A2)$$

$$\qquad\qquad - \frac{30\,913}{1\,835\,008}\eta^2 + \frac{235\,925}{1\,769\,472}\eta^3,$$

$$p_7^{t_c} = \left(-\frac{188\,516\,689}{433\,520\,640} - \frac{28\,099}{57\,344}\eta + \frac{122\,659}{1\,290\,240}\eta^2\right)\pi.$$

All symbols are defined as in Sec. IV. A similar expression can be found for the derivatives with respect to the two mass parameters. The two expansions

$$\partial_\mu\phi = \frac{3}{4\eta\mu} \sum_{k=0}^{7} p_k^\mu \tau^{(5-k)/8}, \qquad (A3)$$

$$\partial_M\phi = \frac{1}{2\eta M} \sum_{k=0}^{7} p_k^M \tau^{(5-k)/8} \qquad (A4)$$

are determined by expansion coefficients

$$p_0^\mu = 1, \qquad p_1^\mu = 0, \qquad p_2^\mu = \frac{18\,575}{24\,192} - \frac{55}{96}\,\eta, \qquad p_3^\mu = -\frac{3}{2}\,\pi,$$

$$p_4^\mu = \frac{9\,275\,495}{6\,193\,152} - \frac{284\,875}{774\,144}\,\eta - \frac{5565}{2048}\,\eta^2, \qquad p_5^\mu = \left(\frac{38\,645}{64\,512} + \frac{5}{256}\,\eta - \frac{38\,645}{64\,512}\log\!\left(\frac{\tau(t)}{\tau(0)}\right)\right)\pi,$$

$$p_6^\mu = \frac{818\,786\,359\,710\,637}{19\,227\,507\,425\,280} - \frac{321}{56}\,C - \frac{159}{40}\,\pi^2 + \frac{321}{448}\log\!\left(\frac{\tau(t)}{256}\right) + \left(-\frac{126\,510\,089\,885}{12\,485\,394\,432} + \frac{2255}{6144}\,\pi^2\right)\eta - \frac{154\,565}{786\,432}\,\eta^2$$
$$+ \frac{5\,898\,125}{1\,769\,472}\,\eta^3,$$

$$p_7^\mu = \left(\frac{942\,583\,445}{260\,112\,384} + \frac{140\,495}{172\,032}\,\eta + \frac{122\,659}{258\,048}\,\eta^2\right)\pi, \qquad p_0^M = 1, \qquad p_1^M = 0, \qquad p_2^M = -\frac{3715}{8064} + \frac{55}{32}\,\eta, \qquad \text{(A5)}$$

$$p_3^M = \frac{3}{2}\,\pi, \qquad p_4^M = -\frac{9\,275\,495}{4\,816\,896} + \frac{284\,875}{258\,048}\,\eta + \frac{9275}{2048}\,\eta^2,$$

$$p_5^M = \left(-\frac{38\,645}{21\,504} - \frac{15}{256}\,\eta + \frac{38\,645}{43\,008}\log\!\left(\frac{\tau(t)}{\tau(0)}\right)\right)\pi$$

$$p_6^M = -\frac{808\,989\,486\,879\,661}{11\,536\,504\,455\,168} + \frac{535}{56}\,C + \frac{53}{8}\,\pi^2 - \frac{535}{448}\log\!\left(\frac{\tau(t)}{256}\right) + \left(\frac{126\,510\,089\,885}{4\,161\,798\,144} - \frac{2255}{2048}\,\pi^2\right)\eta + \frac{463\,695}{1\,835\,008}\,\eta^2$$
$$- \frac{8\,257\,375}{1\,769\,472}\,\eta^3$$

$$p_7^M = -\left(\frac{188\,516\,689}{28\,901\,376} + \frac{140\,495}{57\,344}\,\eta + \frac{122\,659}{258\,048}\,\eta^2\right)\pi.$$

[1] B. Willke *et al.*, Classical Quantum Gravity **19**, 1377 (2002).
[2] LIGO Scientific Collaboration, arXiv:0711.3041.
[3] F. Acernese *et al.*, Classical Quantum Gravity **24**, S381 (2007).
[4] LIGO Scientific Collaboration, Phys. Rev. Lett. **94**, 181103 (2005).
[5] LIGO Scientific Collaboration, Phys. Rev. D **77**, 022001 (2008).
[6] LIGO Scientific Collaboration, Phys. Rev. D **76**, 082003 (2007).
[7] F. Beauville *et al.*, J. Phys.: Conf. Ser. **32**, 212 (2006).
[8] E. Gustafson, D. Shoemaker, K. Strain, and R. Weiss, http://www.ligo.caltech.edu/docs/T/T990080-00.pdf.
[9] LISA study teams, LISA-Final Technical Report (2000).
[10] J. Crowder and N. J. Cornish, Phys. Rev. D **70**, 082004 (2004).
[11] E. Racine and C. Cutler, Phys. Rev. D **76**, 124033 (2007).
[12] C. Cutler and J. Harms, Phys. Rev. D **73**, 042001 (2006).
[13] E. S. Phinney *et al.*, NASA Mission Concept Study (2003).
[14] B. Allen, arXiv:gr-qc/9604033.
[15] T. L. Smith, M. Kamionkowski, and A. Cooray, Phys. Rev. D **73**, 023504 (2006).
[16] L. A. Boyle and A. Buonanno, arXiv:0708.2279.
[17] A. J. Farmer and E. S. Phinney, Mon. Not. R. Astron. Soc. **346**, 1197 (2003).
[18] R. Schneider, V. Ferrari, S. Matarrese, and S. F. Portegies Zwart, Mon. Not. R. Astron. Soc. **324**, 797 (2001).
[19] V. Kalogera, R. Narayan, D. N. Spergel, and J. H. Taylor, Astrophys. J. **556**, 340 (2001).
[20] K. Belczynski, V. Kalogera, and T. Bulik, Astrophys. J. **572**, 407 (2002).
[21] N. J. Cornish and L. J. Rubbo, Phys. Rev. D **67**, 022001 (2003).
[22] N. J. Cornish and T. B. Littenberg, Phys. Rev. D **76**, 083006 (2007).
[23] M. Tinto, F. B. Estabrook, and J. W. Armstrong, Phys. Rev. D **65**, 082003 (2002).
[24] M. Vallisneri, Phys. Rev. D **71**, 022001 (2005).
[25] M. Tinto, F. B. Estabrook, and J. W. Armstrong, Phys. Rev. D **69**, 082001 (2004).
[26] M. Vallisneri, J. Crowder, and M. Tinto, Classical Quantum Gravity **25**, 065005 (2008).
[27] M. Vallisneri, Phys. Rev. D **72**, 042003 (2005).
[28] J. N. Franklin, SIAM Rev. **7**, 68 (1965).
[29] N. J. Cornish and E. K. Porter, Phys. Rev. D **75**, 021301(R) (2007).
[30] J. Crowder and N. J. Cornish, Phys. Rev. D **75**, 043008 (2007).
[31] B. Efron, Ann. Stat. **3**, 1189 (1975).
[32] S.-I. Amari, Ann. Stat. **10**, 357 (1982).
[33] A. F. S. Mitchell, Int. Stat. Rev. **56**, 1 (1988).
[34] L. Blanchet, G. Faye, B. R. Iyer, and B. Joguet, Phys. Rev. D **65**, 061501(R) (2002).
[35] C. Ungarelli and A. Vecchio, Phys. Rev. D **63**, 064030 (2001).
[36] L. S. Finn, Phys. Rev. D **46**, 5236 (1992).

[37] C. Cutler and É. É. Flanagan, Phys. Rev. D **49**, 2658 (1994).
[38] M. Vallisneri, Phys. Rev. D **77**, 042001 (2008).
[39] C. Cutler and B. F. Schutz, Phys. Rev. D **72**, 063006 (2005).
[40] B. Haible and R. B. Kreckel, http://www.ginac.de/CLN/.
[41] I. N. Imam and L. M. Lamont, in Proceedings of Energy and Information Technologies in the Southeast, 1989, p. 421.
[42] T. A. Prince, M. Tinto, S. L. Larson, and J. W. Armstrong, Phys. Rev. D **66**, 122002 (2002).

[43] É. É. Flanagan, Phys. Rev. D **48**, 2389 (1993).
[44] N. Seto and A. Cooray, Phys. Rev. D **70**, 123005 (2004).
[45] N. Seto, Phys. Rev. D **73**, 063001 (2006).
[46] R. Balasubramanian, B. S. Sathyaprakash, and S. V. Dhurandhar, Phys. Rev. D **53**, 3033 (1996).
[47] L. A. Wainstein and V. D. Zubakov, *Extraction of Signals from Noise* (Dover, New York, 1962).
[48] M. Tinto, F. B. Estabrook, and J. W. Armstrong (unpublished).