Parameter estimation methods for analyzing overlapping gravitational wave signals in the third-generation detector era

Justin Janquart[1,2],* Tomasz Baka[1,2], Anuradha Samajdar[3], Tim Dietrich[3,4], and Chris Van Den Broeck[1,2]

[1] *Nikhef, Science Park 105, 1098 XG Amsterdam, The Netherlands*

[2] *Institute for Gravitational and Subatomic Physics (GRASP), Department of Physics,*
Utrecht University, Princetonplein 1, 3584 CC Utrecht, The Netherlands

[3] *Institut für Physik und Astronomie, Universität Potsdam, Haus 28,*
Karl-Liebknecht-Str. 24/25, 14476, Potsdam, Germany and

[4] *Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, Potsdam, Germany*

(Dated: November 3, 2022)

In the coming years, third-generation detectors such as the Einstein Telescope and the Cosmic Explorer will enter the network of ground-based gravitational-wave detectors. Their current design predicts a significantly improved sensitivity band with a lower minimum frequency than existing detectors. This, combined with the increased arm length, leads to two major effects: the detection of more signals and the detection of longer signals. Both will result in a large number of overlapping signals. It has been shown that such overlapping signals can lead to biases in the recovered parameters, which would adversely affect the science extracted from the observed binary merger signals. In this work, we analyze overlapping binary black hole coalescences with two methods to analyze multi-signal observations: *hierarchical subtraction* and *joint parameter estimation*. We find that these methods enable a reliable parameter extraction in most cases and that joint parameter estimation is usually more precise but comes with higher computational costs.

I. INTRODUCTION

The observation of gravitational waves (GWs) originating from compact binary coalescences (CBCs) is now done routinely with the advanced LIGO [1] and advanced Virgo [2] detectors, with tens of detections reported in the O3 GW catalog [3]. The observations from this new information channel has had major impacts in fundamental physics [4], astrophysics [5], and cosmology [6]. Moreover, the possibility to detect electromagnetic counterparts for binary neutron stars (BNSs), such as GW170817 [7], has opened the perspective to do new multi-messenger studies. Further upgrades to the second generation detectors, as well as the addition of new detectors such as KAGRA [8–11] and LIGO India [12], will lead to the observation of numerous events in the coming years. Moreover, going from these second-generation (2G) detectors to third-generation (3G) detectors, such as Einstein Telescope (ET) [13, 14] and Cosmic Explorer (CE) [15–17], will lead to a major jump in the detection rate as well as in the duration of the signal due to the combined global increase in sensitivity and the major enhancements for lower frequencies [18]. In turn, this will lead to a high probability of CBC signals to overlap in the 3G detectors [19–23].

Previous works have studied the impact of such overlapping signals on data analysis when recovering one of the two signals and neglecting the presence of the other [20–24]. These works employed different techniques but all have the same basic conclusions: bias can occur in various scenarios, and is most likely when the signals merge close to each other. In Ref. [21], using a Fisher matrix approach, the authors show that BNSs are less correlated in overlapping signals so that their bias becomes important only for close merger times (< 0.1 s), while the correlation between binary black holes (BBHs) is more important, meaning that the bias can happen for larger differences between the merger times. They then proceed to parameter estimation (PE) for overlapping BBHs in a LIGO-Virgo network varying some of the parameters, showing the appearance of biases for merger times close to each other. In [23], the authors use Fisher matrices to study the bias that can occur in the parameters in both overlapping BBHs and BNSs, also finding that the bias becomes more important for short differences in the time of arrival. In [22], the authors focus on a LIGO Voyager scenario, looking at the biases not only based on their difference in merger times, but also as a function of other parameters, such as the sky location. For overlapping BBHs, they find a more important bias for closer merger times. However, they show that the observed bias for a given difference in merger time can change substantially depending on the sky location of the two events. Moreover, two overlapping BBHs can be mistaken for one strongly precessing BBH. Furthermore, the authors suggest that for overlapping BNS and BBH signals no major bias will occur due to the different durations of the two signals. This is corroborated by the analyses done in [20], where three scenarios are analyzed: two overlapping BBH signals, two overlapping BNS signals and the overlap of a BBH of varying masses with a BNS. In the latter case, the authors find that there is hardly any effect on the BNS parameter estimation, probably due to the difference in the number of cycles present in-band for the this signal. However, in this scenario, the BBH can be affected by significant bias, especially when

* j.janquart@uu.nl

the BBH has high component masses. The bias mostly disappears when the merger times are separated by more than two seconds. For two overlapping BBHs, if the total masses and hence the durations, are different, the PE is done relatively well. However, others [21–23] have shown that, if the two BBHs have similar source properties, biases can be present. Finally, for two overlapping BNSs, it appears that the signal with the highest SNR is relatively well recovered in all cases, but not necessarily the quieter signal.

One problem that is not covered by previous studies is the effect of confusion noise on PE. Indeed, the high rate of the events and their duration will make for very few periods without signal in-band for the 3G detectors [20]. As a consequence, it will be very difficult to estimate the noise present in the detectors, and additional biases can occur due to a mismodelling of the noise [24][1]. A demonstration of the effect of this noise on matched filtering and how the PSD could be computed are presented in [25].

The presence of biases when signals merge close to each other and the relative occurrence of such scenarios based on the estimated rates shows that PE methods will have to be adapted to be suited for the 3G cases. Indeed, biases in the parameters estimated for the CBCs would impact any direct science case for the CBCs (such as measuring their mass distribution and rate, or testing general relativity [26]) and also some indirectly related ones, such as the search for primordial black holes since this requires the subtraction of the foreground sources [27–32].

In this work, we look at two possible methods to analyze overlapping signals. *Hierarchical subtraction* (HS), where we analyze one signal (typically the loudest), then subtract the maximum-likelihood template before analyzing the second one. However, if an important bias happens when analyzing the first signal, the parameters of both events might be biased. Optionally, one can also perform a third run, subtracting the maximum-likelihood parameters for the second event and re-analyzing the dominant signal to reduce the bias in its recovery. Still, this is not guaranteed to lead to unbiased results. Therefore, we also implement a *joint parameter estimation* (JPE) framework, where the two signals are analyzed at the same time to account for the entire model. In principle, this should be the most complete model one can use. Due to the high dimensionality of the parameter space, combined with the large duration of the signals in the 3G detectors, this framework is substantially slower than HS[2]. It would be close to impossible to follow the predicted rates for an ET and CE network us-

ing a simple JPE like the one used in this work. Such constraints could be alleviated by using recently developed techniques, like relative binning [33–35], adaptive frequency resolution [36], or machine learning [37, 38]. One could also count on the development of more powerful computational methods, such as quantum computing [39] but it is difficult to have an idea of the state of such methods by the time the 3G detectors get online. Still, it is important to start preparing for the future of 3G detectors now, hence, it is important to start looking at the parameter estimation of overlapping signals to have the bases to build upon. Due to the limited computational resources, this work focuses on the parameter estimation of two overlapping BBH signals.

This article is structured as follows. In Sec. II, we explain the two methods applied to perform the analysis of the overlapping signals. In the next section, we explain the setup of the analyses, while in Sec. IV we present the results of the analyses. Finally, Sec. V provides conclusions and outlook.

II. DESCRIPTION OF THE METHODS

When performing GW data analysis on CBC signals, our objective is to find the posterior probability density function (PDF) of the binaries' parameters (θ): $p(\theta|d, \mathcal{H}_s)$, where d represents the data, and $\mathcal{H}_s$ is the hypothesis under which we work (e.g. there is a GW signal in the data). Using Bayes' theorem [40]:

$$p(\theta|d, \mathcal{H}_s) = \frac{p(d|\theta, \mathcal{H}_s)p(\theta|\mathcal{H}_s)}{p(d|\mathcal{H}_s)} \,, \qquad (1)$$

where $p(\theta|\mathcal{H}_s)$ is the prior on the binary parameters, and $p(d|\theta, \mathcal{H}_s)$ is the likelihood for the data given a set of parameters. Finally, $p(d|\mathcal{H}_s)$ is called the evidence and represents a normalization factor for the posterior probability density:

$$p(d|\mathcal{H}_s) = \int d\theta p(d|\theta, \mathcal{H}_s)p(\theta|\mathcal{H}_s) \,. \qquad (2)$$

For GW inference, $d(t)$ is the output of the interferometers, which can be seen as made of a noise component $n(t)$ and, under the signal hypothesis, a GW component $h(t)$:

$$d(t) = n(t) + h(t) \,. \qquad (3)$$

In our scenario, the GW component can consist of one or more signals. In the latter case, $h(t) = \sum_{i=1}^{N} h_i(t)$, where $h_i(t)$ is the representation of each individual GW, and N is the total number of GW signals present in the data stretch under consideration.

---

[1] We will not consider this in this work as we will only look at overlapping binary signals.

[2] For our experiments, the two frameworks were run on the same cluster using 16 *Intel(R) Xeon(R) Gold 6152* CPUs. The average run time for JPE is 23.8 days, while for HS, the first run took an average of 6.3 days, the second run an average of 4.3 days, and the last run took an average run-time of 6.1 days. So, on average, JPE takes 7 more days to complete than HS if we perform the

---

3 runs. If one is satisfied with the two first runs, the difference between the two approaches goes up to about 2 weeks.

Assuming the noise to be Gaussian, the likelihood of having the data $d(t)$ given the presence of a GW component $h(t)$, is given by the proportionality

$$p(d|\theta, \mathcal{H}_s) \propto \exp\left[-\frac{1}{2}\langle d - h(\theta)|d - h(\theta)\rangle\right], \quad (4)$$

where

$$\langle a|b\rangle = 4\Re \int_{f_{\min}}^{f_{\max}} \frac{\tilde{a}^*(f)b(f)}{S_n(f)}df. \quad (5)$$

In this expression, $\tilde{a}(f)$ refers to the Fourier transform of the function $a(t)$, $*$ is the complex conjugate, and $S_n(f)$ is the power spectral density (PSD). $f_{\min}$ and $f_{\max}$ represent respectively the lower and the upper frequency chosen for the analysis[3].

A. Joint Parameter estimation

When the noise component in Eq. (3) is made of multiple signals, the likelihood in Eq. (4) becomes

$$p(d|\theta, \mathcal{H}_s)$$
$$\propto \exp\left[-\frac{1}{2}\left\langle d - \sum_{i=1}^{N} h(\theta_i)\middle|d - \sum_{i=1}^{N} h(\theta_i)\right\rangle\right], \quad (6)$$

where we just expanded the expression for $h(t)$ compared to the previous expression. Here, N is the total number of signals, and θ_i represents the set of parameters describing the ith GW signal so that $\theta = \{\theta_1, \ldots, \theta_i, \ldots, \theta_N\}$ in this case.

Using Eq. (6) gives rise to the method of joint parameter estimation, where we jointly look for $\{\theta_i\}_{i=1,\ldots,N}$. Since each θ_i is a set of 15 parameters for spinning BBHs, a set of 16 parameters for NSBHs, and a set of 17 parameters for BNSs[4], it means that, if we have X BBHs, Y NSBHs, and Z BNSs (so $N = X + Y + Z$), the parameter space has $15X + 16Y + 17Z$ parameters to explore. In the end, this means that as soon as we consider 2 signals, the parameter space grows to at least 30 dimensions, which is already challenging with our traditional methods, showing the difficulty to analyze several signals jointly.

---

[3] Usually, $f_{\min}$ is chosen above the noise levels, and $f_{\max}$ is the Nyquist frequency.
[4] Typically, a BBH is described by two mass parameters, 6 spin parameters, a distance parameter, the inclination, 2 parameters for the sky location, the merger time, the phase of coalescence, and the polarization angle. Usually, for each neutron star present in the system, one also add a tidal deformability, however, for BNSs the dimentionality could increase even further if higher-order tidal contributions, spin-induced quadrupole effects, or resonant effects are also taken into account.

In addition, formally, when parameter estimation is performed, one can also add the noise-related calibration parameters, which would further increase the dimensionality of the parameter space.

In addition, we can assume that there is some uncertainty on the total number of signals N in the data. In this case, it is also possible to sample over N and the signal types, effectively allowing for any number of signals to be present in a given data stretch.

The problem of joint analysis of several signals has already been looked at in other contexts, such as the characterization of the nearly monochromatic signals coming from white dwarfs in LISA [41], BNSs in the Big Bang Observer [42], or supermassive black holes in pulsar timing array searches [43]. However, the different characteristics of the signal looked for in these various context makes the methods different from one case to the other.

In this work, we will only consider the possibility to have two signals in the data. So, we write the data as

$$d(A, B, t) = h_A(t) + h_B(t) + n(t), \quad (7)$$

where we just denote the signals by A and B, without any importance on which signal is A and which signal is B.

In this case, the likelihood takes the particular form

$$p(d(A, B)|\theta, \mathcal{H}_s)$$
$$\propto \exp\left[-\frac{1}{2}\left\langle d(A, B) - h_A(\theta_A) - h_B(\theta_B)\right|\right.$$
$$\left.d(A, B) - h_A(\theta_A) - h_B(\theta_B)\right\rangle\right]. \quad (8)$$

This is just the explicit form of Eq. (6) for two signals.

In principle, if the sources are of the same nature, the labels A and B are interchangeable during the sampling, making the likelihood symmetric in two events. In our algorithm, we do not impose any conditions on the parameters to break this symmetry. As a consequence, this is something that needs to be done in a post-processing step, as we need to assign drawn samples to the correct event. Sometimes, not accounting for this condition leads to bimodalities. In this work, we use a time ordering condition, taking the samples for event A to be those that arrive first in time, and the samples for event B to be those arriving second in time. In future work, this condition could directly be imposed in the algorithm by having a conditional prior such that the time of arrival of one event is always smaller than the other. We also note that the condition can be imposed on other parameters.

B. Hierarchical subtraction

In hierarchical subtraction, the idea is not to fit the two signals at once, but to rather do a combination of individual signal analyses and subtraction of best-fit parameters. Therefore, we start by running a single event parameter estimation analysis on the data $d(A, B)$ to get the characteristics of the dominant signal. If we label by

A the loudest signal, we can denote the best-fit parameters (typically the maximum likelihood parameters) $\hat{\theta}_A$, and the waveform corresponding to this signal

$$\hat{h}_A(t) = h(t, \hat{\theta}_A). \tag{9}$$

Using this, we can get the data for the signal B given signal A by subtracting the best-fit template

$$d(B|A) = d(A, B) - \hat{h}_A(t), \tag{10}$$

which is an approximate data strain for the second event in the data since the maximum likelihood parameters used to model the first event are prone to errors, with errors coming from the modeling itself but also from the presence of the second event when characterizing the first one.

We can then analyze $d(B|A)$ to get the parameters for event B, leading to a posterior distribution for the two events. In principle, if the bias on the first is not too important, then the posteriors on the second event should also be fine. However, this approach is less robust than the joint parameter estimation, where we correctly account for the presence of several events.

In Ref. [24], the authors also suggest a way to correct the bias due to the individual characterization of two signals[5]. Once the two signals have been analyzed separately, we can use the two best-fitting posteriors to evaluate the bias made in the model reconstruction for each signal. The estimated biases can then be applied as a correction factor to the best-fitting parameters. We could then redo the subtraction of each event and analyze it again, but now with a subtracted signal that is closer to the real one, reducing the possible bias in the recovered posterior. Though this method is attractive, it requires multiple parameter estimation runs, which are expensive in a 3G detector context, and the computation of the biases also requires solving a combination of Fisher matrices and numerical derivatives, making it a non-trivial operation.

III. SETUP OF THE ANALYSES

Due to the computational resources that would be required to analyze 3G signals, we focus on overlapping BBHs with masses in $[30, 60]M_\odot$[6]. We use a network of detectors made of one triangular Einstein Telescope with 10 km arm-lengths, and a Cosmic explorer detector located at the LIGO-Hanford position and with 40km arm-lengths. We generate stationary Gaussian noise from the detectors' PSDs, where, for the ET, we use the ET-D PSD [13, 14], and for the CE, we use the projected PSD from Refs. [15, 16]. We then inject two simulated BBH signals into the artificial noise. For this study, we take a lower cutoff on the signals of 20 Hz. A representation of the waveforms obtained by the addition of two BBHs can be found in Fig. 1. One sees that the final signal has a non-trivial shape, illustrating the risk of biases when not accounting for the presence of two signals. In addition, one also sees that depending on the relative SNR of the signals, the observed bias is different. For the case where one signal has a significantly higher SNR than the other ($\sim \times 2$, top panel), the quieter signal will somewhat bias the signal, but the observed waveform will resemble mostly the loudest signal. On the other hand, for signals with close SNRs (bottom panel), we see that the deformation of the signal can be more complicated, without really having a dominant signal (except for the fraction of a second where one signal has merged and the other is still merging). Based on these observations, one can expect hierarchical subtraction to be more effective when one signal clearly dominates over the other.

In our study, we simulate 55 such mergers. To produce these high-mass signals, we sample the component masses from the Power-law + Peak distribution from [3] but keep only the systems that fulfill the mass requirement. In addition, the events are sampled in redshift according to the merger rate density reconstructed from Oguri's fit [44]. The sky location is drawn to be uniform on the sky, and the spin parameters are picked from an isotropic distribution. For overlapping signal events, the coalescence time of the first event is drawn from a uniform distribution spanning over an entire year, while the second event is placed 0.1 second later. An overview of the functions used to make the binaries and the priors used for the analysis is given in Table I.

Since the SNR of the signals can reach hundreds to thousands in an ET and CE network, and such high values make the computation time even longer, we decided to rescale the SNR to take values constrained between 8 and 50. This is done by adjusting the luminosity distance. However, since we expect the SNR ratio between the events to play a role, we try to keep this ratio as close as possible to the original one. So, if the loudest signal has a value above 50, we rescale it to take a value between 45 and 50 (this value is drawn randomly from a uniform distribution). Then, we rescale the quieter signal with the same factor. If this value is below 8, then we choose a new scaling factor to bring the SNR back between 8 and 13 (once more using a uniform distribution). Each system is then analyzed once without additional noise, and once injected in Gaussian noise generated from the PSDs.

For the different runs, we choose fixed priors for the various parameters. The right column of Table. I gives an overview of the priors used for the different parameters. In particular, we take a uniform prior on chirp mass ($\mathcal{M}_c = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}$) and mass ratio ($q = \frac{m_2}{m_1}$), with

---

[5] In their paper, they also account for the possible confusion background due to the sum of all the mergers going on in the background.

[6] The higher masses are chosen so as to not have a signal with a too long duration while still enabling overlap for the difference in times of arrival used in this work.

Parameter	Population generation	Prior
Primary component mass	PowerLaw + Peak [3]	/
Mass ratio	PowerLaw + Peak [3]	$\mathcal{U}(0.1, 1)$
Chirp mass	/	$\mathcal{U}(4, 200)$
Redshift	Oguri's fit [44] + rescaling	/
Luminosity distance	/	Uniform comoving volume [1, 100] Gpc
Spin amplitude 1	$\mathcal{U}(0, 1)$	$\mathcal{U}(0, 1)$
Spin amplitude 2	$\mathcal{U}(0, 1)$	$\mathcal{U}(0, 1)$
Tilt angle 1	Uniform in sine	Uniform in sine
Tilt angle 2	Uniform in sine	Uniform in sine
Spin vector azimuthal angle	$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$
Precession angle about angular momentum	$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$
Inclination angle	Uniform in sine	Uniform in sine
Wave polarization	$\mathcal{U}(0, \pi)$	$\mathcal{U}(0, \pi)$
Phase of coalescence	$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$
Right ascension	$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$
Declination	Uniform in cosine	Uniform in cosine
Time of coalescence	Uniform over a year (second precision)	$\mathcal{U}(t_{inj} - 0.1, t_{inj} + 0.1)$

TABLE I. Overview of the functions used to generate the different parameters for the BBH population and the priors used for the parameter estimation recovery.

bounds of $[4, 200]M_{\odot}$, and $[0.1, 1]$. We also take a uniform in comoving volume prior for the luminosity distances, with bounds going from 1 to 100 Gpc. These priors are adapted to cover any possible signal present in our set of data. The priors for the other parameters correspond to the usual priors taken for BBHs. When doing the JPE and HS runs, the priors are the same for the two events, and we do not add any conditions related to the signals (for example time ordering of the signals or enforcing one signal to be heavier than the other).

We note here that an alternative approach is to use narrower priors informed by the results of low-latency searches [45, 46], which very likely could only be applied for a couple of parameters such as the chirp mass and the coalescence time. In [46], the authors show that matched filtering pipelines and unmodelled searches can pick up overlapping signals with reasonable accuracy. They also suggest some enhancements to make the pipelines even more suited for the challenge of overlapping signals detection. In addition, they show that for most of the overlapping signals, the error on the chirp mass is not very much biased compared to the non-overlapping case, even if for the occurrences where the signals are very close in time, the error increases. Therefore, using such searches to set narrower priors is a realistic alternative. However, currently, they also seem to contain risks, as an increased difference in the value recovered for some parameters can happen, and taking too narrow a prior could lead to the exclusion of the actual value from the prior. In the end, extra developments would be needed to make sure that using these results to narrow down the initial priors is viable.

To have a basis of comparison, we also do the parameter estimation of the individual signals. This is done by using the same priors as the one explained above but injecting only one of the two signals in the noise.

All the parameter estimation runs are performed using BILBY [47] with the DYNESTY [48] sampler. For the JPE runs, we added our own adapted joint likelihood in the package to keep a consistent framework.

IV. RESULTS AND DISCUSSIONS

In this section, we show the results of the different approaches. We first compare the HS approach with single parameter estimation (SPE). In the latter, we only inject one of the two signals and perform PE on it. Then, we compare JPE with SPE and HS. Here, we focus on the results for the analyses performed with noise. The conclusion in the no-noise case are similar and can be found in Appendix A. For all the figures presented in this work, when plotting individual event results, we represent by a dot and label as "loud" events those that are the loudest in the pair, while we represent by a triangle and label as "quiet" the quieter ones.

A. Hierachical subtraction

We start with discussing the result of the analysis for the HS approach. In almost all of the HS runs, the first PE stage picked up the signal with higher SNR. Nevertheless there were two cases where it instead picked up the quieter signal; these were instances where the SNRs of the injected signals were close to each other. In such a case, the signal picked first is not the same with and without noise. A representative example of the posteriors can be found in Fig. 2. While the widths of the distributions
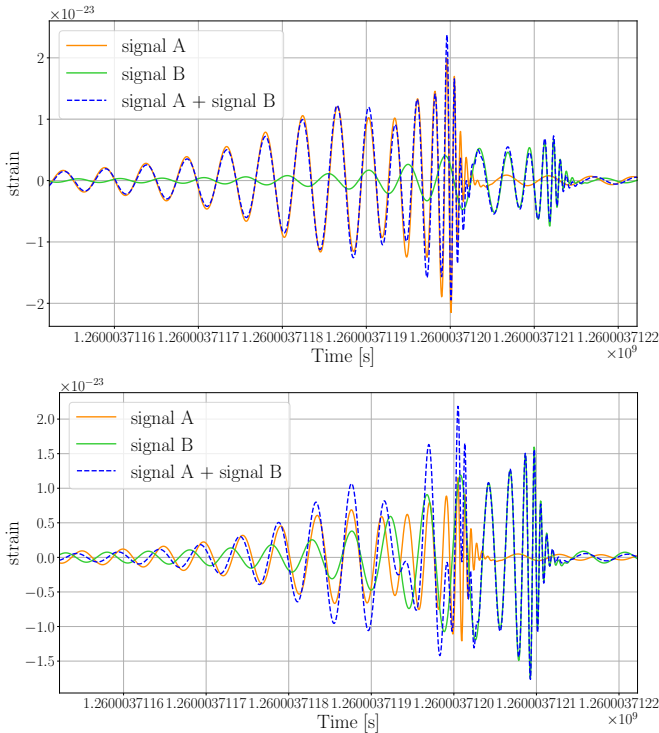
FIG. 1. *Top:* Representation of two overlapping BBHs and the underlying signals for signals with different SNRs (SNR$_A$ = 46.1 and SNR$_B$ = 22.2). In this case, the sum of the two signals is mostly dominated by the loudest signal, while the effect of the quieter signal takes up only after the loudest signal has merged. *Bottom:* Representation of two overlapping BBHs and the underlying signals for signals with similar SNRs (SNR$_A$ = 34.3 and SNR$_B$ = 30.2). In this scenario, there is no signal dominating over the other for the entire duration of the event and the overlapping signals have more complicated features.

match closely the ones obtained with the SPE runs, in most cases they shows bias in the recovered parameters.

When comparing the HS and the SPE recovery of the parameters, one sees that the HS recovery is nearly always biased. However, this bias seems to be more pronounced when there is also noise in the injection. For example, in Fig. 2, one sees that the zero-noise HS recovery is close to the one for SPE with noise, while the recovery with noise is off. This shows that the mismodelling of the noise (due to an additional event) is present, as one could expect from previous works on biases in overlapping signals [20–22, 24]. However, this does not necessarily mean that all parameters are off, and the first signal's characteristics can be recovered.

In Fig. 3, we can see mismatches between the injected and recovered waveforms, comparing HS and SPE cases, with and without noise. The mismatch is defined as $1 - \bar{M}$, where $\bar{M}$ is the match between the waveforms, defined as

$$\bar{M} = \frac{\langle h_1 | h_2 \rangle}{\sqrt{\langle h_1 | h_1 \rangle \langle h_2 | h_2 \rangle}}, \quad (11)$$

where $\langle . | . \rangle$ is the noise weighted product defined in Eq.(5). The mismatch represents the dissimilarity between two waveforms. High values mean a major disagreement between the two waveforms, and smaller values mean that the waveforms agree well. In Fig. 3, we see that the average mismatch throughout the detections is always low, below 0.02. As expected, the presence of another signal leads to worse waveform recovery for HS compared to SPE (most points are below the diagonal). The worst recovery of the signals occurs for overlapping signals with similar SNRs. Note that zero-noise case (see Appendix A) shows a clearer difference between SPE and HS recoveries. It is expected to have a larger difference in this case because the effect of the unmodelled signal is stronger when there is no noise since it is the only source of uncertainty in the signal.

By looking at the difference between the median of the recovered posterior and the injected value normalized by the injected value, we can quantify the bias in the recovery ($\Delta_{\mathcal{M}_c} = \frac{|\mathcal{M}_c^{\text{inj}} - \mathcal{M}_c^{\text{rec}}|}{\mathcal{M}_c^{\text{inj}}}$, where "rec" stands for recovered and "inj" stands for injected.). This is represented in Fig. 4 for the chirp mass recovery with noise and is representative for all the parameters. HS shows higher bias in 71% of the louder events and 51% of the quieter ones. As expected, the bias is larger for HS compared to SPE. For the first recovered signal, since we have an unmodeled signal present in the detectors, the noise properties are not properly modeled. Therefore, a larger bias is generally observed for the louder signal compared to the secondary one. However, when strong biases are present for the first signal, it can reverberate on the second, also leading to biases for this event.

Finally, Fig. 5 shows how the widths of the 90% confidence intervals of the posteriors for HS compare to SPE, normalized by the injected value ($\delta_{\mathcal{M}_c} = \frac{\sigma_{\mathcal{M}_c}}{\mathcal{M}_c^{\text{inj}}}$, where $\sigma_{\mathcal{M}_c}$ represent the width of the 90% confidence interval). We observe that the widths of the distribution are consistent between the two, even though the recovery is biased.

It is of interest to compare how doing successive steps of parameter estimation affects the results. After reconstructing the quieter of the two signals, we subtract it from the initial data and do PE again. In principle, it should result in the better recovery of the louder signal than the original PE run.

Fig. 6 shows the mismatch for the recovered dominant waveform after the 1st and 3rd HS run. The match after the 3rd run is better in 62% of the cases, compared with 50% expected if the procedure had no effect at all. This small effect is also observed on the bias plot – see Fig. 7 – where the 3rd HS step leads to better results in the same proportions. Even if the results get better for some events, it also leads to worse recoveries for other cases, and only few of the other events have comparable results between the first and third HS steps. Therefore, it does not seem like simply applying successive HS steps converges to unbiased posteriors. More sophisticated approaches appear to be needed, like an estimation of the
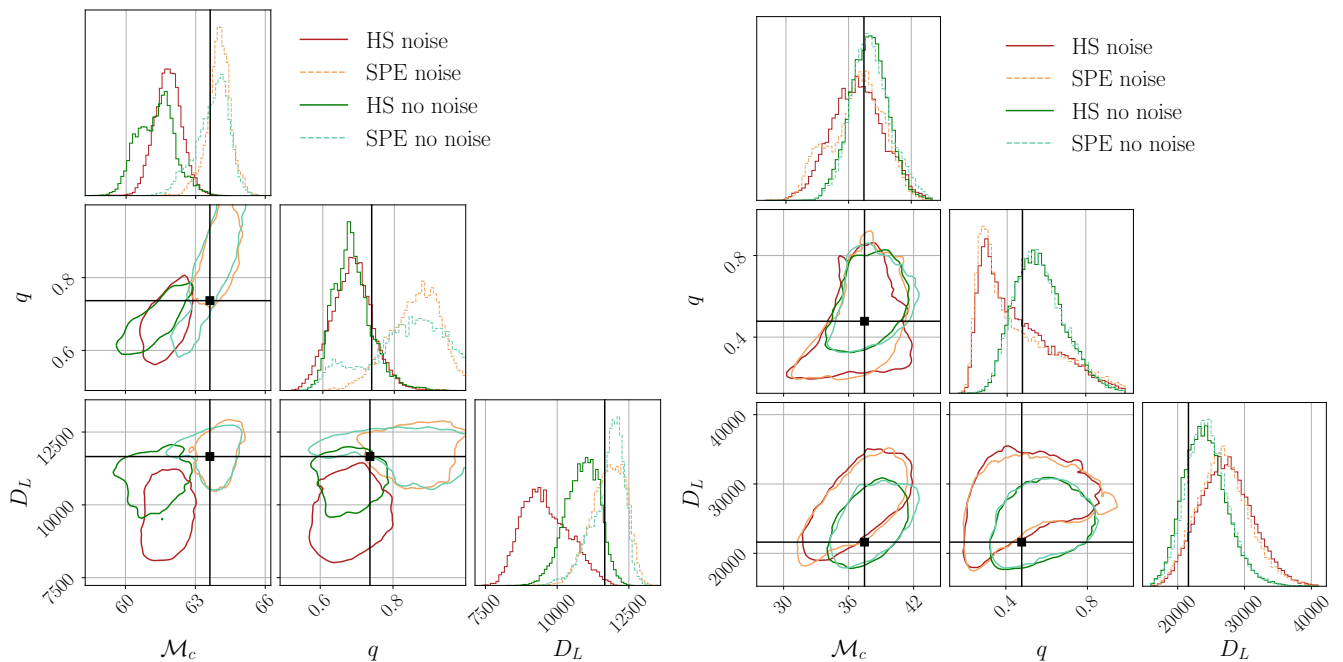
FIG. 2. Comparison of the posteriors for the HS and the SPE methods with and without noise. We represent the chirp mass ($\mathcal{M}_c$), the mass ratio (q), and the luminosity distance (D_L). *Left:* Case where HS is worse than SPE. HS posteriors are significantly shifted compared both to the injected value and SPE posteriors. *Right:* Case where HS is close to SPE, with equivalent posteriors recovered in the two cases.
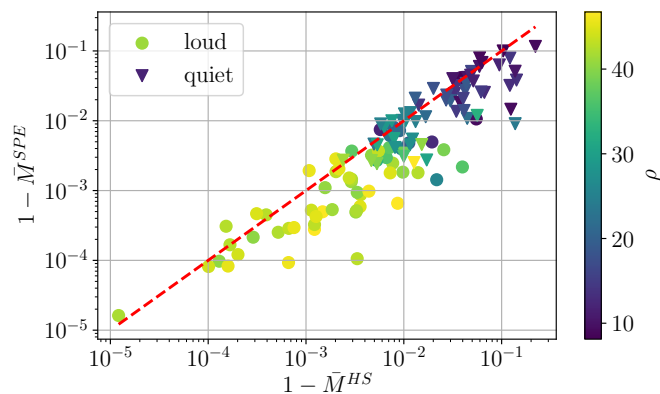

FIG. 3. Mismatch for the SPE analysis versus the HS analysis, with noise. The red dashed line represents the diagonal where the mismatch is the same. HS is worse at signal recovery, as most of the points fall below the diagonal.
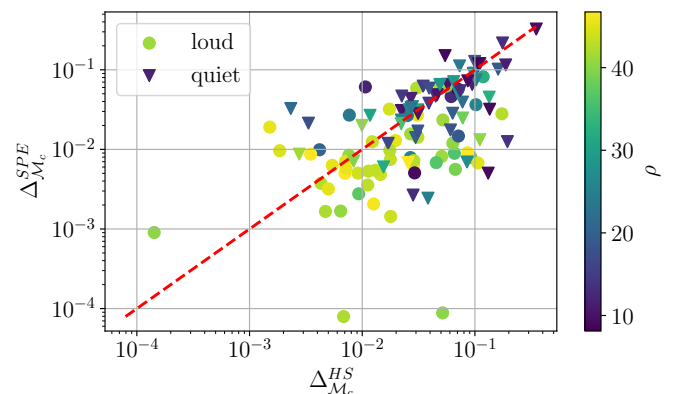
FIG. 4. Representation of the bias of the recovered parameters, with noise for HS and SPE. Plotted is the difference between injected value and the median of the recovered value, normalized by the injected value. HS tends to give higher bias than SPE.

bias as suggested in [24] and briefly explained in Section II B.

In the end, our runs for hierarchical subtraction confirm what has been seen in previous research [20–22]: doing PE for one signal neglecting the other can lead to significant biases when the two signals merge very close to each other. In addition, we have also shown that once the first signal is subtracted, analyzing the second one with parameter estimation is less prone to bias, even if the subtraction of the first event is not perfect.

B. Joint parameter estimation

In what follows, we will discuss the results obtained from the analysis of the overlapping signals using JPE. We focus on the results obtained after having done the time ordering of the samples described in Sec. II, as these are the samples where effectively one set of posteriors is matched with one signal, and the other set is matched with the other signal.
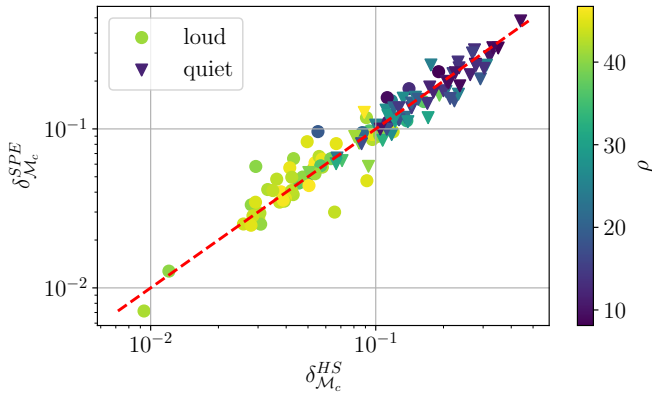
8

FIG. 5. Comparison of the normalized width of the 90% confidence interval for the chirp mass for the HS recovery and the SPE recovery. We see that width of the recovered distribution is largely unaffected by the presence of another signal.
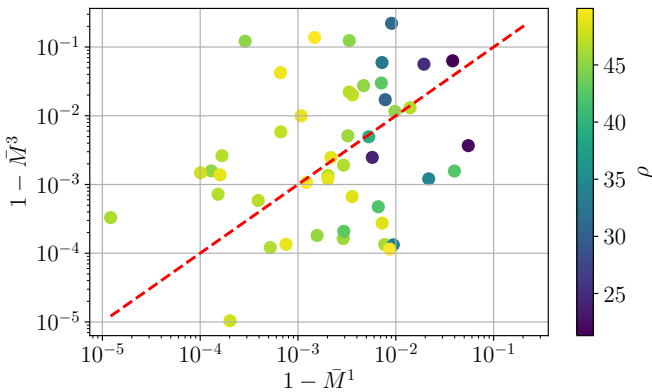

FIG. 6. Comparison of the mismatch for the recovered waveform between the 1st (horizontal axis) and 3rd (vertical axis) steps of HS. The mismatch after the 3rd step is lowered in 62% of the cases.
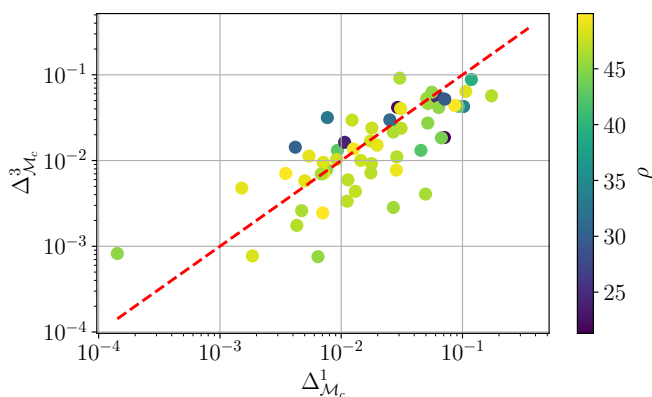

FIG. 7. Comparison of the bias in recovered chirp mass between the 1st (horizontal axis) and the 3rd (vertical axis) HS steps. The recovered bias after the 3rd step is lowered in 62% of the cases.

For the JPE recovery, more diverse scenarios are possible. We show 3 main cases in Fig. 8: one where the recovery is equivalent to the one from SPE, one where JPE has smaller bounds on the posteriors for one or more parameters and one where JPE has trouble fitting the signal properly and biases can occur. This seems to indicate that the behavior of the sampler will be different depending on the specific interaction between the overlapping signals. In some cases, the interference pattern between the two signals brings more information, while in others, it leads to more confusion. Typically, we get an increasing error on the recovery as the two overlapping BBH signals get closer in characteristics. In such a case, the sampling may lead to a systematic offset so that we get a narrower posterior but with the injected value out of the 90% confidence interval, or a widening of the posteriors.

For all the events, we compare the mismatch between the maximum likelihood waveform for the event in the JPE scenario with the maximum likelihood for the event in the SPE and HS cases. This is represented in Fig. 9 for the noise cases. Independent of the presence of noise, we find that the match is smaller for JPE than for HS but larger than for SPE. This is what one would expect since JPE accounts for the presence of the two signals and so should lead to smaller biases. However, fitting simultaneously two signals is more complex than analyzing a single signal. Therefore, the SPE measures remain a better representation of the injected signal.

As before, we present the normalized distance between injection and recovery in Fig. 10 for the noise case. When comparing the bias for the JPE case against the SPE case, we find that 45% of the events have a lower bias for JPE than for SPE. On the other hand, when we compare with HS, we find 65% of the events with a lower bias. This confirms that JPE is better than HS to find the injected signals (when two signals are present at 0.1s of each other in the data). This is indeed what one would expect, as JPE takes care of the mismodeling of the noise but leads to an increased complexity during the analysis.

For the spread in recovered posteriors, contrary to what one had for the HS approach, the normalized width of the 90% confidence interval does not align itself on the diagonal. Indeed, since we have more varying scenarios, with larger or tighter posteriors in some cases, the spread can be significantly larger or tighter in the JPE case compared to the SPE scenario. In addition, since there is no significant difference for this quantity between HS and SPE, the relation between JPE and HS is the same as between JPE and SPE. The increased discrepancy between the two approaches is represented in Fig. 11. Nevertheless, the posteriors are evenly distributed above and below the diagonal representation, showing that, on average, the width of the posteriors is not significantly bigger in one method or the other.

Based on our JPE results we see that, when two signals are close to each other, using an adapted likelihood as the one presented in Eq. (8) enables one to extract in-
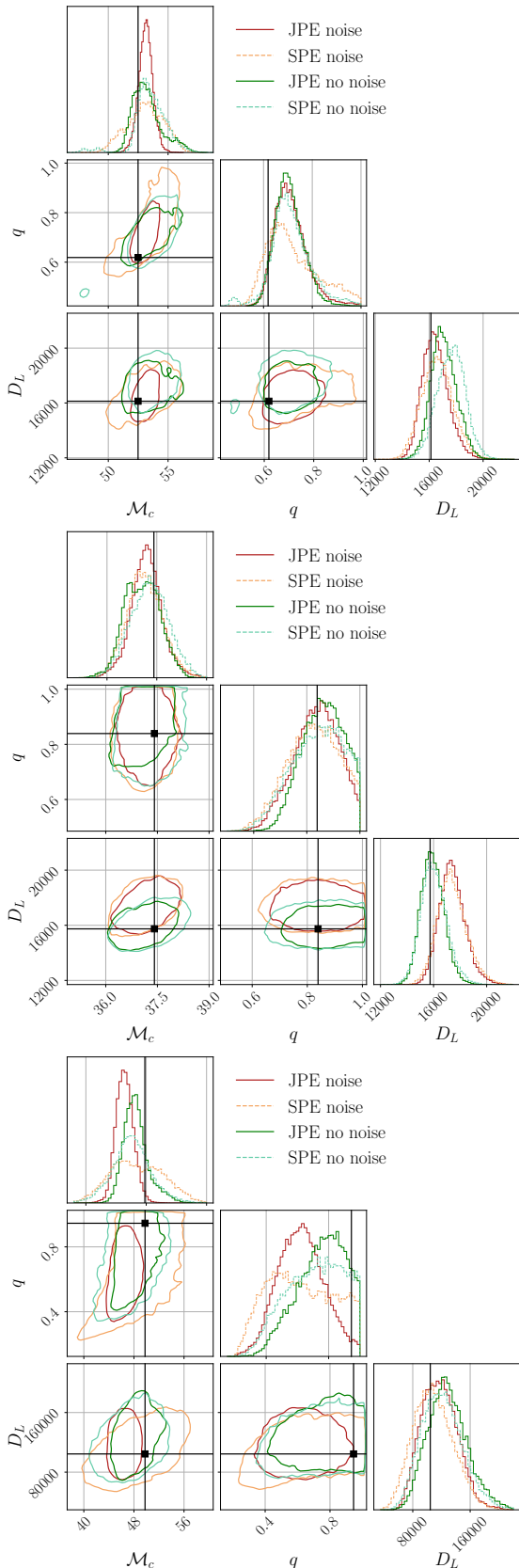
FIG. 8. Comparison of the posteriors recovered for JPE and SPE with and without noise for different types of recovery. *Top:* the posteriors recovered with JPE for the noise case is narrower for the chirp mass and the mass ratio compared to the single PE case, while the posterior for the luminosity distance is narrower for JPE in the zero-noise case. *Middle:* JPE and SPE are very close to each other, with equivalent recovery in the two cases. *Bottom:* representation of a case where the JPE recovery is worse than for SPE. We get narrower posteriors, but the peak is shifted out of the 90% confidence region.

formation from the interference between the two signals, and, in addition, to get better results than by analyzing both signals sequentially. However, this comes with the drawback that the computational time and the complexity of the problem are increased, making the approach a bit less stable as can be seen by the wider variations in the widths of the posteriors.

V. CONCLUSIONS

In this work, we have presented two ways to perform parameter estimation for two overlapping signals: hierarchical subtraction and joint parameter estimation. We have implemented them and compared them to the usual single-parameter estimation.

For the method of hierarchical subtraction, we analyze the first signal, then subtract it (more precisely, the maximum likelihood signal) from the data before analyzing the second signal still left in the data. On the other hand, joint parameter estimation consists of analyzing the two signals simultaneously. We note that since the likelihood is symmetric in the two signals, when they have the same nature, we need to add a post-processing step to have samples corresponding to each event. For this purpose, we order the samples in time.

We have applied both methods to a population of binary black hole mergers to show the feasibility of the two approaches and their respective drawbacks. For hierarchical subtraction, as already mentioned in the literature, we showed that analyzing one signal while neglecting the other can lead to bias in the recovered posteriors because of the erroneous noise representation. However, even when some bias occurs in the first signal, it does not necessarily impact the recovery of the second signal. We have also shown that there is no significant broadening of the posteriors compared to usual single-parameter estimation approaches. Therefore, it shows that hierarchical subtraction suffers from biases due to overlapping signals for signals measured at very close merger times, which can lead to an incorrect inference of the parameters.

For joint parameter estimation, we have shown that the bias in recovered posteriors is smaller than for hierarchical subtraction while remaining higher than for single parameter estimation. This is understood as joint parameter estimation has a correct noise representation for the two signals, which hierarchical subtraction lacks. On the other hand, solving the likelihood for joint parameter estimation means we need to explore a 30-dimensional parameter space, making the analysis more complex and computationally challenging than single parameter estimation. However, the recovered width of the posteriors is, on average, the same as for single parameter estimation. But while the average is the same, we are confronted with posteriors that can be narrower or broader in joint parameter estimation compared to single parameter estimation. This also shows that the method can extract
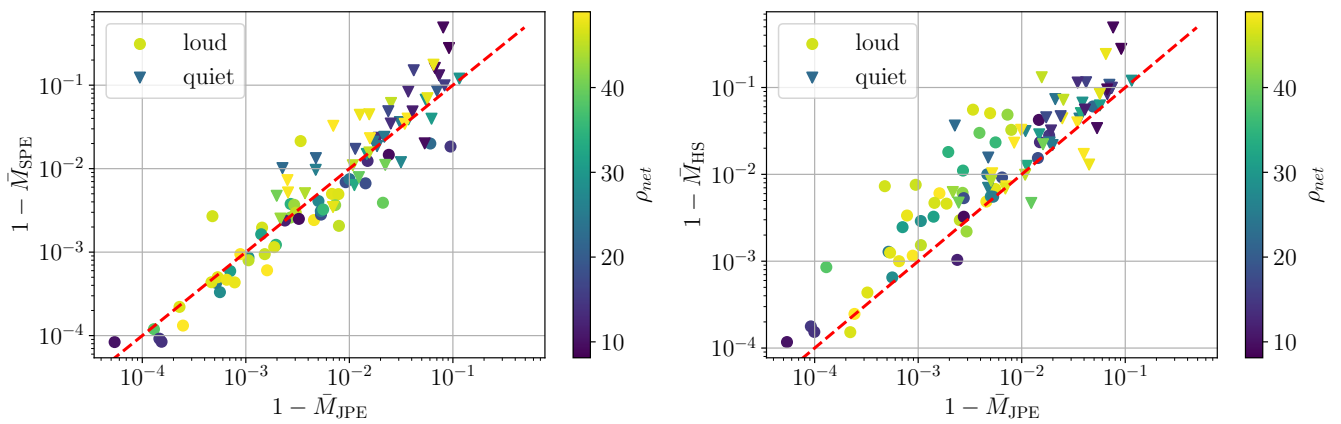
FIG. 9. *Left:* comparison of the mismatch for the JPE and the SPE cases with noise. *Right:* comparison of the mismatch for JPE and HS with noise. We see that overall, the mismatch is higher for JPE than for SPE, while it is lower than for the HS case. This is expected since JPE accounts for the two events in the data, which is better than neglecting one but more complicated than having only one signal present in the data and fitting that signal.
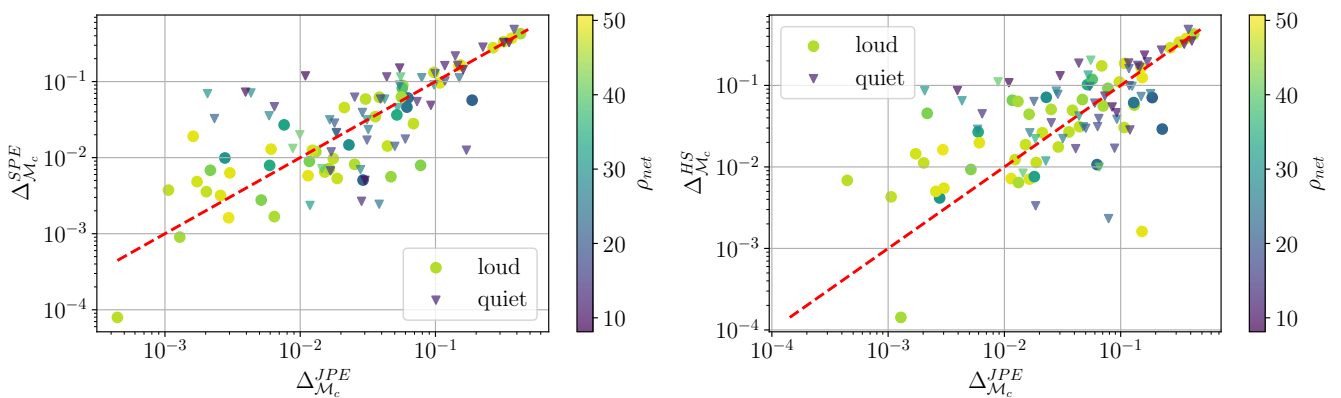

FIG. 10. *Left:* comparison of the bias of the recovered posterior for the chirp mass for JPE and SPE method. *Right:* comparison of the bias of the recovered posteriors for the chirp mass for JPE and HS. The two plots indicate that the bias is better in JPE than in HS, due to the better modeling of the noise, while it is still better in the SPE case, where the noise is well modeled and the problem at hand has a reduced complexity.

information from the interplay between the two signals in some cases (when the two signals have different characteristics) or can have more trouble (when the signals have close characteristics).

Overall, our results indicate that common techniques for a joint likelihood approach are not yet at their best, and several options to make the sampling more efficient could be possible for future work. For example, one could impose the time ordering (or chirp mass hierarchy) directly during the sampling, by imposing that the arrival time of one event is smaller than that of another event. This would prevent the sampler from confusing the two events and enable them to converge more easily. Another possibility could be to use narrower priors motivated by the output of the search pipelines. This is possible for the chirp mass and the time of arrival but still contains risk as the search pipelines themselves can get biased by overlapping signals [46].

One of the major issues with the methods suggested here is the computational time required, as the data analysis takes up to a few months for overlapping binary black hole mergers. This would make it extremely hard, if not impossible, to keep up with the detection rate of the 3G-detector network. However, methods exist to speed up traditional parameter estimation methods, such as relative binning [33–35] or adaptive frequency resolution [36]. These methods could be adapted to overlapping signals in future work to reduce the computational time, enabling one to analyze other types of systems and to use a lower minimum frequency to get closer to the real 3G scenario. A totally different approach that could help in the analysis of such signals in the future is machine learning, where major progress has been made in the parameter inference for single compact binary colescences [37, 49]. In the future, one could think of adapting these methods to overlapping signals.
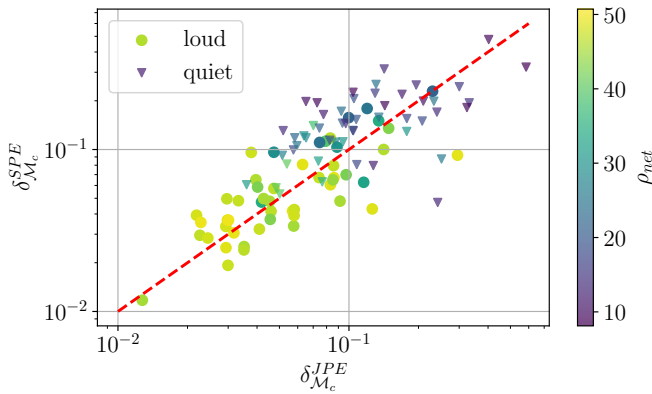
FIG. 11. Comparison between the normalized width of the 90% confidence interval for JPE and SPE for the noise case. Since the spreads are very close for SPE and HS, the same relation is established for the JPE and HS comparison. We see that there are larger differences here between the two approaches but that globally, one is not better than the other as we have about 50% of the events above the diagonal and the same proportion below.

To this end, we believe that this work makes a first step towards the analysis of overlapping compact binary coalescences, which will be crucial to analyze gravitational wave data in the third-generation detector era, as overlaps will become quite common.

ACKNOWLEDGMENTS

[1] T. L. S. Collaboration, Classical and Quantum Gravity **32**, 074001 (2015).

[2] F. Acernese *et al.* (VIRGO), Class. Quant. Grav. **32**, 024001 (2015), arXiv:1408.3978 [gr-qc].

[3] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2111.03606 [gr-qc].

[4] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2112.06861 [gr-qc].

[5] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2111.03634 [astro-ph.HE].

[6] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2111.03604 [astro-ph.CO].

[7] B. P. Abbott *et al.*, Astrophys. J. Lett. **848**, L12 (2017), arXiv:1710.05833 [astro-ph.HE].

[8] K. Somiya (KAGRA), Class. Quant. Grav. **29**, 124007 (2012), arXiv:1111.7185 [gr-qc].

[9] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto (KAGRA), Phys. Rev. D **88**, 043007 (2013), arXiv:1306.6747 [gr-qc].

[10] T. Akutsu *et al.* (KAGRA), Nature Astron. **3**, 35 (2019), arXiv:1811.08079 [gr-qc].

[11] T. Akutsu, M. Ando, K. Arai, *et al.*, arXiv e-prints , arXiv:2005.05574 (2020), arXiv:2005.05574 [physics.ins-det].

[12] B. Iyer, T. Souradeep, C. Unnikrishnan, S. Dhurandhar, S. Raja, and A. Sengupta, "LIGO-India Tech. rep. ," https://dcc.ligo.org/LIGO-M1100296/public (2011).

[13] M. Punturo *et al.*, Classical and Quantum Gravity **27**, 194002 (2010).

[14] S. Hild *et al.*, Class. Quant. Grav. **28**, 094013 (2011), arXiv:1012.0908 [gr-qc].

[15] D. Reitze *et al.*, Bull. Am. Astron. Soc. **51**, 035 (2019), arXiv:1907.04833 [astro-ph.IM].

[16] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, K. Ackley, and C. Adams, Classical and Quantum Gravity **34**, 044001 (2017).

[17] T. Regimbau, M. Evans, N. Christensen, E. Katsavounidis, B. Sathyaprakash, and S. Vitale, Phys. Rev. Lett. **118**, 151105 (2017).

[18] B. Sathyaprakash *et al.*, Class. Quant. Grav. **29**, 124013 (2012), [Erratum: Class.Quant.Grav. 30, 079501 (2013)], arXiv:1206.0331 [gr-qc].

[19] T. Regimbau and S. A. Hughes, Phys. Rev. D **79**, 062002 (2009), arXiv:0901.2958 [gr-qc].

[20] A. Samajdar, J. Janquart, C. Van Den Broeck, and T. Dietrich, Phys. Rev. D **104**, 044003 (2021), arXiv:2102.07544 [gr-qc].

[21] E. Pizzati, S. Sachdev, A. Gupta, and B. Sathyaprakash, Phys. Rev. D **105**, 104016 (2022), arXiv:2102.07692 [gr-qc].

[22] P. Relton and V. Raymond, Phys. Rev. D **104**, 084039 (2021), arXiv:2103.16225 [gr-qc].

[23] Y. Himemoto, A. Nishizawa, and A. Taruya, Phys. Rev. D **104**, 044010 (2021), arXiv:2103.14816 [gr-qc].

[24] A. Antonelli, O. Burke, and J. R. Gair, Mon. Not. Roy. Astron. Soc. **507**, 5069 (2021), arXiv:2104.01897 [gr-qc].

[25] S. Wu and A. H. Nitz, (2022), arXiv:2209.03135 [astro-ph.IM].

[26] Q. Hu and J. Veitch, (2022), arXiv:2210.04769 [gr-qc].

[27] S. Sachdev, T. Regimbau, and B. S. Sathyaprakash, Phys. Rev. D **102**, 024051 (2020).

[28] A. Sharma and J. Harms, Phys. Rev. D **102**, 063009 (2020).

[29] S. Biscoveanu, C. Talbot, E. Thrane, and R. Smith, Phys. Rev. Lett. **125**, 241101 (2020).

[30] B. Zhou, L. Reali, E. Berti, M. Çalişkan, C. Creque-Sarbinowski, M. Kamionkowski, and B. S.

Sathyaprakash, (2022), arXiv:2209.01310 [gr-qc].

[31] B. Zhou, L. Reali, E. Berti, M. Çalişkan, C. Creque-Sarbinowski, M. Kamionkowski, and B. S. Sathyaprakash, (2022), arXiv:2209.01221 [gr-qc].

[32] L. Reali, A. Antonelli, R. Cotesta, S. Borhanian, M. Çalişkan, E. Berti, and B. S. Sathyaprakash, (2022), arXiv:2209.13452 [gr-qc].

[33] B. Zackay, L. Dai, and T. Venumadhav, (2018), arXiv:1806.08792 [astro-ph.IM].

[34] L. Dai, T. Venumadhav, and B. Zackay, (2018), arXiv:1806.08793 [gr-qc].

[35] N. Leslie, L. Dai, and G. Pratten, Phys. Rev. D **104**, 123030 (2021), arXiv:2109.09872 [astro-ph.IM].

[36] S. Morisaki, Phys. Rev. D **104**, 044062 (2021), arXiv:2104.07813 [gr-qc].

[37] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, Phys. Rev. Lett. **127**, 241103 (2021), arXiv:2106.12594 [gr-qc].

[38] M. J. Williams, J. Veitch, and C. Messenger, (2021), arXiv:2102.11056 [gr-qc].

[39] S. Gao, F. Hayes, S. Croke, C. Messenger, and J. Veitch, Phys. Rev. Res. **4**, 023006 (2022), arXiv:2109.01535 [quant-ph].

[40] J. Veitch and A. Vecchio, Phys. Rev. **D81**, 062003 (2010), arXiv:0911.3820 [astro-ph.CO].

[41] T. Littenberg, N. Cornish, K. Lackeos, and T. Robson, Phys. Rev. D **101**, 123021 (2020), arXiv:2004.08464 [gr-qc].

[42] C. Cutler and J. Harms, Phys. Rev. D **73**, 042001 (2006).

[43] A. Petiteau, S. Babak, A. Sesana, and M. de Araújo, Phys. Rev. D **87**, 064036 (2013).

[44] M. Oguri, Mon. Not. Roy. Astron. Soc. **480**, 3842 (2018), arXiv:1807.02584 [astro-ph.CO].

[45] T. Regimbau *et al.*, Phys. Rev. D **86**, 122001 (2012), arXiv:1201.3563 [gr-qc].

[46] P. Relton, A. Virtuoso, S. Bini, V. Raymond, I. Harry, M. Drago, C. Lazzaro, A. Miani, and S. Tiwari, (2022), arXiv:2208.00261 [gr-qc].

[47] G. Ashton *et al.*, Astrophys. J. Suppl. **241**, 27 (2019), arXiv:1811.02042 [astro-ph.IM].

[48] J. S. Speagle, MNRAS **493**, 3132 (2020), arXiv:1904.02180 [astro-ph.IM].

[49] A. Kolmus, G. Baltus, J. Janquart, T. van Laarhoven, S. Caudill, and T. Heskes, Phys. Rev. D **106**, 023032 (2022), arXiv:2111.00833 [gr-qc].

Appendix A: Zero-noise results

In this section, we show results for the posteriors obtained when the injections are analyzed without noise. The conclusion drawn from these plots are the same as in the noise case, which suggests that our findings are robust, due to sampling effects, and not induced by the random noise added to the data.

1. Hierarchical Subtraction

In this section, we present the complementary zero-noise result for the HS method.

Fig. 12 represent the mismatch for HS versus SPE for the zero-noise case. As for the noise case, HS leads to higher mismatches, meaning that the recovered (maximum likelihood) parameters are a worse representation of the injected signals.
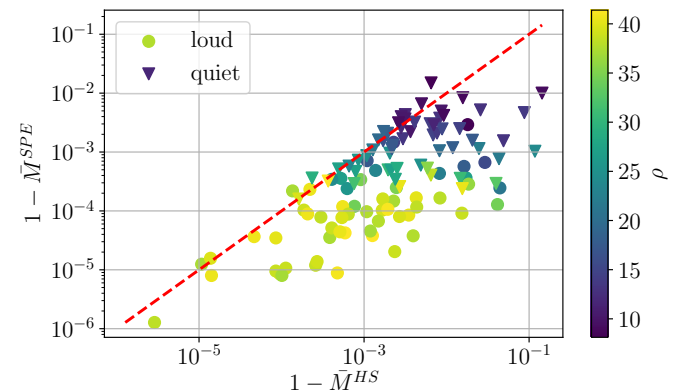

FIG. 12. Mismatch for the SPE analysis versus HS analysis without noise. The red dashed line represents the diagonal where the mismatch is the same. The difference in waveform recovery between SPE and HS is more pronounced than in the case with noise, and it is clear that the recovery is worsened when using the HS approach.

Fig. 13 represent the bias for the chirp mass for HS versus SPE without noise. Here HS shows higher bias for 74% of the louder events and for 57% of the quieter ones. There is no significant difference compared with the injections into noise, and HS is still prone to more biases.

Fig. 14 represents the normalized width of the posteriors for HS versus SPE. Similarly to the analysis with noise, the widths of the distributions are very close to each other.

2. Joint Parameter Estimation

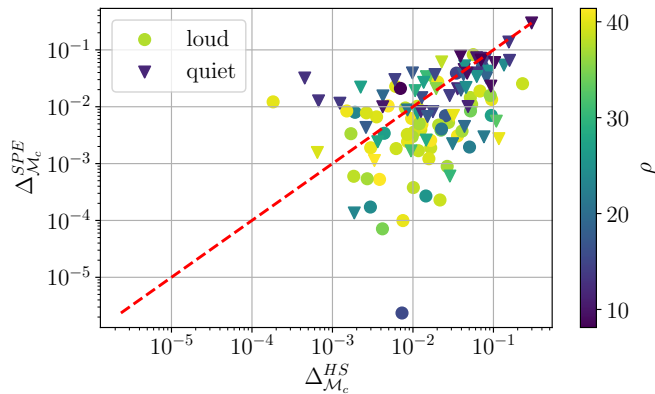In this section, we present the complementary zero-noise results for the JPE method.

FIG. 13. Representation of the bias of the recovered chirp mass without noise. Plotted is the difference between injected value and the median of the recovered value, normalized by the injected value for HS and SPE. One sees that bias is more important for HS.
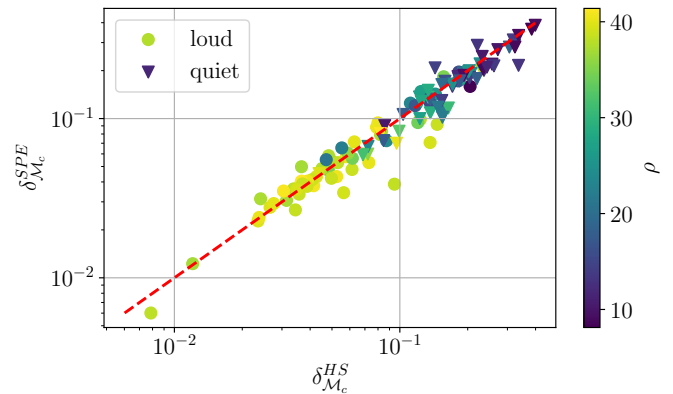

FIG. 14. Comparison of the normalized width of the 90% confidence interval for the chirp mass for the HS recovery and the SPE recovery, with zero-noise. We see that width of the recovered distribution is largely unaffected by the presence of another signal.

Fig. 15 represent the mismatch between the JPE and SPE methods (left) and the JPE and the HS methods (right). This also shows that JPE method leads to a better representation of the data than the HS method, but that the increased complexity of the problem leads to a decrease in the accuracy of the recovery. Still, the mismatch values are relatively low in the two cases.

Fig. 16 represent the comparison between the biases between JPE and SPE, and JPE and HS for the zero-noise case. Here, one also has a larger bias for the JPE than for SPE (39% of the events have smaller bias for JPE), and a larger bias for HS as for JPE (57% of the events have a smaller bias for JPE).

Fig. 17 represents the normalized width of the posteriors for JPE versus SPE (which is comparable to JPE versus HS since HS versus SPE has widths aligning along the diagonal). There is more variance in this plot than for SPE vs HS. This is because the JPE method is significantly different from SPE, and we have a bigger variety of posteriors. Indeed, for JPE, we sometimes get broader posteriors but also tighter once, depending on the characteristic of the two signals present in the data. However, the points are evenly distributed the two sides of the diagonal, showing that, on average, none of the methods has widened posteriors compared to the other.
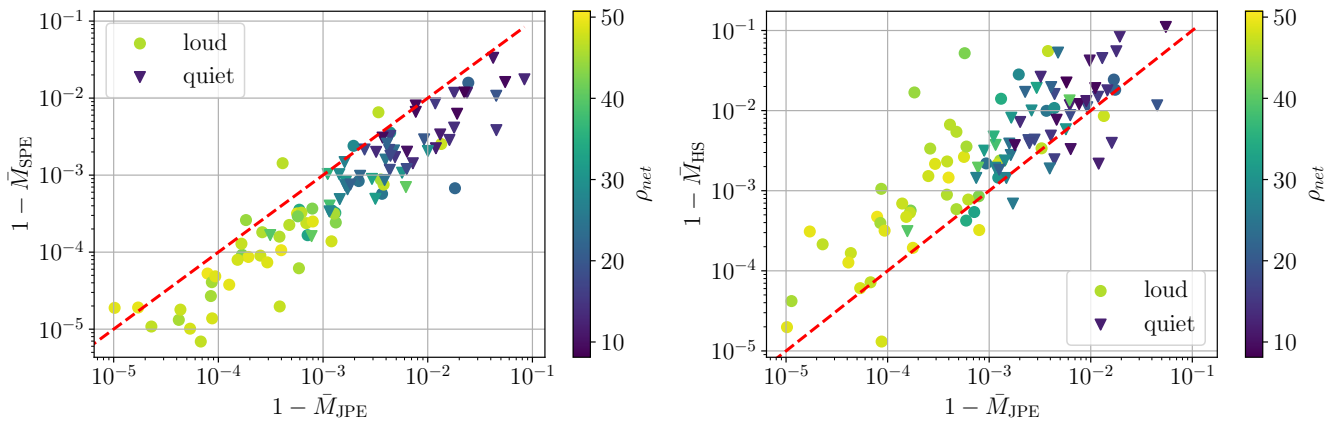
FIG. 15. *Left:* comparison of the mismatch for the recovered events for the two events for JPE and SPE without noise. *Right:* comparison of the mismatch for JPE and HS without noise. Overall, the mismatch is higher for JPE than for SPE, while it is lower than for the HS approach. This is expected since JPE accounts for the two events in the data, which is better than neglecting one but more complicated than having only signal present in the data and fitting that signal.
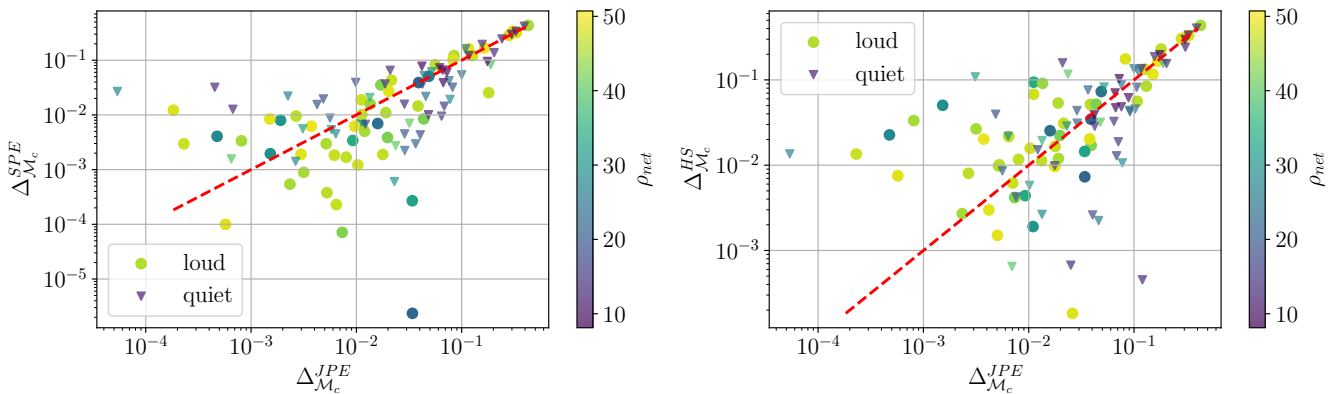

FIG. 16. *Left:* comparison of the bias of the recovered posterior for the chirp mass for JPE and SPE method in the zero-noise case. *Right:* comparison of the bias of the recovered posteriors for the chirp mass for JPE and HS in the zero-noise case. The two plots indicate that the bias is lower for JPE than for HS, due to the better modeling of the noise, while it is still better in the SPE case, where the noise is well modeled and the problem at hand has a reduced complexity.
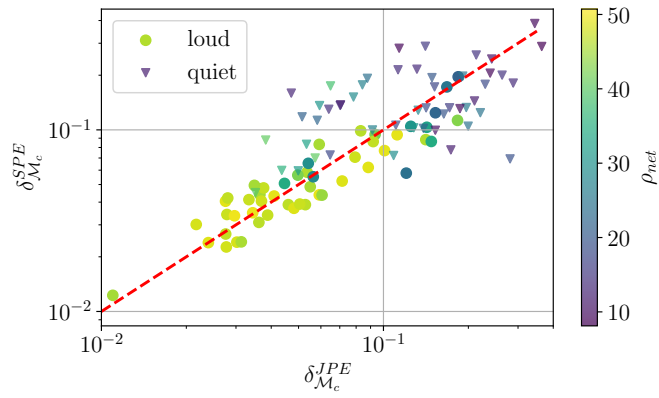
FIG. 17. Comparison between the normalized width of the 90% confidence interval for JPE and SPE in absence of noise. Since the spreads are very close for SPE and HS, the same relation is established for the JPE and HS comparison. We see that the there are larger difference here between the two approaches but that globally, one is not better than the other as we have about 50% of the events above the diagonal and the same proportion below.