

The Detection of Black Hole Encounter Gravitational Waves Interpreted as Binary Black Hole Using Machine Learning

Weichangfeng Guo,^{1,*} Ik Siong Heng,^{2,†} Daniel Williams,^{2,‡} and Hunter Gabbard^{2,§}

¹*Department of Astronomy, Beijing Normal University, Beijing 100875, China*

²*SUPA, School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom*

(Dated: April 20, 2021)

Black hole(BH) encounter has been regarded as an important sources of gravitational waves(GWs), of which the formation processes is a gravitational radiation driven capture (GR capture). The BH binaries that form this way have significant eccentricities, and lack a long inspiral phase. When a binary black hole(BBH) merger system reaches such a high mass that its inspiral portion is below the frequencies that the detectors are sensitive, it may appear similar to encounters in current detectors. Therefore, future GW events may be misclassified. We inject large-scale encounter waveforms using Minke and adopt **Vitamin** network pre-trained on non-spinning BBH mergers for Bayesian parameter estimates. We adjust the parameters of the injections (especially mass and distance), so that their posteriors obtained by **Vitamin** appear to be similar to the ones for a BBH signal. Then we run Bilby on the same injections to see how they can mimic high mass binary black holes merger. Here we show that with suitable parameters, an encounter performs very similar to BBH signal in Bilby's posterior. Compared with the non-spinning model, BBH injection using maximums of posterior obtained by the spinning model can better imitate the encounter.

* Corresponding author: 201921160007@mail.bnu.edu.cn

† ik.heng@glasgow.ac.uk

‡ daniel.williams@glasgow.ac.uk

§ hunter.gabbard@gmail.com

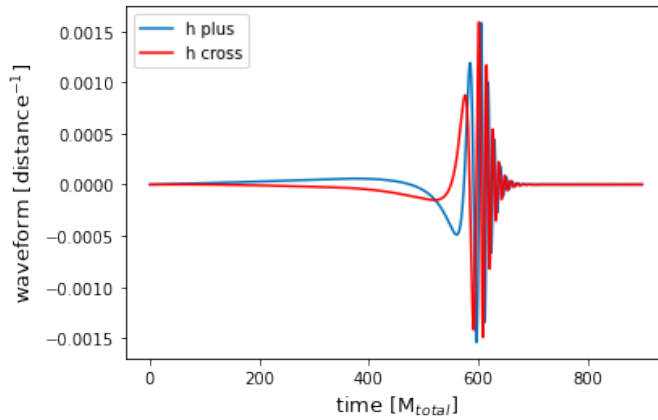


FIG. 1: A parabolic encounter waveform from [199] for a system with mass-ratio $q = 1$.

I. INTRODUCTION

Binary black holes (BBH) merger has been regarded as one of the most promising sources of gravitational waves (GWs) [1]. One of the formation processes of the compact BBH is a gravitational radiation driven capture (GR capture). Capture takes place when the amount of energy radiated by gravitational waves during the encounter becomes larger than the initial orbital energy. One of the consequences of GR captures is the possibility of forming very eccentric binaries. Note that most BBHs are expected to be circularized before the merging due to the loss of the orbital energies during the inspiral phase. Therefore the analyses of GW data and the parameter estimations of GW sources have focused on the binaries with circular orbits. Eccentricity has not been detected in the O1 and O2 observing runs of LIGO/Virgo. However, one BBH merger event in O3a, GW190521, has been shown to be possibly highly eccentric. Encounter waveforms lack a long inspiral phase. For compact binary coalescences (CBC), higher total mass makes lower frequency. When the system reaches such a high mass that its inspiral portion is below the frequencies that the detectors are sensitive, high-mass BBH events may appear similar to encounters in current detectors. Therefore, future GW events that lack the inspiral portion may be misclassified, and this eventuality should be considered in waveforms which seem to lack a clear inspiral. These results are principally low-frequency sources, and as such are ideal candidates for both ET, which aims to achieve much greater low-frequency sensitivity than current detectors, but also for decihertz detectors, such as DECIGO.

The outline of this paper is as follows. In section II, we introduce BH encounter waveforms and the data production with specific distributions. In section ??, we show the math method for data analysis and the physics model we have adopted, while a machine learning approach that we applied in this case are described. The main results of our parameter estimation about BH encounter are presented in section IV. And in section V the application of our research is discussed.

II. DATA

GW astronomy is developing rapidly, so far the astrophysical sources of GWs have already been known and studied, such as continuous wave sources (e.g. rotating neutron star), stochastic backgrounds, CBC (e.g. binary black hole and binary neutron star merger), and bursts, which are unmodelled and poorly modelled transient sources. For burst signals, there is a type which has not been widely considered: encounters between black holes, and the potential capture of one black hole by another which results in a BBH coalescence event. This type of events is proved to be detectable with the current generation of detectors in galaxies beyond the Local Group, and to be observed by burst search algorithms.

introduce BBH merger, 3 portions.

The waveforms for this study are taken from GW driven capture simulations [2]. The parabolic approximation assuming that the orbits are only weakly hyperbolic is adopted, which is applicable for non-spinning pairs of BHs with relative velocity up to $10 \sim 20$ % of the speed of light. One of these waveforms is plotted in figure ref.

In general, BH encounter's amplitude is ~ 2 orders of magnitude smaller than BBH merger's of the same total mass and distance.

Besides, due to a varying sensitivity of an interferometer has to sources across the sky, the location of the source

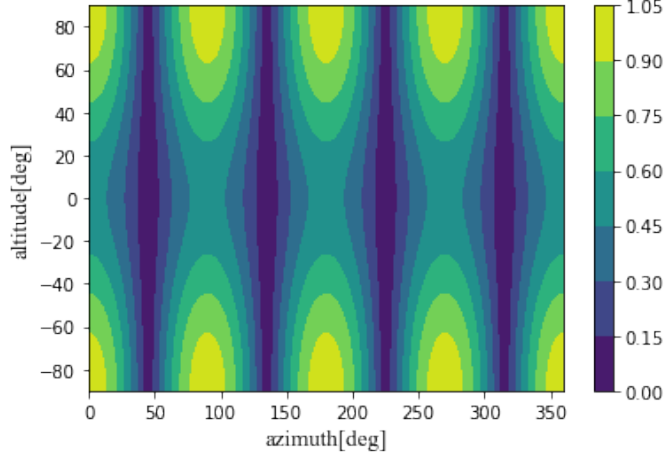


FIG. 2: The normalised antenna pattern F_+ .

also has a great influence on our observation about signal. As we know, the overall measured strain, $h(t)$ in a detector from a GW with components (h_+, h_\times) is:

$$h(t) = F_+(t)h_+(t) + F_\times(t)h_\times(t), \quad (1)$$

where F_+ and F_\times are the antenna pattern functions for the two polarizations of the GW.

$$F_+ = \frac{1}{2}(1 + \sin^2 \delta) \cos 2\alpha \cos 2\psi - \sin \delta \sin 2\alpha \sin 2\psi, \quad (2)$$

$$F_\times = \frac{1}{2}(1 + \sin^2 \delta) \cos 2\alpha \sin 2\psi - \sin \delta \sin 2\alpha \cos 2\psi, \quad (3)$$

where δ is the altitude relative to the plane of the detector, α is the azimuth relative to one of the arms, ϕ is , and ψ is the polarisation angle of the GW, which corresponds to the rotation of the basis vectors defining the polarisations of the GW compared to the detector. F_+ is plotted in figure ??, which depicts the four regions of low sensitivity. Therefore, we want to generate encounters across the sky, and encounters in the conventional prior range, it is necessary to generate a large number of samples for selection.

We use **Minke**, a mock data challenge engine for burst searches, which is designed to produce large-scale signals with numerical relativity (NR) derived waveforms. The process for making injection using **Minke** are as follows.

1. Set injection parameters. The distributions of source parameters and number of samples are specified.
2. The signal can then be generated using **LALSimulation**.
3. The signal is convolved with the antenna pattern at each detector, which gives the antenna response.
4. Finally the injection is combined with detector noise.

The detectors we use are H1, L1, and V1. The distribution of waveform parameters of step 1 are shown in Table I. For 5 waveforms with mass ratio $q = 1, 2, 4, 8, 16$, each has a combination of 100 locations (ra, dec, ψ), thus the data set contains 500 samples.

Since the observed masses of the stellar mass BHs span from about 5 to 30 M_\odot as evident from the X-ray binaries and GW150914, encounters between unequal mass BHs would be more common than equal mass ones. We focus on this kind of cases more.

III. METHODS

The current method of parameter estimation for the GW signal is Bayesian inference, such as **Bilby**, a modular python package, while it requires a lot of computing power. We use **Vitamin**, a recently proposed network for BBH

TABLE I: The distribution of parameters used on injecting BH encounter signals

Parameter	min	max	units
m_{total}	150		M_{\odot}
d_L	5,10,15	^a	Mpc
ra	0	2π	rad.
dec	$-\pi/2$	$\pi/2$	rad.
ψ	0	2π	rad.

^a Additionally m_2 is constrained such that $m_2 < m_1$.

signal based on conditional variational autoencoder (CVAE), which has been proven to produce samples describing the posterior distribution by 6 orders of magnitude faster than the former approach. We use **Bilby** to inject BBH signal and train the given prior parameter space of the **Vitamin** to be wider. Then, we adjust the parameters of the injected BH encounter waveforms (especially mass and distance), so that their posteriors obtained by **Vitamin** appear to be similar to the ones for a BBH signal. Then we run **Bilby** on the same injections of encounters to see how they can mimic high mass binary black holes merger.

A. Bayesian inference

(This part needs to be supplemented)

According to Bayes' Theorem, a posterior probability distribution on a set of parameters, conditional on the measured data, can be represented as

$$p(x|y, H) \propto p(y|x, H)p(x, H), \quad (4)$$

where x are the parameters, y is the observed data, H is the model, $p(x|y, H)$ is the posterior, $p(y|x, H)$ is the likelihood, and $p(x, H)$ is the prior on the parameters. The constant of proportionality, which we omit here, is $p(y)$, the probability of our data, known as the Bayesian evidence or the marginal likelihood. We typically ignore $p(y)$ since it is a constant and for parameter estimation purposes we are only interested in the shape of the posterior.

Due to the size of the parameter space typically encountered in GW parameter estimation and the volume of data analysed, we must stochastically sample the parameter space in order to estimate the posterior. Sampling is done using a variety of techniques including Nested Sampling and Markov chain Monte Carlo methods. The primary software tools used by the LIGO parameter estimation analysis are **LALInference** and **Bilby**, which offer multiple sampling methods. **Bilby** is a Bayesian inference library for GW astronomy.

We also calculate the Bayes factors K of Bayesian Inference under non-spinning BBH-template and spinning BBH-template, to make model comparison. The Bayes factor is defined as

$$K = \frac{p(y|x, H_1)}{p(y|x, H_2)} = \frac{\int p(x_1|H_1)p(y|x_1, H_1)dx_1}{\int p(x_2|H_2)p(y|x_2, H_2)dx_2} \quad (5)$$

where H_1, H_2 are two different models. In this paper H_2 is considered as a null hypothesis. Hence $K > 1$ means that H_1 is more strongly supported by the data than a null hypothesis.

B. Waveform Model

We search BH encounters using **IMRPhenomPv2**, a phenomenological model for GWs from precessing BBH systems, adopting an effective single-spin description to model effects from spin precession. It includes all the three portions of the process. In most of the inspiral, the binary orbits at a speed far below the relativistic velocity, and the waveform of this part is produced by the post-Newton(PN) approximation with a good effect. For the inspiral's last few cycles, as well as the merger and ringdown phases, waveforms are generated by numerical relativity(NR). Finally, through well-matched 'hybridisation', these PN waveforms and NR waveforms are connected. The resulting hybridised waveforms are then parameterised in the Fourier domain.

The phenomenological waveform in the Fourier domain takes the form

$$u(f) = \mathcal{A}_{\text{eff}}(f) e^{i\Psi_{\text{eff}}(f)}. \quad (6)$$

for $\mathcal{A}_{\text{eff}}(f)$ the amplitude of the waveform in the frequency domain, which can be written as a piecewise function

$$\mathcal{A}_{\text{eff}}(f) = C \begin{cases} (f/f_{\text{merge}})^{-7/6} & \text{if } f < f_{\text{merge}} \\ (f/f_{\text{merge}})^{-2/3} & \text{if } f_{\text{merge}} \leq f < f_{\text{ring}} \\ w \mathcal{L}(f, f_{\text{ring}}, \sigma) & \text{if } f_{\text{ring}} \leq f < f_{\text{cut}} \end{cases} \quad (7)$$

where f_{merge} , f_{ring} , f_{cut} are respectively the initial merger frequency, initial ringdown frequency, and the cutoff frequency of the template. \mathcal{L} is a Lorentzian distribution of width σ , and w is a normalisation constant describing the quasi-normal mode frequencies, and C is a numerical constant.

$$\mathcal{L}(f, f_{\text{ring}}, \sigma) = \left(\frac{1}{2\pi} \right) \frac{\sigma}{(f - f_{\text{ring}})^2 + \sigma^2/4}, \quad (8)$$

for $\Psi_{\text{eff}}(f)$ the effective phase, which is an expansion in powers of f .

$$\Psi_{\text{eff}}(f) = 2\pi f t_0 + \phi_0 + \psi_0 f^{-5/3} + \psi_2 f^{-1} + \psi_3 f^{-2/3} + \psi_4 f^{-1/3} + \psi_6 f^{1/3}, \quad (9)$$

where t_0 is the time of arrival, ϕ_0 is the frequency-domain phase offset, and $\psi = \{\psi_0, \psi_2, \psi_3, \psi_4, \psi_6\}$ are the “*phase parameters*”, that is the set of phenomenological parameters describing the phase of the waveform.

When we use the model to inject BBH signal for following machine learning training, the injection parameters includes the component masses m_1, m_2 , the luminosity distance d_L , the sky position ra and dec , the binary inclination Θ_{jn} , the GW polarisation angle ψ , the time of coalescence t_0 , the phase at coalescence ϕ_0 , the primary spin magnitude a_1 , the secondary spin magnitude a_2 , the primary tilt angle θ_1 , the secondary tilt angle θ_2 , the azimuthal angle between the two component spins $\Delta\phi$, and the azimuthal angle between the total binary angular momentum and the orbital angular momentum ϕ_{JL} . The last six spin parameters are omitted and set to zeros when producing training data and are not involved in the current VItamin model.

the optimal signal to noise ratio (SNR) of a GW detection is defined as

$$\rho_{\text{opt}} = 2 \left[\int_{f_{\text{min}}}^{f_{\text{max}}} \frac{|h^2(f)|}{S_h(f)} df \right]^{\frac{1}{2}} \quad (10)$$

where $h(f)$ is the Fourier transform of the GW signal, $S_h(f)$ is the one-sided noise spectral density in units of Hz^{-1} , and $f_{\text{min}} \leq f \leq f_{\text{max}}$ correspond to the frequency band of the instrument, e.g. $(f_{\text{min}}, f_{\text{max}}) \approx (10, 10^4) \text{ Hz}$ for Advanced LIGO. It can be seen that the contribution to the SNR from a frequency bin where the noise PSD is high is smaller than from a bin where the noise PSD is low. Thus, an optimal SNR takes into account the nature of the noise PSD.

Optimal SNR of injection in each detector was calculated. Usually an SNR of 8 is a standard threshold which a signal must exceed to be considered a detection. Virgo’s sensitivity is kind of inferior to Ligo, of which the optimal SNR is allowed to be slightly lower, but also greater than 5.

C. VItamin Search

The Bayesian inference approach is optimally sensitive but consumes expensive computing resource. For one GW signal in this study, a typical Bayesian estimation using *Bilby* takes 2 to 5 hours. As a comparison, a recent neural network model, **VItamin**, costs only ~ 10 seconds dealing with that, which is 3 orders of magnitude faster than the existing inference technology. **VItamin** is a conditional variational auto encoder (CVAE) [] pre-trained on BBH signals that can return Bayesian posterior probability estimates. The training procedure need only be performed once for a given prior parameter space, which takes $\mathcal{O}(1)$ day to complete, and the resulting trained machine can then quickly generate samples describing the posterior distribution, which is proved to achieve the same quality of results as the trusted benchmark analyses used within the LIGO-Virgo Collaboration.

Currently **VItamin** omits the 6 additional parameters() required to model the spin of each BBH component mass.produces posteriors on 7 parameters ($m_1, m_2, d_L, t_0, \Theta_{jn}$, ra and dec). Both ϕ_0 and ψ are internally marginalized out. **VItamin** allows users to train network with customized prior.

The training prior ranges of the network were deliberated simultaneously with the injection parameters of encounter waveform. On the one hand, since the encounters are expected to act like high-mass BBH in the estimation, the total

TABLE II: The uniform prior boundaries and fixed parameter values used on the BBH signal parameters for the benchmark and the CVAE analyses

Parameter	min	max	units	prior
$m_{1,2}$	30	160	M_\odot	uniform
d_L	1000	3000	Mpc	uniform
t_0	0.15	0.35	seconds	uniform
ra	0	2π	rad.	uniform
dec	$-\pi/2$	$\pi/2$	rad.	cosine
Θ_{jn}	0	π	rad.	sine
spins	0		-	-
epoch	1126259642		GPS time	-
detector network	H1, L1, V1			

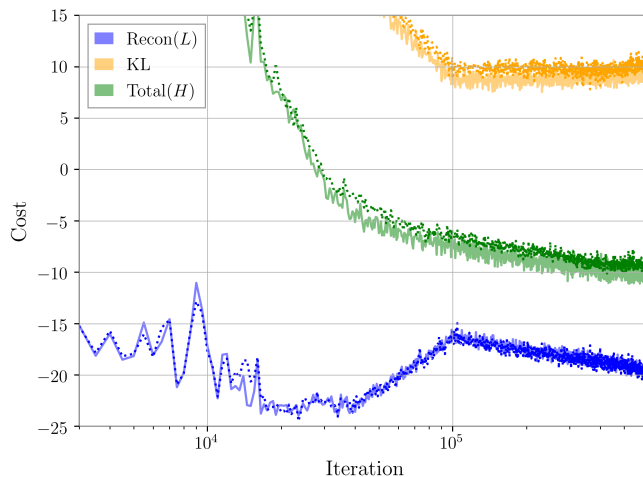


FIG. 3: The cost as a function of training iteration.

mass which determines the frequency should be large enough. In this study we set the total mass to $150 M_\odot$. On the other hand, according to Nyquist sampling theorem, the maximum frequency component of the signal is required to be lower than half of the sampling rate. For a binary system with equal mass in a circular orbit, when the slow inspiral phase ends and the stars begin to interact and merge, the frequency reaches the magnitude of the maximum. Correspondingly, the last stable orbit (LSO) frequency can be estimated as

$$f_{LSO} \sim 4400 \left(\frac{M_\odot}{m_{total}} \right) \text{ Hz} \quad (11)$$

where m_{total} is the total mass. Even though it's an expression for binary who has a circular orbit, it provides a remarkably good order of magnitude prediction. For binary systems with asymmetric component masses the situation gets more complicated [], but for simplicity we adopt Eq. (11) also. Considering $m_{total}=150M_\odot$, then we have $f_{LSO} \sim 29.3\text{Hz}$. It's much lower than half of the **VItamin** sampling rate, which is constrained to be 256 Hz at present.

2 $m_1, m_2, d_L, \text{ra, dec, } \Theta_{jn}, \psi$, the time of coalescence t_0, ϕ_0 . For each parameter we use a uniform prior with the exception of the declination and inclination parameters for which we use priors uniform in $\cos(\text{dec})$ and $\sin \Theta_{jn}$ respectively. The corresponding prior ranges are defined in Table II.

We produce BBH signals using **IMRPhenomPv2** with a minimum cutoff frequency of 20Hz. For each input test waveform we run the benchmark analysis using multiple sampling algorithms available within **Bilby**. For each run and sampler $\mathcal{O}(10^4)$ samples are extracted from the posterior on the 9 physical parameters.

We track the performance of the network during training via loss curves: a time-series duration of 1 second for training data.

For five encounter waveforms with different mass-ratio, the length of the injected signal ranges from 226 to 344. Therefore we manually padded/cut the signal, to fit the 256Hz and one second sampling constraint of **VItamin**. At

the same time we tried to put signal's peak as centered as possible, in this case the Bayesian inference has the best effect. That is, for a signal with a peak on the left side, if its length is greater than 256, we cut more on the right end. Otherwise, we pad more from the left end with zeros.

To make network function properly, signals were processed in the same way as the training data. During this period, whitening process was adopted for raw signals to treat the noise problem and normalise the data for input to the neural network. After whitened, the extra noise at low frequencies and at the spectral lines were suppressed, the weak signals in the most sensitive band could be better seen. In addition, Gaussian white noise of unit variance and zero mean was added to whitened signal, where we obtained the final input data.

The input signal that **VI**tamin would expect is

$$h_{VI\text{tamin}}(t) = h_{\text{whitened}}(t) + n(t) \quad (12)$$

where $n(t)$ is Gaussian white noise of unit variance and zero mean, and the whitened signal

$$h_{\text{whitened}}(f) = \frac{h(f)}{S_h(f)} \quad (13)$$

We used ALIGO zero detuning, high power design sensitivity ASD for H1 and L1, advanced Virgo ASD for V1. During Fast Fourier transform, windowing function wasn't added, since both ends of the signal are zeros and the effect of spectral energy leakage was little. In addition, truth lines function are added to **VI**tamin so that any biases are clear to see from the output corner plot.

Once the processed signals were input to **VI**tamin sampler, it produced posteriors for all 500 samples. Depending on the sky location of the sources, some signals looked too close or too far. For each waveform, we adjusted the distance so that most of the signals were within the training distance prior range, by this means proper distances were acquired for injection in Table I. Thus, the distances for waveforms with different mass ratio q was appropriately set to ensure that most of the signals have a suitable distance in the view of **VI**tamin. We looked for the ones with sharp peaks in m_1 , m_2 , and d_L , and the sky location of which were also required to be close to the truths. For each waveform, we selected at least one signal, and made BBH parameter estimation (PE) runs using spinning and non-spinning models respectively, during which Dynesty sampler was adopted, the explicit time, and phase marginalizations were turned on to improve convergence. We used the maximum from posteriors had been inferred by the two models to inject BBH. At last, two kinds of BBH signals were plotted together with the corresponding encounter to see the imitation effect of **IMRPhenomPv2** template on the latter type of GWs.

IV. RESULTS

V. DISCUSSIONS

we've established that are scenarios in which parameter estimation of a black hole encounter could be interpreted incorrectly as a binary black hole with a very different set of physical parameters, and that therefore it would be valuable to be vigilant to this possibility when a high-mass BBH system is identified in an analysis. We point that, the maximum a posteriori result from the BBH-template PE has a considerable difference to the true injected values of the encounter waveform.

bayes factor indicates that BBH model is decisive for BH encounter signals.

spinning BBH-template is a more complex model than non-spinning one due to six free parameters, thus it can model the data more closely. The ability of Bayes factors to take this into account

(conclusion)

(application)

The most massive BHs dominate the detection rate of future ground based GW detectors with $\sim (10 - 10^3)\xi_{30}$ events expected per year, where ξ_{30} is a measure of the expectation value for the square of number densities near the SMBH. The overall detection rate is sensitive to the number fraction of BHs as well as the maximum mass of BHs. If BH encounter and BBH merger events are classified effectively, we can get a more accurate population of BHs, and then will be able to constrain both the average star formation properties and upper mass of BHs in galactic nuclei.

In this study, limited by computational cost, **VI**tamin network training has not taken account of the spins of BBH. When modeling the data, the six additional parameters are expected to play an important role, allowing a further precise search. In addition, the component masses prior range of **VI**tamin can be expanded and cover more BH encounter samples. We will return to this subject in future work.

TABLE III: Parameters and optimal SNRs of BH encounter signals with a BBH-like posterior

No.	mass-ratio q	$M_{total}(M_{\odot})$	$d_L(\text{Mpc})$	location(rad.)			ρ_{opt}		
				ra	dec	ψ	H1	L1	V1
1	1	150	15	2.94	0.84	0.19	5.29	3.43	14.09
2	1	150	15	0.42	-0.53	0.92	5.15	7.13	12.52
3	2	150	10	5.83	-0.04	0.86	11.29	9.54	11.62
4	4	150	5	5.70	-0.44	1.82	9.44	9.57	15.13
5	8	150	5	0.01	0.41	3.03	13.29	11.07	3.85
6	16	150	5	0.01	0.41	3.03	13.29	11.07	3.85

TABLE IV: Bias of BH encounter's parameters using BBH model for Bayesian Inference.

No.	$m_1(\text{M}_{\odot})$			$m_2(\text{M}_{\odot})$			$d_L(\text{Mpc})$			location(rad.)									Bayes factor K		
										ra			dec			ψ			non-spinning	spinning	ratio
1	75	75.01	75.60	75	74.22	73.30	15	2579	2622	2.94	1.85	3.00	0.84	0.98	0.75	0.19	1.64	1.54	$24.34^{\pm 0.17}$	$24.98^{\pm 0.16}$	1.03
2	75	71.88	72.63	75	70.91	69.12	15	2633	2847	0.42	1.61	1.61	-0.53	0.83	0.84	0.92	1.21	1.64	$20.81^{\pm 0.17}$	$21.68^{\pm 0.18}$	1.04
3	100	79.56	98.60	50	76.88	68.48	10	3963	3162	5.83	4.87	4.89	-0.04	-0.98	-1.01	0.86	1.06	1.49	$38.08^{\pm 0.17}$	$120.16^{\pm 0.20}$	3.16
4	120	84.39	79.69	30	84.18	78.57	5	2523	2187	5.70	1.72	4.62	-0.44	1.00	-1.13	1.82	2.13	2.02	$53.77^{\pm 0.18}$	$62.59^{\pm 0.20}$	1.16
5	133.33	91.90	91.16	16.67	90.91	79.83	5	4332	3413	0.01	3.89	3.69	0.41	-0.94	-0.73	3.03	1.54	1.57	$38.33^{\pm 0.16}$	$46.03^{\pm 0.17}$	1.20
6	141.18			8.82																	

VI. ACKNOWLEDGMENTS

We especially thank

Appendix A: Appendix

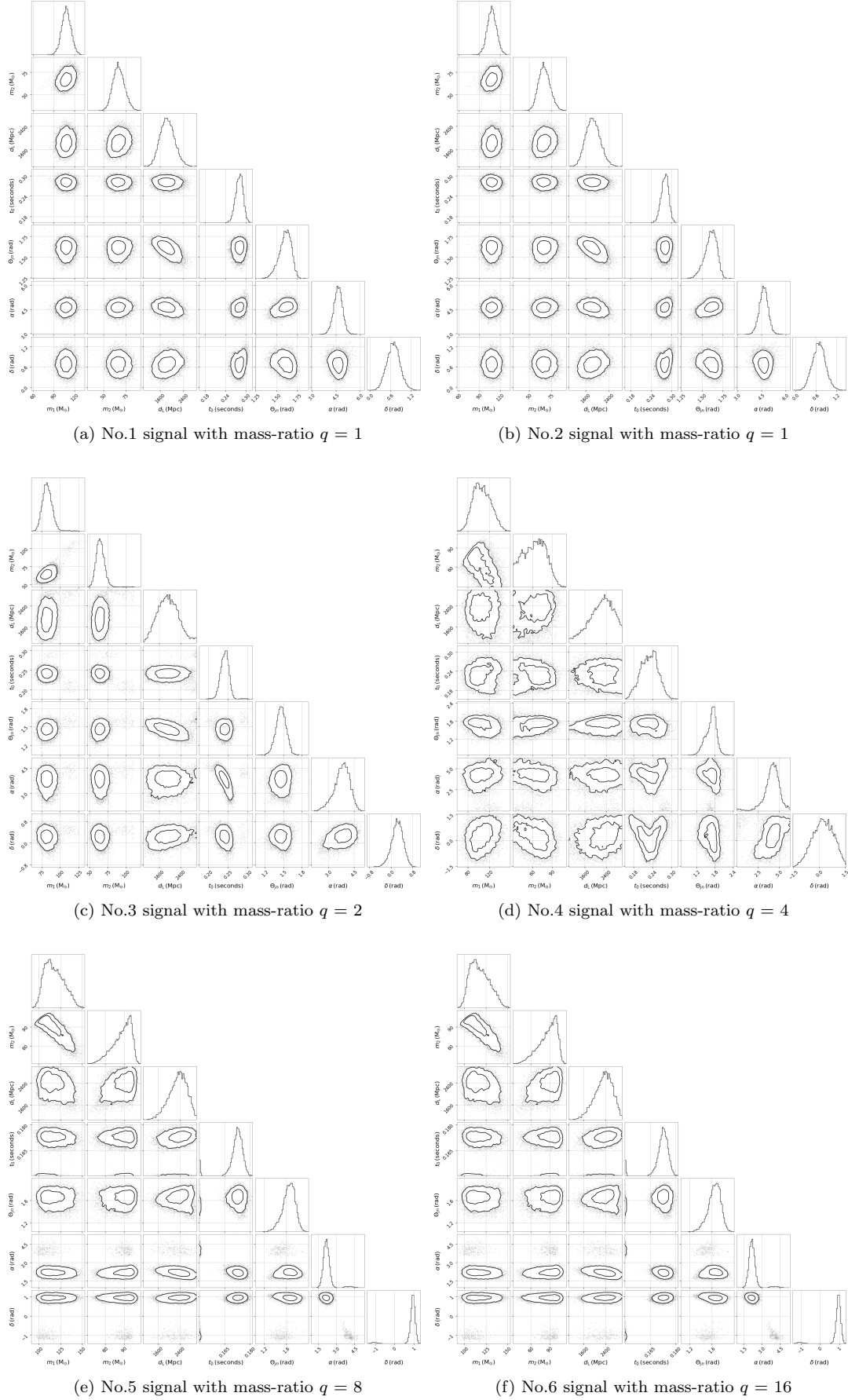


FIG. 4: The posterior distributions of BH encounter given by Vitamin network.

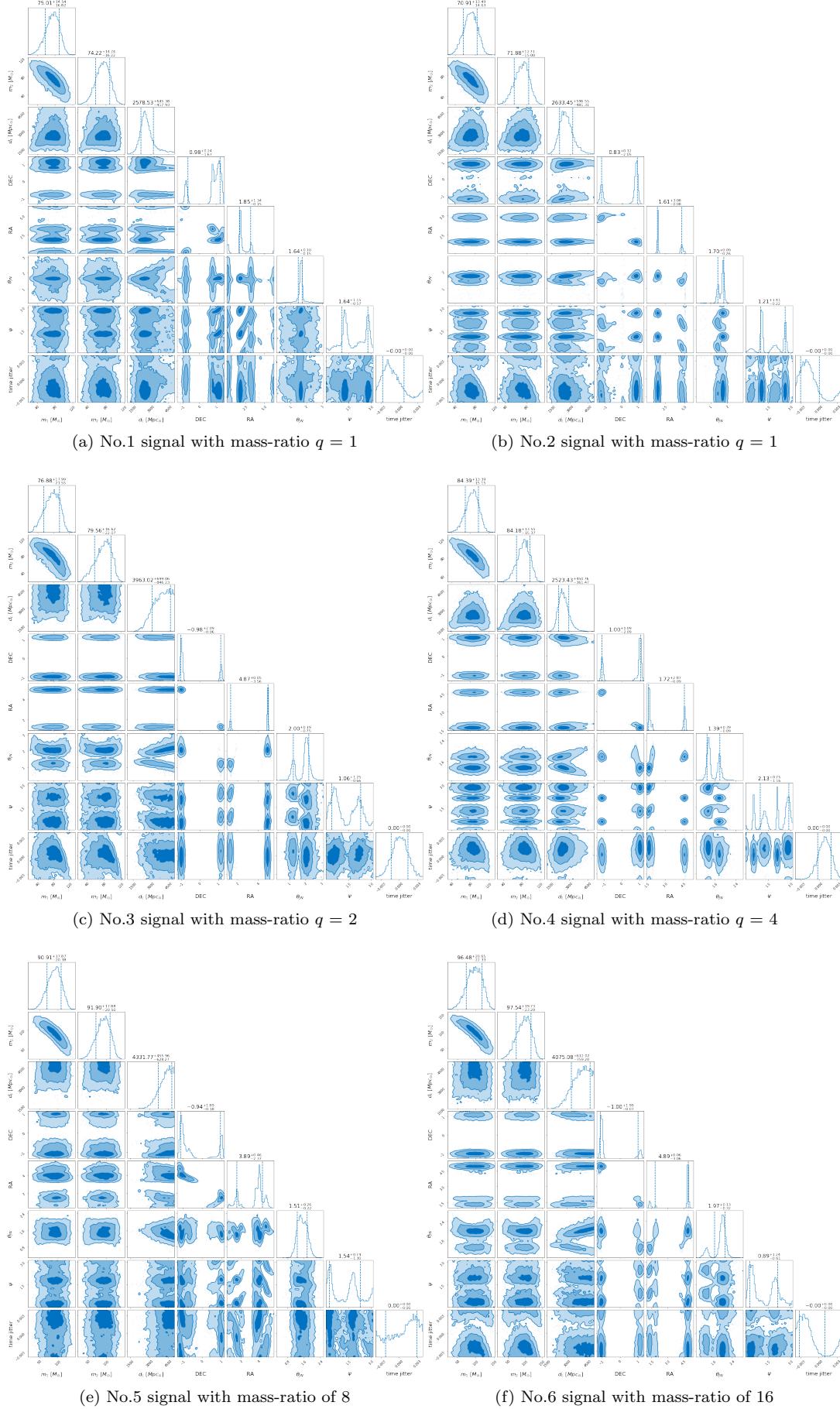


FIG. 5: The posterior distributions of BH encounter using non-spinning BBH model for Bayesian Inference.

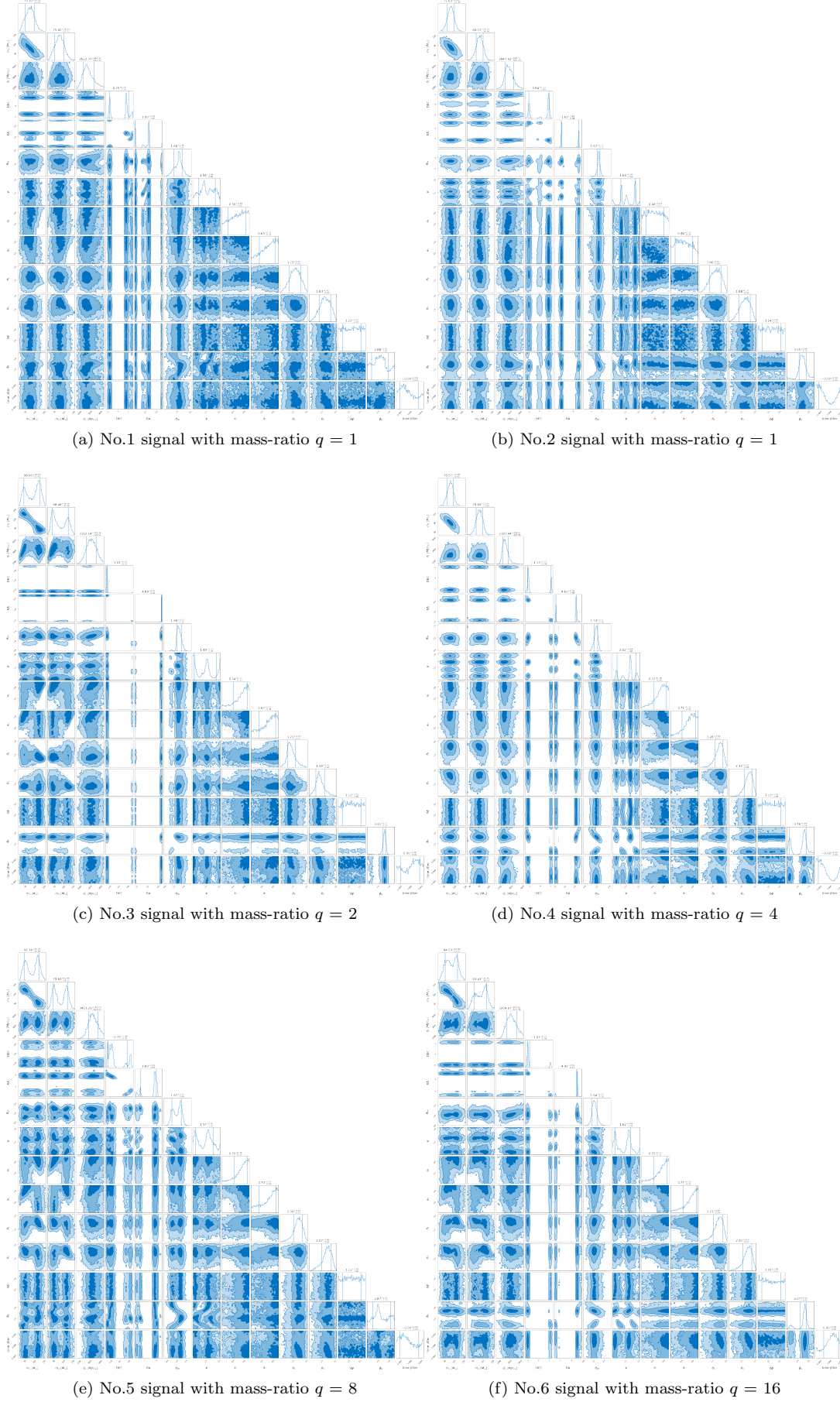


FIG. 6: The posterior distributions of BH encounter using spinning BBH model for Bayesian Inference.

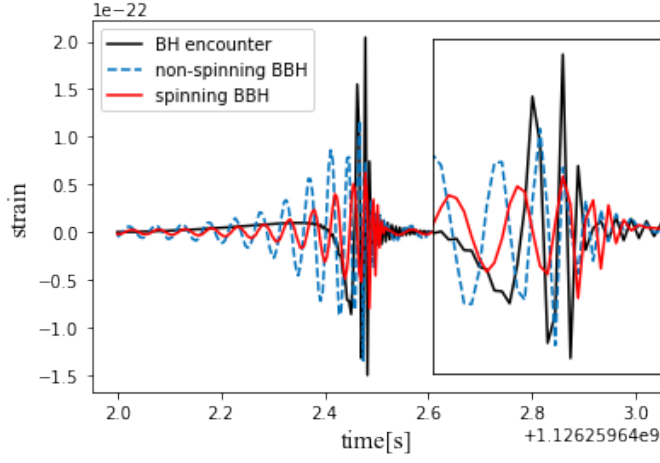
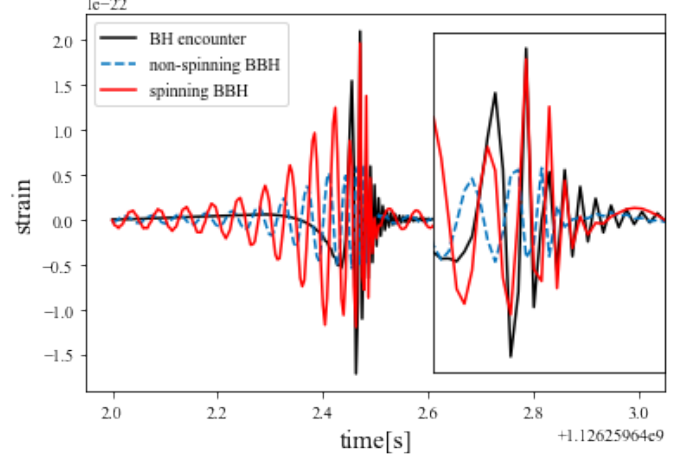
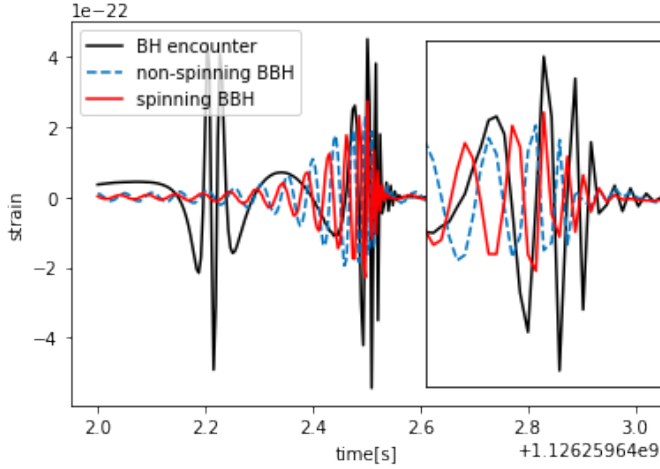
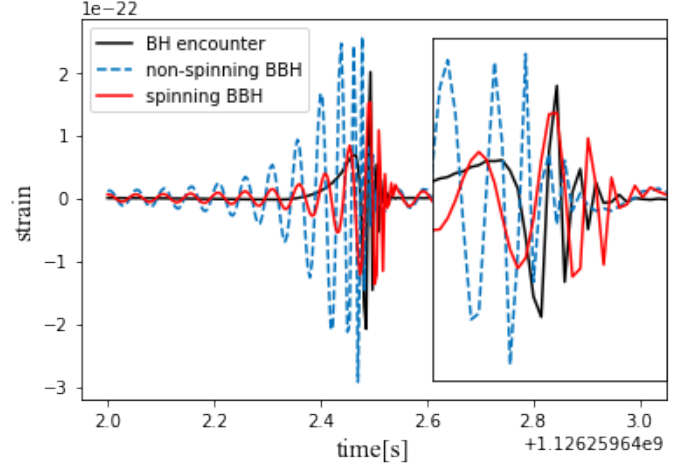
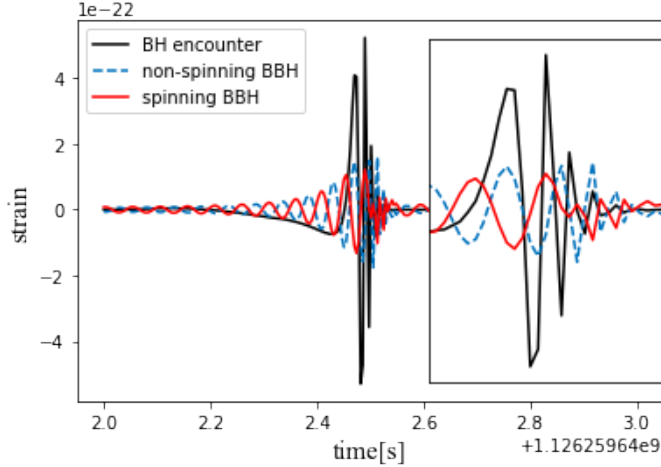
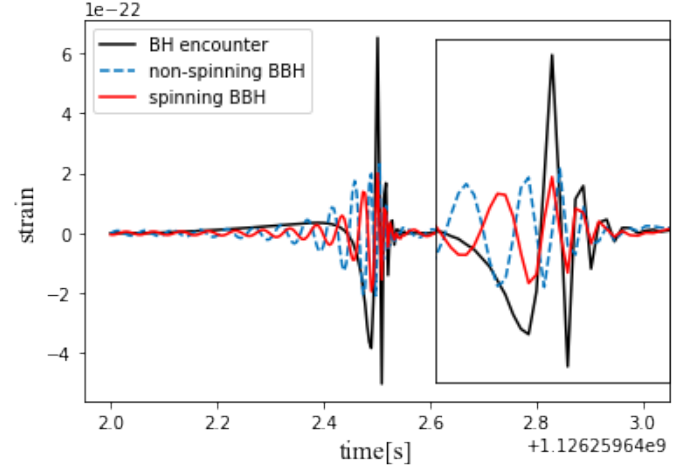
(a) No.1 signal with mass-ratio $q = 1$ (b) No.2 signal with mass-ratio $q = 1$ (c) No.3 signal with mass-ratio $q = 2$ (d) No.4 signal with mass-ratio $q = 4$ (e) No.5 signal with mass-ratio $q = 8$ (f) No.6 signal with mass-ratio $q = 16$

FIG. 7: The comparison of the critical impact parameters from different methods.